Springer Proceedings in Mathematics & Statistics

Dan Crisan
Ben Hambly
Thaleia Zariphopoulou *Editors*

Stochastic Analysis and Applications 2014

Springer

Springer Proceedings in Mathematics & Statistics

Volume 100

Springer Proceedings in Mathematics & Statistics

This book series features volumes composed of selected contributions from workshops and conferences in all areas of current research in mathematics and statistics, including operation research and optimization. In addition to an overall evaluation of the interest, scientific quality, and timeliness of each proposal at the hands of the publisher, individual contributions are all refereed to the high quality standards of leading journals in the field. Thus, this series provides the research community with well-edited, authoritative reports on developments in the most exciting areas of mathematical and statistical research today.

More information about this series at http://www.springer.com/series/10533

Dan Crisan · Ben Hambly
Thaleia Zariphopoulou
Editors

Stochastic Analysis and Applications 2014

In Honour of Terry Lyons

Springer

Editors
Dan Crisan
Department of Mathematics
Imperial College
London
UK

Thaleia Zariphopoulou
Department of Mathematics and IROM,
McCombs School of Business
University of Texas
Austin, TX
USA

Ben Hambly
Mathematical Institute
University of Oxford
Oxford
UK

ISSN 2194-1009 ISSN 2194-1017 (electronic)
Springer Proceedings in Mathematics & Statistics
ISBN 978-3-319-11291-6 ISBN 978-3-319-11292-3 (eBook)
DOI 10.1007/978-3-319-11292-3

Library of Congress Control Number: 2014957408

Mathematics Subject Classification (2010): 60H10, 60H15, 60H30

Springer Cham Heidelberg New York Dordrecht London

Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media (www.springer.com)

Preface

Stochastic Analysis is the branch of mathematics that deals with the analysis of dynamical systems affected by noise. It emerged as a core area of mathematics in the late twentieth century and has subsequently developed into an important theory with a wide range of powerful and novel tools, and with impressive diverse applications within and beyond mathematics. As so many systems are profoundly affected by stochastic fluctuations, it is not surprising that the array of applications of Stochastic Analysis is vast and touches many aspects of life.

This volume includes articles from some of the main contributors to the recent progress in stochastic analysis, and provides a snapshot of the current state of the area and its ongoing developments. It constitutes the proceedings of the conference on "Stochastic Analysis and Applications" held at the University of Oxford and the Oxford-Man Institute during 23–27 September 2013. The conference honoured the 60th birthday of Professor Terry Lyons FLSW FRSE FRS, Wallis Professor of Mathematics, University of Oxford. Terry Lyons is one of the leaders of the field of stochastic analysis. His introduction of the notion of rough path has revolutionized the field in both theory and applications.

A Biographical Account of Terry Lyons

Terence John Lyons was born in May 1953 in London. He went to Trinity College, Cambridge as an undergraduate from 1972–1976. His doctoral studies were in Oxford under the supervision of Richard Hayden. During this period he became interested in stochastic analysis and in particular some of the work of Paul Malliavin. In his early development he was influenced by Ted Gamelin and Henry McKean.

After completing his D.Phil. he held a Junior Research Fellowship at Jesus College, Oxford before taking up a Hedrick visiting professorship at UCLA for 1981–1982. He returned to the UK to take up a lectureship at Imperial College, London which he held until 1985. At that time he was appointed to the Colin

MacLaurin chair at the University of Edinburgh, where he took a turn as Head of Department. He moved back to a chair at Imperial College in 1993 but still maintained links with Edinburgh to encourage the group that he had started there. In 1993 he was awarded an EPSRC Senior Fellowship and used this to develop the theory of rough paths. In 2000 he moved to the Wallis Professorship of Mathematics at the University of Oxford and a Professorial Fellowship at St Anne's College.

Over the course of his career he has been awarded many prizes and honours. He was awarded the Rollo Davidson prize in 1985, a London Mathematical Society Junior Whitehead prize in 1986. He was awarded the Polya prize of the LMS in 2000. He was elected a fellow of the Royal Society of Edinburgh in 1987, the Royal Society in 2002 and the Learned Society of Wales in 2011. He has an Honorary degree from the University Paul Sabatier, Toulouse as well as Honorary fellowships at the Universities of Aberystwyth and Cardiff. He became the first director of the Wales Institute of Mathematical and Computational Sciences in 2008, which he held until 2011 when he became the director of the Oxford-Man Institute of Quantitative Finance. In 2013 he became president of the London Mathematical Society. He has been an invited speaker at the International Congress of Mathematicians, the Schramm lecturer for the Institute of Mathematical Statistics as well as delivering expository lecture courses such as the Summer School in Probability at St Flour.

Terry has been very much concerned with providing for the next generation of mathematicians. From 1988–2000 he managed a series of three EU grants bringing together the leading European institutions in the field, allowing the development of a generation of stochastic analysts. He has also been a great proponent of mathematics in bringing new ideas and techniques from academia to industry and, in particular, its transformative power in tackling the challenges thrown up by the growth and complexity of many aspects of modern society. He has written ground breaking papers on potential theory, Dirichlet forms, Markov chains, numerical analysis, filtering and mathematical finance. However, it is his pathwise view of integration, developed as the theory of rough paths, that has had a profound effect both within and outside the field of stochastic analysis. It has led to deep and powerful results in stochastic differential equations and stochastic partial differential equations. In particular it provided the initial tools that were built on by Martin Hairer in the work that led to the award of a Fields Medal in 2014. Terry Lyons' introduction of the signature of a path as a tool to provide an effective description of a path and the way that it acts has already provided new insights into efficient extraction of information from data.

In his personal life Terry married Barbara in 1975 and has two children, Barnaby, born 1981, and Josephine, born 1983, and currently one grandchild, born 2014.

Throughout his career, Terry Lyons supervised a large number of Ph.D. students. The following is the list of students that have completed or are currently working on their Ph.D. under his supervision:

Babbar, Katia	Boutaib, Youness	Caruana, Michael
Chang, C.Y.	Chevyrev, Ilya	Crisan, Dan
Dickinson, Andrew S.	Fawcett, Thomas	Flint, Guy
Gaines, Jessica	Gyurko, Greg	Hitchcock, David
Hoff, Ben	Janssen, Arend	Liang, Gechun
Litterer, Christian	Lunt, John	Ni, Hao
Pan, Wei	Penrose, Mathew	Potter, Chris
Rapoo (Sipilainen), Eeva	Skipper, Max	Smith, Adam
Tchernychova, Maria	Victoir, Nicolas	Williams, David
Xu, Weijun	Yam, Phillip	Yang, Danyu

The Contents of This Volume

The first chapter starts with the contribution of Shigeki Aida. In it, the author continues his previous work on the study of the strong convergence of Wong-Zakai approximations of the solution to the reflecting stochastic differential equations. In this chapter, he proves the strong convergence under weaker assumptions on the domain. The first main theorem shows the convergence when the domain is convex. The estimate of the order of the convergence is the same as that given in the previous work. The second main theorem establishes the convergence when the domain is not convex, but satisfies certain additional conditions.

The contribution of Dominique Bakry contains a description of symmetric diffusion operators where the spectral decomposition is given through a family of orthogonal polynomials. In dimension one, this reduces to the case of Hermite, Laguerre and Jacobi polynomials. In higher dimension, some basic examples arise from compact Lie groups. The author gives a complete description of the bounded sets on which such operators may live and a classification of those sets when the polynomials are ordered according to their usual degrees.

The contribution of Erich Baur and Jean Bertoin discusses old and new results related to the destruction of a random recursive tree (RRT), in which its edges are cut one after the other in a uniform random order. In particular, the authors study the number of steps needed to isolate or disconnect certain distinguished vertices when the size of the tree tends to infinity. New probabilistic explanations are given in terms of the so-called cut-tree and the tree of component sizes, which both encode different aspects of the destruction process. Finally, the authors establish the connection to Bernoulli bond percolation on large RRTs and present recent results on the cluster sizes in the supercritical regime.

The contribution of René Carmona and Francois Delarue discusses the Master Equation for large population equilibriums. The authors use a simple N-player stochastic game with idiosyncratic and common noises to introduce the concept of the Master Equation, originally proposed by Lions in his lectures at the Collège de France. They highlight the stochastic nature of the limit distributions of the states of the players due to the fact that the random environment does not average out in the limit, and recast the Mean Field Game (MFG) paradigm in a set of coupled Stochastic Partial Differential Equations. The first one is a forward stochastic Kolmogorov equation giving the evolution of the conditional distributions of the states of the players given the common noise. The second equation has the form of a stochastic Hamilton Jacobi Bellman (HJB) equation providing the solution of the optimization problem when the flow of conditional distributions is given. Being highly coupled, the system reads as an infinite dimensional Forward Backward Stochastic Differential Equation (FBSDE). Uniqueness of a solution and its Markov property lead to the representation of the solution of the backward equation (i.e. the value function of the stochastic HJB equation) as a deterministic function of the solution of the forward Kolmogorov equation, function which is usually called the decoupling field of the FBSDE. The (infinite dimensional) PDE satisfied by this decoupling field is identified with the master equation. Finally the authors show that this equation can be derived for other large populations equilibriums like those given by the optimal control of McKean-Vlasov stochastic differential equations.

The contribution of Thomas Cass, Martin Clark and Dan Crisan revisits the filtering equations. The problem of nonlinear filtering has engendered a surprising number of mathematical techniques for its treatment. A notable example is the change-of-probability-measure method introduced by Kallianpur and Striebel to derive the filtering equations and the Bayes-like formula that bears their names. More recent work, however, has generally preferred other methods. In this chapter, the authors reconsider the change-of-measure approach to the derivation of the filtering equations and show that many of the technical conditions present in previous work can be relaxed. The filtering equations are established for general Markov signal processes that can be described by a martingale problem formulation. Two specific applications are treated.

The contribution of Ana Bela Cruzeiro and Remi Lassalle discusses the stochastic least action principle for the Navier-Stokes equation. The authors extend the class of stochastic processes allowed to represent solutions of the Navier-Stokes equation on the two-dimensional torus to certain non-Markovian processes, which they call admissible. More precisely, they provide a criterion for the associated mean velocity field to solve this equation. Due to the fluctuations of the shift, a new term of pressure appears which is of purely stochastic origin. The authors also provide an alternative formulation of this least action principle by means of transformations of measure. Within this approach, the action is a function of the law of the processes, while the variations are induced by some translations on the space of the divergence-free vector fields. Due to the renormalization in the definition of the cylindrical Brownian motion, this action is only related to the relative entropy by an inequality. However it is shown that, if the high frequency modes are cut, this new approach provides a least action principle for the Navier-Stokes equation based on the relative entropy.

The contribution of Sandy Davie studies the dyadic method of Komlós, Major and Tusnády (KMT), which is a powerful way of constructing simultaneous normal approximations to a sequence of partial sums of i.i.d. random variables. The author uses a version of this KMT method to obtain a first-order approximation in a Vaserstein metric to solutions of vector SDEs under a mild nondegeneracy condition using an easily implemented numerical scheme.

The contribution of Joscha Diehl, Peter Friz and Harald Oberhauser studies partial differential equations driven by rough paths. This is a continuation of the authors' earlier work on the subject motivated by the Lions-Souganidis theory of viscosity solutions for SPDEs. The authors continue and complement the previous (uniqueness) results with general existence and regularity statements. Much of this is transformed to questions for deterministic parabolic partial differential equations in viscosity sense. On a technical level, the authors establish a refined parabolic theorem of sums which may be useful in its own right.

The contribution of Yidong Dong and Ronnie Sircar discusses time-inconsistent portfolio investment problems. The explicit results for the classical Merton optimal investment/consumption problem rely on the use of constant risk aversion parameters and exponential discounting. However, many studies have suggested that individual investors can have different risk aversions over time, and they discount future rewards less rapidly than exponentially. While state-dependent risk aversion and nonexponential type (e.g. hyperbolic) discounting align more with real life behaviour and household consumption data, they have tractability issues and make the problem time-inconsistent. In their contribution, Dong and Sircar analyse the cases where these problems can be closely approximated by time-consistent ones. Using asymptotic approximations, they are able to characterize the equilibrium strategies explicitly in terms of corrections to solutions for the base problems with constant risk aversion and exponential discounting. The authors also explore the effects of hyperbolic discounting under proportional transaction costs.

The contribution of David Elworthy discusses decompositions of diffusion operators and related couplings. Results by Cranston, Greven and Feng-Yu Wang on relationships between coupling and shift coupling, and harmonic functions and space time harmonic functions are reviewed. These lead to extensions of a result by Freire on the separate harmonicity of bounded harmonic functions on certain product manifolds. The extensions are to situations where a diffusion operator is decomposed into the sum of two other commuting diffusion operators. This is shown to arise for a class of foliated Riemannian manifolds with totally geodesic leaves. A form of skew product decomposition of Brownian motions on these foliated manifolds is obtained, as are gradient estimates in leaf directions. Relationships between stochastic completeness of the manifold itself and stochastic completeness of its leaves are established. Baudoin and Garafola's "sub-Riemannian manifolds with transverse symmetries" are shown to be examples.

The contribution of Hans Föllmer and Claudia Klüppelberg studies a mathematical consistency problem motivated by the interplay between local and global risk assessment in a large financial network. In analogy to the theory of Gibbs measures in Statistical Mechanics, they focus on the structure of global convex risk measures which are consistent with a given family of local conditional risk measures. Going beyond the locally law-invariant (and hence entropic) case, the authors show that a global risk measure can be characterized by its behaviour on a suitable boundary field. In particular, a global risk measure may not be uniquely determined by its local specification, and this can be seen as a source of "systemic risk" in analogy to the appearance of phase transitions in the theory of Gibbs measures. The proof combines the spatial version of Dynkin's method for constructing the entrance boundary of a Markov process with a certain nonlinear extension of backwards martingale convergence.

In their contribution, Masatoshi Fukushima and Hiroshi Kaneko discuss the Villat's kernels and BMD Schwarz kernels in Komatu-Loewner equations. The classical Loewner differential equation for simply connected domains is attracting new attention since Schramm launched in 2000 the stochastic Loewner evolution (SLE) based on it. The Loewner equation itself has been extended to various canonical domains of multiple connectivity after the works by Komatu in 1943 and 1950, but the Komatu-Loewner (K-L) equations have been derived rigorously only in the left derivative sense. In a recent work, Chen, Fukushima and Rhode prove that the K-L equation for the standard slit domain is a genuine ODE by using a probabilistic method together with an SDE method, and that the right-hand side of the equation admits an expression in terms of the complex Poisson kernel of the Brownian motion with darning (BMD). In the present paper, K-L equations for the annulus and circularly slit annili are investigated. For the annulus, they establish a K-L equation as a genuine ODE possessing a normalized Villat's kernel on its right-hand side by using a variant of the Carathéodory convergence theorem for annuli indicated by Komatu. This method is also used to obtain the same K-L equation in the right derivative sense on annulus for a more general family of growing hulls that satisfies a specific right continuity condition usually adopted in the SLE theory. Villat's kernel is then identified with a BMD Schwarz kernel for the annulus. Finally, the authors derive K-L equations for circularly slit annuli in terms of their normalized BMD Schwarz kernels, but only in the left derivative sense when at least one circular slit is present.

Tomoyuki Ichiba and Ioannis Karatzas study the unfolding of the Skorokhod reflection of a continuous semimartingale, in a possibly skewed manner, into another continuous semimartingale on an enlarged probability space according to the excursion-theoretic methodology of Prokaj. This is done in terms of a skew version of the Tanaka equation, whose properties are studied in some detail. The result is used to construct a system of two diffusive particles with rank-based characteristics and skew-elastic collisions. Unfoldings of conventional reflections are also discussed, as are examples involving skew Brownian Motions and skew Bessel processes.

David Nualart contributes with a survey of some recent developments in the applications of Malliavin calculus combined with Stein's method to derive central limit theorems for random variables on a finite sum of Wiener chaos. Starting from the fourth moment theorem by Nualart and Peccati, the author discusses several related topics such as conditions for the convergence in total variation, absolute continuity of probability laws and uniform convergence of densities under suitable nondegeneracy assumptions. The fact that the random variables belong to a fixed

Wiener chaos (or to a finite sum of Wiener chaos) will play a fundamental role in the results. Normal approximation on a finite Wiener chaos.

Zhenjie Ren, Nizar Touzi and Jianfeng Zhang provide an overview of the recently developed notion of viscosity solutions of path-dependent partial differential equations. The authors review the Crandall-Ishii notion of viscosity solutions, so as to motivate the relevance of the definition in the path-dependent case. The authors focus on the well-posedness theory of such equations. In particular, they provide a simple presentation of the current existence and uniqueness arguments in the semilinear case and review the stability property of this notion of solutions, including the adaptation of the Barles-Souganidis monotonic scheme approximation method. The results rely crucially on the theory of optimal stopping under nonlinear expectation. In the dominated case, we provide a self-contained presentation of all required results. The fully nonlinear case is more involved and is addressed elsewhere.

Marta Sanz-Sole and Andre Suess study logarithmic asymptotics of the densities of SPDEs driven by spatially correlated noise. The authors consider the family of stochastic partial differential equations indexed by a parameter $\varepsilon \in (0,1]$,

$$Lu^\varepsilon(t,x) = \varepsilon\sigma(u^\varepsilon(t,x))\dot{F}(t,x) + b(u^\varepsilon(t,x)),$$

$(t,x) \in (0,T] \times \mathbb{R}^d$ with suitable initial conditions. In this equation, L is a second-order partial differential operator with constant coefficients, σ and b are smooth functions and $\dot{F}$ is a Gaussian noise, white in time and with a stationary correlation in space. Let $p^\varepsilon_{t,x}$ denote the density of the law of $u^\varepsilon(t,x)$ at a fixed point $(t,x) \in (0,T] \times \mathbb{R}^d$. The authors study the existence of $\lim_{\varepsilon\downarrow 0} \varepsilon^2 \log p^\varepsilon_{t,x}(y)$ for a fixed $y \in \mathbb{R}$. The results apply to classes of stochastic wave equations with $d \in \{1,2,3\}$ and stochastic heat equations with $d \geq 1$.

Terry Lyons' List of Publications to 2013

Ni, Hao; Lyons, Terry J. Expected signature of Brownian Motion up to the first exit time from a bounded domain, to appear Ann. Probab.

Cass, Tom; Lyons, Terry J. Evolving communities with individual preferences, to appear Proc. London Math. Soc.

Lyons, Terry J.; Xu, Weijun, A uniform estimate for rough paths. Bull. Sci. Math. 137 (2013), 867–879.

Lyons, Terry J.; Yang, Danyu, The partial sum process of orthogonal expansions as geometric rough process with Fourier series as an example—an improvement of Menshov-Rademacher theorem. J. Funct. Anal. 265 (2013), 3067–3103.

Cass, Thomas; Litterer, Christian; Lyons, Terry, Integrability and tail estimates for Gaussian rough differential equations. Ann. Probab. 41 (2013), 3026–3050.

Litterer, C.; Lyons, Terry, High order recombination and an application to cubature on Wiener space. Ann. Appl. Probab. 22 (2012), 1301–1327.

Cass, Thomas; Litterer, Christian; Lyons, Terry, Rough paths on manifolds. New trends in stochastic analysis and related topics, 33–88, Interdiscip. Math. Sci., 12, World Sci. Publ., Hackensack, NJ, 2012.

Litterer, C.; Lyons, T, Introducing cubature to filtering. The Oxford handbook of nonlinear filtering, 768–796, Oxford Univ. Press, Oxford, 2011.

Liang, Gechun; Lyons, Terry; Qian, Zhongmin, Backward stochastic dynamics on a filtered probability space. Ann. Probab. 39 (2011), 1422–1448.

Gyurkó, Lajos Gergely; Lyons, Terry J., Efficient and practical implementations of cubature on Wiener space. Stochastic analysis 2010, 73–111, Springer, Heidelberg, 2011.

Lyons, Terence J., Kiyoshi Itô (1915–2008). Probab. Theory Related Fields 148 (2010), 1–4.

Gyurkó, Lajos Gergely; Lyons, Terry, Rough paths based numerical algorithms in computational finance. Mathematics in finance, 17–46, Contemp. Math., 515, Amer. Math. Soc., Providence, RI, 2010.

Hambly, Ben; Lyons, Terry, Uniqueness for the signature of a path of bounded variation and the reduced path group. Ann. of Math. (2) 171 (2010), 109–167.

Levin, Daniel; Lyons, Terry, A signed measure on rough paths associated to a PDE of high order: results and conjectures. Rev. Mat. Iberoam. 25 (2009), 971–994.

Litterer, Christian; Lyons, Terry, Cubature on Wiener space continued. Stochastic processes and applications to mathematical finance, 197–217, World Sci. Publ., Hackensack, NJ, 2007.

Hara, Keisuke; Lyons, Terry, Smooth rough paths and the applications. Stochastic processes and applications to mathematical finance, 115–125, World Sci. Publ., Hackensack, NJ, 2007.

Hara, Keisuke; Lyons, Terry, Smooth rough paths and applications for Fourier analysis. Rev. Mat. Iberoam. 23 (2007), 1125–1140.

Lyons, Terry; Victoir, Nicolas, An extension theorem to rough paths. Ann. Inst. H. Poincaré Anal. Non Linéaire 24 (2007), 835–847.

Lyons, Terry J.; Caruana, Michael; Lévy, Thierry, Differential equations driven by rough paths. Lectures from the 34th Summer School on Probability Theory held in Saint-Flour, July 6–24, 2004. With an introduction concerning the Summer School by Jean Picard. Lecture Notes in Mathematics, 1908. Springer, Berlin, 2007. xviii+109 pp.

Li, Xiang-Dong; Lyons, Terry J., Smoothness of Itô maps and diffusion processes on path spaces. I. Ann. Sci. École Norm. Sup. (4) 39 (2006), 649–677.

Lyons, Terry J.; Sidorova, Nadia, On the radius of convergence of the logarithmic signature. Illinois J. Math. 50 (2006), 763–790.

Lyons, Terry J.; Yam, Phillip S. C., On Gauss-Green theorem and boundaries of a class of Hölder domains. J. Math. Pures Appl. (9) 85 (2006), 38–53.

Friz, P.; Lyons, T.; Stroock, D., Lévy's area under conditioning. Ann. Inst. H. Poincaré Probab. Statist. 42 (2006), 89–101.

Lejay, Antoine; Lyons, Terry, On the importance of the Lévy area for studying the limits of functions of converging stochastic processes. Application to homogenization. Current trends in potential theory, 63–84, Theta Ser. Adv. Math., 4, Theta, Bucharest, 2005.

Lyons, Terry, Systems controlled by rough paths. European Congress of Mathematics, 269–281, Eur. Math. Soc., Zürich, 2005.

Lyons, Terry; Victoir, Nicolas, Cubature on Wiener space. Stochastic analysis with applications to mathematical finance. Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci. 460 (2004), 169–198.

Lyons, Terry; Qian, Zhongmin, System control and rough paths. Oxford Mathematical Monographs. Oxford Science Publications. Oxford University Press, Oxford, 2002. x+216 pp.

Lyons, T. J. System control and rough paths. Numerical methods and stochastics (Toronto, ON, 1999), 91–99, Fields Inst. Commun., 34, Amer. Math. Soc., Providence, RI, 2002.

Crisan, D.; Lyons, T., Optimal filtering on discrete sets. Numerical methods and stochastics (Toronto, ON, 1999), 21–27, Fields Inst. Commun., 34, Amer. Math. Soc., Providence, RI, 2002. Numerical methods and stochastics. Proceedings of the workshop held in Toronto, ON, April 20–23, 1999. Edited by T. J. Lyons and

T. S. Salisbury. Fields Institute Communications, 34. American Mathematical Society, Providence, RI, 2002. vi+121 pp.

Crisan, Dan; Lyons, Terry, Minimal entropy approximations and optimal algorithms. Monte Carlo Methods Appl. 8 (2002), 343–355.

Bass, R. F.; Hambly, B. M.; Lyons, T. J. Extending the Wong-Zakai theorem to reversible Markov 4 processes. J. Eur. Math. Soc. (JEMS) 4 (2002), 237–269.

Ledoux, M.; Lyons, T.; Qian, Z., Lévy area of Wiener processes in Banach spaces. Ann. Probab. 30 (2002), 546–578.

Lyons, Terry; Smith, Adam, Uncertain volatility. Risk Magazine, (September 1999), 106–109.

Lyons, Terry J., Derivatives as tradable assets. Seminario de Matematica Financiera MEFF-UAM, 2, (1999) 213–232.

Lyons, Terry; Zeitouni, Ofer, Conditional exponential moments for iterated Wiener integrals. Ann. Probab. 27 (1999), 1738–1749.

Crisan, D.; Del Moral, P.; Lyons, T. J., Interacting particle systems approximations of the Kushner- Stratonovitch equation. Adv. in Appl. Probab. 31 (1999), 819–838.

Crisan, D.; Lyons, T., A particle approximation of the solution of the Kushner-Stratonovitch equation. Probab. Theory Related Fields 115 (1999), 549–578.

Crisan, D.; Del Moral, P.; Lyons, T., Discrete filtering using branching and interacting particle systems. Markov Process. Related Fields 5 (1999), 293–318.

Lyons, Terry; Stoica, Lucretiu, The limits of stochastic integrals of differential forms. Ann. Probab. 27 (1999), 1–49.

Lyons, Terry J., Differential equations driven by rough signals. Rev. Mat. Iberoamericana 14 (1998), 215–310.

Lyons, Terry; Qian, Zhongmin, Flow of diffeomorphisms induced by a geometric multiplicative functional. Probab. Theory Related Fields 112 (1998), 91–119.

Crisan, Dan; Gaines, Jessica; Lyons, Terry, Convergence of a branching particle method to the solution of the Zakai equation. SIAM J. Appl. Math. 58 (1998), 1568–1590.

Lunt, John; Lyons, T. J.; Zhang, T. S., Integrability of functionals of Dirichlet processes, probabilistic representations of semigroups, and estimates of heat kernels. J. Funct. Anal. 153 (1998), 320–342.

Hambly, B. M.; Lyons, T. J., Stochastic area for Brownian motion on the Sierpinski gasket. Ann. Probab. 26 (1998), 132–148.

Lyons, T. J.; Qian, Z. M., Calculus of variation for multiplicative functionals. New trends in stochastic analysis (Charingworth, 1994), 348–374, World Sci. Publ., River Edge, NJ, 1997.

Lyons, Terry; Qian, Zhongmin, Stochastic Jacobi fields and vector fields induced by varying area on path spaces. Probab. Theory Related Fields 109 (1997), 539–570.

Crisan, Dan; Lyons, Terry, Nonlinear filtering and measure-valued processes. Probab. Theory Related Fields 109 (1997), 217–244.

Lyons, Terry; Qian, Zhongmin, Flow equations on spaces of rough paths. J. Funct. Anal. 149 (1997), 135–159.

Gaines, J. G.; Lyons, T. J., Variable step size control in the numerical solution of stochastic differential equations. SIAM J. Appl. Math. 57 (1997), 1455–1484.

Lyons, T. J.; Qian, Z. M., A class of vector fields on path spaces. J. Funct. Anal. 145 (1997), 205–223.

Lyons, T. J.; Qian, Z. M., Calculus for multiplicative functionals, Itô's formula and differential equations. Itô's stochastic calculus and probability theory, 233–250, Springer, Tokyo, 1996.

Lyons, Terry; Zhang, Tusheng, Convergence of non-symmetric Dirichlet processes. Stochastics Stochastics Rep. 57 (1996), 159–167.

Lyons, T. J.; Röckner, M.; Zhang, T. S., Martingale decomposition of Dirichlet processes on the Banach space $C_0[0; 1]$. Stochastic Process. Appl. 64 (1996), 31–38.

Lyons, Terry; Stoica, Lucretiu, On the limit of stochastic integrals of differential forms. Stochastic processes and related topics (Siegmundsberg, 1994), 61–66, Stochastics Monogr., 10, Gordon and Breach, Yverdon, 1996.

Lyons, T. J., Uncertain volatility and the risk-free synthesis of derivatives. Appl. Math. Finance 2 (1995), 117–133

Albeverio, S.; Lyons, T.; Rozanov, Yu., Boundary conditions for stochastic evolution equations with an extremely chaotic source. (Russian) Mat. Sb. 186 (1995), 3–20; translation in Sb. Math. 186 (1995), 1693–1709.

Lyons, Terry J., The interpretation and solution of ordinary differential equations driven by rough signals. Stochastic analysis (Ithaca, NY, 1993), 115–128, Proc. Sympos. Pure Math., 57, Amer. Math. Soc., Providence, RI, 1995.

Lyons, Terry, Differential equations driven by rough signals. I. An extension of an inequality of L. C. Young. Math. Res. Lett. 1 (1994), 451–464.

Gaines, J. G.; Lyons, T. J., Random generation of stochastic area integrals. SIAM J. Appl. Math. 54 (1994), 1132–1146.

Lyons, T. J.; Zhang, T. S., Decomposition of Dirichlet processes and its application. Ann. Probab. 22 (1994), 494–524.

Lyons, T. J.; Zhang, T. S., Note on convergence of Dirichlet processes. Bull. London Math. Soc. 25 (1993), 353–356.

Duplantier, B.; Lawler, G. F.; Le Gall, J.-F.; Lyons, T. J., The geometry of the Brownian curve. Bull. Sci. Math. 117 (1993), 91–106.

Lyons, Terry, Random thoughts on reversible potential theory. Summer School in Potential Theory (Joensuu, 1990), 71–114, Joensuun Yliop. Luonnont. Julk., 26, Univ. Joensuu, Joensuu, 1992.

Erëmenko, A. E.; Lyons, T. J., Finely open sets in the limit set of a finitely generated Kleinian group. Approximation by solutions of partial differential equations (Hanstholm, 1991), 61–67, NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci., 365, Kluwer Acad. Publ., Dordrecht, 1992.

Lyons, Terry; Röckner, Michael, A note on tightness of capacities associated with Dirichlet forms. 6 Bull. London Math. Soc. 24 (1992), 181–184.

Lyons, Terry, Instability of the conservative property under quasi-isometries. J. Differential Geom. 34 (1991), 483–489.

Lyons, Terry, On the nonexistence of path integrals. Proc. Roy. Soc. London Ser. A 432 (1991), 281–290.

Hayman, W. K.; Lyons, T. J., Bases for positive continuous functions. J. London Math. Soc. (2) 42 (1990), 292–308.

Lyons, T. J.; Zheng, W. A., On conditional diffusion processes. Proc. Roy. Soc. Edinburgh Sect. A 115 (1990), 243–255.

Lyons, T. J.; Zheng, W. A., Diffusion processes with nonsmooth diffusion coefficients and their density functions. Proc. Roy. Soc. Edinburgh Sect. A 115 (1990), 231–242.

Lyons, Terry, A synthetic proof of Makarov's law of the iterated logarithm. Bull. London Math. Soc. 22 (1990), 159–162.

Lyons, Terence J.; Zheng, Wei An, A crossing estimate for the canonical process on a Dirichlet space and a tightness result. Colloque Paul Lévy sur les Processus Stochastiques (Palaiseau, 1987). Astérisque No. 157–158 (1988), 249–271.

Lyons, T. J., Reversible diffusion processes on manifolds. Proceedings of the 1st World Congress of the Bernoulli Society, Vol. 1 (Tashkent, 1986), 297–305, VNU Sci. Press, Utrecht, 1987.

Lyons, T. J., What you can do with n observations. Geometrization of statistical theory (Lancaster, 1987), 209–218, ULDM Publ., Lancaster, 1987.

Lyons, Terry, Instability of the Liouville property for quasi-isometric Riemannian manifolds and reversible Markov chains. J. Differential Geom. 26 (1987), 33–66.

Lyons, Terry J., The critical dimension at which quasi-every Brownian path is self-avoiding. Adv. in Appl. Probab. 1986, suppl., 87–99.

Barnett, Chris; Lyons, Terry, Stopping noncommutative processes. Math. Proc. Cambridge Philos. Soc. 99 (1986), 151–161.

Lyons, T. J.; Reuter, G. E. H., On exponential bounds for solutions of second order differential equations. Bull. London Math. Soc. 17 (1985), 139–143.

Hayman, Walter K.; Kershaw, Donald; Lyons, Terry J., The best harmonic approximant to a continuous function. Anniversary volume on approximation theory and functional analysis (Oberwolfach, 1983), 317–327, Internat. Schriftenreihe Numer. Math., 65, Birkhäuser, Basel, 1984.

Lyons, T. J.; MacGibbon, K. B.; Taylor, J. C., Projection theorems for hitting probabilities and a theorem of Littlewood. J. Funct. Anal. 59 (1984), 470–489.

Lyons, Terry; Sullivan, Dennis, Function theory, random paths and covering spaces. J. Differential Geom. 19 (1984), 299–323.

Lyons, Terry, Finely harmonic functions need not be quasi-analytic. Bull. London Math. Soc. 16 (1984), 413–415.

Lyons, Terry J., An application of fine potential theory to prove a Phragmén-Lindelöf theorem. Ann. Inst. Fourier (Grenoble) 34 (1984), 63–66.

Lyons, T. J.; McKean, H. P., Winding of the plane Brownian motion. Adv. in Math. 51 (1984), 212–225.

Lyons, Terry, A simple criterion for transience of a reversible Markov chain. Ann. Probab. 11 (1983), 393–402.

Gamelin, T. W.; Lyons, T. J., Jensen measures for R(K). J. London Math. Soc. (2) 27 (1983), 317–330.

Lyons, Terry J., Cones of lower semicontinuous functions and a characterisation of finely hyperharmonic functions. Math. Ann. 261 (1982), 293–297.

Lyons, T. J., Finely holomorphic functions. Aspects of contemporary complex analysis (Proc. NATO Adv. Study Inst., Univ. Durham, Durham, 1979), pp. 451–459, Academic Press, London- New York, 1980.

Lyons, T. J., A theorem in fine potential theory and applications to finely holomorphic functions. J. Funct. Anal. 37 (1980), 19–26.

Lyons, T. J., Finely holomorphic functions. J. Funct. Anal. 37 (1980), 1–18.

Lyons, Terry, A definition of BMOp for an abstract harmonic space and a John-Nirenberg theorem. Bull. London Math. Soc. 12 (1980), 127–129.

Stochastic Analysis and Applications 23–27 September 2013 Conference Timetable

Monday, September 23, 2013

08:30–09:20	Registration and coffee/tea
09:20–09:30	Opening Remarks, Dan Crisan (Imperial)
09:30–10:30	René Carmona (Princeton) *Mean Field Games in Random Environment*
10:30–11:00	Refreshment Break
11:00–12:00	Nizar Touzi (École Polytechnique, Paris) *Viscosity Solutions of Path-Dependent PDEs: The Elliptic Case*
12:00–13:00	Ronnie Sircar (Princeton) *Optimal Investment with Transaction Costs and Stochastic Volatility*
13:00–14:00	Lunch, Dining Hall
14:00–15:00	Hans Föllmer (HU Berlin) *Spatial Risk Measures: Local Specification, Aggregation, Phase Transition*
15:00–15.30	Refreshment Break
15:30–16:30	Ioannis Karatzas (Columbia) *Explosions and Arbitrage*
16:30–17:30	Nicolas Victoir (J.P. Morgan, New York) *A Particle Algorithm for Precise Smile Calibration*
18:00–19.15	Drinks Reception at Mathematical Institute
19.30	Dinner at St. Anne's

Tuesday, September 24, 2013

08:30–09:00	Coffee/tea
09:00–10:00	Dominique Bakry (Universite Paul Sabatier, Toulouse) *Orthogonal Polynomials and Diffusions in Dimension larger than 1*
10:00–10:30	Refreshment Break

10:30–11:30	Shigeki Aida (Tohoku) *Wong-Zakai Approximation of Solutions to Reflecting Stochastic Differential Equations on Domains in Euclidean Spaces*
11:30–12.30	Gérard Ben Arous (Courant Institute, NYU) *Scaling limits for random walks in heavy tailed random environments*
12:30–13:30	Lunch
13:30–14:30	Ana Bela Cruzeiro (IST, University of Lisbon) *Forward-Backward Stochastic Differential Systems and the Navier-Stokes Equations*
14:30–15:30	David Elworthy (University of Warwick) *Coupling Sets, Decompositions of Diffusions, and Bounded Harmonic Functions*
15:30–16:00	Refreshment Break
16:00–17:00	Peter Friz (Technische Universität Berlin) (*Rough*) *Pathwise Stochastic Analysis: Old and New*
17:00–18:00	Masatoshi Fukushima (Osaka University) *On Stochastic Komatu-Loewner Evolutions and a BMD Domain Constant*
19:00	Dinner, St. Anne's College

Wednesday, September 25, 2013

08:30–09:00	Coffee/Tea
09:00–10:00	Martin Barlow (University of British Columbia) *Quenched and averaged invariance principles for the random conductance model*
10:00–11:00	Refreshment Break
11:00–12:00	Jean Bertoin (I-Math, University of Zürich) *Local times for functions with finite variation*
12:00–13:00	Michael Röckner (University of Bielefeld) *Extinction results for the stochastic total variation flow*
13:00–13:40	Lunch
13:40–17:15	Session for Terry's Former Students
13:40–14:05	Mathew Penrose (Bath)
14:05–14:30	John Lunt (Rabobank)
14:30–14:55	Andrew Dickinson (JP Morgan)
14:55–15:20	Chris Potter (HSBC)
15:20–15:35	Refreshment Break
15:35–16:00	Katia Babbar (Lloyds)
16:00–16:25	Arend Janssen (Amplitude Capital)
16:25–16:50	Christian Litterer (Imperial)
16:50–17:15	Gechun Liang (Kings College)
17:25–18:45	Thank You Drinks Reception, OMI, Eagle House
19:00	Dinner, St. Anne's College

Thursday, September 26, 2013

08:30–09:00	Coffee/Tea
09:00–10:00	Yves Le Jan (Université Paris Sud) *Markov Loops and Coalescence Processes*
10:00–11:00	Refreshment Break
11:00–12:00	David Nualart (University of Kansas) *Convergence of Densities for Random Variables on a Finite Wiener Chaos*
12:00–13:00	Etienne Pardoux (Aix-Marseille University) *Continuous State Branching Process with Competition*
13:00–14:00	Lunch
14:00–15:00	Ofer Zeitouni (Weizmann Institute/NYU) *Extremes for Log-Correlated Gaussian Fields*
15:00–16:00	Refreshment Break
16:00–17:00	Sandy Davie (University of Edinburgh) *KMT theory applied to approximation of solutions of SDE*
17:00–18:00	Marta Sanz-Solé (Universitat de Barcelona) *Asymptotics of the Densities of SPDEs Driven by Spatially Correlated Noise*
19:00–21:00	Conference Gala Dinner, The Ruth Deech building, St. Anne's College

Friday, September 27, 2013

08:30–09:00	Coffee/Tea
09:00–10:00	Ben Hambly (University of Oxford) *Diffusions on Critical Random Clusters in the Diamond Lattice*
10:00–10.30	Refreshment Break
10:30–11:30	Martin Hairer (University of Warwick) *Dynamics near Criticality*
11:30–12:30	Alison Etheridge (University of Oxford) *The Spatial Lambda-Fleming-Viot Process*
12.30–12:45	Closing Remarks

Acknowledgments

The editors would like to thank all the authors for contributing to the volume in a timely manner and the anonymous referees for their constructive criticisms of the contributions. We would like to thank Springer and, in particular Catriona Byrne, for their strong support for this project. We also thank Marina Reizakis for her continuous assistance and patience throughout the entire editorial process and Frank Holzwarth for putting up with the continuous stream of questions and blunders from Dan and the other Springer Online Conference Service users.

In their capacity as organizers of the meeting, Dan Crisan, Ben Hambly and Thaleia Zariphopoulou would like to thank all the speakers and participants for their contributions to making it a successful workshop as well as St Anne's College and the Mathematical Institute for their hospitality. The organizers would also like to thank Alison Broadby and Wendy Ball for providing the administrative support for the meeting. Special thanks are due to Petrona Winton for setting the whole venture in motion. Finally, the meeting would not have been possible without the funding we received from the London Mathematical Society, the Oxford-Man Institute and the stochastic analysis fund at the Mathematical Institute and we thank them for their generosity.

Contents

Wong-Zakai Approximation of Solutions to Reflecting Stochastic Differential Equations on Domains in Euclidean Spaces II

Shigeki Aida

Abstract The strong convergence of Wong-Zakai approximations of the solution to the reflecting stochastic differential equations was studied in [2]. We continue the study and prove the strong convergence under weaker assumptions on the domain.

Keywords Wong-Zakai approximation · Reflecting SDE

1 Introduction

Wong-Zakai approximations of solutions of stochastic differential equations (=SDEs) were studied by many researchers, e.g. [13, 15, 27]. In the case of reflecting SDEs, Doss and Priouret [5] studied the Wong-Zakai approximations when the boundary is smooth. Actually, the unique existence of strong solutions of reflecting SDEs were proved for domains whose boundary may not be smooth by Tanaka [26], Lions-Sznitman [16] and Saisho [23]. In their studies, the standard conditions, (A), (B), (C) and admissibility condition, on the domain for reflecting SDEs were introduced and the unique existence of strong solutions were proved under the conditions either (A) and (B) hold or the domain is convex in [23, 26]. We explain the conditions (A), (B), (C) in the next section. There were studies on Wong-Zakai approximations in such cases, e.g., [20–22] for convex domains and [7] for domains satisfying admissibility condition as well as conditions (A), (B), (C). When the domain is convex, Ren and Xu [22] proved that Wong-Zakai approximations converge to the true solution in probability in the setting of stochastic variational inequality. In [2], the strong convergence of Wong-Zakai approximations was proved under the conditions (A), (B), (C). We note that Zhang [28] proved the strong convergence of Wong-Zakai approximations in the setting of [7] independent of [2]. The aim of this paper is to

This research was partially supported by Grant-in-Aid for Scientific Research (B) No. 24340023.

S. Aida (✉)
Mathematical Institute, Tohoku University, Sendai 980-8578, Japan
e-mail: aida@math.tohoku.ac.jp

D. Crisan et al. (eds.), *Stochastic Analysis and Applications 2014*,
Springer Proceedings in Mathematics & Statistics 100,
DOI 10.1007/978-3-319-11292-3_1

prove the strong convergence of Wong-Zakai approximation under the conditions either (A) and (B) hold or the domain is convex following the proof in [2]. Note that our proof in the case of convex domains is different from [22] and we give an estimate of the order of convergence.

The paper is organized as follows. In Sect. 2, we recall conditions of the boundary and state the main theorems. The first main theorem (Theorem 2.2) shows the strong convergence of Wong-Zakai approximations when the domain is convex. The estimate of the order of the convergence is the same as given in [2]. The second main theorem (Theorem 2.3) is concerned with the convergence of Wong-Zakai approximations in the case where the domain satisfies the conditions (A) and (B). We prove main theorems in Sects. 3 and 4.

2 Preliminaries and Main Theorems

Let D be a connected domain in $\mathbb{R}^d$. The following conditions can be found in [23]. In [2], we used the conditions (A), (B), (C) on D. In this paper, we will use (B′) too. The set $\mathcal{N}_x$ of inward unit normal vectors at $x \in \partial D$ is defined by

$$\mathcal{N}_x = \cup_{r>0}\mathcal{N}_{x,r},$$
$$\mathcal{N}_{x,r} = \left\{\boldsymbol{n} \in \mathbb{R}^d \mid |\boldsymbol{n}| = 1,\, B(x - r\boldsymbol{n}, r) \cap D = \emptyset\right\},$$

where $B(z, r) = \{y \in \mathbb{R}^d \mid |y - z| < r\}$, $z \in \mathbb{R}^d$, $r > 0$.

Definition 2.1 (A) (uniform exterior sphere condition). There exists a constant $r_0 > 0$ such that

$$\mathcal{N}_x = \mathcal{N}_{x,r_0} \neq \emptyset \quad \text{for any } x \in \partial D. \tag{2.1}$$

(B) There exist constants $\delta > 0$ and $\beta \geq 1$ satisfying: for any $x \in \partial D$ there exists a unit vector l_x such that

$$(l_x, \boldsymbol{n}) \geq \frac{1}{\beta} \qquad \text{for any } \boldsymbol{n} \in \cup_{y \in B(x,\delta)\cap\partial D}\mathcal{N}_y. \tag{2.2}$$

(B′) (uniform interior cone condition) There exist $\delta > 0$ and $0 \leq \alpha < 1$ such that for any $x \in \partial D$ there exists a unit vector l_x such that

$$C(y, l_x, \alpha) \cap B(x, \delta) \subset \bar{D} \quad \text{for any } y \in B(x, \delta) \cap \partial D,$$

where $C(y, l_x, \alpha) = \{z \in \mathbb{R}^d \mid (z - y, l_x) \geq \alpha|z - y|\}$.

(C) There exists a C_b^2 function f on $\mathbb{R}^d$ and a positive constant γ such that for any $x \in \partial D$, $y \in \bar{D}$, $\boldsymbol{n} \in \mathcal{N}_x$ it holds that

$$(y - x, \boldsymbol{n}) + \frac{1}{\gamma}\left((Df)(x), \boldsymbol{n}\right)|y - x|^2 \geq 0. \tag{2.3}$$

Note that if D is a convex domain, the condition (A) holds for any r_0 and the condition (C) holds for $f \equiv 0$. Also we can prove that the condition (B′) implies condition (B) with the same δ and $\beta = (1 - \alpha^2)^{-1/2}$ by noting that $\boldsymbol{n}_y \in \mathcal{N}_{y,r}$ is equivalent to

$$\left(z - y, \boldsymbol{n}_y\right) + \frac{1}{2r}|y - z|^2 \geq 0 \quad \text{for any } z \in \bar{D}.$$

Further, if D is a convex domain in $\mathbb{R}^2$ or a bounded convex domain in any dimensions, then the condition (B) holds. This is stated in [26]. Before considering reflecting SDE, let us explain the Skorohod problem on the multidimensional domain D for which $\mathcal{N}_x \neq \emptyset$ for all $x \in \partial D$. Let $w = w(t)$ $(0 \leq t \leq T)$ be a continuous path on $\mathbb{R}^d$ with $w(0) \in \bar{D}$. The pair of paths (ξ, ϕ) on $\mathbb{R}^d$ is a solution of a Skorohod problem associated with w if the following properties hold.

(i) $\xi = \xi(t)$ $(0 \leq t \leq T)$ is a continuous path in $\bar{D}$ with $\xi(0) = w(0)$.
(ii) It holds that $\xi(t) = w(t) + \phi(t)$ for all $0 \leq t \leq T$.
(iii) $\phi = \phi(t)$ $(0 \leq t \leq T)$ is a continuous bounded variation path on $\mathbb{R}^d$ such that $\phi(0) = 0$ and

$$\phi(t) = \int_0^t \mathbf{n}(s) d\|\phi\|_{[0,s]} \tag{2.4}$$

$$\|\phi\|_{[0,t]} = \int_0^t 1_{\partial D}(\xi(s)) d\|\phi\|_{[0,s]}. \tag{2.5}$$

where $\mathbf{n}(t) \in \mathcal{N}_{\xi(t)}$ if $\xi(t) \in \partial D$.

In the above, the notation $\|\phi\|_{[s,t]}$ stands for the total variation norm of $\phi(u)$ $(0 \leq s \leq u \leq t \leq T)$.

Let us consider reflecting SDEs. Let $\sigma \in C_b^2(\mathbb{R}^d \to \mathbb{R}^d \otimes \mathbb{R}^n)$ and $b \in C_b^1(\mathbb{R}^d \to \mathbb{R}^d)$. Let $\Omega = C([0, \infty) \to \mathbb{R}^n; \omega(0) = 0)$ and P be the Wiener measure on Ω. Let $B(t, \omega) = \omega(t)$ $(\omega \in \Omega)$ be the canonical realization of Brownian motion. We consider the reflecting SDE on $\bar{D}$:

$$X(t, x, \omega) = x + \int_0^t \sigma(X(s, x, \omega)) \circ dB(s, \omega) + \int_0^t b(X(s, x, \omega))ds + \Phi(t, \omega), \tag{2.6}$$

where $\circ dB(s)$ denotes the Stratonovich integral. We use the notation $(\mathrm{SDE})_{\sigma,b}$ to indicate this equation. Note that this usage is different from that in [2] but I think there are no confusion. The solution $(X(t), \Phi(t))$ to this equation is nothing but a solution of the Skorohod problem associated with

$$Y(t) = x + \int_0^t \sigma(X(s, x, \omega)) \circ dB(s, \omega) + \int_0^t b(X(s, x, \omega))ds.$$

As explained in the Introduction, if either the condition "(i) D is a convex domain" or the condition "(ii) D satisfies the conditions (A) and (B)" holds, then the strong solution $X(t)$ to (2.6) exists uniquely. These are due to Tanaka [26] for (i) and Saisho [23] for (ii). See also [6, 16]. Let X^N be the Wong-Zakai approximation of X. That is, X^N is the solution to the reflecting differential equation driven by continuous bounded variation paths:

$$\begin{aligned} X^N(t, x, \omega) = x &+ \int_0^t \sigma(X^N(s, x, \omega))dB^N(s, \omega) \\ &+ \int_0^t b(X^N(s, x, \omega))ds + \Phi^N(t, \omega), \end{aligned} \tag{2.7}$$

where

$$B^N(t) = B(t^N_{k-1}) + \frac{\Delta_N B_k}{\Delta_N}(t - t^N_{k-1}) \qquad t^N_{k-1} \le t \le t^N_k, \tag{2.8}$$

$$\Delta_N B_k = B(t^N_k) - B(t^N_{k-1}), \qquad \Delta_N = \frac{T}{N}, \qquad t^N_k = \frac{kT}{N}. \tag{2.9}$$

We may denote t^N_k and Δ_N by t_k and Δ respectively. The solution X^N uniquely exists under conditions (A) and (B) on D. See, e.g., [2, 23]. Under the convexity assumption of D too, the solution X^N uniquely exists by the results in [26]. In the convex case, we can check the existence in the following different way. More generally we consider a reflecting differential equation driven by a continuous bounded variation path w_t:

$$x_t = x_0 + \int_0^t \sigma(x_s)dw_s + \int_0^t b(x_s)ds + \Phi(t) \quad x_t \in \bar{D}. \tag{2.10}$$

The definition of the solution to this equation is similar to that of the equation previously discussed. Let $D_R = B(x_0, R) \cap D$. Then conditions (A) and (B) hold on D_R and the solution, say, x^R_t to reflecting differential equation on D_R exists. Moreover by Lemma 2.4 in [2], $\|x^R\|_{[0,T]} \le 2(\sqrt{2}+1)(\|\sigma\|_\infty \|w\|_{[0,T]} + \|b\|_\infty T)$, where $\|w\|_{[0,T]}$ denotes the total variation of $w(t)$ $(0 \le t \le T)$ as we already explained and $\|\sigma\|_\infty$ and $\|b\|_\infty$ denotes the sup-norm of the operator norm and the Euclidean norm of σ and b respectively. Thus, we have $\max_{0\le t\le T} |x^R(t) - x_0| \le$

$2(\sqrt{2}+1)(\|\sigma\|_\infty\|w\|_{[0,T]}+\|b\|_\infty T)$ and we can apply the result in the case where (A) and (B) hold. Now we are in a position to state our main theorems.

Theorem 2.2 *Assume D is convex. Then, for any $0<\theta<1$, we have*

$$\max_{0\le t\le T} E\left[|X^N(t)-X(t)|^2\right] \le C_\theta \cdot \Delta_N^{\theta/2} \tag{2.11}$$

$$E\left[\max_{0\le t\le T}|X^N(t)-X(t)|^2\right] \le C_{T,\theta}\Delta_N^{\theta/6}. \tag{2.12}$$

Theorem 2.3 *Assume the conditions* (A) *and* (B) *hold. Then for any $\varepsilon>0$, we have*

$$\lim_{N\to\infty} P\left(\max_{0\le t\le T}|X^N(t)-X(t)|\ge\varepsilon\right)=0. \tag{2.13}$$

Remark 2.4 We refer the readers to [2, 24, 25] for Euler approximations. Rough path analysis clarifies the meaning of Wong-Zakai approximations. We refer the readers for basic results of rough path analysis to [3, 11, 12, 17–19] and for Wong-Zakai approximations of rough differential equations driven by fractional Brownian motions to [4, 9, 10, 14]. Note that reflecting differential equations driven by rough paths are defined and the existence and estimates of the solutions are studied in the author's recent paper [1]. See also [8] for reflecting differential equations driven by fractional Brownian motions whose Hurst parameter are greater than 1/2.

3 Convex Domains

In this section, we prove Theorem 2.2. Below, we use the notation

$$\|w\|_{\infty,[s,t]}=\max_{s\le u\le v\le t}|w(u)-w(v)|.$$

The notation $\|w\|_{[s,t]}$ was already defined in Sect. 2. We can prove the following in the same way as in the proof of Lemma 2.3 in [2].

Lemma 3.1 *Assume conditions* (A) *and* (B) *hold. Let w be a q-variation continuous path such that*

$$|w(t)-w(s)|\le\omega(s,t)^{1/q} \qquad 0\le s\le t\le T \tag{3.1}$$

where $q\ge1$ and ω is a control function. That is, $\omega(s,t)$ is a nonnegative continuous function of (s,t) with $0\le s\le t\le T$ satisfying $\omega(s,u)+\omega(u,t)\le\omega(s,t)$ for all $0\le s\le u\le t\le T$. Then the local time ϕ of the solution to the Skorohod problem associated with w has the following estimate.

$$\|\phi\|_{[s,t]} \le \beta \left(\left\{ \delta^{-1} G(\|w\|_{\infty,[s,t]}) + 1 \right\}^q \omega(s,t) + 1 \right) \\ \times \left(G(\|w\|_{\infty,[s,t]}) + 2 \right) \|w\|_{\infty,[s,t]}, \tag{3.2}$$

where

$$G(a) = 4\left\{1 + \beta \exp\left\{\beta\left(2\delta + a\right)/(2r_0)\right\}\right\} \exp\left\{\beta\left(2\delta + a\right)/(2r_0)\right\}. \tag{3.3}$$

The above estimate is one of key for the proof in [2]. Since the unbounded convex domains in $\mathbb{R}^d$ ($d \ge 3$) may not satisfy the condition (B), we cannot use this estimate. However, it is possible to estimate the total variation $\|\phi\|_{[s,t]}$ by $\|w\|_{\infty,[s,t]}$ together with the sup-norm of ξ since we can give an estimate for the numbers β and δ in the condition (B) for bounded convex domains.

Lemma 3.2 *Let D be a convex domain in $\mathbb{R}^d$. Let $x_0 \in D$ and assume that there exists $R_0 > 0$ such that $\overline{B(R_0, x_0)} \subset D$. Let $R \ge R_0$ and define $D_R = D \cap B(R, x_0)$. The Condition* (B) *holds for the bounded convex domain D_R with $\delta = R_0/2$ and $\beta = \left(1 + \left(\frac{2R}{R_0}\right)^2\right)^{1/2}$.*

Proof We prove the condition (B′). Let $x \in \partial D_R$. Let l_x be the unit vector in the direction from x to x_0. Let $S(x_0)$ be a $d-1$ dimensional ball which is the slice of the ball $\overline{B(R_0, x_0)}$ by a hyperplane $H(x_0)$ that passes through x_0 and is orthogonal to l_x. Let $\alpha = \frac{R}{\sqrt{R^2 + (R_0/2)^2}}$. Then for any point $y \in B(\delta, x)$, it holds that $C(y, l_x, \alpha) \cap H(x_0) \subset S(x_0)$. Hence for any $y \in B(\delta, x) \cap \partial D_R$, $C(y, l_x, \alpha) \cap B(x, \delta) \subset \overline{D_R}$ which implies condition (B′).

Lemma 3.3 *Let D be a convex domain. Let $x_0 \in D$ and assume that there exists $R_0 > 0$ such that $\overline{B(R_0, x_0)} \subset D$. Let $w(t)$ $(0 \le t \le T)$ be a continuous q-variation path with the control function ω on $\mathbb{R}^d$ with $w(0) \in \bar{D}$ and $q \ge 1$. Assume that there exists a solution (ξ, ϕ) to the Skorohod problem associated with w. Then it holds that*

$$\|\phi\|_{[s,t]} \le 10 \left[\left\{ 16 R_0^{-1} \left(1 + 4R_0^{-2} \|\xi - x_0\|_{\infty,[0,T]}^2\right)^{1/2} + 1 \right\}^q \omega(s,t) + 1 \right] \\ \times \left(1 + 4R_0^{-2} \|\xi - x_0\|_{\infty,[0,T]}^2\right) \|w\|_{\infty,[s,t]}. \tag{3.4}$$

Proof Note that ξ is the solution of the Skorohod problem associated with w on $\overline{D \cap B(x_0, R)}$, where $R = \|\xi - x_0\|_{\infty,[0,T]}$. This domain satisfies (B) with the constants δ and β specified in the above lemma. In the lemma, letting $r_0 \to \infty$, G reads

$$G(a) = 4\left\{1 + \sqrt{1 + (2R_0^{-1}R)^2}\right\}. \tag{3.5}$$

By applying Lemma 3.1, we complete the proof.

To prove Theorem 2.2, we need moment estimates for increments of X^N and Φ^N.

Lemma 3.4 *Assume D is a convex domain. For the Wong-Zakai approximation X^N, we define*

$$Y^N(t,x,\omega)=x+\int_0^t \sigma(X^N(s,x,\omega))dB^N(s,\omega)+\int_0^t b(X^N(s,x,\omega))ds. \quad (3.6)$$

(1) *For all $p\ge 1$, we have*

$$E[\|Y^N\|^{2p}_{\infty,[s,t]}]\le C_p|t-s|^p. \quad (3.7)$$

(2) *Let $t_{k-1}\le s<t\le t_k$. Then we have for all $p\ge 1$,*

$$E[|X^N(t)-X^N(s)|^{2p}\mid \mathcal{F}_{t_{k-1}}]\le C_p|t-s|^p, \quad (3.8)$$

$$\|\Phi^N\|_{[s,t]}\le C\left(|\Delta B_k|\frac{t-s}{\Delta}+(t-s)\right), \quad (3.9)$$

where C_p and C are positive constants.

Proof These assertions can be proved by the same way as the proof of Lemma 4.3 and 4.4 in [2]. We assumed the condition (B) in those lemmas but we can argue in the same way since Skorohod equation associated with the continuous bounded variation path is uniquely solved under the convexity of D.

Lemma 3.5 *Assume D is convex. Let $p\ge 2$ be an integer. For $0\le s\le t\le T$, we have*

$$E\left[|X(t)-X(s)|^p\right]\le C_p|t-s|^{p/2}, \quad (3.10)$$

$$E\left[|X^N(t)-X^N(s)|^p\right]\le C_p|t-s|^{p/2}, \quad (3.11)$$

$$E\left[\|\Phi^N\|^p_{[s,t]}\right]\le C_p|t-s|^{p/2}, \quad (3.12)$$

where C_p is a positive number independent of N.

Proof Let $\tau_R=\inf\{t>0\mid X(t,x,w)\notin B(x,R)\}$ and $X^{\tau_R}(t)=X(t\wedge\tau_R)$. For (3.10), it suffices to prove $E[|X^{\tau_R}(t)-X^{\tau_R}(s)|^p|]\le C_p|t-s|^{p/2}$ for all even positive integers p and $0\le s\le t\le T$, where C_p is independent of R. We prove this by an induction on p. Let $\tilde{b}=b+\frac{1}{2}\mathrm{tr}(D\sigma)(\sigma)$. By the Ito formula,

$$|X^{\tau_R}(t)-X^{\tau_R}(s)|^2=2\int_{s\wedge\tau_R}^{t\wedge\tau_R}(X^{\tau_R}(u)-X^{\tau_R}(s),\sigma(X^{\tau_R}(u))dB(u))$$

$$+2\int_{s\wedge\tau_R}^{t\wedge\tau_R}(X^{\tau_R}(u)-X^{\tau_R}(s),\tilde{b}(X^{\tau_R}(u)))du$$
$$+\int_{s\wedge\tau_R}^{t\wedge\tau_R}\mathrm{tr}\left((\sigma\,{}^t\sigma)(X^{\tau_R}(u))\right)du$$
$$+2\int_{s\wedge\tau_R}^{t\wedge\tau_R}(X^{\tau_R}(u)-X^{\tau_R}(s),d\Phi(u)). \qquad (3.13)$$

Noting the non-positivity of the term containing Φ which follows from the convexity of D and taking the expectation, we have

$$E\left[|X^{\tau_R}(t)-X^{\tau_R}(s)|^2\right]\le C\int_s^t E\left[|X^{\tau_R}(u)-X^{\tau_R}(s)|^2\right]du+C(t-s) \qquad (3.14)$$

which implies $E[|X^{\tau_R}(t)-X^{\tau_R}(s)|^2]\le C(t-s)$. Let $p\ge 4$ and suppose the inequality holds for $p-2$.

$$|X^{\tau_R}(t)-X^{\tau_R}(s)|^p$$
$$=p\int_{s\wedge\tau_R}^{t\wedge\tau_R}|X^{\tau_R}(u)-X^{\tau_R}(s)|^{p-2}(X^{\tau_R}(u)-X^{\tau_R}(s),\sigma(X^{\tau_R}(u))dB(u))$$
$$+p\int_{s\wedge\tau_R}^{t\wedge\tau_R}|X^{\tau_R}(u)-X^{\tau_R}(s)|^{p-2}(X^{\tau_R}(u)-X^{\tau_R}(s),\tilde{b}(X^{\tau_R}(u)))du$$
$$+\frac{p}{2}\int_{s\wedge\tau_R}^{t\wedge\tau_R}|X^{\tau_R}(u)-X^{\tau_R}(s)|^{p-2}\mathrm{tr}\left((\sigma\,{}^t\sigma)(X^{\tau_R}(u))\right)du$$
$$+\frac{1}{2}p\,(p-2)\int_{s\wedge\tau_R}^{t\wedge\tau_R}|X^{\tau_R}(u)-X^{\tau_R}(s)|^{p-4}|{}^t\sigma(X^{\tau_R}(u)\left(X^{\tau_R}(u)-X^{\tau_R}(s)\right)|^2du$$
$$+p\int_{s\wedge\tau_R}^{t\wedge\tau_R}|X^{\tau_R}(u)-X^{\tau_R}(s)|^{p-2}(X^{\tau_R}(u)-X^{\tau_R}(s),d\Phi(u)).$$

Hence we have

$$\begin{aligned} E\left[|X^{\tau_R}(t)-X^{\tau_R}(s)|^p\right] &\le C_p\bigg(\int_s^t E\left[|X^{\tau_R}(u)-X^{\tau_R}(s)|^{p-2}\right] \\ &\qquad + E\left[|X^{\tau_R}(u)-X^{\tau_R}(s)|^{p-1}\right]\bigg)du \\ &\le C_p\bigg(\int_s^t E\left[|X^{\tau_R}(u)-X^{\tau_R}(s)|^{p-2}\right] \\ &\qquad + E\left[|X^{\tau_R}(u)-X^{\tau_R}(s)|^{p}\right]\bigg)du \end{aligned}$$

which implies

$$\begin{aligned} E\left[|X^{\tau_R}(t)-X^{\tau_R}(s)|^p\right] &\le C_p e^{C_p(t-s)}\int_s^t E\left[|X^{\tau_R}(u)-X^{\tau_R}(s)|^{p-2}\right]du \\ &\le C_p(t-s)^{p/2}. \end{aligned} \tag{3.15}$$

This proves (3.10). Next we prove (3.11). Again, is is sufficient to prove the case where p is an even number. We prove this by an induction on p similarly to (3.10). By Lemma 2.4 in [2], we have $E[\|X^N\|^p_{[0,T]}] < \infty$ for any $p \ge 1$. We consider the case where $p = 2$. Let $s = t_l < t_m = t$. By the chain rule,

$$\begin{aligned} |X^N(t)-X^N(s)|^2 &= 2\int_s^t (X^N(u)-X^N(s), \sigma(X^N(u))dB^N(u)) \\ &\quad + 2\int_s^t (X^N(u)-X^N(s), b(X^N(u)))du \\ &\quad + 2\int_s^t (X^N(u)-X^N(s), d\Phi^N(u)) \\ &\le 2\int_s^t (X^N(u)-X^N(s), \sigma(X^N(u))dB^N(u)) \\ &\quad + 2\int_s^t (X^N(u)-X^N(s), b(X^N(u)))du =: I_1 + I_2, \end{aligned} \tag{3.16}$$

where we have used the non-positivity of the third term which follows from the convexity of D. We estimate I_1, I_2. We have

$$I_1 = \sum_{k=l+1}^{m} 2 \int_{t_{k-1}}^{t_k} \left(X^N(u) - X^N(s), \sigma(X^N(u)) \frac{\Delta B_k}{\Delta} \right) du. \tag{3.17}$$

$$\begin{aligned}
I_{1,k} &:= \int_{t_{k-1}}^{t_k} (X^N(u) - X^N(s), \sigma(X^N(u)) \frac{\Delta B_k}{\Delta}) du \\
&= \left(X^N(t_{k-1}) - X^N(s), \sigma(X^N(t_{k-1})) \Delta B_k \right) \\
&\quad + \int_{t_{k-1}}^{t_k} \left(X^N(u) - X^N(t_{k-1}), \sigma(X^N(t_{k-1})) \frac{\Delta B_k}{\Delta} \right) du \qquad (3.18) \\
&\quad + \int_{t_{k-1}}^{t_k} \left(X^N(t_{k-1}) - X^N(s), \left(\sigma(X^N(u)) - \sigma(X^N(t_{k-1})) \right) \frac{\Delta B_k}{\Delta} \right) \\
&\quad + \int_{t_{k-1}}^{t_k} \left(X^N(u) - X^N(t_{k-1}), \left(\sigma(X^N(u)) - \sigma(X^N(t_{k-1})) \right) \frac{\Delta B_k}{\Delta} \right) du
\end{aligned}$$

By Lemma 3.4 (2),

$$\begin{aligned}
E\left[I_{1,k}\right] &\le C\left(1 + E[|X^N(t_{k-1}) - X^N(s)|]\right)\Delta \qquad (3.19) \\
&\le C\left(\int_{t_{k-1}}^{t_k} \left(E[|X^N(u) - X^N(s)|^2] + 1 \right) du \right).
\end{aligned}$$

Thus, we obtain

$$E[|X^N(t) - X^N(s)|^2] \le C\left((t-s) + \int_s^t E[|X^N(u) - X^N(s)|^2] du \right). \tag{3.20}$$

Again by noting Lemma 3.4 (2), we see that (3.20) holds for any $0 \le s \le t \le T$. Applying the Gronwall inequality, we get the inequality (3.11) with $p = 2$. Let $p \ge 4$. Let $s = t_l < t_m = t$. By the chain rule,

$$\begin{aligned}
|X^N(t) - X^N(s)|^p = p \int_s^t & |X^N(u) - X^N(s)|^{p-2} \\
& \times (X^N(u) - X^N(s), \sigma(X^N(u)) dB^N(u))
\end{aligned}$$

$$
\begin{aligned}
&+ p\int_s^t |X^N(u)-X^N(s)|^{p-2}(X^N(u)-X^N(s), b(X^N(u)))du \\
&+ p\int_s^t |X^N(u)-X^N(s)|^{p-2}(X^N(u)-X^N(s), d\Phi^N(u)) \\
\le\ & p\int_s^t |X^N(u)-X^N(s)|^{p-2} \qquad (3.21)\\
&\times (X^N(u)-X^N(s), \sigma(X^N(u))dB^N(u)) \\
&+ p\int_s^t |X^N(u)-X^N(s)|^{p-2}(X^N(u)-X^N(s), b(X^N(u)))du \\
=:\ & J_1 + J_2,
\end{aligned}
$$

where we have used the non-positivity of the third term which follows from the convexity of D. By noting $|X^N(u)-X^N(s)|^{p-1} \le \frac{1}{2}\left(|X^N(u)-X^N(s)|^p + |X^N(u)-X^N(s)|^{p-2}\right)$ and by the assumption of induction, we have

$$
E[J_2] \le C(t-s)^{p/2} + \int_s^t E[|X^N(u)-X^N(s)|^p]du. \tag{3.22}
$$

For J_1, we have

$$
J_1 = \sum_{k=l+1}^m p\int_{t_{k-1}}^{t_k} |X^N(u)-X^N(s)|^{p-2}\left(X^N(u)-X^N(s), \sigma(X^N(u))\frac{\Delta B_k}{\Delta}\right)du. \tag{3.23}
$$

$$
\begin{aligned}
&\int_{t_{k-1}}^{t_k} |X^N(u)-X^N(s)|^{p-2}(X^N(u)-X^N(s), \sigma(X^N(u))\frac{\Delta B_k}{\Delta})du \\
&= \int_{t_{k-1}}^{t_k} |X^N(u)-X^N(s)|^{p-2}\left(X^N(t_{k-1})-X^N(s), \sigma(X^N(t_{k-1}))\frac{\Delta B_k}{\Delta}\right)du \\
&\quad + \int_{t_{k-1}}^{t_k} |X^N(u)-X^N(s)|^{p-2}\left(X^N(u)-X^N(t_{k-1}), \sigma(X^N(t_{k-1}))\frac{\Delta B_k}{\Delta}\right)du
\end{aligned}
$$

$$
\begin{aligned}
&+\int_{t_{k-1}}^{t_k} |X^N(u)-X^N(s)|^{p-2} \\
&\times\left(X^N(t_{k-1})-X^N(s), \left(\sigma(X^N(u))-\sigma(X^N(t_{k-1}))\right)\frac{\Delta B_k}{\Delta}\right) \\
&+\int_{t_{k-1}}^{t_k} |X^N(u)-X^N(s)|^{p-2} \\
&\times\left(X^N(u)-X^N(t_{k-1}), \left(\sigma(X^N(u))-\sigma(X^N(t_{k-1}))\right)\frac{\Delta B_k}{\Delta}\right)du \\
=\ & J^k_{1,1}+J^k_{1,2}+J^k_{1,3}+J^k_{1,4}.
\end{aligned} \tag{3.24}
$$

We have

$$
J^k_{1,1}=J^k_{1,1,1}+J^k_{1,1,2}+J^k_{1,1,3}+J^k_{1,1,4} \tag{3.25}
$$

where

$$
\begin{aligned}
J^k_{1,1,1}=\int_{t_{k-1}}^{t_k}\Bigg\{\int_{t_{k-1}}^{u}&(p-2)|X^N(r)-X^N(s)|^{p-4} \\
&\times\left(X^N(r)-X^N(s), \sigma(X^N(r))\frac{\Delta B_k}{\Delta}\right)dr\Bigg\} \\
&\times\left(X^N(t_{k-1})-X^N(s), \sigma(X^N(t_{k-1}))\frac{\Delta B_k}{\Delta}\right)du,
\end{aligned} \tag{3.26}
$$

$$
\begin{aligned}
J^k_{1,1,2}=\int_{t_{k-1}}^{t_k}&\left\{\int_{t_{k-1}}^{u}(p-2)|X^N(r)-X^N(s)|^{p-4}\left(X^N(r)-X^N(s), b(X^N(r))\right)dr\right\} \\
&\times\left(X^N(t_{k-1})-X^N(s), \sigma(X^N(t_{k-1}))\frac{\Delta B_k}{\Delta}\right)du,
\end{aligned} \tag{3.27}
$$

$$
\begin{aligned}
J^k_{1,1,3}=\int_{t_{k-1}}^{t_k}&\left\{\int_{t_{k-1}}^{u}(p-2)|X^N(r)-X^N(s)|^{p-4}\left(X^N(r)-X^N(s), d\Phi^N(r)\right)\right\} \\
&\times\left(X^N(t_{k-1})-X^N(s), \sigma(X^N(t_{k-1}))\frac{\Delta B_k}{\Delta}\right)du
\end{aligned} \tag{3.28}
$$

$$
J^k_{1,1,4}=\int_{t_{k-1}}^{t_k}\left|X^N(t_{k-1})-X^N(s)\right|^{p-2}\left(X^N(t_{k-1})-X^N(s), \sigma(X^N(t_{k-1}))\frac{\Delta B_k}{\Delta}\right)du
$$

By the estimate for $p = 2$ and Lemma 3.4 (2), we have

$$\begin{aligned} E[J^k_{1,1,1}] \le C_p E\left[|X^N(t_{k-1}) - X^N(s)|^{p-2}\right]\Delta \\ + \int_{t_{k-1}}^{t_k}\int_{t_{k-1}}^{u} E\Big[|X^N(r) - X^N(t_{k-1})|^{p-3}||X^N(t_{k-1}) \\ -X^N(s)|\left(\frac{|\Delta B_k|}{\Delta}\right)^2\Big] drdu \\ \le C_p E\left[|X^N(t_{k-1}) - X^N(s)|^{p-2}\right]\Delta + C(t_{k-1} - s)^{1/2}\Delta^{(p-1)/2}. \end{aligned} \tag{3.29}$$

Noting that for any $a > 0$, $\sum_{k=l+1}^m (t_{k-1}-s)^a\Delta \le \int_s^t (u-s)^a du \le (t-s)^{a+1}/(a+1)$ and using the assumption of induction,

$$\begin{aligned} E[\sum_{k=l+1}^{m} J^k_{1,1,1}] &\le C \sum_{k=l+1}^{m}\left\{(t_{k-1} - s)^{(p-2)/2}\Delta + (t_{k-1} - s)^{1/2}\Delta^{(p-1)/2}\right\} \\ &\le C(t - s)^{p/2}. \end{aligned} \tag{3.30}$$

Similarly,

$$E[J^k_{1,1,2}] \le C_p E[|X^N(t_{k-1}) - X^N(s)|^{p-2}]\Delta^{3/2} + C(t_{k-1} - s)^{1/2}\Delta^{p/2}. \tag{3.31}$$

$$\begin{aligned} E[J^k_{1,1,3}] \le C_p E\Big[|X^N(t_{k-1}) - X^N(s)|^{p-2} E\left[\|\Phi^N\|_{[t_{k-1},t_k]}|\Delta B_k||\mathcal{F}_{t_{k-1}}\right]\Big] \\ + C_p E\Big[|X^N(t_{k-1}) - X^N(s)| E\Big[\max_{t_{k-1}\le r\le t_k} |X^N(r) \\ -X^N(t_{k-1})|^{p-3}\|\Phi^N\|_{t_{k-1},t_k}|\Delta B_k|\,|\,\mathcal{F}_{t_{k-1}}\Big]\Big] \\ \le C_p E[|X^N(t_{k-1}) - X^N(s)|]^{p-2}\Delta + (t_{k-1} - s)^{1/2}\Delta^{(p-1)/2}. \end{aligned} \tag{3.32}$$

A similar estimate holds for $J^k_{1,1,4}$. Thus, we have $\sum_{i=2}^4 E\left[\sum_{k=l+1}^m J^k_{1,1,i}\right] \le C(t - s)^{p/2}$. We consider the terms $J^k_{1,i}$ $(2 \le i \le 4)$.

$$\begin{aligned} E[J^k_{1,2}] \le C\Delta^{-1}\int_{t_{k-1}}^{t_k} E\Big[|X^N(t_{k-1}) - X^N(s)|^{p-2} \\ \times E\left[|X^N(u) - X^N(t_{k-1})||\Delta B_k||\mathcal{F}_{t_{k-1}}\right]\Big]du + C\Delta^{p/2} \\ \le CE[|X^N(t_{k-1}) - X^N(s)|^{p-2}]\Delta + C\Delta^{p/2}. \end{aligned} \tag{3.33}$$

$$\begin{aligned} E[J^k_{1,3}] \le C\Delta^{-1} \int_{t_{k-1}}^{t_k} & E\Big[|X^N(t_{k-1}) - X^N(s)|E \\ & \times \Big[|X^N(u) - X^N(t_{k-1})|^{p-1}|\Delta B_k||\mathcal{F}_{t_{k-1}}\Big]\Big]du \\ & + C\Delta^{-1} \int_{t_{k-1}}^{t_k} E\Big[|X^N(t_{k-1}) - X^N(s)|^{p-1}E \\ & \times \Big[|X^N(u) - X^N(t_{k-1})||\Delta B_k||\mathcal{F}_{t_{k-1}}\Big]\Big]du \\ & \le CE[|X^N(t_{k-1}) - X^N(s)|]\Delta^{p/2} + CE[|X^N(t_{k-1}) - X^N(s)|^{p-1}]\Delta. \end{aligned} \tag{3.34}$$

$$\begin{aligned} E[J^k_{1,4}] \le C\Delta^{-1} \int_{t_{k-1}}^{t_k} & E\Big[|X^N(t_{k-1}) - X^N(s)|^{p-2}E \\ & \times \Big[|X^N(u) - X^N(t_{k-1})|^2|\Delta B_k||\mathcal{F}_{t_{k-1}}\Big]\Big]du \\ & + C\Delta^{-1} \int_{t_{k-1}}^{t_k} E\Big[|X^N(u) - X^N(t_{k-1})|^p|\Delta B_k|\Big]du \\ & \le CE[|X^N(t_{k-1}) - X^N(s)|^{p-2}]\Delta^{3/2} + C\Delta^{(p+1)/2}. \end{aligned} \tag{3.35}$$

Hence

$$E\left[|X^N(t) - X^N(s)|^p\right] \le C(t-s)^{p/2} + \int_s^t E\left[|X^N(u) - X^N(s)|^p\right]du. \tag{3.36}$$

By using (3.8), we see that (3.36) holds for any $0 \le s \le t \le T$. By the Gronwall inequality, we get the desired inequality for p and we complete the proof of (3.11). The estimate (3.7) and the Garsia-Rodemich-Rumsey estimate imply the L^r-boundedness of the Hölder norm with exponent $1/2-\varepsilon$ of Y^N for any $r \ge 1$ and $0 < \varepsilon < 1/2$. Hence, (3.12) follows from Lemma 3.3 and (3.11).

Thanks to the above estimates, we can prove the first main theorem as in [2].

Proof (Proof of Theorem 2.2) Let $X^N_E(t)$ be the Euler approximation of X. That is, $X^N_E(0) = x$ and X^N_E is the solution to the Skorohod equation:

$$X_E^N(t) = X_E^N(t_{k-1}^N) + \sigma(X_E^N(t_{k-1}^N))(B(t) - B(t_{k-1}^N)) + \tilde{b}(X_E^N(t_{k-1}^N))(t - t_{k-1}^N) \\ + \Phi_E^N(t) - \Phi_E^N(t_{k-1}^N) \qquad t_{k-1}^N \le t \le t_k^N, \tag{3.37}$$

where $\Phi_E^N(t) - \Phi^N(t_{k-1}^N)$ is the local time term and $\tilde{b} = b + \frac{1}{2}\mathrm{tr}(D\sigma)(\sigma)$. By a similar argument to (3.10) and (3.12), we obtain

$$E[\|X_E^N\|_{\infty,[s,t]}^{2p}] \le C_p|t-s|^p, \tag{3.38}$$

$$E\left[\|\Phi_E^N\|_{[s,t]}^{2p}\right] \le C_p|t-s|^p. \tag{3.39}$$

Hence by the same proof as in [2], we obtain there exists $C_p > 0$ such that

$$E\left[\max_{0\le t\le T} |X_E^N(t) - X(t)|^{2p}\right] \le C_p \Delta_N^p \tag{3.40}$$

By these estimates and Lemma 3.5, we can prove the desired estimates as in the same way in [2]. The proof is simpler than that in [2] because $f \equiv 0$ when D is convex.

4 General Domains Satisfying Conditions (A) and (B)

In this section, we prove Theorem 2.3. The following observation which can be found in Lemma 5.3 in [23] is crucial for our purpose.

Lemma 4.1 *Assume* (A) *and* (B) *are satisfied on D. Let $\gamma = 2r_0\beta^{-1}$. Then for each $z_0 \in \partial D$ we can find a function $f \in C_b^2(\mathbb{R}^d)$ satisfying* (2.3) *for any $x \in B(z_0, \delta) \cap \partial D$, $y \in \bar{D}$ and $\boldsymbol{n} \in \mathcal{N}_x$. Moreover the sup-norms $\|D^k f\|_\infty$ $(k = 0, 1, 2)$ are bounded by some constant independent of z_0.*

It is stated in Lemma 5.3 in [23] that the conclusion in the above proposition holds for $y \in B(z_0, \delta) \cap \bar{D}$. However, it is obvious to see the same conclusion holds for any $y \in \bar{D}$. Thanks to this proposition, we can localize the problem. Let us choose a positive number $\delta' < \delta/2$. For any $z \in \bar{D}$, if $B(z, \delta') \cap \partial D \neq \emptyset$, then there exists $z_0 \in \partial D$ such that $\overline{B(z, \delta')} \subset B(z_0, \delta)$. Next, let χ be a C^∞ function on $\mathbb{R}^d$ such that $\chi(x) = 1$ for x with $|x| \le \delta'/2$, $\chi(x) = 0$ for x with $|x| \ge 2\delta'/3$. Let $z \in \bar{D}$ and define

$$\sigma^z(x) = \sigma(x)\chi(x - z), \qquad b^z(x) = b(x)\chi(x - z) \qquad x \in \mathbb{R}^d. \tag{4.1}$$

We denote the solution and the Wong-Zakai approximation to $(\mathrm{SDE})_{\sigma^z, b^z}$ with the starting point x by $X^z(t, x, \omega)$ and $X^{N,z}(t, x, \omega)$ respectively. By the uniqueness of strong solutions, we have

(i) $X^z(t, x, \omega) = X^{N,z}(t, x, \omega) = x$ for all $x \in B(z, 2\delta'/3)^c$

(ii) If $x \in B(z, 2\delta'/3)$, then both $X^z(t, x, \omega)$ and $X^{N,z}(t, x, \omega)$ belong to $B(z, 2\delta'/3)$ for all t and N.

We need a continuous dependence of solutions of reflecting SDE with respect to the starting point as in the following. Below, we state it for the particular case $\mathrm{SDE}_{\sigma^z, b^z}$ but it is easy to extend the result to more general situations.

Lemma 4.2 *Assume (A) and* (B) *hold on* D.

(1) *For any* $p \geq 1$ *and* $x, y \in \bar{D}$, *we have*

$$E\left[\max_{0\leq t\leq T} |X^z(t,x) - X^z(t,y)|^p\right] \leq C_p |x-y|^p. \tag{4.2}$$

The constant C_p *is independent of* z.

(2) *Let* $0 < \theta < 1$. *There exists a positive constant* $C_{T,\theta}$ *such that for any* $x, z \in \bar{D}$, *we have*

$$E\left[\max_{0\leq t\leq T} |X^{N,z}(t,x) - X^z(t,x)|^2\right] \leq C_{T,\theta}\Delta_N^{\theta/6}. \tag{4.3}$$

(3) *Let* $x \in B(z, \delta'/2)$. *Let* $\tau(\omega)$ *and* $\sigma(\omega)$ *be the exit time of* $X(t, x, \omega)$ *and* $X^z(t, x, \omega)$ *respectively from* $B(z, \delta'/2)$. *Then* $\tau(\omega) = \sigma(\omega)$ P*-a.s.* ω *and* $X(t, x, \omega) = X^z(t, x, \omega)$ $(0 \leq t \leq \tau(\omega))$.

Proof (1) If x or y belongs to $B(z, 5\delta'/6)^c$, the assertion is true because of (i) and (ii) above. Therefore we may assume $x, y \in B(z, 5\delta'/6)$. It is sufficient to consider the case where $B(z, \delta') \cap \partial D \neq \phi$. Then we can pick a point $z_0 \in B(z, \delta') \cap \partial D$ such that $B(z, \delta') \subset B(z_0, \delta)$. Let f be a function in Lemma 4.1 associated with z_0. Let

$$\begin{aligned} Z^z(t) = X^z(t,x) - X^z(t,y), \quad \rho^z(t) = e^{-\frac{2}{\gamma}(f(X^z(t,x)) + f(X^z(t,y)))}, \\ k^z(t) = \rho^z(t)|Z^z(t)|^2. \end{aligned} \tag{4.4}$$

In the calculation below, we omit the superscript z in the notation X^z, and so on. Let $\tilde{b} = b + \frac{1}{2}\mathrm{tr}(D\sigma)(\sigma)$. By the Ito formula,

$$\begin{aligned} dk(t) = \rho(t)\Bigg\{ & 2\Big(Z(t), (\sigma(X(t,x)) - \sigma(X(t,y)))\,dB(t)\Big) \\ & + 2\left(Z(t), \tilde{b}(X(t,x)) - \tilde{b}(X(t,y))\right) dt \\ & + \|\sigma(X(t,x)) - \sigma(X(t,y))\|_{H.S.}^2 dt \Bigg\} \\ & + 2\rho(t)\,(Z(t), d\Phi(t,x) - d\Phi(t,y)) \\ & - \frac{2\rho(t)}{\gamma}|Z(t)|^2 \Big\{((Df)(X(t,x)), d\Phi(t,x)) + ((Df)(X(t,y)), d\Phi(t,y))\Big\} \end{aligned}$$

$$
\begin{aligned}
&- \frac{2\rho(t)}{\gamma} |Z(t)|^2 \Big\{ ((Df)(X(t)), \sigma(X(t,x)) dB(t)) \\
&\qquad\qquad\qquad + ((Df)(X(t,y)), \sigma(X(t,y)) dB(t)) \Big\} \\
&+ R(t) dt,
\end{aligned} \tag{4.5}
$$

where

$$
\begin{aligned}
R(t) = {} & \frac{4\rho(t)}{\gamma} \Big((Df)(X(t,x)), \sigma(X(t,x))^t \left(\sigma(X(t,x)) - \sigma(X(t,y))\right) (Z(t)) \Big) dt \\
& + \frac{4\rho(t)}{\gamma} \big((Df)(X(t,y)), \sigma(X(t,y))^t \left(\sigma(X(t,x)) - \sigma(X(t,y))\right) (Z(t)) \big) dt \\
& - \frac{2\rho(t)}{\gamma} |Z(t)|^2 \Big(\Big((Df)(X(t,x)), \tilde{b}(X(t,x)) \Big) dt \\
& \qquad\qquad + \Big((Df)(X(t,y)), \tilde{b}(X(t,y)) \Big) dt \Big) \\
& - \frac{\rho(t)}{\gamma} |Z(t)|^2 \Big\{ \mathrm{tr}(D^2 f)(X(t,x)) \left(\sigma(X(t,x))\cdot, \sigma(X(t,x))\cdot\right) \\
& \qquad\qquad + \mathrm{tr}(D^2 f)(X(t,y)) \left(\sigma(X(t,y))\cdot, \sigma(X(t,y))\cdot)\right) \Big\} dt \\
& + \frac{2\rho(t)}{\gamma^2} \| (Df)(X(t,x))(\sigma(X(t,x))) \\
& \qquad\qquad + (Df)(X(t,y))(\sigma(X(t,y))) \|^2 |Z(t)|^2 dt.
\end{aligned} \tag{4.6}
$$

Let us take a look at the second and third terms of (4.6). This term is not equal to 0 when $X(t,x)$ or $X(t,y)$ hits ∂D. By the property of f, these terms are negative. Taking this into account and using the Burkholder-Davis-Gundy inequality, we estimate L^p-norm of $\max_{0 \le t \le T'} k(t)$ $(0 \le T' \le T)$, where $p \ge 2$. Similarly to the proof of Theorem 3.1 in [2] and Lemma 3.1 in [16], we have

$$
E[\max_{0 \le t \le T'} k(t)^p] \le C_p |x-y|^{2p} + C'_p \int_0^{T'} E\left[\max_{0 \le s \le t} k(s)^p\right] dt \tag{4.7}
$$

which implies the desired result.

We prove (2). When $x \notin B(z, 2\delta'/3)$, $X^z(t,x,\omega) = X^{N,z}(t,x,\omega) = x$ for all t, N. So we assume $x \in B(z, 2\delta'/3)$. If $B(z,\delta') \cap \partial D = \emptyset$, by the properties (i) and (ii), $X^{N,z}(t,x)$ and $X^z(t,x)$ never hits the boundary of D. Hence the classical Wong-Zakai theorem implies the assertion. Suppose $B(z,\delta') \cap \partial D \neq \emptyset$. Then there exists $z_0 \in \partial D$ such that $\overline{B(z,\delta')} \subset B(z_0,\delta)$. In [2], (4.3) is proved under the conditions (A), (B) and (C) on D. By Lemma 3.1, the condition (C) holds locally in some sense. Also, $X^{N,z}(t,x), X^z(t,x) \in B(z, 2\delta'/3)$. However, we cannot conclude that the proof in [2] works in the present case too. Because, there, first, we proved that the Euler approximation converges to the true solution in Theorem 3.1 and, second, the difference of the Euler approximation and the Wong-Zakai approximation converges to 0 in Lemma 4.6 in [2]. In the present case, the Euler approximation solution may exit from $x \in B(z, 2\delta'/3)$ and reach the boundary of D outside $B(z_0, \delta)$ even if $B(z, \delta')$. However, such a probability is small and we can prove (4.3). Let us show it more precisely. Let $X_E^{N,z}(t, x)$ be the Euler approximation of the solution to $(\mathrm{SDE})_{\sigma^z, b^z}$ with the starting point x associated with the partition $\{kT/N\}_{k=0}^N$ and $\Phi_E^{N,z}(t, x)$ be the associated local time term. See (3.37) for the definition of the Euler approximation. Let N be a sufficiently large number such that $\|b\|_\infty \Delta_N$ is small. Then by the estimate (3.2), we have

$$
\begin{aligned}
&P\left(\left\{\text{There exists a time } t \in [0, T] \text{ such that } X_E^{N,z}(t, x) \in B(z, \delta')^c\right\}\right)\\
&\le P\left(\max_{1\le k\le N} \|B\|_{\infty,[(k-1)T/N, kT/N]} \ge \varepsilon\delta'\right) \qquad (4.8)\\
&\le P\left(\|B\|_{\mathcal{H},\theta} > \varepsilon\delta' \left(\frac{N}{T}\right)^{\theta}\right) \le \exp\left(-C(\varepsilon\delta')^2 \left(\frac{N}{T}\right)^{2\theta}\right)
\end{aligned}
$$

where ε is a small positive number and $\| \|_{\mathcal{H},\theta}$ denotes the Hölder norm with exponent θ (θ <1/2). Thus, combining (4.8), and the moment estimates in Lemma 2.8 and Lemma 3.2 in [2] for X, Φ, $X_E^{N,z}$, $\Phi_E^{N,z}$, by a similar calculation to the proof of Theorem 3.1, we obtain

$$
E\left[\max_{0\le t\le T'} |X_E^{N,z}(t, x) - X(t, x)|^{2p}\right] \qquad (4.9)
$$

$$
\le C_T \Delta_N^p + e^{-C(N/T)^{2\theta}} + C_T \int_0^{T'} E\left[\max_{0\le s\le t} |X_E^{N,z}(s, x) - X(s, x)|^{2p}\right] ds \qquad (4.10)
$$

which implies $E[\max_{0\le t\le T} |X_E^{N,z}(t, x) - X(t, x)|^{2p}] \le C_T \Delta_N^p$. Similarly, the key of the proof of Lemma 4.6 in [2] is the non-positivity of the sum of second and third terms involving local times Φ^N and Φ_E^N in (4.49) in [2]. For $(\mathrm{SDE})_{\sigma^z, b^z}$ too, the corresponding term involving $\Phi^{N,z}$ is non-positive. For the term $\Phi_E^{N,z}$, by the same reasoning as in (4.10), we have

$$
\begin{aligned}
E\Bigg[\int_{t_{k-1}}^{t_k} \Bigg\{\rho^{N,z}(t) \left(\tilde{Z}^{N,z}(t), d\Phi_E^{N,z}(t)\right) - \frac{\rho^{N,z}(t)}{\gamma} |\tilde{Z}^{N,z}(t)|^2\\
\left((Df)(X_E^{N,z}(t)), d\Phi_E^{N,z}(t)\right)\Bigg\}\Bigg] \le C_T e^{-C(N/T)^{2\theta}}, \qquad (4.11)
\end{aligned}
$$

where $\rho^{N,z}(t) = \exp\left(-\frac{2}{\gamma}\left(f(X_E^{N,z}(t,x)) + f(X^{N,z}(t,x))\right)\right)$ and $\tilde{Z}^{N,z} = X_E^{N,z} - X^{N,z}$. Consequently, in a similar way to the proof of Lemma 4.6 in [2], we obtain for any $0 < \theta < 1$

$$\max_{0\le k\le N} E\left[|X^{N,z}(t_k^N) - X_E^{N,z}(t_k^N)|^2\right] \le C_\theta \cdot \Delta_N^{\theta/2} \tag{4.12}$$

$$E\left[\max_{0\le t\le T} |X^{N,z}(t) - X^z(t)|^2\right] \le C_{T,\theta}\Delta_N^{\theta/6}. \tag{4.13}$$

The assertion (3) can be proved by the same argument as in the proof of Lemma 5.5 in [23].

Proof (Proof of Theorem 2.3) Let $x \in \bar{D}$ and P_x denote the probability law of the process $X(t,x)$ $(0 \le t \le T)$ which exists on $C([0,T] \to \bar{D}; w(0) = x)$. Let $c(t)$ $(0 \le t \le T)$ be a point of the support of P_x and

$$U_r(c) = \left\{\omega \;\middle|\; \max_{0\le t\le T} |X(t,x,\omega) - c(t)| \le r\right\}. \tag{4.14}$$

It is sufficient to prove that for any $\varepsilon > 0$ and c

$$\lim_{N\to\infty} P\left(\left\{\max_{0\le t\le T} |X(t,x) - X^N(t,x)| \ge \varepsilon\right\} \cap U_{\delta'/4}(c)\right) = 0. \tag{4.15}$$

Let us define a subset of increasing numbers $\{s_0, \dots, s_K\} \subset \{t_k^N\}_{k=0}^N$ so that $s_0 = 0$ and $s_k = \max\{t_l^N \ge s_{k-1} \mid \max_{s_{k-1}\le t\le t_l^N} |c(t) - c(s_{k-1})| \le \delta'/8\}$. For any c, if N is sufficiently large, then the set on the RHS in the definition of s_k is not empty and $s_K = T$. Note that the set $\{s_k\}$ and K may depend on N but $\limsup_{N\to\infty} K < \infty$. We prove by an induction on $1 \le k \le K$ that for any $\varepsilon > 0$

$$\lim_{N\to\infty} P\left(\left\{\max_{0\le t\le s_k} |X(t,x) - X^N(t,x)| \ge \varepsilon\right\} \cap U_{\delta'/4}(c)\right) = 0. \tag{4.16}$$

First, we prove the case $k = 1$. Let $s_1^* = \max\{t \mid \max_{0\le s\le t} |c(s) - x| \le \delta'/8\}$. Clearly, $s_1 \le s_1^*$ and $s_1 \to s_1^*$ as $N \to \infty$. We prove

$$\lim_{N\to\infty} P\left(\left\{\max_{0\le t\le s_1^*} |X(t,x) - X^N(t,x)| \ge \varepsilon\right\} \cap U_{\delta'/4}(c)\right) = 0. \tag{4.17}$$

By Lemma 4.2 (3), we have

$$\begin{aligned} &P\left(\left\{\max_{0\le t\le s_1^*} |X(t,x) - X^N(t,x)| \ge \varepsilon\right\} \cap U_{\delta'/4}(c)\right) \\ &\quad = P\left(\left\{\max_{0\le t\le s_1^*} |X^x(t,x) - X^N(t,x)| \ge \varepsilon\right\} \cap U_{\delta'/4}(c)\right) \end{aligned}$$

$$
\begin{aligned}
&= P\Biggl(\left\{\max_{0\le t\le s_1^*}|X^x(t,x)-X^N(t,x)|\ge\varepsilon\right\}\\
&\qquad\qquad \cap\left\{\max_{0\le t\le s_1^*}|X^x(t,x)-X^{N,x}(t,x)|\ge\delta'/8\right\}\cap U_{\delta'/4}(c)\Biggr)\\
&\quad + P\Biggl(\left\{\max_{0\le t\le s_1^*}|X^x(t,x)-X^N(t,x)|\ge\varepsilon\right\}\\
&\qquad\qquad \cap\left\{\max_{0\le t\le s_1^*}|X^x(t,x)-X^{N,x}(t,x)|\le\delta'/8\right\}\cap U_{\delta'/4}(c)\Biggr)\\
&\le P\left(\left\{\max_{0\le t\le s_1^*}|X^x(t,x)-X^{N,x}(t,x)|\ge\delta'/8\right\}\right)\\
&\quad + P\left(\left\{\max_{0\le t\le s_1^*}|X^x(t,x)-X^{N,x}(t,x)|\ge\varepsilon\right\}\right).
\end{aligned}
\tag{4.18}
$$

Here we have used that for ω satisfying $\max_{0\le t\le s_1^*}|X^{N,x}(t,x,\omega)-x|\le\delta'/2$, $X^N(t,x,\omega)=X^{N,x}(t,x,\omega)$ holds for $0\le t\le s_1^*$. The estimate (4.18) and Lemma 4.2 (2) implies the case $k=1$. We prove (4.16) in the case of $k+1$ assuming the case of k. Let

$$
V_{\eta,k}=\left\{\omega\;\middle|\;\max_{0\le t\le s_k}|X(t,x)-X^N(t,x)|\le\eta\right\}. \tag{4.19}
$$

It suffices to prove

$$
\limsup_{\eta\to 0}\limsup_{N\to\infty}P\left(\left\{\max_{s_k\le t\le s_{k+1}}|X(t,x)-X^N(t,x)|\ge\varepsilon\right\}\cap U_{\delta'/4}(c)\cap V_{\eta,k}\right)=0. \tag{4.20}
$$

Note that for $t\ge s_k$,

$$
\begin{aligned}
X(t,x,\omega)&=X(t-s_k,X(s_k,x,\omega),\tau_k\omega),\\
X^N(t,x,\omega)&=X^N(t-s_k,X^N(s_k,x,\omega),\tau_k\omega),
\end{aligned}
\tag{4.21}
$$

where $(\tau_k\omega)(t)=\omega(t+s_k)$. This identity follows from the uniqueness of strong solutions and $B^N(t,\tau_k\omega)=B^N(s_k+t,\omega)$ for all k and $t\ge 0$. Hence,

$$
\begin{aligned}
&P\left(\left\{\max_{s_k\le t\le s_{k+1}}|X(t,x)-X^N(t,x)|\ge\varepsilon\right\}\cap U_{\delta'/4}(c)\cap V_{\eta,k}\right)\\
&\quad\le P\Biggl(\left\{\max_{0\le s\le s_{k+1}-s_k}|X(s,X(s_k),\tau_k\omega)-X(s,X^N(s_k),\tau_k\omega)|\ge\varepsilon/2\right\}
\end{aligned}
$$

$$
\begin{aligned}
&\qquad \cap U_{\delta'/4}(c) \cap V_{\eta,k}\Big) \\
&\quad + P\bigg(\Big\{\max_{0\le s\le s_{k+1}-s_k} |X(s, X^N(s_k), \tau_k\omega) - X^N(s, X^N(s_k), \tau_k\omega)| \ge \varepsilon/2\Big\} \\
&\qquad \cap U_{\delta'/4}(c) \cap V_{\eta,k}\bigg) \\
&:= I_1 + I_2, \qquad (4.22)
\end{aligned}
$$

where we have written $X(s_k) = X(s_k, x, \omega)$ and $X^N(s_k) = X^N(s_k, x, \omega)$ for simplicity. By Lemma 4.2 (1) and the Chebyshev inequality, we have $I_1 \le 4\varepsilon^{-2}C_2\eta^2$. Let

$$
\begin{aligned}
W_{k,\delta',\eta} &= \Big\{\max_{0\le s\le s_{k+1}-s_k} |X(s, X(s_k), \tau_k\omega) - X(s, X^N(s_k), \tau_k\omega)| \le \delta'/16\Big\} \\
&\qquad \cap U_{\delta'/4}(c) \cap V_{\eta,k}, \\
I_3 &= P\bigg(\Big\{\max_{0\le s\le s_{k+1}-s_k} |X(s, X^N(s_k), \tau_k\omega) - X^N(s, X^N(s_k), \tau_k\omega)| \ge \varepsilon/2\Big\} \\
&\qquad \cap W_{k,\delta',\eta}\bigg).
\end{aligned}
$$

To prove $\limsup_{\eta\to 0} \limsup_{N\to\infty} I_2 = 0$, it suffices to show $\limsup_{N\to\infty} I_3 = 0$ for sufficiently small η. We explain the reason. By Lemma 4.2 (3),

$$
\begin{aligned}
&P\bigg(\Big\{\max_{0\le s\le s_{k+1}-s_k} |X(s, X(s_k), \tau_k\omega) - X(s, X^N(s_k), \tau_k\omega)| \ge \delta'/16\Big\} \cap U_{\delta'/4}(c) \\
&\qquad \cap \Big\{|X(s_k) - X^N(s_k)| \le 2\Big\}\bigg) \\
&\le P\bigg(\Big\{\max_{0\le s\le s_{k+1}-s_k} |X^{c(s_k)}(s, X(s_k), \tau_k\omega) - X^{c(s_k)}(s, X^N(s_k), \tau_k\omega)| \ge \delta'/16\Big\} \\
&\qquad \cap U_{\delta'/4}(c) \cap \Big\{|X(s_k) - X^N(s_k)| \le 2\Big\}\bigg). \qquad (4.23)
\end{aligned}
$$

By Lemma 4.2 (1), this probability goes to 0 as $N \to \infty$ by the assumption of the induction. Now we estimate I_3. For $\omega \in W_{k,\delta',\eta}$, we have

$$
|X(s, X^N(s_k, x, \omega), \tau_k\omega) - c(s + s_k)| \le \frac{5\delta'}{16} \quad 0 \le s \le s_{k+1} - s_k \qquad (4.24)
$$

and so

$$
|X(s, X^N(s_k, x, \omega), \tau_k\omega) - c(s_k)| \le \frac{7\delta'}{16} \quad 0 \le s \le s_{k+1} - s_k. \qquad (4.25)
$$

Here we consider $(\mathrm{SDE})_{\sigma^{c(s_k)}, b^{c(s_k)}}$, where the driving path is $\tau_k\omega$. Then, for any $\omega \in W_{k,\delta',\eta}$, by (4.25) and Lemma 4.2 (3),

$$X(s, X^N(s_k, x, \omega), \tau_k\omega) = X^{c(s_k)}(s, X^N(s_k, x, \omega), \tau_k\omega) \qquad 0 \le s \le s_{k+1} - s_k. \tag{4.26}$$

Hence, by a similar argument to the case $k = 1$, we can prove $\limsup_{N\to\infty} I_3 = 0$ which completes the proof.

Remark 4.3 As explained in the above proof, we estimated the difference $X^N - X^N_E$ in Lemma 4.6 in [2]. However, it is easy to check that we can estimate the difference $X^N - X$ in a similar way to the proof of $X^N - X^N_E$ and obtain $\max_{0\le t\le T} E[|X^N(t) - X(t)|^2] \le C_{T,\theta}\Delta_N^{\theta/2}$ in the setting in [2]. In the proofs of Theorems 2.2 and 2.3 too, we can directly estimate the difference $X^N - X$ in the convex case and $X^{N,z} - X^z$ similarly. By noting this, actually, we do not need to use the Euler approximation in the above proofs too. Also, we note that Zhang [28] proved that the difference $X^N - X$ converges to 0 without using the Euler approximation under stronger assumptions than those in [2].

References

1. S. Aida, Reflecting rough differential equations, arXiv:1311.6104
2. S. Aida, K. Sasaki, Wong-Zakai approximation of solutions to reflecting stochastic differential equations on domains in Euclidean spaces. Stoch. Process. Appl. **123**(10), 3800–3827 (2013)
3. A.M. Davie, Differential equations driven by rough paths: an approach via discrete approximations, Appl. Math. Res. Express. AMRX 2007, no. 2, Art. ID abm009, 40
4. A. Deya, A. Neuenkirch, S. Tindel, A Milstein-type scheme without Lévy area terms for SDEs driven by fractional Brownian motion. Ann. Inst. Henri. Poincaré Probab. Stat. **48**(2), 518–550 (2012)
5. H. Doss, P. Priouret, Support d'un processus de réflexion. Z. Wahrsch. Verw. Gebiete **61**(3), 327–345 (1982)
6. P. Dupuis, H. Ishii, On Lipschitz continuity of the solution mapping to the Skorokhod problem, with applications. Stoch. Stoch. Rep. **35**(1), 31–62 (1991)
7. L.C. Evans, D.W. Stroock, An approximation scheme for reflected stochastic differential equations. Stoch. Process. Appl. **121**(7), 1464–1491 (2011)
8. M. Ferrante, C. Rovira, Stochastic differential equations with non-negativity constraints driven by fractional Brownian motion. J. Evol. Equ. **13**, 617–632 (2013)
9. P. Friz, H. Oberhauser, Rough path limits of the Wong-Zakai type with a modified drift term. J. Funct. Anal. **256**(10), 3236–3256 (2009)
10. P. Friz, S. Riedel, Convergence rates for the full Gaussian rough paths, arXiv:1108.1099
11. P. Friz, N. Victoir, *Multidimensional Stochastic Processes as Rough Paths Theory and Applications*, Cambridge Studies in Advanced Mathematics, 120 (Cambridge University Press, Cambridge, 2010)
12. M. Gubinelli, Controlling rough paths. J. Funct. Anal. **216**(1), 86–140 (2004)
13. I. Gyöngy, P.R. Stinga, Rate of convergence of Wong-Zakai approximations for stochastic partial differential equations. Seminar on Stochastic Analysis, Random Fields and Applications VII, Progress in Probability **67**, 95–130 (2013)

14. Y. Hu, D. Nualart, Rough path analysis via fractional calculus. Trans. Am. Math. Soc. **361**(5), 2689–2718 (2009)
15. N. Ikeda, S. Watanabe, *Stochastic Differential Equations and Diffusion Processes*, North-Holland Mathematical Library, 24 (North-Holland Publishing Co., Amsterdam, 1981)
16. P.L. Lions, A.S. Sznitman, Stochastic differential equations with reflecting boundary conditions. Comm. Pure Appl. Math. **37**(4), 511–537 (1984)
17. T. Lyons, Differential equations driven by rough signals. Rev. Mat. Iberoamer. **14**, 215–310 (1998)
18. T. Lyons, Z. Qian, *System Control and Rough Paths*, Oxford Mathematical Monographs (Oxford University Press, Oxford, 2002)
19. T. Lyons, M. Caruana, T. Lévy, *Differential Equations Driven by Rough Paths*, Lecture Notes in Mathematics, 1908 (Springer, Berlin, 2007)
20. R. Pettersson, Wong-Zakai approximations for reflecting stochastic differential equations. Stoch. Anal. Appl. **17**(4), 609–617 (1999)
21. J. Ren, S. Xu, A transfer principle for multivalued stochastic differential equations. J. Funct. Anal. **256**(9), 2780–2814 (2009)
22. J. Ren, S. Xu, Support theorem for stochastic variational inequalities. Bull. Sci. Math. **134**(8), 826–856 (2010)
23. Y. Saisho, Stochastic differential equations for multi-dimensional domain with reflecting boundary. Probab. Theory Relat. Fields **74**(3), 455–477 (1987)
24. L. Słomiński, On approximation of solutions of multidimensional SDEs with reflecting boundary conditions. Stoch. Process. Appl. **50**(2), 197–219 (1994)
25. L. Słomiński, Euler's approximations of solutions of SDEs with reflecting boundary. Stoch. Process. Appl. **94**(2), 317–337 (2001)
26. H. Tanaka, Stochastic differential equations with reflecting boundary condition in convex regions. Hiroshima Math. J. **9**(1), 163–177 (1979)
27. E. Wong, M. Zakai, On the relation between ordinary and stochastic differential equations. Int. J. Eng. Sci. **3**, 213–229 (1965)
28. T-S. Zhang, Strong convergence of Wong-Zakai approximations of reflected SDEs in a multidimensional general domain, 2014, doi:10.1007/s11118-014-9394-9, Potential analysis

Symmetric Diffusions with Polynomial Eigenvectors

Dominique Bakry

Abstract We describe symmetric diffusion operators where the spectral decomposition is given through a family of orthogonal polynomials. In dimension one, this reduces to the case of Hermite, Laguerre and Jacobi polynomials. In higher dimension, some basic examples arise from compact Lie groups. We give a complete description of the bounded sets on which such operators may live. We then provide in dimension 2 a classification of those sets when the polynomials are ordered according to their usual degrees.

Keywords Orthogonal polynomials · Diffusion operators · Random matrices

Classifications 60B15 · 60B20 · 60G99 · 43A75 · 14H50

1 Introduction

Symmetric diffusion operators and their associated heat semigroups play a central rôle in the study of continuous Markov processes, and also in differential geometry and partial differential equations. The analysis of the associated heat or potential kernels have been considered from many points of view, such as short and long time asymptotics, upper and lower bounds, on the diagonal and away from it, convergence to equilibrium, for example. All these topics had been deeply investigated during the past half century, see [3, 11, 27] for example. Unfortunately, there are very few examples where computations are explicit.

The spectral decomposition may provide one approach to the heat kernel study and the analysis of convergence to equilibrium, especially when the spectrum is discrete. Once again, there are very few models where this spectral decomposition is at hand, either for the explicit expressions of the eigenvalues or the eigenvectors. The aim of this survey is to present a family of models where this spectral decomposition is

D. Bakry (✉)
Institut de Mathématiques de Toulouse, Université Paul Sabatier,
118 Route de Narbonne, 31062 Toulouse, France
e-mail: dominique.bakry@gmail.com

D. Crisan et al. (eds.), *Stochastic Analysis and Applications 2014*,
Springer Proceedings in Mathematics & Statistics 100,
DOI 10.1007/978-3-319-11292-3_2

completely explicit. Namely, we shall require the eigenvectors to be polynomials in a finite number of variables. We are then dealing with orthogonal polynomials with respect to the reversible measure of this diffusion operator. Once again, orthogonal polynomial have been thoroughly investigated in many aspects, going back to the early works of Legendre, Chebyshev, Markov and Stieltjes, see [12, 22, 26] e.g.

To be more precise, we shall consider some open connected set $\Omega \subset \mathbb{R}^d$, with piecewise smooth boundary $\partial\Omega$ (say at least piecewise $\mathcal{C}^1$, and may be empty), and some probability measure $\mu(dx)$ on Ω, with smooth positive density measure $\rho(x)$ with respect to the Lebesgue measure and such that polynomials are dense in $\mathcal{L}^2(\mu)$.

A diffusion operator L on Ω (but we shall also consider such objects on smooth manifolds with no further comment) is a linear second order differential operator with no 0-order terms, therefore written as

$$\mathrm{L}(f) = \sum_{ij} g^{ij}(x)\partial^2_{ij} f + \sum_i b^i(x)\partial_i f, \tag{1.1}$$

such that at every point $x \in \Omega$, the symmetric matrix $(g^{ij}(x))$ is non negative. For simplicity, we shall assume here that this matrix is always non degenerate in Ω (it may, and will in general, be degenerate at the boundary $\partial\Omega$). This is an ellipticity assumption, which will be in force throughout. We will also assume that the coefficients $g^{ij}(x)$ and $b^i(x)$ are smooth. We are mainly interested in the case where L is symmetric on $\mathcal{L}^2(\mu)$ when restricted to smooth functions compactly supported in Ω. On the generator L, this translates into the following relation between the coefficients of the operator L and the density $\rho(x)$ of the reversible measure with respect to the Lebesgue measure

$$\mathrm{L}(f) = \frac{1}{\rho}\sum_i \partial_i\Big(\rho \sum_j g^{ij}\partial_j f\Big), \tag{1.2}$$

as is readily seen using integration by parts in Ω, see [3].

We are also interested in the case when $\mathcal{L}^2(\mu)$ admits a complete orthonormal basis $P_q(x)$, $q \in \mathbb{N}$, of polynomials such that $\mathrm{L}(P_q) = -\lambda_q P_q$, for some real (indeed non negative) parameters λ_q. This is equivalent to the fact that there exists an increasing sequence $\mathcal{P}_n$ of finite dimensional subspaces of the set $\mathcal{P}$ of polynomials such that $\cup_n \mathcal{P}_n = \mathcal{P}$ and such that L maps $\mathcal{P}_n$ into itself.

When this happens, we have a spectral decomposition of L in this basis, and, when a function $f \in \mathcal{L}^2(\mu)$ is written as $f = \sum_q c_q P_q$, then $\mathrm{L}(f) = \sum_q -\lambda_q c_q P_q$, such that any expression depending only on the spectral decomposition may be analyzed easily.

Our aim is to describe various families of such situations, which will be referred to as polynomial models. In dimension d, many such models may be constructed with various techniques: Lie groups, root systems, Hecke algebras, etc., see [1, 2, 7, 9, 10, 12–14, 16–22], among many other possible references. Introducing weighted degrees, we shall analyse the situation where the operator maps for any $k \in \mathbb{N}$

the space $\mathcal{P}_k$ of polynomials with degree less than or equal to k into itself. For bounded sets Ω with regular boundaries, this leads to an algebraic description of the admissible boundaries for such sets. In dimension 2 and for the usual degree, we give a complete description of all the admissible models (reducing to 11 families up to affine transformation). We then present some other models, with other degrees, or in larger dimension, with no claim for exhaustivity.

This short survey is organized as follows. In Sect. 2, we present a brief tour of the dimension 1 case, where the classical families of Hermite, Laguerre and Jacobi polynomials appear. We provide some geometric description for the Jacobi case (which will serve as a guide in higher dimension) together with various relations between those three families. In Sect. 3 we describe some basic notions concerning the symmetric diffusion operators, introducing in particular the operator Γ and the integration by parts formula. We also introduce the notion of image, or projection, of operators, a key tool towards the construction of polynomial models. In Sect. 4, we describe the Laplace operators on spheres, $SO(n)$ and $SU(n)$, and we provide various projections arising from these examples leading to polynomials models, in particular for spectral measures in the Lie group cases. Section 5 describes the general case, with the introduction of the weighted degree models. In particular, when the set Ω is bounded, we provide the algebraic description of the sets on which such polynomial models may exist. In particular, we show that the boundaries of those admissible sets Ω must lie in some algebraic variety, satisfying algebraic restrictions. The description of the sets lead then to the description of the measures and the associated operators. Section 6 is a brief account of the complete classification for the ordinary degree of those bounded models in dimension 2. This requires a precise analysis of the algebraic nature of the boundary which is only sketched in this paper. Section 7 provides some examples of 2 dimensional models with weighted degrees, and is far from exhaustive, since no complete description is valid at the moment. Section 8 proposes some new ways (apart from tensorization) to construct higher dimensional models from low dimension ones. Finally, we give in Sect. 9 the various pictures corresponding to the 2 dimensional models described in Sect. 6.

2 Dimension 1 Case: Jacobi, Laguerre and Hermite

In dimension one, given a probability measure μ for which the polynomials are dense in $\mathcal{L}^2(\mu)$, there is, up to the choice of a sign, a unique family (P_n) of polynomials with $\deg(P_n) = n$ and which is an orthonormal basis in $\mathcal{L}^2(\mu)$. It is obtained through the Gram-Schmidt orthonormalization procedure of the sequence $\{1, x, x^2, \ldots\}$. When does such a sequence consists of eigenvectors of some given diffusion operator of the form $L(f)(x) = a(x)\partial^2 f + b(x)\partial f$ (a Sturm-Liouville operator)? This had been described long ago (see e.g. [4, 23]), and reduces up to affine transformation to the classical cases of Jacobi, Hermite and Laguerre polynomials.

Those three families play a very important rôle in many branches of mathematics and engineering (mainly in statistics and probability for the Hermite family and in fluid mechanics for the Jacobi one), and we refer to the huge literature on them for further information. We briefly recall these models, one the different possible intervals $\mathcal{I} \subset \mathbb{R}$, up to affine transformations.

1. $\mathcal{I} = \mathbb{R}$. The measure μ is Gaussian centered: $\mu(dx) = \frac{e^{-x^2/2}}{\sqrt{2\pi}}\, dx$. The associated polynomials are the Hermite polynomials (H_n). They are eigenvectors of the Ornstein–Uhlenbeck operator

$$\mathcal{H} = \frac{d^2}{dx^2} - x\frac{d}{dx} \quad \mathcal{H}(H_n) = -nH_n.$$

2. $\mathcal{I} = \mathbb{R}_+^*$ The measure is the gamma measure $\mu_a(dx) = C_a x^{a-1} e^{-x}\, dx, a > 0$. The associated polynomials are the Laguerre polynomials $L_n^{(a)}$, and the associated operator is the Laguerre operator

$$\mathcal{L}_a = x\frac{d^2}{dx^2} + (a-x)\frac{d}{dx}, \quad \mathcal{L}_a(L_n^{(a)}) = -nL_n^{(a)}.$$

3. $\mathcal{I} = (-1, 1)$. The measure is the beta measure $\mu_{a,b}(dx) = C_{a,b}(1-x)^{a-1}(1+x)^{b-1}\, dx,\ a, b > 0$. The associated polynomials are the Jacobi polynomials $(J_n^{(a,b)})$ and the associated Jacobi operator is

$$\mathcal{J}_{a,b} = (1-x^2)\frac{d^2}{dx^2} - \big(a-b+(a+b)x\big)\frac{d}{dx}, \quad \mathcal{J}_{a,b} J_n^{(a,b)} = -n(n+a+b)J_n^{(a,b)}.$$

The Jacobi family contains the ultraspherical (or Gegenbauer) family (when $a{=}b$) with as particular cases the Legendre polynomials $a = b = 0$, the Chebyshev of the first and second kind ($a = b = -1/2$ and $a = b = 1/2$ respectively), which appear, after renormalization, when writing $\cos(n\theta) = P_n(\cos\theta)$ (first kind) and $\sin(n\theta) = \sin(\theta)Q_n(\cos\theta)$ (second kind). The first two families (Hermite and Laguerre) appear as limits of the Jacobi case. For example, when we chose $a = b = n/2$ and let then n go to ∞, and scale the space variable x into $x/\sqrt{n}$, the measure $\mu_{a,a}$ converges to the Gaussian measure, the Jacobi polynomials converge to the Hermite ones, and $\frac{2}{n}\mathcal{J}_{a,a}$ converges to $\mathcal{H}$.

In the same way, the Laguerre case is obtained from the Jacobi one fixing b, changing x into $\frac{2x}{a} - 1$, and letting a go to infinity. Then $\mu_{a,b}$ converges to μ_b, and $\frac{1}{a}\mathcal{J}_{a,b}$ converges to $\mathcal{L}_b$.

Also, when a is a half-integer, the Laguerre operator may be seen as the image of the Ornstein–Uhlenbeck operator in dimension d. Indeed, as the product of one dimensional Ornstein–Uhlenbeck operators, the latter has generator $H_d = \Delta - x.\nabla$. It's reversible measure is $e^{-|x|^2/2}dx/(2\pi)^{d/2}$, it's eigenvectors are the products $Q_{k_1}(x_1)\dots Q_{k_d}(x_d)$, and the associated process $X_t = (X_t^1, \dots, X_t^d)$, is formed of independent one dimensional Ornstein-Uhlenbeck processes. Then, if one considers $R(x) = |x|^2$, then one may observe that, for any smooth function $F : \mathbb{R}_+ \mapsto \mathbb{R}$,

$$\mathcal{H}_d\big(F(R)\big) = 2\mathcal{L}_a(F)(R),$$

where $a = d/2$. In the probabilist interpretation, this amounts to observe that if X_t is a d-dimensional Ornstein–Uhlenbeck process, then $|X_{t/2}|^2$ is a Laguerre process with parameter $a = d/2$.

In the same way, as we shall see below in Sect. 4, when $a = b = (d-1)/2$, $\mathcal{J}_{a,a}$ may be seen as the Laplace operator $\Delta_{\mathbb{S}^{d-1}}$ on the unit sphere $\mathbb{S}^{d-1}$ in $\mathbb{R}^d$ acting on functions depending only on the first coordinate (or equivalently on functions invariant under the rotations leaving $(1, 0, \dots, 0)$ invariant), and a similar interpretation is valid for $J_{p/2,q/2}$ for integers p and q. This interpretation comes from Zernike and Brinkman [8] and Braaksma and Meulenbeld [6] (see also Koornwinder [15]). Jacobi polynomials also play a central role in the analysis on compact Lie groups. Indeed, for (a, b) taking the various values of $(q/2, q/2)$, $((q-1)/2, 1)$, $(q-1, 2)$, $(2(q-1), 4)$ and $(4, 8)$ the Jacobi operator $\mathcal{J}_{a,b}$ appears as the radial part of the Laplace-Beltrami (or Casimir) operator on the compact rank 1 symmetric spaces, that is spheres, real, complex and quaternionic projective spaces, and the special case of the projective Cayley plane (see Sherman [25]).

3 Basics on Symmetric Diffusions

Diffusion operators are associated with diffusion processes (that is continuous Markov processes) through the requirement that, if (X_t) is the Markov process with associated generator L, then for any smooth function f, $f(X_t) - \int_0^t \mathrm{L}(f)(X_s)ds$ is a local martingale, see [3]. Here, we are mainly interested in such diffusion operators which are symmetric with respect to some probability measure μ. In probabilistic terms, this amounts to require that when the law of X_0 is μ, then not only at any time the law of X_t is still μ, but also, for any time $t > 0$, the law of $(X_{t-s}, 0 \leq s \leq t)$ is the same as the law of $(X_s, 0 \leq s \leq t)$ (that is why this measure is often called the reversible measure).

For any diffusion operator as such described in Eq. (1.1), we may define the carré du champ operator Γ as the following bilinear application

$$\Gamma(f, g) = \frac{1}{2}\Big(\mathrm{L}(fg) - f\mathrm{L}(g) - g\mathrm{L}(f)\Big), \tag{3.1}$$

defined say for smooth functions defined on $\Omega \subset \mathbb{R}^d$. From formula (1.2), it is readily seen that

$$\Gamma(f, g) = \sum_{ij} g^{ij} \partial_i f \partial_j g, \tag{3.2}$$

such that $\Gamma(f, f) \geq 0$. As already mentioned, we restrict ourselves to the case where the matrix (g^{ij}) is everywhere positive definite. Then, the inverse matrix (g_{ij}) defines a Riemannian metric. By abuse of language, we shall refer to the matrix (g^{ij}) (or equivalently to the operator Γ) as the metric of the operator, although formally it should be called a co-metric. Then, Riemannian properties of this metric may carry important information on the models under study. Typically, most of the models described below in Sect. 6 have constant curvature.

The operator L satisfies the following chain rule (or change of variable formula). For any k-uple $f = (f_1, \ldots, f_k)$ and any smooth function $\Phi : \mathbb{R}^k \mapsto \mathbb{R}$,

$$\mathrm{L}\big(\Phi(f)\big) = \sum_i \partial_i \Phi(f)\mathrm{L}(f_i) + \sum_{ij} \partial^2_{ij}\Phi(f)\Gamma(f_i, f_j). \tag{3.3}$$

This allows us to compute $\mathrm{L}\big(\Phi(f)\big)$ as soon as we know $\mathrm{L}(f_i)$ and $\Gamma(f_i, f_j)$. This is in particular the case for the coordinates $x^i : \Omega \mapsto \mathbb{R}$, where $\Gamma(x^i, x^j) = g^{ij}(x)$ and $\mathrm{L}(x^i) = b^i(x)$, recovering then the form given in Eq. (1.1). But we may observe that the family $f = (f_i, i = 1 \ldots k)$ in the previous formula does not need to be a coordinate system (that is a diffeomorphism from Ω into some new set Ω_1 between d-dimensional manifolds or open sets in $\mathbb{R}^d$). There may be more function (f_i) than required ($k > d$) or less ($k < d$). This remark will play an important rôle in the sequel.

In general, when looking at such operators, one considers first the action of the operator on smooth compactly supported function on Ω. Since we want to work on polynomials, it is better to enlarge the set of functions we are working on, for example the set of $\mathcal{L}^2(\mu)$ functions which are smooth and compactly supported in a neighborhood of Ω, referred to below just as "smooth functions".

The operator is symmetric in $\mathcal{L}^2(\mu)$ when, for any pair (f, g) of smooth functions

$$\int_\Omega \mathrm{L}(f)\, g\, d\mu = \int_\Omega f\, \mathrm{L}(g)\, d\mu. \tag{3.4}$$

Usually, for this to be true, one should require one of the functions f or g to be compactly supported in Ω, or ask for some boundary conditions on f and g, such as Dirichlet or Neuman. However, in the case of polynomial models, the operator will be such that no boundary conditions will be required for Eq. (3.4) to hold. More precisely, at any regular point of the boundary, we shall require the unit normal vector to belong to the kernel of the matrix (g^{ij}). Under such assumption, the symmetry Eq. (3.4) is satisfied whenever f and g are smooth functions.

If we observe that $\mathrm{L}(\mathbf{1}) = 0$ (where $\mathbf{1}$ denotes the constant function), then applying (3.4) to $\mathbf{1}$ shows that $\int \mathrm{L}(f)\, d\mu = 0$, and therefore, applying the definition of Γ and integrating over Ω (provided $fg \in \mathcal{L}^2(\mu)$), that

$$\int \mathrm{L}(f)\, g\, d\mu = -\int \Gamma(f, g)\, d\mu = \int_\Omega f\, \mathrm{L}(g)\, d\mu, \tag{3.5}$$

which is called the integration by parts formula.

The fact that $\Gamma(f, f) \geq 0$ and the previous formula (3.5) shows that, for any function f, $\int_\Omega f\,\mathrm{L}(f)\,d\mu \leq 0$, and therefore all eigenvalues are non positive.

The operator is entirely determined by the knowledge of Γ and μ, as is obvious from formula (1.2), and the datum (Ω, Γ, μ) is called a Markov triple in the language of [3], to which we refer for more details about the general description of such symmetric diffusion operators.

As already mentioned, we want to analyse those situations such that the set $\mathcal{P}$ of polynomials is dense in $\mathcal{L}^2(\mu)$ (and, being an algebra, all polynomials will be automatically in any $\mathcal{L}^p(\mu)$ for any $p > 1$), and such that there exists some Hilbert basis (P_n) of $\mathcal{L}^2(\mu)$ with elements in $\mathcal{P}$ such that $\mathrm{L}(P_n) = -\lambda_n P_n$. Since we also require that any polynomial is a finite linear combination of the P_n's, we see that the set $\mathcal{P}_n = \{\sum_0^n \mu_k P_k\}$ is an increasing sequence of finite dimensional linear subspaces of $\mathcal{P}$ such that $\mathrm{L} : \mathcal{P}_n \mapsto \mathcal{P}_n$, and $\cup_n \mathcal{P}_n = \mathcal{P}$.

Conversely, if there exists such an increasing sequence $\mathcal{P}_n$ of finite dimensional linear subspaces of $\mathcal{P}$ such that $\cup_n \mathcal{P}_n = \mathcal{P}$ satisfying $\mathrm{L} : \mathcal{P}_n \mapsto \mathcal{P}_n$, then we may find a sequence (P_n) which is an orthonormal basis of $\mathcal{L}^2(\mu)$ and eigenvectors of L. Indeed, the restriction of L to the finite dimensional subspace $\mathcal{P}_n$ is symmetric when we provide it with the Euclidean structure inherited from the $\mathcal{L}^2(\mu)$ structure, and therefore may be diagonalized in some orthonormal basis. Repeating this in any space $\mathcal{P}_n$ provides the full sequence of polynomial orthogonal vectors.

It may be worth to observe that when this happens, the set of polynomials is an algebra dense in $\mathcal{L}^2(\mu)$ and stable under L and the associated heat semigroup $Q_t = \exp(t\mathrm{L})$. When this happens, it is automatically dense in the $\mathcal{L}^2(\mu)$ domain of L, for the domain topology, and the set of polynomial will be therefore a core for our operator (see [3] for more details).

From now on, we shall denote by (P_n) such a sequence of eigenvectors, with $\mathrm{L}(P_n) = -\lambda_n P_n$ (and we recall that $\lambda_n \geq 0$). Since $\mathrm{L}(\mathbf{1}) = 0$, we may always chose $P_0 = \mathbf{1}$, and $\lambda_0 = 0$. In general, this eigenvalue is simple, in particular in the elliptic case. Indeed, thanks to the integration by parts formula (3.5), any function f such that $\mathrm{L}(f) = 0$ satisfies $\int \Gamma(f, f)d\mu = 0$, from which $\Gamma(f, f) = 0$. If ellipticity holds (but also under much weaker requirements), then this implies that f is constant.

As mentioned in the introduction, one is often interested in the heat semigroup Q_t associated with L, that is the linear operator $\exp(t\mathrm{L})$, defined through the fact that $Q_t P_n = e^{-\lambda_n t} P_n$, or equivalently by the fact $F(x, t) = Q_t(f)(x)$ satisfies the heat equation $\partial_t F = \mathrm{L}_x(F)$, with $F(x, 0) = f(x)$.

This heat semigroup may be represented (at least at a formal level) as

$$Q_t(f)(x) = \int_\Omega f(y) q_t(x, y) d\mu(y),$$

where the heat kernel $q_t(x, y)$ may be written

$$q_t(x, y) = \sum_n e^{-\lambda_n t} P_n(x) P_n(y), \tag{3.6}$$

provided that the series $\sum_n e^{-2\lambda_n t}$ is convergent (in practise and in all our models, this will always be satisfied). From this we see that a good knowledge on λ_n and P_n provides information of this heat kernel. However, it happens (thanks to the positivity of Γ) that P_t preserves positivity, (and of course $P_t(\mathbf{1}) = 1$), which is equivalent to the fact that $p_t(x, y)\mu(dy)$ is a probability measure for any $t > 0$ and any $x \in \Omega$, in particular $p_t(x, y) \geq 0$. This is not at all obvious from the representation (3.6). Therefore, this representation (3.6) of $p_t(x, y)$ does not carry all the information about it.

It is worth to observe the following, which will be the basic tool for the construction of our polynomial models. Start from an operator L (defined on some manifold Ω_1 or some open set in it), symmetric under some probability measure μ. Assume that we may find a set of functions $f = (f_i)_{i=1,\ldots k}$, that we consider as a function $f : \Omega \mapsto \Omega_1 \subset \mathbb{R}^k$, and smooth functions B^i and G^{ij} mapping Ω_1 to $\mathbb{R}$ such that

$$L(f_i) = B^i(f), \Gamma(f_i, f_j) = G^{ij}(f).$$

Then, for any smooth function $F : \Omega_1 \mapsto \mathbb{R}$, and thanks to formula (3.3), one has

$$\mathrm{L}\big(F(f)\big) = \mathrm{L}_1(F)(f),$$

where

$$\mathrm{L}_1(F) = \sum_{ij} G^{ij}\partial^2_{ij}F + \sum_i B^i\partial_i F. \tag{3.7}$$

This new diffusion operator L_1 is said to be the image of L under f. In probabilistic terms, the image $Y_t = f(X_t)$ of the diffusion process X_t associated with L under f is still a diffusion process, and it's generator is L_1. Moreover, if L is symmetric with respect to μ, L_1 is symmetric with respect of the image measure ν of μ under f.

This could (and will) be an efficient method to determine the density ρ_1 of this image measure with respect to the Lebesgue measure, through the use of formula (1.2).

4 Examples: Spheres, $SO(d)$, $SU(d)$

We now describe a few natural examples leading to polynomial models. The first basic remark is that, given a symmetric diffusion operator L described as before by a Markov triple (Ω, Γ, μ), it is enough to find a set $\{X_1, \ldots, X_k\}$ of functions such that for any $i = 1, \ldots, k$, $\mathrm{L}(X_i)$ is a degree 1 polynomial in the variables X_j, and for any pair (i, j), $i, j = 1, \ldots, k$, $\Gamma(X_i, X_j)$ is a degree 2 polynomial in those variables. Indeed, if such happens, then the image L_1 of L under $X = \{X_1, \ldots, X_k\}$ given in (3.7) is a symmetric diffusion operator with reversible measure μ_1, where μ_1 is the image of μ under X. Thanks to formula (3.3), L_1 preserves for each $n \in \mathbb{N}$ the set $\mathcal{P}_n$ of polynomials in the variables (X_i) with total degree less than n. One may then diagonalize L_1 in $\mathcal{P}_n$, and this leads to the construction of a $\mathcal{L}^2(\mu_1)$ orthonormal basis formed with polynomials.

Moreover, given polynomial models, we may consider product models, which are again polynomial models, and from them consider projections. Indeed, given two polynomial models described by triples $(\Omega_i, \Gamma_i, \mu_i)$ as in Sect. 3, we may introduce on $\Omega_1 \times \Omega_2$ the measure $\mu_1 \otimes \mu_2$, and the sum operator $\Gamma_1 \oplus \Gamma_2$, acting on functions $f(x, y)$ and $h(x, y)$ as

$$\big(\Gamma_1 \otimes \mathrm{Id} \oplus \mathrm{Id} \otimes \Gamma_2\big)(f, h)(x, y) = \sum_{ij} g_1^{ij} \partial_{x_i} f \partial_{x_j} h + \sum_{kl} g_2^{kl}(y) \partial_{yk} f \partial_{yl} h.$$

Once we have some polynomial models $(\Omega_1, \Gamma_1, \mu_1)$ and $(\Omega_2, \Gamma_2, \mu_2)$, then $(\Omega_1 \times \Omega_2, \Gamma_1 \oplus \Gamma_2, \mu_1 \otimes \mu_2)$ is again a polynomial model. At the level of processes, if (X_t^i) is associated with $(\Omega_i, \Gamma_i, \mu_i)$, and are chosen independent, then the process (X_t^1, X_t^2) is associated with this product. This kind of tensorisation procedure constructs easily new dimensional models in higher dimension once we have some in low dimension. The polynomials $R_{i,j}(x, y)$, $x \in \Omega_1$, $y \in \Omega_2$ associated with the product are then just the tensor products $P_i(x) Q_j(y)$ of the polynomials associated with each of the components.

Moreover, one may consider quotients in these products to construct more polynomial models, as we did to pass from the one dimensional Ornstein-Uhlenbeck operator to the Laguerre operator.

The easiest example to start with is the Laplace operator $\Delta_{\mathbb{S}}$ on the unit sphere $\mathbb{S}^{d-1} \subset \mathbb{R}^d$. This operator may be naively described as follows: considering some smooth function f on $\mathbb{S}^{d-1}$, we extend it in a neighborhood of $\mathbb{S}^{d-1}$ into a function which is independent of the norm, that is $\hat{f}(x) = f(\frac{x}{\|x\|})$, where $\|x\|^2 = \sum_i x_i^2$, for $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$. Then, we consider $\Delta(\hat{f})$, where Δ is the usual Laplace operator in $\mathbb{R}^d$, and restrict $\Delta(\hat{f})$ on $\mathbb{S}^d$. This is $\Delta_{\mathbb{S}}(f)$.

An easy exercise shows that, for the functions x_i which are the restriction to $\mathbb{S}^{d-1}$ of the usual coordinates in $\mathbb{R}^d$, then

$$\Delta_{\mathbb{S}}(x_i) = -(d-1)x_i, \;\; \Gamma_{\mathbb{S}}(x_i, x_j) = \delta_{ij} - x_i x_j.$$

The uniform measure on the sphere (that is the unique probability measure which is invariant under rotations) is the reversible measure for $\Delta_{\mathbb{S}}$. The system of functions $(x_1, \ldots, x_d)$ is not a coordinate system, since those functions are linked by the relation $\sum_i x_i^2 = 1$. For example, one sees from the previous formulae that $\Delta_{\mathbb{S}}(\|x\|^2) = \Gamma(\|x\|^2, \|x\|^2) = 0$ on $\mathbb{S}^{d-1}$ (an good way to check the right value for λ when imposing $\mathrm{L}(x_i) = -\lambda x_i$ with the same Γ operator).

But the system $(x_1, \ldots, x_{d-1})$ is a system of coordinates for say the upper half sphere. We may observe that the operator indeed projects onto $(x_1, \ldots, x_{d-1})$ into an elliptic operator in the unit ball $\mathbb{B}^{d-1} = \{\|x\| < 1\}$, with exactly the same relations for $\mathrm{L}(x_i)$ and $\Gamma(x_i, x_j)$. The image operator (that is the Laplace operator in this system of coordinates) is then

$$\mathrm{L} = \sum_{ij} (\delta_{ij} - x_i x_j)\partial^2_{ij} - (d-1)\sum_i x_i \partial_i,$$

and from the formula (1.2), is is easy to determine the density measure (up to a multiplicative constant), which is $\rho(x) = (1 - \|x\|^2)^{-1/2}$, which happens here to be $\det(g^{ij})^{-1/2}$ (as is always the case for Laplace operators). We see then that this provides an easy way to compute the density of the uniform measure on the sphere under this map $\mathbb{S}^{d-1} \mapsto \mathbb{B}^{d-1}$, which projects the upper half sphere onto the unit ball.

One may also observe that if $M = (M_{ij})$ is a matrix (with fixed coefficients), and if $y_i = \sum_j M_{ij} x_j$, then $\Delta(y_i) = -(d-1)y_i$ and $\Gamma(y_i, y_j) = (MM^t)_{ij} - y_i y_j$. Then, when M is orthogonal, the image measure of $\Delta_{\mathbb{S}}$ under $x \mapsto Mx$ is $\Delta_{\mathbb{S}}$ itself, which tells us that the Laplace operator is invariant under orthogonal transformations.

We may also consider the projection π from $\mathbb{S}^{d-1}$ to $\mathbb{B}_p$ for $p < d-1 : \pi : (x_1, \ldots, x_d) \mapsto (x_1, \ldots, x_p)$, which provides the same operator as before, except that now we are working on the unit ball $\mathbb{B}^p \subset \mathbb{R}^p$ and $L(x_i) = -(d-1)x_i$, where the parameter d is no longer correlated with the dimension p of the ball. We may as well consider the generic operator L_λ on the unit ball in $\mathbb{R}^p$ with $\Gamma(x_i, x_j) = \delta_{ij} - x_i x_j$ and $\mathrm{L}(x_i) = -\lambda x_i$, where $\lambda > p-1$. Is is readily checked that it has symmetric measure density $C_{p,d}(1 - \|x\|^2)^{\frac{\lambda-p-1}{2}}$. As a consequence, the image measure of the sphere $\mathbb{S}^{d-1}$ onto the unit ball through this projection has density $C_{d,p}(1 - \|x\|^2)^{(d-p-2)/2}$. It is worth to observe that when λ converges to $p-1$, the measure converges to the uniform measure on the boundary of $\mathbb{B}^p$, that is $\mathbb{S}^{p-1}$, and the operator converges to the operator on $\mathbb{S}^{p-1}$.

When we chose $p = 1$, we recover the symmetric Jacobi polynomial model in dimension 1 with parameters $a = b = (d-1)/2$.

For these operators, we see that, in terms of the variables (x_i), g^{ij} are polynomials with total degree 2 and $L(x_i)$ are degree 1. Therefore, in view of the chain rule formula (3.3), we see that the operator L such defined maps the space $\mathcal{P}_n$ of polynomials with total degree less than n into itself, and this provides a first family of polynomial models.

One may also still consider the unit sphere in $\mathbb{R}^d$, and choose integers such that $p_1 + \cdots + p_k = d$. Then, setting $P_0 = 0$, $P_i = p_1 + \cdots + p_i$, consider the functions $X_i = \sum_{P_{i-1}+1}^{P_i} x_j^2$, for $i = 1, \ldots k-1$. The image of the sphere under this application is the simplex $\mathcal{D}_{k-1} = \{X_i \geq 0, \sum_1^{k-1} X_i \leq 1\}$. We have $\Delta_{\mathbb{S}}(X_i) = 2(p_i - dX_i)$, and $\Gamma(X_i, X_j) = 4X_i(\delta_{ij} - X_i X_j)$. The operator $\Delta_{\mathbb{S}}$ thus projects on the simplex, with

$$G^{ij} = 4X_i(\delta_{ij} - X_j), \;\; B^i = 2(p_i - dX_i),$$

and provides again a polynomial model on it. The reversible measure is easily seen to have density $C \prod_1^{k-1} X_i^{r_i}(1 - \sum_1^{k-1} X_i)^{r_k}$, with $r_i = \frac{p_i - 2}{2}$, $ri = 1, \ldots, k$, which a Dirichlet measure on the simplex $\mathcal{D}_{k-1}$. This measure is then seen as the image of the uniform measure on the sphere through this application $X : \mathbb{S}^{d-1} \mapsto \mathcal{D}_{k-1}$. The general Dirichlet density measure $\rho_{r_1,\ldots,r_k}(X) = X_1^{r_1} \ldots X_{k-1}^{r_{k-1}}(1 - \sum_i X_i)^{r_k}$, with real parameters $r_i > -1, i = 1, \dots, k$, also produces (with the same Γ operator) a new family of polynomial models. We may play a around with the Dirichlet measures for half integer parameters. For example, the image in a Dirichlet $\rho_{r_1,\dots,r_k}$ measure for the image through $(X_1, X_2, \dots, X_{k-1}) \mapsto (X_1 + X_2, X_3, \dots, X_{k-1})$ is a Dirichlet measure with parameters $\rho_{r_1+r_2,r_3,\dots,r_k}$, which is obvious from the sphere interpretation (that is when $a_i = \frac{p_i-2}{2}$ for some integers p_i), and extends easily to the general class. The same procedure is still true at the level of the operators (or the associated stochastic processes).

Once again, when considering the case $k = 2$, we get an operator on $[0, 1]$. Changing X into $2X - 1$, we get an operator on $[-1, 1]$, which, up to the scaling factor 4, is the asymmetric Jacobi model with parameters $a = r_1/2, b = (d - r_1)/2$.

There are many other ways to produce polynomial models from spheres, (and we shall provide some later in dimension 2, see Sect. 7). But we want to show another very general way to construct some. Let us consider some semi-simple compact Lie group $\mathbb{G}$ (such as $SO(n)$, $SU(n)$, $Sp(n)$, etc.). On those groups, there exist a unique invariant probability measure (that is invariant under $x \mapsto gx$ and $x \mapsto xg$, for any $g \in \mathbb{G}$). This is the Haar measure. There also exists a unique (up to a scaling constant) diffusion operator $\Delta_{\mathbb{G}}$ which is also invariant under left and right action: this means that if, for a smooth function $f : \mathbb{G} \mapsto \mathbb{R}$, one defines for any $g \in \mathbb{G}$, $L_g(f)(x) = f(gx)$, then $\Delta_{\mathbb{G}}(L_g(f)) = L_g(\Delta_{\mathbb{G}} f)$, and the same is true for the right action $R_g(f)(x) = f(xg)$. This operator is called the Laplace (or Casimir) operator on the group $\mathbb{G}$. Assume then that the Lie group $\mathbb{G}$ is represented as a group of matrices, as is the case for the natural presentation of the natural above mentioned Lie groups. The Lie algebra $\mathcal{G}$ of $\mathbb{G}$ is the tangent space at the origin, and to any element A of the Lie algebra, that is a matrix (A_{ij}), we may associate some vector field R_A on the group through $X_A(f)(g) = \partial_t f(ge^{tA})_{t=0}$. If we write $g = (g_{ij})$ and consider a function $f(g_{ij}) : \mathbb{G} \mapsto \mathbb{R}$, then

$$X_A(g) = \sum_{ijk} g_{ik} A_{kj} \partial_{g_{ij}} f,$$

and therefore X_A preserves the set $\mathcal{P}_n$ of polynomials with total degree n in the variables (g_{ij}). Now, the Casimir operator may be written as $\sum_i X_{A_i}^2$, where the A_i form an orthonormal basis for some natural quadratic form on the Lie algebra $\mathcal{G}$ called the Killing form. This operator also preserves the set $\mathcal{P}_n$. Unfortunately, those "coordinates" g_{ij} are in general linked by algebraic relations, and may not serve as a true coordinate system on the group. However, we may then describe the operator $\Delta_{\mathbb{G}}$ through it's action on those functions $g_{ij} : \mathbb{G} \mapsto \mathbb{R}$.

Without further detail, consider the group $SO(n)$ of orthogonal matrices with determinant 1. Let m_{ij} be the entries of a matrix in $SO(n)$, considered as functions on the group. We have

$$\Delta_{SO(n)}(m_{ij}) = -(n-1)m_{ij}, \ \Gamma_{SO(n)}(m_{kl}, m_{qp}) = \delta_{kq}\delta_{lp} - m_{kp}m_{ql}.$$

We now show some projections of this operator. Let $\mathcal{M}(p,q)$ be the space of $p\times q$ matrices. Select p lines and q columns in the matrix $g\in SO(n)$, (say the first ones), and consider the map $\pi : SO(n)\mapsto \mathcal{M}(p,q)$ which to $M\in SO(n)$ associates the extracted matrix $N=(m_{ij})$, $1\leq i\leq p$, $1\leq j\leq q$. From the form of $\Delta_{SO(n)}(m_{ij})$ and $\Gamma_{SO(n)}(m_{ij},m_{kl})$, it is clear that the operator projects on $\mathcal{M}(p,q)$ through π. It may happen (whenever $p+q>n$) that the image is contained in a sub-manifold (indeed an algebraic variety) of $\mathcal{M}(p,q)$. But we have nevertheless a new diffusion operator on this image, and the associated process is known as the matrix Jacobi process. It is worth to observe that if $p=n$ and $q=1$, this is nothing else than the spherical operator $\Delta_{\mathbb{S}}$ in $\mathbb{S}^{n-1}$. In general, whenever $p+q\leq n$, this process lives of the symmetric domain $\{NN^*\leq \mathrm{Id}\}$, and has a reversible measure with density ρ with respect to the Lebesgue measure which is $\det(\mathrm{Id}-NN^*)^{(n-p-q-1)/2}$, which is easily seen from formula (1.2). We may also now fix p and q and consider n as a parameter, and we obtain a family of polynomial processes on this symmetric domain a long as $p+q<n+1$.

One may play another game and consider the image of the operator on the spectrum. More precisely, given the associated process $X_t\in SO(n)$, one looks at the process obtained on the eigenvalues of X_t (that is the spectral measure of X_t). This process is again a diffusion process, for which we shall compute the generator. To analyze the spectral measure (that is the eigenvalues up to permutations of them), the best is to look at the characteristic polynomial $P(X)=\det(M-X\mathrm{Id})=\sum_{i=1}^n a_iX^i$. Then we want to compute $\Delta_{SO(n)}(a_i)$ and $\Gamma_{SO(n)}(a_i,a_j)$.

For a generic matrix $M=(M_{ij})$, Cramer's formulae tells us that, on the set where M is invertible, $\partial_{M_{ij}}\log(\det(M))=M^{-1}_{ji}$ and $\partial^2_{M_{ij},M_{kl}}\log(\det(M))=-M^{-1}_{jk}M^{-1}_{li}$.

From this, and using the chain rule formula, we get that

$$\begin{aligned}\Delta_{SO(n)}\log P(X)= &-(n-1)\mathrm{trace}\,(M(M-X\mathrm{Id})^{-1}\\ &-\mathrm{trace}\Big((M-X\mathrm{Id})^{-1}(M^t-X\mathrm{Id})^{-1}\Big)\\ &+\Big(\mathrm{trace}\,M(M-X\mathrm{Id})^{-1}\Big)^2.\end{aligned}$$

and

$$\begin{aligned}\Gamma\Big(\log(P(X)),\log(P(Y))\Big)= &\ \mathrm{trace}\,\Big((M-X\mathrm{Id})^{-1}(M^t-Y\mathrm{Id})^{-1}\Big)\\ &-\mathrm{trace}\,\Big(M^2(M-X\mathrm{Id})^{-1}(M-Y\mathrm{Id})^{-1}\Big).\end{aligned}$$

But

$$\begin{cases}\mathrm{trace}\,(M(M-X\mathrm{Id})^{-1}=n-X\frac{P'}{P}(X),\\ \mathrm{trace}\,\Big((M-X\mathrm{Id})^{-1}(M^t-Y\mathrm{Id})^{-1}\Big)=\frac{1}{1-XY}\Big(\frac{1}{Y}\frac{P'}{P}(\frac{1}{Y})-X\frac{P'}{P}(X)\Big)\\ \mathrm{trace}\,\Big(M^2(M-X\mathrm{Id})^{-1}(M-Y\mathrm{Id})^{-1}\Big)=n+\frac{1}{X-Y}\Big(X^2\frac{P'}{P}(X)-Y^2\frac{P'}{P}(Y)\Big).\end{cases}$$

One may use the fact that $M \in SO(n)$ to see that $P(\frac{1}{X}) = (-X)^{-n}P(X)$, so that $\frac{1}{Y}\frac{P'}{P}(\frac{1}{Y}) = n - Y\frac{P'}{P}(Y)$.

In the end, we see that

$$\Delta_{SO(n)}(P) = -(d-1)XP' + X^2P'',$$

$$\Gamma_{SO(n)}\big(P(X), P(Y)\big) = \frac{XY}{1-XY}\Big(nP(X)P(Y) \\ + \frac{(1-X^2)P(Y)P'(X) - (1-Y^2)P(X)P'(Y)}{X-Y}\Big).$$

Since $\Delta_{SO(n)}P(X) = \sum_i \Delta_{SO(n)}(a_i)X^i$ and

$$\Gamma_{SO(n)}\big(P(X), P(Y)\big) = \sum_{ij} \Gamma_{SO(n)}(a_i, a_j)X^iY^j,$$

we see from the action of $\Delta_{SO(n)}$ and $\Gamma_{SO(n)}$ on $P(X)$ that, in terms of the variables (a_k), $\Delta_{SO(n)}(a_i)$ are degree one polynomials and $\Gamma_{SO(n)}(a_i, a_j)$ are degree two. Therefore, from the same argument as before, the operator $\Delta_{SO(n)}$ projects on it's spectrum into a polynomial model in the variables a_i.

The same is true (with similar computations) for the spectra of NN^*, where N is the extracted matrix in $\mathcal{M}(p,q)$ described above (corresponding to the matrix Jacobi process), and for many more models.

Similarly, one may also look at the special unitary group $SU(n)$, where the coordinates $(z_{ij} = x_{ij} + \mathbf{i}y_{ij})$ are the entries of the matrix. Using complex coordinates, one has then to consider $\Delta_{SU(n)}(z_{ij})$ and $\Gamma(z_{ij}, z_{kl})$ and $\Gamma(z_{ij}, \bar{z}_{kl})$ in order to recover the various quantities corresponding to the variables x_{ij} and y_{ij} (using the linearity of Δ and the bilinearity of Γ). We have (up to some normalization)

$$\begin{cases} \Delta_{SU(n)}(z_{ij}) = -(n^2-1)z_{ij}, \\ \Gamma_{SU(n)}(z_{ij}, z_{kl}) = z_{ij}z_{kl} - nz_{il}z_{kj}, \\ \Gamma_{SU(n)}(z_{ij}, \bar{z}_{kl}) = n\delta_{ik}\delta_{jl} - z_{ij}\bar{z}_{kl}) \end{cases}$$

The same remark as before applies about the extracted matrices, and also, with the same method, we get for the characteristic polynomial

$$\begin{cases} \Delta_{SU(n)}(P) = -(n^2-1)XP' + (n+1)X^2P'', \\ \Gamma_{SU(n)}(P(X), P(Y)) = XY\Big(P'(X)P'(Y) + \frac{n}{X-Y}\big(P'(X)P(Y) - P'(Y)P(X)\big)\Big), \\ \Gamma_{SU(n)}(P(X), \bar{P}(Y)) = \frac{1}{1-XY}\Big(nP(X)\bar{P}(Y) - \bar{Y}\bar{P}'(Y)P(X) - XP'(X)\bar{P}(Y)\Big). \end{cases} \tag{4.1}$$

Once again, the Casimir operator on $SU(n)$ projects onto a polynomial model in the variables of the characteristic polynomial.

5 The General Case

As we briefly showed in Sect. 4, there are many models for orthogonal polynomials and they are quite hard to describe in an exhaustive way. We propose in this Section a more systematic approach. Recall that we are considering probability measures μ on $\mathbb{R}^d$, on some open connected set $\Omega \subset \mathbb{R}^d$ for which the set $\mathcal{P}$ of polynomials are dense in $\mathcal{L}^2(\mu)$. Recall that it is enough for this to hold that there exists some $\epsilon > 0$ such that $\int e^{\epsilon \|x\|} d\mu < \infty$.

The first thing to do is to describe some ordering of the polynomials. For this, we chose a sequence $a = (a_1, \ldots, a_d)$ of positive integers and say that the degree of a monomial $x_1^{p_1} x_2^{p_2} \ldots x_d^{p_d}$ is $a_1 p_1 + \cdots a_d p_d$. Then, the degree $\deg_a(P)$ of a polynomial $P(x_1, \ldots, x_d)$ is the maximum value of the degree of it's monomials (such a degree is usually referred as a weighted degree). Then, the space $\mathcal{P}_n$ of polynomials with $\deg_a(P) \le n$ is finite dimensional. Moreover, $\mathcal{P}_n \subset \mathcal{P}_{n+1}$, and $\cup_n \mathcal{P}_n = \mathcal{P}$.

To chose a sequence of orthogonal polynomials (P_k) for μ, we chose at each step n some orthonormal basis in the orthogonal complement $\mathcal{H}_n$ of $\mathcal{P}_{n-1}$ in $\mathcal{P}_n$. This space in general has a large dimension (increasing with n), and there is therefore not an unique choice of such an orthonormal basis.

We are then looking for diffusion differential operators L (with associated Γ operator) on Ω, such that L admits such a sequence (P_n) as eigenvectors, with real eigenvalues . The operator L will be then automatically essentially self-adjoint on $\mathcal{P}$. The first observation is that for each n, L maps $\mathcal{P}_n$ into itself.

In particular, for each coordinate, $\mathrm{L}(x_i) \in \mathcal{P}_{a_i}$ and for each pair of coordinates (x_i, x_j), $\Gamma(x_i, x_j) \in \mathcal{P}_{a_i + a_j}$. Then, writing $\mathrm{L} = \sum_{ij} g^{ij} \partial_{ij} + \sum_i b^i \partial_i$, we see that

$$g^{ij} \in \mathcal{P}_{a_i + a_j}, \; b^i \in \mathcal{P}_{a_i} \tag{5.1}$$

Moreover, under the conditions (5.1), we see from the chain rule formula (3.3) that for each $n \in \mathbb{N}$, L maps $\mathcal{P}_n$ into itself. Provided it is symmetric on $\mathcal{P}_n$ for the Euclidean structure induced from $\mathcal{L}^2(\mu)$, we will then be able to derive an orthonormal basis formed with eigenvectors for it.

Once we have a polynomial model with a given choice of degrees $(a_1, \ldots, a_d)$, say, in the variables $(x_i, i = 1, \ldots, d)$, and as soon as one may find polynomials $X_1, \ldots, X_k$ in the variables x_i with $\deg_a(X_i) = b_i$, as soon as $\mathrm{L}(X_i)$ and $\Gamma(X_i, X_j)$ are polynomials in those variables X_i, then we get a new model in the variables (X_i) (provided however that the ellipticity requirement is satisfied), with new degrees $b = (b_1, \ldots, b_k)$. Indeed, from the chain rule (3.3), one sees that the image operator

L_1 maps the set $\mathcal{Q}_n$ of polynomials in the variables (X_i) with degree $\deg_b \leq n$ into itself.

The next task is then to describe the sets Ω on which such a choice for L and μ is possible. For that, we restrict our attention for those $\Omega \subset \mathbb{R}^d$ which are bounded with piecewise $\mathcal{C}^1$ boundaries. We call those sets admissible sets.

Then, we have the main important characterization

Theorem 5.1

1. *If Ω is an admissible set, then $\partial\Omega$ is included into an algebraic variety, with* $\deg_a \leq 2\sum_i a_i$.
2. *Let Q be the reduced equation of ∂_Ω. Ω is admissible if and only if there exist some $g^{ij} \in \mathcal{P}_{a_i+a_j}$ and $L_i \in \mathcal{P}_{a_i}$ such that*

$$\forall i = 1, \ldots, d, \sum_j g^{ij}\partial_j Q = L_i Q, \ (g^{ij}) \textit{ with non negative in } \Omega. \tag{5.2}$$

   When such happens, Q divides $\det(g)$.
3. *Let $Q = Q_1, \ldots, Q_k$ the decomposition of Q into irreducible factors. If* (5.2) *is satisfied, then any function $\rho(x) = Q_1^{r_1}, \ldots, Q_k^{r_k}$ is an admissible density measure for the measure μ, provided $\int_\Omega \rho(x)dx < \infty$. When $Q =$* $\det(g)$*, then there are no other measures.*
4. *For any solution (g^{ij}) of* (5.2)*, and any μ as before, setting $\Gamma(f, g) = \sum_{ij} g^{ij}\partial_i\partial_j g$, then the triple (Ω, Γ, μ) is an admissible solution for the operator* L.

Remark 5.2 Observe that Eq. (5.2) may be rewritten as $\Gamma(\log Q, x_i) = L_i$.

Proof We shall not give the full details of the proof here (see [5] for a complete proof), and just describe the main ideas.

Suppose we have a polynomial model with coefficients g^{ij}, b^i on Ω, with polynomial functions g^{ij} and b^i satisfying the above requirements on their degrees.

The first thing to observe is that if L is diagonalizable of $\mathcal{P}_n$ for each $n \in \mathbb{N}$, then for each polynomial pair (P, Q)

$$\int_\Omega \mathrm{L}(P)\, Q\, d\mu = \int_\Omega P\, \mathrm{L}(Q)\, d\mu. \tag{5.3}$$

This extends easily to any pair (f, g) of smooth functions compactly supported in Ω, so that the description (1.2) holds true. Moreover, Ω being bounded, and the coefficients g^{ij} and b^i being polynomials, formula (5.3) extends further to every pair (f, g) of smooth functions, not necessarily with support in Ω. Using Stokes formula, (and the regularity of the boundary of Ω), this imposes that, for any pair of smooth function (f, h), $\int_{\partial\Omega} \sum_{ij} f\partial_i h g^{ij} n_j dx = 0$, where (n_j) is the normal tangent vector at the boundary $\partial\Omega$. Therefore, this implies that, for any i, $\sum_j g^{ij} n_j = 0$ on the boundary, so that (n_j) is in the kernel of (g) at the boundary. This implies in turn that the boundary lies inside the algebraic set $\{\det(g) = 0\}$.

Therefore, $\partial\Omega$ is included in some algebraic variety. For any regular point $x \in \partial\Omega$ consider an irreducible polynomial Q_1 such that, in a neighborhood $\mathcal{V}$ of x, the boundary is included in $\{Q_1 = 0\}$. Then, (n_j) is parallel to $\partial_j Q_1$, so that $\sum_j g^{ij}\partial_j Q_1 = 0$ on $\mathcal{V} \cap \{Q_1 = 0\}$. From Hilbert's Nullstellensatz, $\sum_j g^{ij}\partial_j Q_1 = L_i Q_1$, for some polynomial L_i.

This being valid for any polynomial Q_1 appearing in the reduced equation of $\partial\Omega$, this is still true for the reduced equation itself (and in fact the two assertions are equivalent).

For a given polynomial Q, if Eq. (5.2) admits a non trivial solution (g^{ij}), then $\partial_i Q$ is in the kernel of (g^{ij}) at every regular point of $\{Q = 0\}$. Then, $\det(g)$ vanishes at that point. Q being reduced, then Q is a factor of $\det(g)$.

Now, the link between $b^i = \mathrm{L}(x_i)$ and $\sum_g^{ij} \partial_j \log \rho$ given in (1.2) shows that, in order for b^i to be a polynomial with degree less than a_i, it is enough (and in fact equivalent) to have $\sum_j g^{ij}\partial_j \log \rho = A_i$, for some polynomial A_i with degree less than a_i. But comparing with Eq. (5.2) shows that it is satisfied for $Q_i^{r_i}$ for any factor Q_i of Q and any parameter r_i. Then, all the condition are satisfied and the model (Ω, Γ, μ) is a polynomial model.

One sees that indeed the problem of determining polynomial models relies entirely on the study of the boundary $\partial\Omega$, at least as far as bounded sets Ω are considered. Given any algebraic curve, and a fixed choice of degrees $(a_1, \ldots, a_k)$ it is an easy task to decide if this curve is a candidate to be (or to belong to) the boundary of some set Ω on which there exist a polynomial model: Eq. (5.2) must have a non trivial solution. This equation is a linear system of equations in the coefficients of the polynomials g^{ij} and L_i, however in general with much more equations than variables.

Moreover, as soon as one has a model on Ω, there exist as we already saw many other models on the same set, with the same (g^{ij}), with measures described with a finite number of parameters, depending on the number of irreducible components in the reduced equation of $\partial\Omega$.

The solutions of Eq. (5.2) provide a set of measures which are admissible. The admissible measures are determined through the requirement that $\sum_j g^{ij}\partial_j \log \rho = A_i$, with $\deg_a(A_i) \leq a_i$, or in other terms $\Gamma(\log \rho, x_i) = A_i$. When the reduced equation of the boundary is $\{\det(g) = 0\}$, then we have described all the measures in Theorem 5.1. But when some factor of $\det(g)$ does not appear in the reduced equation of $\partial\Omega$, it is nor excluded that those factor may provide some other admissible measure (see [5]). However, in dimension two and for the usual degree, where we are able do provide a complete description of all possible models, this situation never appears and we wonder if this may appear at all.

The fact that the boundary is included into $\{\det(g)\} = 0$ allows to restrict in general to one of the connected components of the complement of this set, so that the metric may never degenerate inside Ω. But it may happen (although we have no example for this) that there exist some solutions of this problem for which the solution (g^{ij}) is not positive inside any bounded region bounded by $\{Q = 0\}$.

But the determination of all possible admissible boundaries (that is the curves for which Eq. (5.2) have a non trivial solution) is a much harder problem. The possibility for an algebraic surface to have a non trivial solution in Eq. (5.2) is a very strong requirement, as we shall see next, and this reduces considerably the possible models.

6 Classification with Usual Degree in Dimension 2

In this Section, we reduce to the two dimensional case, and moreover to the usual degree $a_1 = a_2 = 1$. In this situation, the problem is then invariant under affine transformation, and this allows to use classical tools of algebraic geometry to reduce the problem. The coefficients (g^{ij}) have at most degree 2 and the boundary maximum degree 4. The main result is the following

Theorem 6.1 *In dimension* 2 *and for the usual degree* $a_1 = a_2 = 1$*, and up to affine transformations, there exist exactly 11 bounded sets* Ω *(with piecewise* $\mathcal{C}^1$ *boundary) corresponding to a polynomial model. Their boundaries are (see pictures in Sect.* 9*): the triangle, the circle, the square, two coaxial parabolas, a parabola, a tangent line and a line parallel to the axis, a parabola and two tangent lines, a cuspidal cubic and a tangent line, a cuspidal cubic and a line passing through the infinite point of the cubic, a nodal cubic, the swallow tail and the deltoid curve.*

In all the models, the only possible values for the measure are the one described in Theorem 5.1*. When the boundary has maximal degree, then the metric* (g^{ij}) *is unique up to scaling, and correspond to a constant curvature metric, either* 0 *or* 1 *(after appropriate scaling).*

There are models (triangle, circle) where the metric (g^{ij}) *is not unique.*

Remark 6.2 The previous assertion is not completely exact. The family described by two axial parabolas are not reducible one to the other under affine transformations. But a simple quadratic transformation does the job.

Proof It is out of scope to give the proof here, which is quite lengthy and technical. But it relies on some simple argument. The main point is to show (using appropriate change of coordinates allowed by the affine invariance of the problem) that there may be no flex point and no flat point on the boundary. That is (in complex variables), that one may no find an analytic branch locally of the form $y = x^3 + o(x^3)$ or $y = x^4 + o(x^4)$. This is done through the local study of Eq. (5.2). Such points correspond to singular points of the dual curve. But there is a balance between the singular points of the curve, of it's dual curve, and the genus of the curve (seen as a compact Riemann surface), known as Plucker's formulae. This allows to show that $\partial\Omega$ must indeed have many singular points, which list is easy to write since the degree is low (here 4). It remains to analyze all the possible remaining cases. See [5] for details.

Observe that the triangle and the circle case where already described in Sect. 4 as image of the two dimensional sphere. But even in this case, Eq. (5.2) produces other metrics (g^{ij}) than the one already described. If one considers a single entry of a $SU(d)$ matrix, then it corresponds to a polynomial model in the unit disk which is one of these exotic metrics on the unit ball in $\mathbb{R}^2$. Typically, on the circle, one may add to the generator $a(x\partial_y - y\partial_x)^2$, which satisfies the boundary condition and corresponds to some extra random rotation in the associated stochastic process. Indeed, all these 11 models may be described, at least for some half-integer values of the parameters appearing in the description of the measure, as the image of the above mentioned models constructed on spheres, $SO(d)$ or $SU(d)$. But there are also, for other values of the measure, some constructions provided by more sophisticated geometric models, in particular root systems in Euclidean spaces. We refer to [5] for a complete description of theses models.

7 Other Models with Other Degrees in Dimension 2

When Ω is not bounded, Eq. (5.2) is not fully justified. If one restricts our attention to the usual degree and to those boundaries which satisfy this condition, then we obtain only the products of the various one dimensional models, and two extra models which are bounded by a cuspidal cubic or a parabola. In this situation, there are some exponential factors in the measures, as happens in the Laguerre case. When there is no boundary at all, it may be proved (although not easily) that the only admissible measures are the Gaussian ones: they correspond to the product of Ornstein-Uhlenbeck operators, but as is the case with the circle, one may add to the metric some rotational term $(x\partial_y - y\partial_x)^2$, which produces new families of orthogonal polynomials.

Beyond this, one may exhibit some examples on various bounded sets Ω with weighted degrees. There is no complete classification in general for such general models at the moment. The reason is that affine invariance is then lost (this is counterbalanced by the fact that some other polynomial change of variables are allowed), but the local analysis made above is no longer valid. To show how rich this new family of models may be, we just present here some examples.

On $SU(3)$, let Z be the trace of the matrix, considered as a function $Z : SU(3) \mapsto \mathbb{C} = \mathbb{R}^2$. Then thanks to the fact that there are 3 eigenvalues belonging to the unit circle and product 1, the characteristic polynomial $\det(M - \mathrm{Id})$ of an $SU(3)$ matrix may be written as $-X^3 + ZX^2 - \bar{Z}X + 1$ such that Z itself encodes the spectral measure. Applying formulae (4.1), and up to a change of Z into $Z/3$ and scaling, one gets

$$\begin{cases} \Gamma(Z, Z) = \bar{Z} - Z^2, \\ \Gamma(\bar{Z}, Z) = 1/2(1 - Z\bar{Z}), \\ \Gamma(\bar{Z}, \bar{Z}) = Z - \bar{Z}^2, \\ \mathrm{L}Z = -4Z, \mathrm{L}\bar{Z} = -4\bar{Z}. \end{cases} \tag{7.1}$$

This corresponds indeed with the deltoid model appearing in Sect. 6. From these formulae, one sees that functions $F(Z, \bar{Z})$ which are symmetric in $(Z, \bar{Z})$ are preserved by the image operator. Setting $S = Z + \bar{Z}$ and $P = Z\bar{Z}$, this leads to the following polynomial model with degree $\deg_S + 2\deg_P$, with

$$\begin{cases} \Gamma(S,S) = 1 + S + P - S^2, \\ \Gamma(S,P) = \frac{1}{2}S - 2P + S^2 - \frac{3}{2}SP, \\ \Gamma(P,P) = P - 3SP - 3P^2 + S^3, \\ \mathrm{L}S = -4S, \mathrm{L}(P) = 1 - 9P. \end{cases}$$

Up to some constant, the determinant of the metric is $(4P - S^2)(4S^3 - 3P^2 - 12SP - 6P + 1)$, and the boundary is the domain which is delimited by a parabola and a cuspidal cubic (a degree 5 curve), which are bi-tangent at their intersection point. This leads to a two-parameter family of measures and associated orthogonal polynomials.

One may also construct more models using discrete symmetry groups. Here are some examples.

We cut the 2-d sphere into n vertical slices along the meridians, and and look for a basis of functions invariant under the reflections around these meridians. Writing the sphere as $x_1^2 + x_2^2 + x_3^2 = 1$, we chose $X = x_3$ and writing in complex notations $x_1 + \mathbf{i}x_2 = z$,we chose $Y = \Re(z^n)$. In polynomials terms $Y = (x_1^2 + x_2^2)^{n/2} P_n(\frac{x_1}{\sqrt{x_1^2 + x_2^2}})$, where P_n is the n-th first kind Chebychef polynomial. For parity considerations on P_n, this is always a polynomial in the variables (x_1, x_2). We have

$$\begin{cases} \Gamma(X,X) = 1 - X^2, \\ \Gamma(X,Y) = -nXY, \\ \Gamma(Y,Y) = n^2\Big((1 - X^2)^{n-1} - Y^2\Big) \\ \mathrm{L}X- = -2X, \ \mathrm{L}(Y) = -n(n+1)Y \end{cases}$$

The boundary equation is then $(1 - X^2)^n - Y^2 = 0$, which is irreducible when n is odd, leading to a one parameter family of measures, and splits into two parts when n is even, leading to a two parameters family. We may in the previous model look at functions of (X^2, Y) adding a new invariance under symmetry around the hyperplane $\{x_3 = 0\}$, or of (X, Y^2), or also of (X^2, Y^2) leading to 3 new families.

There are other ways to construct such two dimensional models. A general idea is to consider some finite sub-group of $SO(3)$, and extract from the axes of the rotation some subfamily V_i which is invariant under the group action. Then, one considers the homogeneous polynomial $P(x, y, z) = \prod_i X \cdot V_i$, where $X = (x, y, z)$ and $X \cdot V$ denotes the scalar product. It is also invariant under the group action. Let m be the degree of this polynomial. With P, one constructs a new polynomial $Q = r^{m+1/2} P(\partial_x, \partial_y, \partial_z) r^{-1/2}$, where $r = (x^2 + y^2 + z^2)^{1/2}$. Q is still homogeneous with degree m, invariant under the group action, and moreover harmonic. Then, one looks for the system $X = Q$, $Y = \Gamma(Q, Q)$ where Γ is the spherical Laplace operator on the unit sphere $\mathbb{S}^2 \subset \mathbb{R}^3$. The action of the spherical operator on the pair (X, Y) may lead to a polynomial system. This is not always the case however. For example, with the symmetry group of the icosahedron, there are 3 homogeneous polynomials P_6, P_{10} and P_{15}, with degrees 6, 10 and 15 which generate all homogeneous polynomials which are invariant under the group action. The technology that we provided works starting from P_6 but not from P_{15}. The associated formulae are too complicated to be given here. For the reader interested in explicit computations (see [24] for more examples), the explicit value of P_6 is as follows (with $c = (1+\sqrt{5})/2$)

$$P_6(x, y, z) = (c^2x^2 - y^2)(c^2y^2 - z^2)(c^2z^2 - x^2).$$

8 Higher Dimensional Models

The technology which allows to describe all the bounded dimensional models for the usual degree is not available in higher dimension, mainly because of the lack of the analogues of Plucker's formulae. The many models issued from Lie group action that we produced so far provide many families, with various degrees. Beyond these explicit constructions, and sticking to the bounded models with the usual degree in dimension 2, it is worth to observe that one may produce new admissible sets by double cover. More explicitly, as soon as we have a model in dimension d, with reduced boundary equation $P(x) = 0$, one may look for models in dimension $d + 1$ with boundary equation $y^2 - P(x) = 0$ (y being the extra one dimensional variable. It turns out that this produces a new model for every two dimensional model which have no cusps or tangent lines as singular points (that is, in our setting, the circle, the triangle, the square, the double parabola and the nodal cubic). The boundary has no longer maximal degree, even if the starting model has, and the metric is not unique in general. Moreover, even in the simplest cases, the curvature is not constant.

We may then pursue the construction adding new dimensions. The reason why this works (together with the obstruction about singular points) remains mysterious. Most of the questions regarding these constructions and others remain open at the moment.

9 Pictures

In this Section, we give the various pictures for the 11 bounded models in dimension 2 with natural degree. We give the reduced equation of the boundary, and we indicate when the metric is unique, up to a scaling factor. When it is unique, we indicate the cases when the curvature is constant, and what is it's sign. It is worth to observe that all the models with maximal degree (here 4) have a unique constant curvature metric (Figs. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 and 11).

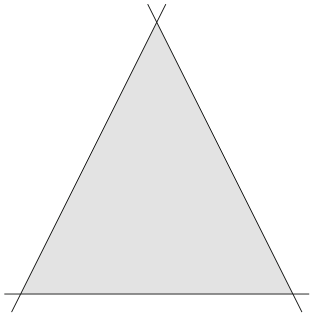


Fig. 1 Triangle: $xy(1 - x - y) = 0$, metric not unique

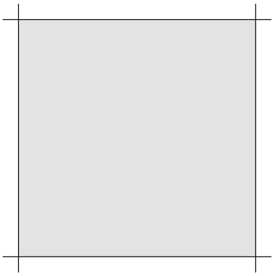


Fig. 2 Square: $(1 - x^2)(1 - y^2) = 0$, one metric, curvature 0

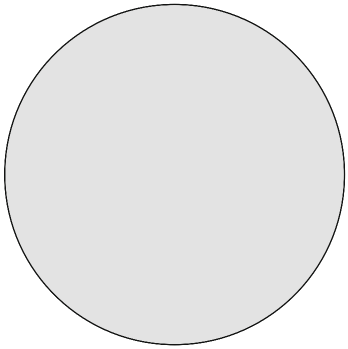


Fig. 3 Circle: $x^2 + y^2 = 1$, metric not unique

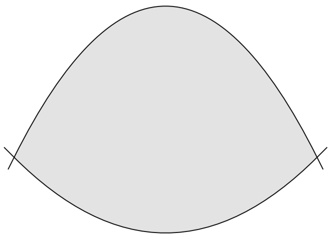


Fig. 4 Double parabola: $(y+1-x^2)(y-1+ax^2) = 0$, one metric, curvature 1

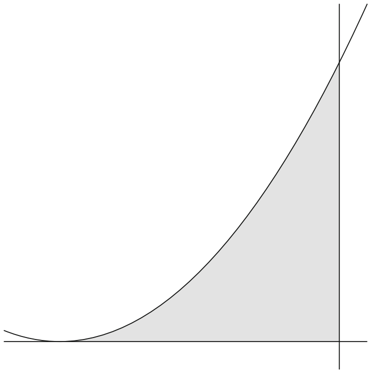


Fig. 5 Parabola with 1 tangent and one secant line: $(y - x^2)y(x - 1) = 0$, one metric, curvature 1

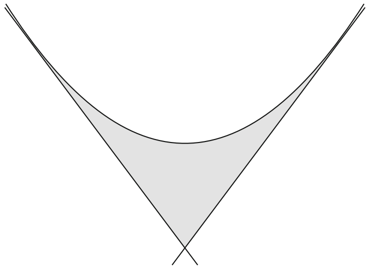


Fig. 6 Parabola with 2 tangents: $(y - x^2)(y + 1 - 2x)(y + 1 + 2x) = 0$, one metric, curvature 0

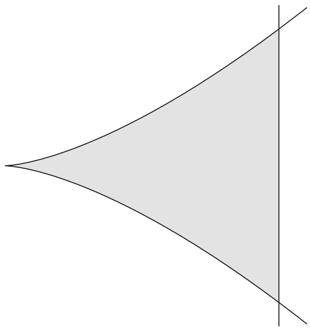


Fig. 7 Cuspidal cubic with secant line: $(y^2 - x^3)(x - 1) = 0$, one metric, curvature 1

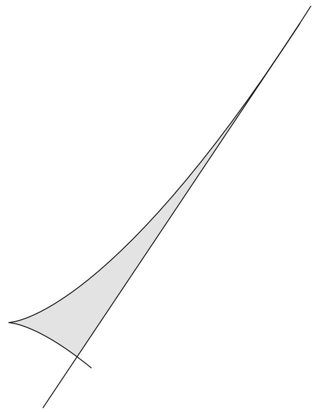


Fig. 8 Cuspidal cubic with tangent line: $(y^2 - x^3)(3x - 2y - 1) = 0$, one metric, curvature 1

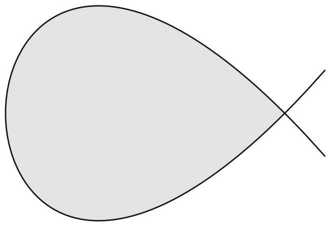


Fig. 9 Nodal cubic: $y^2 - x^2(1-x) = 0$, one metric, non constant curvature

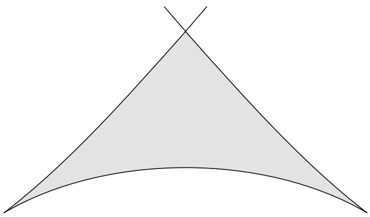


Fig. 10 Swallow tail: $4x^2 - 27x^4 + 16y - 128y^2 - 144x^2y + 256y^3 = 0$, one metric, curvature 1

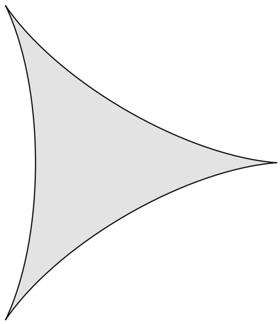


Fig. 11 Deltoid: $(x^2+y^2)^2 + 18(x^2+y^2) - 8x^3 + 24xy^2 - 27 = 0$, one metric, curvature 0

References

1. J. Arvesù, J. Coussement, W. Van Assche, Some discrete multiple orthogonal polynomials, Preprint, www.wis.kuleuven.ac.be/applied/intas/RomeJon.pdf,2001
2. R. Askey, Jacobi polynomials. I. New proofs of Koornwinder's laplace type integral representation and Bateman's bilinear sum. SIAM J. Math. Anal. **5**, 119–124 (1974). MR0385197 (52 #6062)
3. D. Bakry, I. Gentil, M. Ledoux, *Analysis and Geometry of Markov Diffusion Operators*, vol. 348, Grund. Math. Wiss. (Springer, Berlin, 2013)
4. D. Bakry, O. Mazet, *Characterization of Markov Semigroups on R Associated to Some Families of Orthogonal Polynomials., Séminaire de Probabilités XXXVII*, vol. 1832, Lecture Notes in Math. (Springer, Berlin, 2003)
5. D. Bakry, S. Orevkov, M. Zani, Orthogonal polynomials and diffusions operators, preprint, (2013), http://arxiv.org/abs/1309.5632
6. B.L.J. Braaksma, B. Meulenbed, Jacobi polynomials as spherical harmonics. Nederl. Akad. Wetensch. Proc. Ser. A 71=Indag. Math. **30**, 384–389 (1968)
7. C.F. Bracciali, T.E. Pérez, M.A. Piñar, Stieltjes functions and discrete classical orthogonal polynomials. Comput. Appl. Math. **32**(3), 537–547 (2013). MR 3120139
8. H.C. Brinkman, F. Zernike, Hypersphärische funkktionen und die in sphärischen bereichen orthogonalen polynome. Nederl. Akad. Wetensch. Proc. **38**, 161–173 (1935)
9. T.S. Chihara, *Mathematics and its Applications*, vol. 13, An introduction to orthogonal polynomials (Gordon and Breach Science Publishers, New York, 1978). MR 0481884 (58 #1979)
10. D. Damanik, P. Munger, W.N. Yessen, Orthogonal polynomials on the unit circle with fibonacci Verblunsky coefficients, II. applications. J. Stat. Phys. **153**(2), 339–362 (2013). MR 3101200
11. E.B. Davies, *Heat Kernels and Spectral Theory*, vol. 92, Cambridge Tracts in Mathematics (Cambridge University Press, Cambridge, 1990). MR1103113 (92a:35035)
12. C. Dunkl, Y. Xu, *Orthogonal Polynomials of Several Variables*, vol. 81, Encyclopedia of Mathematics and its Applications (Cambridge University Press, Cambridge, 2001)
13. G.J. Heckman, Root sytems and hypergeometric functions i. Compos. Math. **64**, 353–373 (1987)
14. G.J. Heckmann, Dunkl operators, Séminaire Bourbaki 828, 1996–97, vol. Astérisque, SMF, pp. 223–246 (1997)
15. T. Koornwinder, Explicit formulas for special functions related to symmetric spaces. Proc. Symp. Pure Math. **26**, 351–354 (1973)
16. T.H. Koornwinder, Orthogonal polynomials in two variables which are eigenfunctions of two algebraically independent partial differential operators I. Nederl. Akad. Wetensch. Proc. Ser. A 77 = Indag. Math. **36**, 48–58 (1974)
17. T.H. Koornwinder, Orthogonal polynomials in two variables which are eigenfunctions of two algebraically independent partial differential operators II. Nederl. Akad. Wetensch. Proc. Ser. A 77 = Indag. Math. **36**, 59–66 (1974)
18. T.H. Koornwinder, Orthogonal polynomials in two variables which are eigenfunctions of two algebraically independent partial differential operators III. Nederl. Akad. Wetensch. Proc. Ser. A 77 = Indag. Math. **36**, 357–369 (1974)
19. T.H. Koornwinder, Orthogonal polynomials in two variables which are eigenfunctions of two algebraically independent partial differential operators IV. Nederl. Akad. Wetensch. Proc. Ser. A 77 = Indag. Math. **36**, 370–381 (1974)
20. H.L. Krall, I.M. Sheffer, Orthogonal polynomials in two variables. Ann. Mat. Pura Appl. **76**, 325–376 (1967)
21. I.G. Macdonald, Symmetric functions and orthogonal polynomials., University Lecture Series, vol. 12, Am. Math. Soc., Providence, RI (1998)
22. I. Orthogonal, Polynomials associated with root systems., Séminaire Lotharingien de Combinatoire, vol. 45, Université Louis Pasteur, Strasbourg (2000)

23. O. Mazet, Classification des semi-groupes de diffusion sur R associés à une famille de polynômes orthogonaux, Séminaire de probabilités XXXI Lecture Notes in Mathematics (J. Azéma et al, ed.), vol. 1655, Springer, pp. 40–53 (1997)
24. B. Meyer, On the symmetries of spherical harmonics. Canadian J. Math. **6**, 135–157 (1954). MR 0059406 (15,525b)
25. T.O. Sherman, The helgason fourier transform for compact riemannian symmetric spaces of rank one. Acta Math. **164**(1–2), 73–144 (1990)
26. G. Szegő, Orthogonal polynomials, fourth ed., Am. Math. Soc., Providence, R.I., Am. Math. Soc., Colloquium Publications, Vol. XXIII (1975)
27. NTh Varopoulos, L. Saloff-Coste, T. Coulhon, *Analysis and Geometry on Groups*, vol. 100, Cambridge Tracts in Mathematics (Cambridge University Press, Cambridge, 1992)

Cutting Edges at Random in Large Recursive Trees

Erich Baur and Jean Bertoin

Abstract We comment on old and new results related to the destruction of a random recursive tree (RRT), in which its edges are cut one after the other in a uniform random order. In particular, we study the number of steps needed to isolate or disconnect certain distinguished vertices when the size of the tree tends to infinity. New probabilistic explanations are given in terms of the so-called cut-tree and the tree of component sizes, which both encode different aspects of the destruction process. Finally, we establish the connection to Bernoulli bond percolation on large RRT's and present recent results on the cluster sizes in the supercritical regime.

Keywords Random recursive tree · Destruction of graphs · Isolation of nodes · Disconnection · Supercritical percolation · Cluster sizes · Fluctuations

1 Introduction

Imagine that we destroy a connected graph by removing or cutting its edges one after the other, in a uniform random order. The study of such a procedure was initiated by Meir and Moon [31]. They were interested in the number of steps needed to isolate a distinguished vertex in a (random) Cayley tree, when the edges are removed uniformly at random from the current component containing this vertex. Later on, Meir and Moon [32] extended their analysis to random recursive trees. The latter

Acknowledgment of support The research of the first author was supported by the Swiss National Science Foundation grant P2ZHP2_15640.

E. Baur
ENS Lyon, Lyon, France
e-mail: erich.baur@math.uzh.ch

Jean Bertoin (✉)
Universität Zürich, Zürich, Switzerland
e-mail: jean.bertoin@math.uzh.ch

D. Crisan et al. (eds.), *Stochastic Analysis and Applications 2014*,
Springer Proceedings in Mathematics & Statistics 100,
DOI 10.1007/978-3-319-11292-3_3

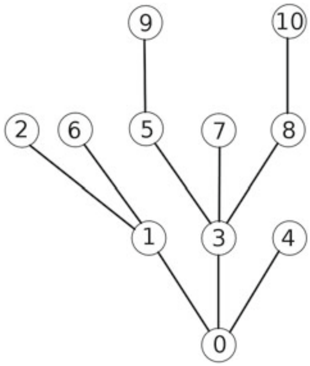


Fig. 1 An increasing tree on the vertex set $\{0, 1, \ldots, 10\}$

form an important family of increasing labeled trees (see Sect. 2 for the definition), and it is the goal of this paper to shed light on issues related to the destruction of such trees.

Mahmoud and Smythe [30] surveyed a multitude of results and applications for random recursive trees. Their recursive structure make them particularly amenable to mathematical analysis, from both a combinatorial and probabilistic point of view. We focus on the probabilistic side. Our main tools include the fundamental splitting property, a coupling due to Iksanov and Möhle [22] and the so-called *cut-tree* (see [9]), which records the key information about the destruction process. The cut-tree allows us to re-prove the results of Kuba and Panholzer [28] on the multiple isolation of nodes. Moreover, we gain information on the number of steps needed to disconnect a finite family nodes.

Finally, we relate the destruction of a random recursive tree to Bernoulli bond percolation on the same tree. We explain some results concerning the sizes of percolation clusters in the supercritical regime, where the root cluster forms the unique giant cluster.

2 Main tools

In this section, we present some basic tools in the study of random recursive trees which will be useful to our purposes.

2.1 The Recursive Construction, Yule Process and Pólya Urn

Consider a finite and totally ordered set of vertices, say V. A tree on V is naturally rooted at the smallest element of V, and is called increasing if and only if the sequence of vertices along a segment from the root to an arbitrary vertex increases (Fig. 1). Most of the time we shall take $V = \{0, 1, \ldots, n\}$, which induces of course no loss of generality. More precisely, it is convenient to introduce the following notion. For an arbitrary totally ordered set V with cardinality $|V| = n + 1$, we call the bijective map from V to $\{0, 1, \ldots, n\}$ which preserves the order, the canonical relabeling of vertices. Plainly the canonical relabeling transforms an increasing tree on V into an increasing tree on $\{0, 1, \ldots, n\}$. Such relabelings enable us to focus on the structure of the rooted tree without retaining specifically the elements of V.

A *random recursive tree* (in short, RRT) on $\{0, 1, \ldots, n\}$ is a tree picked uniformly at random amongst all the increasing trees on $\{0, 1, \ldots, n\}$; it shall be denoted henceforth by T_n. In particular, T_n has n edges and size (i.e. number of vertices) $|T_n| = n+1$. The terminology stems from the easy observation that a version of T_n can be constructed by the following simple recursive random algorithm in which vertices are incorporated one after the other. The vertex 1 is naturally connected by an edge to the root 0, then 2 is connected either to 0 or to 1 with equal probability $1/2$, and more generally, the parent of the vertex i is chosen uniformly at random amongst $0, 1, \ldots, i-1$ and independently of the other vertices. This recursive construction is a close relative to the famous Chinese Restaurant construction of uniform random permutations (see, for instance, Sect. 3.1 in Pitman [34]), and in particular the number of increasing trees of size $n + 1$ equals $n!$

Another useful observation is that this recursive construction can be interpreted in terms of the genealogy of a Yule process. Recall that a Yule process describes the evolution in continuous time of a pure birth process in which each individual gives birth to a child at unit rate and independently of the other individuals. We label individuals in the increasing order of their birth times, the ancestor receiving by convention the label 0. If we let the process evolve until the population reaches size $n + 1$, then its genealogical tree, that is the tree where individuals are viewed as vertices and edges connect children to their parent, is clearly a RRT. Here is an application to percolation on T_n which will be useful later on.

Lemma 1 *Perform a Bernoulli bond percolation on T_n with parameter $0 < p < 1$ (i.e. each edge of T_n is deleted with probability $1 - p$, independently of the other edges), and let $C_n^0(p)$ denote the size of the cluster containing the root. Then*

$$\lim_{n\to\infty} n^{-p} C_n^0(p) = C^0(p) \qquad \textit{in distribution,}$$

where $C^0(p) > 0$ a.s. is some random variable.

Proof We view T_n as the genealogical tree of a standard Yule process $(\mathcal{Y}_s)_{s\geq 0}$ up to time $\rho_n = \inf\{s \geq 0 : \mathcal{Y}_s = n + 1\}$. It is well-known that the process $\mathrm{e}^{-s}\mathcal{Y}_s$ is a martingale which converges a.s. to some random variable W with the exponential distribution, and it follows that

$$\lim_{n\to\infty} n^{-1}\mathrm{e}^{\rho_n} = 1/W \qquad \text{a.s.}$$

In this setting, performing a Bernoulli bond percolation can be interpreted as superposing neutral mutations to the genealogical tree, namely each child is a clone of its parent with probability p and a mutant with a new genetic type with probability $1 - p$, independently of the other children. Neutrality means that the rate of birth does not depend on the genetic type. Then the process $(\mathcal{Y}_s(p))_{s \geq 0}$ of the number of individuals with the same genetic type as the ancestor is again a Yule process, but now with birth rate p. As a consequence

$$\lim_{s \to \infty} \mathrm{e}^{-ps} \mathcal{Y}_s(p) = W(p) \qquad \text{a.s.},$$

where $W(p)$ denotes another exponentially distributed random variable. We then observe that

$$C_n^0(p) = \mathcal{Y}_{\rho_n}(p) \sim W(p)\mathrm{e}^{p\rho_n} \sim W(p)W^{-p}n^p,$$

which completes the proof. □

Plainly, the recursive construction can also be interpreted in terms of urns, and we conclude this section by exemplifying this connection. Specifically, the size of the root cluster $C_n^0(p)$ in the above lemma can be identified as the number of red balls in the following Pólya-Hoppe urn. Start with one red ball which represents the root of the tree. A draw is effected as follows: (i) Choose a ball at random from the urn, observe its color, and put the ball back to the urn. (ii) If its color was red, add a red ball to the urn with probability p, and add a black ball to the urn with probability $1 - p$. If its color was black, add another black ball to the urn. Then, after n draws, the number of red balls is given by $C_n^0(p)$, and in this way, Lemma 1 yields a limit theorem for the proportion of red balls.

The choice $p = 1$ in this urn scheme corresponds to the usual Pólya urn. Here, if one starts with one red ball and k black balls, then the number of red balls after $n - k$ draws is distributed as the size of the subtree T_n^k of a RRT T_n that stems from the vertex k. It is well-known from the theory of Pólya urns that this number follows the beta-binomial distribution with parameters $(n - k, 1, k)$. Moreover,

$$\lim_{n \to \infty} n^{-1}|T_n^k| = \beta(1, k) \qquad \text{in distribution,} \tag{1}$$

where $\beta(1, k)$ is a beta$(1, k)$-distributed random variable. We will use this fact several times below.

2.2 The Splitting Property

The *splitting property* (also called randomness preserving property) reveals the fractal nature of RTT's: roughly speaking, if one removes an edge from a RRT, then the two subtrees resulting from the split are in turn, conditionally on their sizes, independent RRT's. This is of course of crucial importance when investigating the destruction of a RRT, as we can then apply iteratively the splitting property when removing the edges uniformly at random and one after the other.

We select an edge of T_n uniformly at random and remove it. Then T_n splits into two subtrees, say τ_n^0 and τ_n^*, where τ_n^0 contains the root 0. We denote by T_n^0 and T_n^* the pair of increasing trees which then result from the canonical relabelings of the vertices of τ_n^0 and τ_n^*, respectively. Introduce also an integer-valued variable ξ with distribution

$$\mathbb{P}(\xi = j) = \frac{1}{j(j+1)}, \qquad j = 1, 2, \dots \tag{2}$$

Proposition 1 (Meir and Moon [32]) *In the notation above,* $|\tau_n^*| = |T_n^*|$ *has the same law as* ξ *conditioned on* $\xi \le n$*, that is*

$$\mathbb{P}(|T_n^*| = j) = \frac{n+1}{nj(j+1)}, \qquad j = 1, 2, \dots, n.$$

Further, conditionally on $|T_n^*| = j$*,* T_n^0 *and* T_n^* *are two independent RRT's with respective sizes* $n - j + 1$ *and* j*.*

Proof There are $nn!$ configurations $(\mathbf{t}, e)$ given by an increasing tree $\mathbf{t}$ on $\{0, 1, \dots, n\}$ and a distinguished edge e. We remove the edge e and then relabel vertices canonically in each of the resulting subtrees. Let us enumerate the configurations that yield a given pair $(\mathbf{t}^0, \mathbf{t}^*)$ of increasing trees on $\{0, 1, \dots, n-j\}$ and $\{0, 1, \dots, j-1\}$, respectively.

Let $k \in \{0, 1, \dots, n-1\}$ denote the extremity of the edge e which is the closest to the root 0 in $\mathbf{t}$, and V^* the set of vertices which are disconnected from k when e is removed. Since $\mathbf{t}$ is increasing, all the vertices in V^* must be larger than k, and since we want $|V^*| = j$, there are $\binom{n-k}{j}$ ways of choosing V^* (note that this is possible if and only if $k \le n - j$). There are a unique increasing tree structure on V^* and a unique increasing tree structure on $\{0, 1, \dots, n\} \backslash V^*$ that yield respectively $\mathbf{t}^*$ and $\mathbf{t}^0$ after the canonical relabelings.

Conversely, given $\mathbf{t}^0$, $\mathbf{t}^*$, $k \in \{0, 1, \dots, n-j\}$ and $V^* \subset \{k+1, \dots, n\}$ with $|V^*| = j$, there is clearly a unique configuration $(\mathbf{t}, e)$ which yields the quadruple $(k, V^*, \mathbf{t}^0, \mathbf{t}^*)$. Namely, relabeling vertices in $\mathbf{t}^0$ and $\mathbf{t}^*$ produces two increasing tree structures τ^0 and τ^* on $\{0, 1, \dots, n\} \backslash V^*$ and V^*, respectively. We let e denote the edge $(k, \min V^*)$ and then $\mathbf{t}$ is the increasing tree obtained by connecting τ^0 and τ^* using e.

It follows from the analysis above that

$$\mathbb{P}(T_n^0 = \mathbf{t}^0, T_n^* = \mathbf{t}^*) = \frac{1}{nn!} \sum_{k=0}^{n-j} \binom{n-k}{j}.$$

Now recall that

$$\sum_{k=0}^{n-j} \binom{n-k}{j} = \sum_{\ell=j}^{n} \binom{\ell}{j} = \binom{n+1}{j+1}$$

to conclude that

$$\mathbb{P}(T_n^0 = \mathbf{t}^0, T_n^* = \mathbf{t}^*) = \frac{n+1}{n(n-j)!(j+1)!} = \frac{n+1}{nj(j+1)} \times \frac{1}{(n-j)!(j-1)!}.$$

Since there are $(n-j)!$ increasing trees with size $n-j+1$ and $(j-1)!$ increasing trees with size j, this yields the claim. □

Remark It can be easily checked that the splitting property holds more generally when one removes a fixed edge, that is the edge connecting a given vertex $k \in \{1, \dots, n\}$ to its parent. Of course, the distribution of the sizes of the resulting subtrees then changes; see the connection to Pólya urns mentioned in the beginning.

2.3 The Coupling of Iksanov and Möhle

The splitting property was used by Meir and Moon [32] to investigate the following random algorithm for isolating the root 0 of a RRT. Starting from T_n, remove a first edge chosen uniformly at random and discard the subtree which does not contain the root 0. Iterate the procedure with the subtree containing 0 until the root is finally isolated, and denote by X_n the number of steps of this random algorithm. In other words, X_n is the number of random cuts that are needed to isolate 0 in T_n.

Iksanov and Möhle [22] derived from Proposition 1 a useful coupling involving an increasing random walk with step distribution given by (2). Specifically, let $\xi_1, \xi_2, \dots$ denote a sequence of i.i.d. copies of ξ and set $S_0 = 0$,

$$S_n = \xi_1 + \cdots + \xi_n. \tag{3}$$

Further, introduce the last time that the random walk S remains below the level n,

$$L(n) = \max\{k \geq 0 : S_k \leq n\}. \tag{4}$$

Corollary 1 (Iksanov and Möhle [22]) *One can construct on the same probability space a random recursive tree T_n together with the random algorithm of isolation of the root, and a version of the random walk S, such that if*

$$T_{n,0}^0 = T_n \supset T_{n,1}^0 \supset \cdots \supset T_{n,X_n}^0 = \{0\} \tag{5}$$

denotes the nested sequence of the subtrees containing the root induced by the algorithm, then $X_n \geq L(n)$ and

$$(|T^0_{n,0} \backslash T^0_{n,1}|, \ldots, |T^0_{n,L(n)-1} \backslash T^0_{n,L(n)}|) = (\xi_1, \ldots, \xi_{L(n)}). \tag{6}$$

Proof Let us agree for convenience that $T^0_{n,j} = \{0\}$ for every $j > X_n$, and first work conditionally on $(|T^0_{n,i}|)_{i \geq 1}$. Introduce a sequence $((\varepsilon_i, \eta_i))_{i \geq 1}$ of independent pairs of random variables such that for each i, ε_i has the Bernoulli law with parameter $1/|T^0_{n,i-1}| = \mathbb{P}(\xi \geq |T^0_{n,i-1}|)$ and η_i is an independent variable distributed as ξ conditioned on $\xi \geq |T^0_{n,i-1}|$. Then define for every $i \geq 1$

$$\xi_i = \begin{cases} |T^0_{n,i-1}| - |T^0_{n,i}| & \text{if } \varepsilon_i = 0 \\ \eta_i & \text{if } \varepsilon_i = 1 \end{cases}$$

and the partial sums $S_i = \xi_1 + \cdots + \xi_i$. Observe that $\varepsilon_i = 1$ if and only if $\xi_i \geq |T^0_{n,i-1}|$, and hence, by construction, there is the identity

$$\min\{i \geq 1 : \varepsilon_i = 1\} = \min\{i \geq 1 : S_i \geq n+1\}.$$

Therefore, (6) follows if we show that $\xi_1, \xi_2 \ldots$ are (unconditionally) i.i.d. copies of ξ. This is essentially a consequence of the splitting property. Specifically, for $j \leq n$, we have

$$\mathbb{P}(\xi_1 = j) = \mathbb{P}(\varepsilon_1 = 0)\mathbb{P}(n+1-|T^0_{n,1}| = j) = \frac{n}{n+1}\mathbb{P}(|T^*_n| = j) = \frac{1}{j(j+1)},$$

where we used the notation and the result in Proposition 1, whereas for $j > n$ we have

$$\mathbb{P}(\xi_1 = j) = \mathbb{P}(\varepsilon_1 = 1)\mathbb{P}(\xi = j \mid \xi \geq n+1) = \frac{1}{j(j+1)}.$$

Next, consider the conditional law of ξ_2 given ξ_1 and $|T^0_{n,1}|$. Of course, $|T^0_{n,1}| \geq n+1-\xi_1$, and this inequality is in fact an equality whenever $\xi_1 \leq n$. We know from the splitting property that conditionally on its size, say $|T^0_{n,1}| = m+1$ with $m \leq n-1$, $T^0_{n,1}$ is a RRT. Therefore Proposition 1 yields again for $j \leq m$

$$\begin{aligned}
&\mathbb{P}\left(\xi_2 = j \mid \xi_1 \text{ and } |T^0_{n,1}| = m+1\right) \\
&\quad = \mathbb{P}\left(\varepsilon_2 = 0 \mid \xi_1 \text{ and } |T^0_{n,1}| = m+1\right)\mathbb{P}(m+1-|T^0_{n,2}| \\
&\quad = j \mid \xi_1 \text{ and } |T^0_{n,1}| = m+1) \\
&\quad = \frac{m}{m+1}\mathbb{P}(|T^*_m| = j) \\
&\quad = \frac{1}{j(j+1)}.
\end{aligned}$$

Similarly for $j > m$

$$\begin{aligned}
&\mathbb{P}(\xi_2 = j \mid \xi_1 \text{ and } |T^0_{n,1}| = m+1) \\
&\qquad = \mathbb{P}(\varepsilon_2 = 1 \mid \xi_1 \text{ and } |T^0_{n,1}| = m+1)\mathbb{P}(\xi = j \mid \xi \geq m+1) \\
&\qquad = \frac{1}{j(j+1)}.
\end{aligned}$$

This shows that ξ_2 has the same distribution as ξ and is independent of ξ_1 and $|T^0_{n,1}|$. Iterating this argument, we get that the ξ_i form a sequence of i.i.d. copies of ξ, which completes the proof. □

3 The Number of Random Cuts Needed to Isolate the Root

Recall the algorithm of isolation of the root which was introduced in the preceding section, and recall that X_n denotes its number of steps for T_n, i.e. X_n is the number of random cuts that are needed to isolate the root 0 in T_n. Meir and Moon [32] used Proposition 1 to investigate the first two moments of X_n and showed that

$$\lim_{n\to\infty} \frac{\ln n}{n} X_n = 1 \qquad \text{in probability.} \tag{7}$$

The problem of specifying the fluctuations of X_n was left open until the work by Drmota et al., who obtained the following remarkable result.

Theorem 1 (Drmota, Iksanov, Möhle and Rösler [16]) *As $n \to \infty$,*

$$\frac{\ln^2 n}{n} X_n - \ln n - \ln \ln n$$

converges in distribution to a completely asymmetric Cauchy variable X with characteristic function

$$\mathbb{E}(\exp(itX)) = \exp\left(it \ln|t| - \frac{\pi}{2}|t|\right), \qquad t \in \mathbb{R}. \tag{8}$$

In short, the starting point of the proof in [16] is the identity in distribution

$$X_n \overset{\text{(d)}}{=} 1 + X_{n-D_n}, \tag{9}$$

where D_n is a random variable with the law of ξ given $\xi \leq n$, and D_n is assumed to be independent of $X_1, \ldots, X_n$. More precisely, (9) derives immediately from the splitting property (Proposition 1). Drmota et al. deduce from (9) a PDE for the generating function of the variables X_n, and then singularity analysis provides the key tool for investigating the asymptotic behavior of this generating function and elucidating the asymptotic behavior of X_n.

Iksanov and Möhle [22] developed an elegant probabilistic argument which explains the unusual rescaling and the Cauchy limit law in Theorem 1. We shall now sketch this argument.

Sketch proof of Theorem 1: One starts observing that the distribution in (2) belongs to the domain of attraction of a completely asymmetric Cauchy variable X whose law is determined by (8), namely

$$\lim_{n\to\infty}\left(n^{-1}S_n - \ln n\right) = -X \qquad \text{in distribution.} \tag{10}$$

Then one deduces from (10) that the asymptotic behavior of the last-passage time (4) is given by

$$\lim_{n\to\infty}\left(\frac{\ln^2 n}{n}L(n) - \ln n - \ln\ln n\right) = X \qquad \text{in distribution,} \tag{11}$$

see Proposition 2 in [22]. This limit theorem resembles of course Theorem 1, and the relation between the two is explained by the coupling of the algorithm of isolation of the root and the random walk S stated in Corollary 1, as we shall now see.

Let the algorithm for isolating the root run for $L(n)$ steps. Then the size of the remaining subtree that contains the root is $n+1-S_{L(n)}$, and as a consequence, there are the bounds

$$L(n) \le X_n \le L(n) + n - S_{L(n)},$$

since at most $\ell - 1$ edge removals are needed to isolate the root in any tree of size ℓ. On the other hand, specializing a renewal theorem of Erickson [17] for the increasing random walk S, one gets that

$$\lim_{n\to\infty} \ln(n - S_{L(n)})/\ln n = U \qquad \text{in distribution,}$$

where U is a uniform [0, 1] random variable. In particular

$$\lim_{n\to\infty}\frac{\ln^2 n}{n}(n - S_{L(n)}) = 0 \qquad \text{in probability.}$$

Thus Theorem 1 follows from (11). □

It should be noted that there exists a vertex version of the isolation algorithm, where one chooses a vertex at random and destroys it together with its descending subtree. The algorithm continues until the root is chosen. Using an appropriate coupling with X_n, one readily shows that the number of random vertex removals $X_n^{(v)}$ needed to destroy a RRT T_n satisfies $(X_n - X_n^{(v)}) = o(n/\ln^2 n)$ in probability. Henceforth, we concentrate on cutting edges.

Remark Weak limit theorems for the number of cuts to isolate the root vertex have also been obtained for other tree models, like conditioned Galton-Watson trees including e.g. uniform Cayley trees and random binary trees (Panholzer [33] and, in greater generality, Janson [24]), deterministic complete binary trees (Janson [23]) and random split trees (Holmgren [20, 21]). More generally, Addario-Berry et al. [1] and Bertoin [6] found the asymptotic limit distribution for the number of cuts required to isolate a fixed number $\ell \geq 1$ of vertices picked uniformly at random in a uniform Cayley tree. This result was further extended by Bertoin and Miermont [11] to conditioned Galton-Watson trees. We point to the remark after Corollary 3 for more on this. Turning back to RRT's, recent generalizations of Theorem 1 were found first by Kuba and Panholzer [27, 28] and then by Bertoin [9], some of which will be discussed in the reminder of this paper.

In [28], Kuba and Panholzer considered the situation when one wishes to isolate the *first* ℓ vertices of a RRT T_n, $0, 1, \dots, \ell-1$, where $\ell \geq 1$ is fixed. In this direction, one modifies the algorithm of isolation of the sole root in an obvious way. A first edge picked uniformly at random in T_n is removed. If one of the two resulting subtrees contains none of the vertices $0, 1, \dots, \ell-1$, then it is discarded forever. Else, the two subtrees are kept. In both cases, one iterates until each and every vertex $0, 1, \dots, \ell-1$ has been isolated, and we write $X_{n,\ell}$ for the number of steps of this algorithm.

The approach of Kuba and Panholzer follows analytic methods similar to the original proof of Theorem 1 by Drmota et al. [16]. We point out here that the asymptotic behavior of $X_{n,\ell}$ can also be deduced from Theorem 1 by a probabilistic argument based on the following elementary observation, which enables us to couple the variables $X_{n,\ell}$ for different values of ℓ. Specifically, we run the usual algorithm of isolation of the root, except that now, at each time when a subtree becomes disconnected from the root, we keep it aside whenever it contains at least one of the vertices $1, \dots, \ell-1$, and discard it forever otherwise. Once the root 0 of T_n has been isolated, we resume with the subtree containing 1 which was set aside, meaning that we run a further algorithm on that subtree until its root 1 has been isolated, keeping aside the further subtrees disconnected from 1 which contain at least one of the vertices $2, \dots, \ell-1$. We then continue with the subtree containing the vertex 2, and so on until each and every vertex $0, 1, \dots, \ell-1$ has been isolated. If we write $X'_{n,\ell}$ for the number of steps of this algorithm, then it should be plain that $X'_{n,\ell}$ has the same law as $X_{n,\ell}$, and further $X_n = X'_{n,1} \leq \dots \leq X'_{n,\ell}$.

We shall now investigate the asymptotic behavior of the increments $\Delta_{n,i} = X'_{n,i+1} - X'_{n,i}$ for $i \geq 1$ fixed. In this direction, suppose that we now remove the edges of T_n one after the other in a uniform random order until the edge connecting the vertex i to its parent is removed. Let τ_n^i denote the subtree containing i that arises at this step.

Lemma 2 *For each fixed* $i \geq 1$,

$$\lim_{n\to\infty} \frac{\ln |\tau_n^i|}{\ln n} = U \qquad \textit{in distribution},$$

where U *is a uniform* $[0, 1]$ *random variable.*

For the moment, let us take Lemma 2 for granted and deduce the following.

Corollary 2 *We have*

$$\lim_{n\to\infty} \frac{\ln \Delta_{n,i}}{\ln n} = U \qquad \textit{in distribution},$$

where U is a uniform [0, 1] *random variable.*

Proof Just observe that $\Delta_{n,i}$ has the same law as the number of cuts needed to isolate the root i of τ_n^i, and recall from an iteration of the splitting property that conditionally on its size, τ_n^i is a RRT. Our statement now follows readily from (7) and Lemma 2. □

Writing $X'_{n,\ell} = X_n + \Delta_{n,1} + \cdots + \Delta_{n,\ell-1}$, we now see from Theorem 1 and Corollary 2 that for each fixed $\ell \geq 1$, there is the weak convergence

$$\lim_{n\to\infty} \left(\frac{\ln^2 n}{n} X'_{n,\ell} - \ln n - \ln\ln n \right) = X \qquad \text{in distribution,} \tag{12}$$

which is Theorem 1 in [28]. We now proceed to the proof of Lemma 2.

Proof Let T_n^i denote the subtree of T_n that stems from the vertex i, and equip each edge e of T_n with a uniform [0, 1] random variable U_e, independently of the other edges. Imagine that the edge e is removed at time U_e, and for every time $0 \leq s \leq 1$, write $T_n^i(s)$ for the subtree of T_n^i which contains i at time s. Hence, if we write $U = U_e$ for e the edge connecting i to its parent, then $\tau_n^i = T_n^i(U)$. Further, since U is independent of the other uniform variables, conditionally on U and T_n^i, τ_n^i can be viewed as the cluster that contains the root vertex i after a Bernoulli bond percolation on T_n^i with parameter $1 - U$. Thus, conditionally on $|T_n^i| = m + 1$ and $U = 1 - p$, $|\tau_n^i|$ has the same law as $C_m^0(p)$ in the notation of Lemma 1.

From (1) we know that $n^{-1}|T_n^i|$ converges in distribution as $n \to \infty$ to a beta variable with parameters $(1, i)$, say β, which is of course independent of U. On the other hand, conditionally on its size, and after the usual canonical relabeling of its vertices, T_n^i is also a RRT (see the remark at the end of Sect. 2). It then follows from Lemma 1 that

$$\lim_{n\to\infty} \frac{\ln |\tau_n^i|}{\ln n} = 1 - U \qquad \text{in probability,}$$

which establishes our claim. □

4 The Destruction Process and Its Tree Representations

Imagine now that we remove the edges of T_n one after the other and in a uniform random order, no matter whether they belong to the root component or not. We call this the *destruction process* of T_n. After n steps, no edges are present anymore and all the vertices have been isolated. In particular, the random variable which counts only the number of edge removals from the root component can be identified with X_n from the previous section.

The purpose of this section is to introduce and study the asymptotic behavior of two trees which can be naturally associated to this destruction process, namely the tree of component sizes and the cut-tree. Furthermore, we give some applications of the cut-tree to the isolation and disconnection of nodes and comment on ordered destruction of a RRT.

4.1 The Tree of Component Sizes

In this part, we are interested in the sizes of the tree components produced by the destruction process. Our analysis will also prove helpful for studying percolation clusters of a RRT in Sect. 5.

The component sizes are naturally stored in a tree structure. As our index set, we use the universal tree

$$\mathcal{U} = \bigcup_{k=0}^{\infty} \mathbb{N}^k,$$

with the convention $\mathbb{N}^0 = \{\emptyset\}$ and $\mathbb{N} = \{1, 2, \dots\}$. In particular, an element $u \in \mathcal{U}$ is a finite sequence of strictly positive integers $(u_1, \dots, u_k)$, and its length $|u| = k$ represents the "generation" of u. The jth child of u is given by $uj = (u_1, \dots, u_k, j)$, $j \in \mathbb{N}$. The empty sequence $\emptyset$ is the root of the tree and has length $|\emptyset| = 0$. If no confusion occurs, we drop the separating commas and write $(u_1 \dots u_k)$ or simply $u_1, \dots, u_k$ instead of $(u_1, \dots, u_k)$. Also, $\emptyset u$ represents the element u.

We define a tree-indexed process $\mathcal{B}^{(n)} = (\mathcal{B}_u^{(n)} : u \in \mathcal{U})$, which encodes the sizes of the tree components stemming from the destruction of T_n. We will directly identify a vertex u with its label $\mathcal{B}_u^{(n)}$. Following the steps of the destruction process, we build this process dynamically starting from the singleton $\mathcal{B}_\emptyset^{(n)} = n + 1$ and ending after n steps with the full process $\mathcal{B}^{(n)}$. More precisely, when the first edge of T_n is removed in the destruction process, T_n splits into two subtrees, say τ_n^0 and τ_n^*, where τ_n^0 contains the root 0. We stress that τ_n^0 is naturally rooted at 0 and τ_n^* at its smallest vertex. The size $|\tau_n^*|$ is viewed as the first child of $\mathcal{B}_\emptyset^{(n)}$ and denoted by $\mathcal{B}_1^{(n)}$. Now first suppose that the next edge which is removed connects two vertices in τ_n^*. Then, τ_n^* splits into two tree components. The size of the component not containing the root of τ_n^* is viewed as the first child of $\mathcal{B}_1^{(n)}$ and denoted by $\mathcal{B}_{11}^{(n)}$. On the other hand, if the second edge which is removed connects two vertices in τ_n^0, then the size of the component not containing 0 is viewed as the second child of $\mathcal{B}_\emptyset^{(n)}$ and denoted by $\mathcal{B}_2^{(n)}$. It should now be plain how to iterate this construction. After n steps, we have in this way defined $n + 1$ variables $\mathcal{B}_u^{(n)}$ with $|u| \le n$, and we extend the definition to the full universal tree by letting $\mathcal{B}_u^{(n)} = 0$ for all the remaining $u \in \mathcal{U}$. We refer to Fig. 2 for an example. The tree components whose sizes are encoded by the elements $\mathcal{B}_u^{(n)}$ with $|u| = k$ are called the components of generation k.

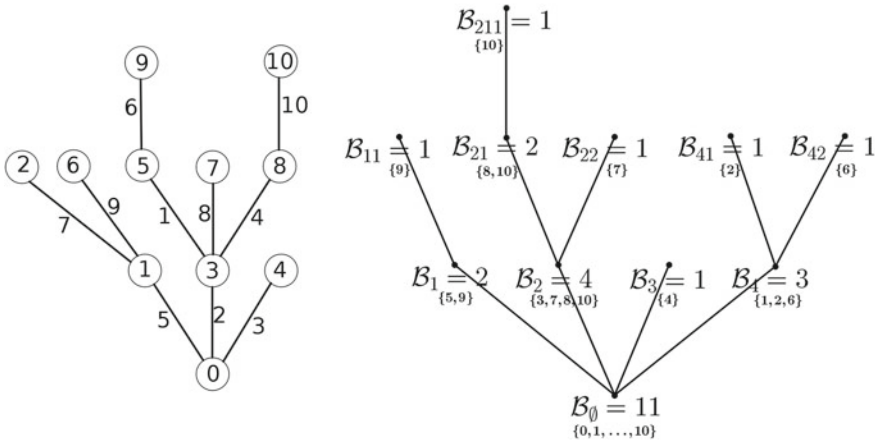


Fig. 2 *Left* A recursive tree with vertices labeled 0,1,…,10. The labels on the edges indicate the order in which they are removed by the destruction process. *Right* The corresponding tree of component sizes, with the vertex sets of the tree components. The elements $\mathcal{B}_u^{(n)}$ of size 0 are omitted

To sum up, every time an edge is removed in the destruction process, a tree component τ_n splits into two subtrees, and we adjoin the size of the subtree which does not contain the root of τ_n as a new child to the vertex representing τ_n. Note that the root $\mathcal{B}_{\emptyset}^{(n)}$ has X_n many nontrivial children, and they represent the sizes of the tree components which were cut from the root one after the other in the algorithm for isolating the root.

We now interpret $\mathcal{B}^{(n)}$ as the genealogical tree of a multi-type population model, where the type reflects the size of the tree component (and thus takes integer values). In particular the ancestor $\emptyset$ has type $n+1$; furthermore, a node u with $\mathcal{B}_u^{(n)} = 0$ corresponds to an empty component and is therefore absent in the population model. We also stress that the type of an individual is always given by the sum of the types of its children plus 1. As a consequence, types can be recovered from the sole structure of the genealogical tree. More precisely, the type of an individual is simply given by the total size of the subtree of the genealogical tree stemming from that individual.

The splitting property of a RRT immediately transfers into a branching property for this population model.

Lemma 3 *The population model induced by the tree of component sizes $\mathcal{B}^{(n)}$ is a multi-type Galton-Watson process starting from one particle of type $n+1$. The reproduction distribution λ_i of an individual of type $i \geq 1$ is given by the law of the sequence of the sizes of the non-root subtrees which are produced in the algorithm for isolating the root of a RRT of size i.*

Even though the coupling of Iksanov and Möhle is not sufficient to fully describe the reproduction law, it nonetheless provides essential information on λ_i in terms of a sequence of i.i.d. copies of ξ. As we will see next, extreme value theory for the i.i.d. sequence then enables us to specify asymptotics of the population model when the type $n+1$ of the ancestor goes to infinity.

To give a precise statement, we rank the children of each individual in the decreasing order of their types. Formally, given the individual indexed by $u \in \mathcal{U}$ has exactly ℓ children of type ≥ 1, we let σ_u be the random permutation of $\{1, \ldots, \ell\}$ which sorts the sequence of types $\mathcal{B}^{(n)}_{u1}, \ldots, \mathcal{B}^{(n)}_{u\ell}$ in the decreasing order, i.e.

$$\mathcal{B}^{(n)}_{u\sigma_u(1)} \geq \mathcal{B}^{(n)}_{u\sigma_u(2)} \geq \ldots \geq \mathcal{B}^{(n)}_{u\sigma_u(\ell)},$$

where in the case of ties, children of the same type are ranked uniformly at random. We extend σ_u to a bijection $\sigma_u : \mathbb{N} \to \mathbb{N}$ by putting $\sigma_u(i) = i$ for $i > \ell$.

We then define the global random bijection $\sigma = \sigma^{(n)} : \mathcal{U} \to \mathcal{U}$ recursively by setting $\sigma(\emptyset) = \emptyset$, $\sigma(j) = \sigma_\emptyset(j)$, and then, given $\sigma(u)$, $\sigma(uj) = \sigma(u)\sigma_{\sigma(u)}(j)$, $u \in \mathcal{U}$, $j \in \mathbb{N}$. Note that σ preserves the parent-child relationship, i.e. children of u are mapped into children of $\sigma(u)$. We simply write $(\mathcal{B}^{(n)\downarrow}_u : u \in \mathcal{U}) = (\mathcal{B}^{(n)}_{\sigma(u)} : u \in \mathcal{U})$ for the process which is ranked in this way.

Now, if the sizes of the components of generation k are normalized by a factor $\ln^k n/n$, we obtain finite-dimensional convergence of $\mathcal{B}^{(n)\downarrow}$ towards the genealogical tree of a continuous-state branching process with reproduction measure $\nu(da) = a^{-2}\mathrm{d}a$ on $(0, \infty)$. More precisely, the limit object is a tree-indexed process $\mathcal{Z} = (\mathcal{Z}_u : u \in \mathcal{U})$ with initial state $\mathcal{Z}_\emptyset = 1$, whose distribution is characterized by induction on the generations as follows.

(a) $\mathcal{Z}_\emptyset = 1$ almost surely;
(b) for every $k = 0, 1, 2, \ldots$, conditionally on $(\mathcal{Z}_v : v \in \mathcal{U}, |v| \leq k)$, the sequences $(\mathcal{Z}_{uj})_{j\in\mathbb{N}}$ for the vertices $u \in \mathcal{U}$ at generation $|u| = k$ are independent, and each sequence $(\mathcal{Z}_{uj})_{j\in\mathbb{N}}$ is distributed as the family of the atoms of a Poisson random measure on $(0, \infty)$ with intensity $\mathcal{Z}_u\nu$, where the atoms are ranked in the decreasing order of their sizes.

Proposition 2 *As $n \to \infty$, there is the convergence in the sense of finite-dimensional distributions,*

$$\mathcal{Z}^{(n)} = \left(\frac{(\ln n)^{|u|}}{n}\mathcal{B}^{(n)\downarrow}_u : u \in \mathcal{U}\right) \Longrightarrow \mathcal{Z}.$$

We only sketch the proof and refer to the forthcoming paper [5] for details. Basically, if $\xi_1, \xi_2, \ldots$ is a sequence of of i.i.d. copies of ξ, then for $a > 0$, the number of indices $j \leq k$ such that $\xi_j \geq an/\ln n$ is binomially distributed with parameters k and $\lceil an/\ln n\rceil^{-1}$. From (11) and Theorem 16.16 in [25] we deduce that for a fixed integer j, the j largest among $\xi_1, \ldots, \xi_{L(n)}$, normalized by a factor $\ln n/n$, converge in distribution to the j largest atoms of a Poisson random measure on $(0, \infty)$ with intensity $\nu(da) = a^{-2}\mathrm{d}a$. Since $n - S_{L(n)} = o(n/\ln^2 n)$ in probability, finite-dimensional convergence of $\mathcal{Z}^{(n)}$ restricted to generations ≤ 1 then follows from (6). Lemma 3 enables us to transport the arguments to the next generations.

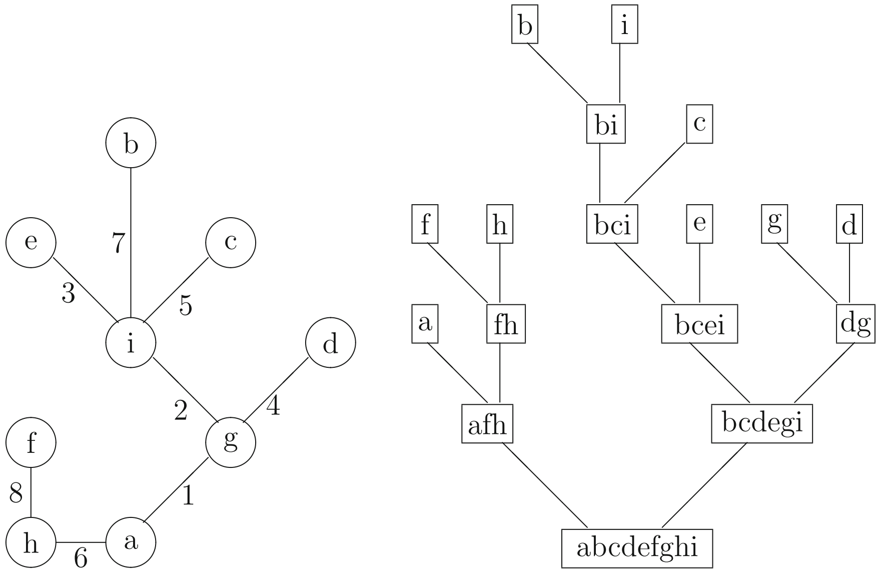


Fig. 3 *Left* Tree T with vertices labeled a,…,i; edges are enumerated in the order of the cuts. *Right* Cut-tree $\mathrm{Cut}(T)$ on the set of blocks recording the destruction of T

4.2 The Cut-Tree

Consider for a while a deterministic setting where T is an arbitrary tree on some finite set of vertices V. Imagine that its edges are removed one after the other in some given order, so at the end of the process, all the vertices of T have been disconnected from each other. We shall encode the destruction of T by a rooted binary tree, which we call the cut-tree and denote by $\mathrm{Cut}(T)$. The cut-tree has internal nodes given by the non-singleton connected components which arise during the destruction, and leaves which correspond to the singletons and which can thus be identified with the vertices in V. More precisely, the root of $\mathrm{Cut}(T)$ is given by V, and when the first edge of T is removed, disconnecting V into, say, V_1 and V_2, then V_1 and V_2 are viewed as the two children of V and thus connected to V by a pair of edges. Suppose that the next edge which is removed connects two vertices in V_1, so removing this second edge disconnects V_1 into, say $V_{1,1}$ and $V_{1,2}$. Then $V_{1,1}$ and $V_{1,2}$ are viewed in turn as the two children of V_1. We iterate in an obvious way, see Fig. 3 for an example.[1]

It should be clear that the number of cuts required to isolate a given vertex v in the destruction of T (as previously, we only count the cuts occurring in the component

[1] For the sake of simplicity, this notation does not record the order in which the edges are removed, although the latter is of course crucial in the definition of the cut-tree. In this part, we are concerned with uniform random edge removal, while in the last part of this section, we look at ordered destruction of a RRT, where edges are removed in the order of their endpoints most distant from the root.

which contains v) corresponds precisely to the height of the leaf $\{v\}$ in $\mathrm{Cut}(T)$. More generally, the number of cuts required to isolate k distinct vertices $v_1, \dots, v_k$ coincides with the total length of the cut-tree reduced to its root and the k leaves $\{v_1\}, \dots, \{v_k\}$ minus $(k-1)$, where the length is measured as usual by the graph distance on $\mathrm{Cut}(T)$. In short, the cut-tree encapsulates all the information about the numbers of cuts needed to isolate any subset of vertices.

We now return to our usual setting, that is T_n is a RRT of size $n+1$, whose edges are removed in a uniform random order, and we write $\mathrm{Cut}(T_n)$ for the corresponding cut-tree. We point out that the genealogical tree of component sizes which was considered in the previous section can easily be recovered from $\mathrm{Cut}(T_n)$. Specifically, the root $\{0, 1, \dots, n\}$ of $\mathrm{Cut}(T_n)$ has to be viewed as the ancestor of the population model, its type is of course $n+1$. Then the blocks of $\mathrm{Cut}(T_n)$ which are connected by an edge to the segment from the root $\{0, 1, \dots, n\}$ to the leaf $\{0\}$ are the children of the ancestor in the population model, the type of a child being given by the size of the corresponding block. The next generations of the population model are then described similarly by an obvious iteration.

The segment of $\mathrm{Cut}(T_n)$ from its root $\{0, 1, \dots, n\}$ to the leaf $\{0\}$ is described by the nested sequence (5), and the coupling of Iksanov and Möhle stated in Corollary 1 expresses the sequence of the block-sizes along the portion of this segment starting from the root and with length $L(n)$, in terms of the random walk S. We shall refer to this portion as the trunk of $\mathrm{Cut}(T_n)$ and denote it by $\mathrm{Trunk}(T_n)$. The connected components of the complement of the trunk, $\mathrm{Cut}(T_n)\backslash\mathrm{Trunk}(T_n)$ are referred to as the branches of $\mathrm{Cut}(T_n)$.

Roughly speaking, it has been shown in [9] that upon rescaling the graph distance of $\mathrm{Cut}(T_n)$ by a factor $n^{-1}\ln n$, the latter converges to the unit interval. The precise mathematical statement involves the notion of convergence of pointed measured metric spaces in the sense of the Gromov-Hausdorff-Prokhorov distance.

Theorem 2 *Endow* $\mathrm{Cut}(T_n)$ *with the uniform probability measure on its leaves, and normalize the graph distance by a factor* $n^{-1}\ln n$*. As* $n \to \infty$*, the latter converges in probability in the sense of the pointed Gromov-Hausdorff-Prokhorov distance to the unit interval* $[0, 1]$ *equipped with the usual distance and the Lebesgue measure, and pointed at* 0.

Providing the background on the Gromov-Hausdorff-Prokhorov distance needed to explain rigorously the meaning of Theorem 2 would probably drive us too far away from the purpose of this survey, so we shall content ourselves here to give an informal explanation. After the rescaling, each edge of $\mathrm{Cut}(T_n)$ has length $n^{-1}\ln n$, and it follows from (11) that the length $n^{-1}\ln n \times L(n)$ of $\mathrm{Trunk}(T_n)$ converges in probability to 1 as $n \to \infty$. Because the trunk is merely a segment, if we equip it with the uniform probability measure on its nodes, then we obtain a space close to the unit interval endowed with the Lebesgue measure. The heart of the argument of the proof in [9] is to observe that in turn, $\mathrm{Trunk}(T_n)$ is close to $\mathrm{Cut}(T_n)$ when n is large, both in the sense of Hausdorff and in the sense of Prokhorov. First, as $\mathrm{Trunk}(T_n)$ is a subset of $\mathrm{Cut}(T_n)$, the Hausdorff distance between $\mathrm{Trunk}(T_n)$ and

$\mathrm{Cut}(T_n)$ corresponds to the maximal depth of the branches of $\mathrm{Cut}(T_n)$, and one thus have to verify that all the branches are small (recall that the graph distance has been rescaled by a factor $n^{-1}\ln n$). Then, one needs to check that the uniform probability measures, respectively on the set of leaves of $\mathrm{Cut}(T_n)$ and on the nodes of $\mathrm{Trunk}(T_n)$, are also close to each other in the sense of the Prokhorov distance between probability measures on a metric space. This is essentially a consequence of the law of large numbers for the random walk defined in (3), namely

$$\lim_{n\to\infty} \frac{S_n}{n\ln n} = 1 \qquad \text{in probability;}$$

see (10).

4.3 Applications

Theorem 2 enables us to specify the asymptotic behavior of the number of cuts needed to isolate randomly chosen vertices of T_n. For a given integer $\ell \geq 1$ and for each $n \geq 1$, let $U_1^{(n)}, \dots, U_\ell^{(n)}$ denote a sequence of i.i.d. uniform variables in $\{0, 1, \dots, n\}$. We write $Y_{n,\ell}$ for the number of random cuts which are needed to isolate $U_1^{(n)}, \dots, U_\ell^{(n)}$. The following corollary, which is taken from [9], is a multi-dimensional extension of Theorem 3 of Kuba and Panholzer [28].

Corollary 3 *As $n \to \infty$, the random vector*

$$\left(\frac{\ln n}{n} Y_{n,1}, \dots, \frac{\ln n}{n} Y_{n,\ell}\right)$$

converges in distribution to

$$(U_1, \max\{U_1, U_2\}, \dots, \max\{U_1, \dots, U_\ell\}),$$

where $U_1, \dots, U_\ell$ are i.i.d. uniform $[0, 1]$ *random variables. In particular, $\frac{\ln n}{n} Y_{n,\ell}$ converges in distribution to a beta$(\ell, 1)$ variable.*

Proof Recall that $U_1^{(n)}, \dots, U_\ell^{(n)}$ are ℓ independent uniform vertices of T_n. Equivalently, the singletons $\{U_1^{(n)}\}, \dots, \{U_\ell^{(n)}\}$ form a sequence of ℓ i.i.d. leaves of $\mathrm{Cut}(T_n)$ distributed according to the uniform law. Let also $U_1, \dots, U_l$ be a sequence of ℓ i.i.d. uniform variables on $[0, 1]$. Denote by $\mathcal{R}_{n,\ell}$ the reduction of $\mathrm{Cut}(T_n)$ to the ℓ leaves $\{U_1^{(n)}\}, \dots, \{U_\ell^{(n)}\}$ and its root $\{0, 1, \dots, n\}$, i.e. $\mathcal{R}_{n,\ell}$ is the smallest subtree of $\mathrm{Cut}(T_n)$ which connects these nodes. Similarly, write $\mathcal{R}_\ell$ for the reduction of I to $U_1, \dots, U_\ell$ and the origin 0. Both reduced trees are viewed as combinatorial trees structures with edge lengths, and Theorem 2 entails that $n^{-1}\ln n\mathcal{R}_{n,\ell}$ converges in distribution to $\mathcal{R}_\ell$ as $n \to \infty$. In particular, focusing on the lengths of those reduced trees, there is the weak convergence

$$\lim_{n\to\infty}\left(\frac{\ln n}{n}|\mathcal{R}_{n,1}|,\ldots,\frac{\ln n}{n}|\mathcal{R}_{n,\ell}|\right)=(|\mathcal{R}_1|,\ldots,|\mathcal{R}_\ell|)\qquad\text{in distribution.}\quad(13)$$

This yields our claim, as plainly $|\mathcal{R}_i| = \max\{U_1,\ldots,U_i\}$ for every $i = 1,\ldots,\ell$. □

Remark The nearly trivial proof of this corollary exemplifies the power of Theorem 2, and one might ask for convergence of the cut-tree for other tree models. In fact, employing the work of Haas and Miermont [19], it has been shown in [6] that if $T_n^{(c)}$ is a uniform Cayley tree of size n, then $n^{-1/2}\mathrm{Cut}(T_n^{(c)})$ converges weakly in the sense of Gromov-Hausdorff-Prokhorov to the Brownian Continuum Random Tree (CRT), see Aldous [2]. Since the total length of the CRT reduced to the root and ℓ i.i.d leaves picked according to its mass-measure follows the $\mathrm{Chi}(2\ell)$-distribution, one readily obtains the statement corresponding to Corollary 3 for uniform Cayley trees ([6] and also, by different means, [1]). Bertoin and Miermont [11] extended the convergence of the cut-tree towards the CRT to the full family of critical Galton-Watson trees with finite variance and conditioned to have size n, in the sense of Gromov-Prokhorov. As a corollary, one obtains a multi-dimensional extension of Janson's limit theorem [24]. Very recently, Dieuleveut [14] proved the analog of [11] for the case of Galton-Watson trees with offspring distribution belonging to the domain of attraction of a stable law of index $\alpha \in (1, 2)$.

With Corollary 3 at hand, we can also study the number $Z_{n,\ell}$ of random cuts which are needed to isolate the ℓ *last* vertices of T_n, i.e. $n-\ell+1,\ldots,n$, where $\ell \geq 1$ is again a given integer. As Kuba and Panholzer [28] proved in their Theorem 2, $Z_{n,\ell}$ has the same asymptotic behavior in law as $Y_{n,\ell}$. The following multi-dimensional version was given in [9], relying on Theorem 2 of [28]. Here we give a self-contained proof of the same statement.

Corollary 4 *As $n \to \infty$, the random vector*

$$\left(\frac{\ln n}{n}Z_{n,1},\ldots,\frac{\ln n}{n}Z_{n,\ell}\right)$$

converges in distribution to

$$(U_1, \max\{U_1,U_2\},\ldots,\max\{U_1,\ldots,U_\ell\}),$$

where $U_1,\ldots,U_\ell$ are i.i.d. uniform $[0, 1]$ *random variables.*

Proof For ease of notation, we consider only the case $\ell = 1$, the general case being similar. The random variable $Z_n = Z_{n,1}$ counts the number of random cuts needed to isolate the vertex n, which is a leaf of T_n. If we write v for the parent of n in T_n, then v is uniformly distributed on $\{0, 1,\ldots,n-1\}$, and it follows that the number Y_n' of cuts needed to isolate v has the same limit behavior in law as $Y_{n-1,1}$. In view of Corollary 3, it suffices therefore to verify that

$$\lim_{n\to\infty} \frac{\ln n}{n} \left(Y_n' - Z_n\right) = 0 \qquad \text{in probability.}$$

We now consider the algorithm for isolating the vertex v. Clearly, the number of steps of this algorithm until the edge e joining v to n is removed is distributed as Z_n. In particular, we obtain a natural coupling between Y_n' and Z_n with $Z_n \leq Y_n'$. Denote by $[0; n]$ the segment of T_n from the root 0 to the leaf n, and write k for the outer endpoint of the first edge from $[0; n]$ which is to be removed by the isolation algorithm. Since $|[0; n]| \sim \ln n$ in probability (see e.g. Theorem 6.17 of [15]), and since the isolation algorithm chooses its edges uniformly at random, the probability that k is equal to n tends to zero. Moreover, with high probability $|[k; n]|$ will still be larger than $(\ln n)^{1/2}$, say. By conditioning on k and repeating the above argument with $[k; n]$ in place of $[0; n]$, we see that we can concentrate on the event that before n is isolated, at least two edges different from e are removed from the segment $[0; n]$. On this event, after the second time an edge from $[0; n]$ is removed, the vertices v and n lie in a tree component which can be interpreted as a tree component of the second generation in the destruction process. As a consequence of Proposition 2, the size of this tree component multiplied by factor $\ln n/n$ converges to zero in probability. Since the size of the component gives an upper bound on the difference $Y_n' - Z_n$, the claim follows. □

As another application of the cut-tree, Theorem 2 allows us to determine the number of cuts $A_{n,\ell}$ which are required to *disconnect* (and not necessarily isolate) $\ell \geq 2$ vertices in T_n chosen uniformly at random. For ease of description, let us assume that the sequence of vertices $U_1^{(n)}, \ldots, U_\ell^{(n)}$ is chosen uniformly at random in $\{0, 1, \ldots, n\}$ without replacement. Note that in the limit $n \to \infty$, it makes no difference whether we sample with or without replacement.

We run the algorithm for isolating the vertices $U_1^{(n)}, \ldots, U_\ell^{(n)}$, with the modification that we discard emerging tree components which contain at most one of these ℓ vertices. We stop the algorithm when $U_1^{(n)}, \ldots, U_\ell^{(n)}$ are totally disconnected from each other, i.e. lie in ℓ different tree components. Write $A_{n,2}$ for the (random) number of steps of this algorithm until for the first time, the vertices $U_1^{(n)}, \ldots, U_\ell^{(n)}$ do no longer belong to the same tree component, further $A_{n,3}$ for the number of steps until for the first time, the ℓ vertices are spread out over three distinct tree components, and so on, up to $A_{n,\ell}$, the number of steps until the ℓ vertices are totally disconnected. We obtain the following result.

Corollary 5 *As $n \to \infty$, the random vector*

$$\left(\frac{\ln n}{n} A_{n,2}, \ldots, \frac{\ln n}{n} A_{n,\ell}\right)$$

converges in distribution to

$$\left(U_{(1,\ell)}, \ldots, U_{(\ell-1,\ell)}\right),$$

where $U_{(1,\ell)} \leq U_{(2,\ell)} \leq \cdots \leq U_{(\ell-1,\ell)}$ *denote the first* $\ell - 1$ *order statistics of an i.i.d. sequence* $U_1, \ldots, U_\ell$ *of uniform* $[0, 1]$ *random variables.*

In particular, $\frac{\ln n}{n} A_{n,2}$ converges in distribution to a beta$(1, \ell)$ random variable, and $\frac{\ln n}{n} A_{n,\ell}$ converges in distribution to a beta$(\ell - 1, 2)$ law.

Proof Since the branches of $\mathrm{Cut}(T_n)$ are asymptotically small compared to the trunk (see e.g. Proposition 1 in [9]), with probability tending to 1 as $n \to \infty$ the ℓ vertices $U_1^{(n)}, \ldots, U_\ell^{(n)}$ are cut from the root component one after the other, i.e. in no stage of the disconnection algorithm, a non-root tree component will contain more than one of the $U_1^{(n)}, \ldots, U_\ell^{(n)}$. On this event, writing again $\mathcal{R}_{n,\ell}$ for the reduction of $\mathrm{Cut}(T_n)$ to the ℓ leaves $\{U_1^{(n)}\}, \ldots, \{U_\ell^{(n)}\}$ and its root $\{0, 1, \ldots, n\}$, the variable $A_{n,i+1} - 1$ is given by the length of the path in $\mathcal{R}_{n,\ell}$ from the root to the ith branch point. Now, if $U_1, \ldots, U_\ell$ and $\mathcal{R}_\ell$ are defined as in the proof of Corollary 3, the distance in $\mathcal{R}_\ell$ from the root 0 to the ith smallest among $U_1, \ldots, U_\ell$ is distributed as $U_{(i,\ell)}$. Together with (13), this proves the claim.

Remark With a proof similar to that of Corollary 4, one sees that the statement of Corollary 5 does also hold if $A_{n,2}, \ldots, A_{n,\ell}$ are replaced by the analogous quantities for disconnecting the ℓ last vertices $n - \ell + 1, \ldots, n$. On the other hand, if one is interested in disconnecting the first ℓ vertices $0, \ldots, \ell - 1$, and if $B_{n,2}, \ldots, B_{n,\ell}$ denote in this case the quantities corresponding to $A_{n,2}, \ldots, A_{n,\ell}$, one first observes the trivial bound

$$B_{n,2} \leq \cdots \leq B_{n,\ell} \leq X_{n,\ell},$$

where $X_{n,\ell}$ is the number of steps needed to isolate $0, 1, \ldots, \ell - 1$. Now, $B_{n,2}$ can be identified with the number of steps in the algorithm for isolating the root until for the first time, an edge connecting one of the vertices $1, \ldots, \ell - 1$ to its parent is removed. By similar means as in the proof of Lemma 2, one readily checks that at this time, the root component has a size of order n^β, with β having a beta$(\ell - 1, 1)$-distribution. In particular, we see that $(X_n - B_{n,2}) = o(n/\ln^2 n)$ in probability, where X_n is the number of steps to isolate the root 0. But by (12), also $(X_n - X_{n,\ell}) = o(n/\ln^2 n)$ in probability. Therefore, the variables $B_{n,i}$ have the same limit behavior in law as X_n, that is as $n \to \infty$, $\frac{\ln^2 n}{n} B_{n,i} - \ln n - \ln \ln n$, $i = 2, \ldots, \ell$, converge all to the same completely asymmetric Cauchy variable X defined by (8).

4.4 Ordered Destruction

Here, we consider briefly another natural destruction procedure of a RRT, where instead of removing edges in a uniform random order, we remove them deterministically in their *natural* order. That is the ith edge of T_n which is removed is now the one connecting the vertex i to its parent, for $i = 1, \ldots, n$.

We first point at the fact that the number of ordered edge removals which are now needed to isolate the root (recall that we only take into account edge removals inside the current subtree containing the root) can be expressed as $d_n(0) = \beta_1 + \cdots + \beta_n$, where $\beta_i = 1$ if the parent of vertex i in T_n is the root 0, and 0 otherwise. That is to say that $d_n(0)$ is the degree of the root. Further the recursive construction entails the β_i are independent variables, such that each β_i has the Bernoulli distribution with parameter $1/i$. As is well-known, it then follows e.g. from Lyapunov's central limit theorem that

$$\lim_{n\to\infty} \frac{d_n(0) - \ln n}{\sqrt{\ln n}} = \mathcal{N}(0, 1) \qquad \text{in distribution.}$$

We refer to Kuba and Panholzer [26] for many more results about the degree distributions in random recursive trees.

We then turn our attention to the cut-tree described in Sect. 4.2, which encodes the ordered destruction of T_n. We write $\mathrm{Cut}^{\mathrm{ord}}(T_n)$ for the latter and observe that the recursive construction of T_n implies that in turn, $\mathrm{Cut}^{\mathrm{ord}}(T_n)$ can also be defined by a simple recursive algorithm. Specifically, $\mathrm{Cut}^{\mathrm{ord}}(T_1)$ is the elementary complete binary tree with two leaves, $\{0\}$ and $\{1\}$, and root $\{0, 1\}$. Once T_n and hence $\mathrm{Cut}^{\mathrm{ord}}(T_n)$ have been constructed, T_{n+1} is obtained by incorporating the vertex $n+1$ and creating a new edge between $n+1$ and its parent U_{n+1}, which is chosen uniformly at random in $\{0, 1, \ldots, n\}$. Note that this new edge is the last one which will be removed in the ordered destruction of T_{n+1}. In terms of cut-trees, this means that the leaf $\{U_{n+1}\}$ of $\mathrm{Cut}^{\mathrm{ord}}(T_n)$ should be replaced by an internal node $\{U_{n+1}, n+1\}$ to which two leaves are attached, namely $\{U_{n+1}\}$ and $\{n+1\}$. Further, any block (internal node) B of $\mathrm{Cut}^{\mathrm{ord}}(T_n)$ with $U_{n+1} \in B$ should be replaced by $B \cup \{n+1\}$. The resulting complete binary tree is then distributed as $\mathrm{Cut}^{\mathrm{ord}}(T_{n+1})$.

If we discard labels, this recursive construction of $\mathrm{Cut}^{\mathrm{ord}}(T_n)$ corresponds precisely to the dynamics of the Markov chain on complete binary trees described e.g. in Mahmoud [29] for Binary Search Trees (in short, BST). We record this observation in the following proposition.

Proposition 3 *The combinatorial tree structure of* $\mathrm{Cut}^{\mathrm{ord}}(T_n)$ *is that of a BST with* $n+1$ *leaves.*

BST have been intensively studied in the literature, see Drmota [15] and references therein, and the combination with Proposition 3 yields a number of precise results about the number of ordered cuts which are needed to isolate vertices in T_n. For instance, the so-called *saturation level* $\bar{H}_n$ in a BST is the minimal level of a leaf, and can then be viewed as the smallest number of ordered cuts after which some vertex of T_n has been isolated. Similarly, the height H_n is the maximal level of a leaf, and thus corresponds to the maximal number of ordered cuts needed to isolate a vertex in T_n. The asymptotic behaviors of the height and of the saturation level of a large BST are described in Theorem 6.47 of Drmota [15], in particular one has

$$\lim_{n\to\infty} \frac{\bar{H}_n}{\ln n} = \alpha_- \quad \text{and} \quad \lim_{n\to\infty} \frac{H_n}{\ln n} = \alpha_+$$

where $0 < \alpha_- < \alpha_+$ are the solutions to the equation $\alpha \ln(2e/\alpha) = 1$. In the same vein, the asymptotic results of Chauvin et al. on the profile of large BST can be translated into sharp estimates for the number of vertices of T_n which are isolated after exactly k ordered cuts (see in particular Theorem 3.1 in [13]).

Finally, let us look at component sizes when edges are removed in their natural order. Compared to uniform random edge removal, the picture is fairly different. Indeed, when removing an edge from T_n picked uniformly at random, the size of the subtree not containing 0 is distributed according to the law of ξ conditioned on $\xi \leq n$. If, in contrast, the first edge to be removed is the edge joining 1 to its parent 0, then we know from (1) that both originating subtrees are of order n. Since the splitting property still holds when we remove a fixed edge, the component sizes again inherit a branching structure. In fact, it is an immediate consequence of the definition that the structure of the tree of component sizes corresponding to the ordered destruction on T_n agrees with the structure of T_n and therefore yields the same RRT of size $n+1$.

5 Supercritical Percolation on RRT's

5.1 Asymptotic Sizes of Percolation Clusters

In Sect. 3 it has become apparent that Bernoulli bond percolation on T_n is a tool to study the sizes of tree components which appear in isolation algorithms. Here, we take in a certain sense the opposite point of view and obtain results on the sizes of percolation clusters using what we know about the sizes of tree components. Throughout this section, we use the term cluster to designate connected components induced by percolation, while we use the terminology tree components for connected components arising from isolation algorithms.

More specifically, the algorithm for isolating the root can be interpreted as a dynamical percolation process in which components that do not contain the root are instantaneously frozen. Imagine a continuous-time version of the algorithm, where each edge of T_n is equipped with an independent exponential clock of some parameter α. When a clock rings, the corresponding edge is removed if and only if it currently belongs to the root component. At time $t > 0$, the root component can naturally be viewed as the root cluster of a Bernoulli bond percolation on T_n with parameter $p = \exp(-\alpha t)$. Moreover, under this coupling each percolation cluster is contained in some tree component which was generated by the isolation process up to time t. In order to discover the percolation clusters inside a non-root tree component T', the latter has to be unfrozen, i.e. additional edges from T' have to be removed. In particular, the percolation cluster containing the root of T' can again be identified as the root component of an isolation process on T', stopped at an appropriate time.

These observations lead in [7] to the study of the asymptotic sizes of the largest and next largest percolation clusters of T_n, when the percolation parameter $p(n)$ satisfies

$$p(n) = 1 - \frac{t}{\ln n} + o(1/\ln n) \qquad \text{for } t > 0 \text{ fixed.} \tag{14}$$

This regime corresponds precisely to the supercritical regime, in the sense that the root cluster is the unique giant cluster, and its complement in T_n has a size of order n, too. Indeed, the height h_n of a vertex u picked uniformly at random in a RRT of size $n+1$ satisfies $h_n \sim \ln n$. Since the probability that u is connected to the root is given by the first moment of $(n+1)^{-1}C_{0,n}$, where $C_{0,n}$ denotes the size of the root cluster, one obtains

$$\mathbb{E}((n+1)^{-1}C_{0,n}) = \mathbb{E}\left(p(n)^{h_n}\right) \sim \mathrm{e}^{-t}.$$

A similar argument shows $\mathbb{E}((n^{-1}C_{0,n})^2) \sim \mathrm{e}^{-2t}$, which proves $\lim_{n\to\infty} n^{-1}C_{0,n} = \mathrm{e}^{-t}$ in $L^2(\mathbb{P})$.

Let us now consider the next largest clusters in the regime (14). We write $C_{1,n}, C_{2,n}, \ldots$ for the sizes of the non-root percolation clusters of T_n, ranked in the decreasing order. We quote from [7] the following limit result.

Proposition 4 *For every fixed integer $j \geq 1$,*

$$\left(\frac{\ln n}{n}C_{1,n}, \ldots, \frac{\ln n}{n}C_{j,n}\right)$$

converges in distribution as $n \to \infty$ towards

$$(x_1, \ldots, x_j),$$

where $x_1 > x_2 > \ldots$ denotes the sequence of the atoms of a Poisson random measure on $(0, \infty)$ with intensity $t\mathrm{e}^{-t}x^{-2}\mathrm{d}x$.

The intensity is better understood as the image of the intensity measure $a^{-2}\mathrm{d}a \otimes \mathrm{e}^{-s}\,\mathrm{d}s$ on $(0,\infty) \times (0,t)$ by the map $(a,s) \mapsto x = \mathrm{e}^{-(t-s)}a$. In fact, from our introductory remarks and Proposition 2 it should be clear that the first coordinate of an atom (a,s) stands for the asymptotic (and normalized) size of the tree component containing the percolation cluster, while the second encodes the time when the component was separated from the root.

Instead of providing more details here, let us illustrate an alternative route to prove the proposition, which was taken in [10] to generalize the results to *scale-free random trees*. These random graphs form a family of increasing trees indexed by a parameter $\beta \in (-1, \infty)$ that grow according to a preferential attachment algorithm, see [3]. In the boundary case $\beta \to \infty$, one obtains a RRT, while in the case $\beta = 0$, the ith vertex is added to one of the first $i-1$ vertices with probability proportional to its current degree. In [10], the connection of scale-free random trees to the genealogy of Yule processes was employed, and it should not come as a surprise that this approach can be adapted to random recursive trees. In fact, the case of RRT's is considerably simpler, since one has not to keep track of the degree of vertices when edges are deleted.

Let us sketch the main changes. Denote by $T(s)$ the genealogical tree of a standard Yule process $(\mathcal{Y}_r)_{r\geq 0}$ at time s. Similar to Sect. 3 of [10], we superpose Bernoulli bond percolation with parameter $p = p(n)$ to this construction. Namely, if a new vertex is attached to the genealogical tree, we delete the edge connecting this vertex to its parent with probability $1-p$. We write $T^{(p)}(s)$ for the resulting combinatorial structure at time s, and $T_0^{(p)}(s)$, $T_1^{(p)}(s), \ldots$ for the sequence of the subtrees at time s, enumerated in the increasing order of their birth times, where we use the convention that $T_j^{(p)}(s) = \emptyset$ if less than j edges have been deleted up to time s. In particular, $T_0^{(p)}(s)$ is the subtree containing the root 0, and $\sum_{i\geq 0} |T_i^{(p)}(s)| = \mathcal{Y}_s$. Furthermore, if $b_i^{(p)}$ denotes the birth time of the ith subtree, then the process $(T_i^{(p)}(b_i^{(p)} + s) : s \geq 0)$ is a Yule process with birth rate p per unit population size, started from a single particle of size 1. By analyzing the birth times as in [10], one readily obtains the analogous statements of Section 2 and 3 there. This leads to another proof of Proposition 4.

Remark As it is shown in the forthcoming paper [5], the approach via Yule processes can be extended further to all percolation regimes $p(n) \to 1$. Moreover, if the entire family of cluster sizes is encoded by a tree structure similar to the tree of component sizes, one can specify the finite-dimensional limit of this "tree of cluster sizes". Details will be given in [5].

5.2 Fluctuations of the Root Cluster

We finally take a closer look at the size of the root cluster $C_{0,n}$ for supercritical percolation with parameter

$$p(n) = 1 - \frac{t}{\ln n}.$$

As we have already discussed, $C_{0,n}$ satisfies a law of large numbers, but as we will point out here, $C_{0,n}$ exhibits non-Gaussian fluctuations. This should be seen in sharp contrast to other graph models, were asymptotic normality of the giant cluster has been established, e.g. for the complete graph on n vertices and percolation parameter c/n, $c > 1$ fixed (Stephanov [37], Pittel [35], Barraez et al. [4]).

For RRT's, the fluctuations can be obtained from a recent result of Schweinsberg [36]. Among other things, he studied how the number of blocks in the Bolthausen-Sznitman coalescent changes over time. The Bolthausen-Sznitman coalescent was introduced in [12] in the context of spin glasses, and Goldschmidt and Martin [18] discovered the following connection to the random cutting of RRT's: Equip each edge of a RRT of size n on the vertex set $\{1, \ldots, n\}$ with an independent standard exponential clock. If a clock rings, delete the corresponding edge, say e, and the whole subtree rooted at the endpoint of e most distant from the root 1. Furthermore, replace the label of the vertex of e which is closer to the root 1, say i, by the label set consisting of i and all the vertex labels of the removed subtree. Then the sets of labels form a partition of $\{1, \ldots, n\}$, which evolves according to the dynamics of the Bolthausen-Sznitman coalescent started from n blocks $\{1\}, \ldots, \{n\}$ (see Proposition 2.2 of [18] for details).

Note that in this framework, the variable X_n counting the number of steps in the algorithm for isolating the root can be interpreted as the number of collision events which take place until there is just one block left.

Theorem 1.7 in [36], rephrased in terms of $C_{0,n}$, now reads as follows.

Theorem 3 (Schweinsberg [36]) *There is the weak convergence*

$$\left(n^{-1}C_{0,n} - \mathrm{e}^{-t}\right) \ln n - t\mathrm{e}^{-t} \ln \ln n \Longrightarrow t\mathrm{e}^{-t}(X - \ln t),$$

where X is a completely asymmetric Cauchy variable whose law is determined by (8).

This statement was re-proved in [8], with a different approach which does not rely on the Bolthausen-Sznitman coalescent. Instead, three different growth phases of a RRT T_n are considered, and the effect of percolation is studied in each of these phases. This approach makes again use of the coupling of Iksanov and Möhle and the connection to Yule processes, providing an intuitive explanation for the correction terms in the statement.

References

1. L. Addario-Berry, N. Broutin, C. Holmgren, Cutting down trees with a Markov chainsaw. Ann. Appl. Probab. **24**(6), 2297–2339 (2014)
2. D. Aldous, The Continuum Random Tree III. Ann. Probab. **21**, 248–289 (1993)
3. A.-L. Barabási, R. Albert, Emergence of scaling in random networks. Science **286**(5439), 509–512 (1999)
4. D. Barraez, S. Boucheron, W. de la Fernandez, Vega, On the fluctuations of the giant component. Comb. Probab. Comput. **9**, 287–304 (2000)
5. E. Baur, Percolation on random recursive trees. Preprint (2014). arXiv:1407.2508
6. J. Bertoin, Fires on trees. Ann. Inst. Henri Poincaré Probab. Stat. **48**, 909–921 (2012)
7. J. Bertoin, Sizes of the largest clusters for supercritical percolation on random recursive trees. Random Struct. Algorithms **44–1**, 1098–2418 (2014)
8. J. Bertoin, On the non-Gaussian fluctuations of the giant cluster for percolation on random recursive trees. Electron. J. Probab. **19**(24), 1–15 (2014)
9. J. Bertoin, The cut-tree of large recursive trees. To appear in Ann. Instit. Henri Poincaré Probab. Stat.
10. J. Bertoin, G. Uribe Bravo, Supercritical percolation on large scale-free random trees. To appear in Ann. Appl. Probab.
11. J. Bertoin, G. Miermont, The cut-tree of large Galton-Watson trees and the Brownian CRT. Ann. Appl. Probab. **23**, 1469–1493 (2013)
12. E. Bolthausen, A.-S. Sznitman, On Ruelle's probability cascades and an abstract cavity method. Comm. Math. Phys. **197–2**, 247–276 (1998)
13. B. Chauvin, T. Klein, J.-F. Marckert, A. Rouault, Martingales and profile of binary search trees. Electron. J. Probab. **10**, 420–435 (2005)

14. D. Dieuleveut,The vertex-cut-tree of Galton-Watson trees converging to a stable tree. To appear in Ann. Appl. Probab.
15. M. Drmota, *Random Trees* (Springer, New York, Vienna, 2009)
16. M. Drmota, A. Iksanov, M. Möhle, U. Rösler, A limiting distribution for the number of cuts needed to isolate the root of a random recursive tree. Random Struct. Algorithms **34–3**, 319–336 (2009)
17. K.B. Erickson, Strong renewal theorems with infinite mean. Trans. Am. Math. Soc. **151**, 263–291 (1970)
18. C. Goldschmidt, J.B. Martin, Random recursive trees and the Bolthausen-Sznitman coalescent. Electron. J. Probab. **10**, 718–745 (2005)
19. B. Haas, G. Miermont, Scaling limits of Markov branching trees with applications to Galton-Watson and random unordered trees. Ann. Probab. **40–6**, 2589–2666 (2012)
20. C. Holmgren, Random records and cuttings in binary search trees. Comb. Probab. Comput. **19–3**, 391–424 (2010)
21. C. Holmgren, A weakly 1-stable distribution for the number of random records and cuttings in split trees. Adv. Appl. Probab. **43–1**, 151–177 (2011)
22. A. Iksanov, M. Möhle, A probabilistic proof of a weak limit law for the number of cuts needed to isolate the root of a random recursive tree. Electron. Comm. Probab. **12**, 28–35 (2007)
23. S. Janson, Random records and cuttings in complete binary trees, in *Mathematics and Computer Science III, Algorithms, Trees, Combinatorics and Probabilities (Vienna 2004)*, ed. by M. Drmota, P. Flajolet, D. Gardy, B. Gittenberger (Basel, Birkhäuser, 2004), pp. 241–253
24. S. Janson, Random cutting and records in deterministic and random trees. Random Struct. Algorithms **29–2**, 139–179 (2006)
25. O. Kallenberg, *Foundations of Modern Probability, Probability and its Applications* (Springer, New York, 2002)
26. M. Kuba, A. Panholzer, On the degree distribution of the nodes in increasing trees. J. Combin. Theory, Series A 114–4, 597–618 (2007)
27. M. Kuba, A. Panholzer, Isolating nodes in recursive trees. Aequ. Math. **76**, 258–280 (2008)
28. M. Kuba, A. Panholzer, Multiple isolation of nodes in recursive trees. OJAC **9**, 26 (2014)
29. H. Mahmoud, *Evolution of Random Search Trees* (Wiley, New York, 1992)
30. H. Mahmoud, R.T. Smythe, A survey of recursive trees. Theor. Probab. Math. Stat. **51**, 1–29 (1994)
31. A. Meir, J.W. Moon, Cutting down random trees. J. Austral. Math. Soc. **11**, 313–324 (1970)
32. A. Meir, J.W. Moon, Cutting down recursive trees. Math. Biosci. **21**, 173–181 (1974)
33. A. Panholzer, Cutting down very simple trees. Quaest. Math. **29–2**, 211–227 (2006)
34. J. Pitman, Combinatorial Stochastic Processes. École d'été de Probabilités de St. Flour, Lecture Notes in Mathematics 1875 (Springer, Berlin, 2006)
35. B. Pittel, On tree census and the giant component in sparse random graphs. Random Struct. Algorithms **1**, 311–342 (1990)
36. J. Schweinsberg, Dynamics of the evolving Bolthausen-Sznitman coalescent. Electron. J. Probab. **17**, 1–50 (2012)
37. V.E. Stepanov, On the probability of connectedness of a random graph $\mathcal{G}_m$ (t). Theory Probab. Appl. **15–1**, 55–67 (1970)

The Master Equation for Large Population Equilibriums

René Carmona and François Delarue

Abstract We use a simple N-player stochastic game with idiosyncratic and common noises to introduce the concept of Master Equation originally proposed by Lions in his lectures at the *Collège de France*. Controlling the limit $N \to \infty$ of the explicit solution of the N-player game, we highlight the stochastic nature of the limit distributions of the states of the players due to the fact that the random environment does not average out in the limit, and we recast the Mean Field Game (MFG) paradigm in a set of coupled Stochastic Partial Differential Equations (SPDEs). The first one is a forward stochastic Kolmogorov equation giving the evolution of the conditional distributions of the states of the players given the common noise. The second is a form of stochastic Hamilton Jacobi Bellman (HJB) equation providing the solution of the optimization problem when the flow of conditional distributions is given. Being highly coupled, the system reads as an infinite dimensional Forward Backward Stochastic Differential Equation (FBSDE). Uniqueness of a solution and its Markov property lead to the representation of the solution of the backward equation (i.e. the value function of the stochastic HJB equation) as a deterministic function of the solution of the forward Kolmogorov equation, function which is usually called the *decoupling field* of the FBSDE. The (infinite dimensional) PDE satisfied by this decoupling field is identified with the *master equation*. We also show that this equation can be derived for other large populations equilibriums like those given by the optimal control of McKean-Vlasov stochastic differential equations. The paper is

Paper presented at the conference "Stochastic Analysis", University of Oxford, September 23, 2013.

René Carmona—Partially supported by NSF: DMS-0806591.

R. Carmona
ORFE, Bendheim Center for Finance, Princeton University, Princeton, NJ 08544, USA
e-mail: rcarmona@princeton.edu

F. Delarue (✉)
Laboratoire J.A. Dieudonné, Université de Nice Sophia-Antipolis, Nice Cedex 02, 06108 Parc Valrose, France
e-mail: Francois.delarue@unice.fr

D. Crisan et al. (eds.), *Stochastic Analysis and Applications 2014*,
Springer Proceedings in Mathematics & Statistics 100,
DOI 10.1007/978-3-319-11292-3_4

written more in the style of a review than a technical paper, and we spend more time motivating and explaining the probabilistic interpretation of the Master Equation, than identifying the most general set of assumptions under which our claims are true.

1 Introduction

In several lectures given at the *Collège de France*, P.L. Lions describes mean-field games by a single equation referred to as the *fundamental equation* or *master equation*. Roughly speaking, this equation encapsulates all the information about the Mean Field Game (MFG) problem into a single equation. The purpose of this paper is to review its theoretical underpinnings and to derive it for general MFGs with common noise.

The master equation is a Partial Differential Equation (PDE) in time, the state controlled by the players (typically an element of a Euclidean space, say $\mathbb{R}^d$), and the probability distribution of this state. While standard differential calculus can be used in the time domain $[0, T]$ and the state space $\mathbb{R}^d$, a special kind of differential calculus needs to be used in the space $\mathcal{P}(\mathbb{R}^d)$ of probability measures. The rules of this special differential calculus are described in Lions' lectures, and explained in the notes Cardaliaguet wrote from these lectures [2]. See also Ref. [3] and its appendix at the end of the paper for useful idiosyncrasies of this calculus.

Here our goal is to emphasize the probabilistic nature of the master equation, as the associated characteristics are (possibly random) paths with values in the space $\mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d)$. Our approach is especially enlightening for mean field games in a random environment (see Sect. 2 for definitions and examples), the simplest instances occurring in the presence of random shocks common to all the players. In that framework, the characteristics are given by the sample paths $((X_t, \mathcal{L}(X_t|W^0)))_{0\le t\le T}$, where $(X_t)_{0\le t\le T}$ are the state equilibrium trajectories as identified by the solution of the mean field game problem, and $(\mathcal{L}(X_t|W^0))_{0\le t\le T}$ denote the state conditional marginal distributions in equilibrium, given the value of the past history of the common noise. Examples of mean field games with a common noise were considered in Refs. [7, 10, 11]. Their theory is developed in the forthcoming paper [5] in a rather general setting.

As in the analysis of standard MFG models, the main challenge is the solution of a system comprising a forward PDE coupled with a backward PDE. However, in the random environment case, both equations are stochastic PDEs (SPDEs). The forward SPDE is a Kolmogorov equation describing the dynamics of the conditional laws of the state given the common noise, and the backward SPDE is a stochastic Hamilton-Jacobi-Bellman equation describing the dynamics of the conditional value function. Our contention is that this couple of SPDEs should be viewed as a Forward Backward Stochastic Differential Equation (FBSDE) in infinite dimension. For with this point of view, if some form of Markov property holds, it is natural to expect that the backward component can be written as a function of the forward component, this function being called the *decoupling field*. In finite dimension, a simple application of Itô's formula shows that when the decoupling field is smooth, it must satisfy a PDE. We use an infinite dimensional version of this argument to derive the master equation. The infinite dimension version of Itô's formula needed for the differential calculus chosen for the space of measures is taken from another forthcoming paper [8], and is adapted to the case of a random environment in the appendix.

While the MFG approach does not require the solution of stochastic equations of the McKean-Vlasov type per se, the required fixed point argument identifies the equilibrium trajectory of the game as a *de facto* solution of such an equation. This suggests that the tools developed for solving MFG problems could be reused toward the solution of optimal control problems for McKean-Vlasov dynamics. In the previous paper [3], we established a suitable version of the stochastic Pontryagin principle for the control of McKean-Vlasov SDEs and highlighted the differences with the version of the stochastic Pontryagin principle used to tackle MFG models. Here we show in a similar way that our derivation of the master equation can be used as well for this type of large population equilibrium problem.

This research agenda, namely deriving the master equation for mean-field games and the control of McKean-Vlasov SDEs, has been considered in Ref. [1] in parallel and independently of our work. Therein, another approach is suggested. It relies on a different interpretation of the master equation, yielding a different equation in the case of the control of McKean-Vlasov SDEs. It also involves a different differential calculus on the space of measures, operating at the level of the densities of the probability distributions whenever they exist. We expand on the similarities and differences between the two sets of results in Sect. 4.7.

The present paper is organized as follows. Mean field games in a random environment are presented in Sect. 2. The problem is formulated in terms of a stochastic forward-backward system in infinite dimension. A specific example, taken from [7], is exposed in Sect. 3. The master equation is derived explicitly in this particular case. In Sect. 4, we propose a systematic approach to the master equation for large population control problems in random environment. We consider both MFGs and the control of McKean-Vlasov dynamics. Another example, taken from [11], is revisited in Sect. 5. In the Appendix, we conclude with a proof of the Itô's chain rule along flows of random measures.

When analyzed within the probabilistic framework of the stochastic maximum principle, MFGs with a common noise lead to the analysis of stochastic differential equations conditioned on the knowledge of some of the driving Brownian motions. These forms of conditioned forward stochastic dynamics are best understood in the framework of Terry Lyons' theory of rough paths. Indeed integrals and differentials with respect to the conditioned paths can be interpreted in the sense of rough paths while the meaning of the others can remain in the classical Itô calculus framework. We thought this final remark was appropriate given the raison d'être of the present volume, and our strong desire to convey our deepest appreciation to the man, and pay homage to the mathematician as a remarkably creative scientist.

2 Mean Field Games in a Random Environment

The basic purpose of mean-field game theory is to analyze asymptotic Nash equilibriums for large populations of individuals with mean-field interactions. This goes back to the independent works of Lasry and Lions [13–15] and Huang, Caines and Malhamé [12].

Throughout the paper, we consider models in which individuals (also referred to as *particles* or *players*) are subject to two sources of noise: an idiosyncratic noise, independent from one individual to another, and a common noise, accounting for the common environment in which the individuals evolve. We decide to model the environment by means of a zero-mean Gaussian white noise field $W^0 = (W^0(\Lambda, B))_{\Lambda,B}$, parameterized by the Borel subsets Λ of a Polish space Ξ and the Borel subsets B of $[0, \infty)$, and such that

$$\mathbb{E}\big[W^0(\Lambda, B)W^0(\Lambda', B')\big] = \nu\big(\Lambda \cap \Lambda'\big)|B \cap B'|,$$

where we use the notation $|B|$ for the Lebesgue measure of a Borel subset of $[0, \infty)$. Here ν is a non-negative measure on Ξ, called the spatial intensity of W^0. Often we shall use the notation W_t^0 for $W^0(\,\cdot\,, [0, t])$, and most often, we shall simply take $\Xi = \mathbb{R}^\ell$.

We now assume that the dynamics in $\mathbb{R}^d$, with $d \geq 1$, of the private state of player $i \in \{1, \dots, N\}$ are given by stochastic differential equations (SDEs) of the form:

$$\begin{aligned} dX_t^i &= b\big(t, X_t^i, \overline{\mu}_t^N, \alpha_t^i\big)dt + \sigma\big(t, X_t^i, \overline{\mu}_t^N, \alpha_t^i\big)dW_t^i \\ &\quad + \int_\Xi \sigma^0\big(t, X_t^i, \overline{\mu}_t^N, \alpha_t^i, \xi\big)W^0(d\xi, dt), \end{aligned} \tag{1}$$

where $W^1, \dots, W^N$ are N independent Brownian motions, independent of W^0, all of them being defined on some filtered probability space $(\Omega, \mathbb{F} = (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$. For simplicity, we assume that $W^0, W^1, \dots, W^N$ are scalar valued, multidimensional analogs can be handled along the same lines. The term $\overline{\mu}_t^N$ denotes the empirical distribution of the individual states at time t:

$$\overline{\mu}_t^N = \frac{1}{N}\sum_{i=1}^N \delta_{X_t^i}.$$

The processes $((\alpha_t^i)_{t\geq 0})_{1\leq i\leq N}$ are progressively-measurable, with values in an open subset A of some Euclidean space. They stand for control processes. The coefficients b, σ and σ^0 are defined accordingly on $[0, T] \times \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \times A(\times \Xi)$ with values in $\mathbb{R}^d$, in a measurable way, the set $\mathcal{P}(\mathbb{R}^d)$ denoting the space of probability measures on $\mathbb{R}^d$ endowed with the topology of weak convergence.

The simplest example of random environment corresponds to a coefficient σ^0 independent of ξ. In this case, the random measure W^0 may as well be independent of the spatial component. In other words, we can assume that $W^0(d\xi, dt) = W^0(dt) = dW_t^0$, for an extra Wiener process W^0 independent of the space location ξ and of the idiosyncratic noise terms $(W^i)_{1\le i\le N}$, representing an extra source of noise which is *common* to all the players.

We should think of $W^0(d\xi, dt)$ as a random noise which is white in time (to provide the time derivative of a Brownian motion) and colored in space (the spectrum of the color being given by the Fourier transform of ν). In fact, if $\Xi = \mathbb{R}^d$ and ν is integrable enough, then a motivating example we should have in mind is as follows. Denoting by δ a mollified version of the delta function (which we treat as the actual point mass at 0 for the purpose of this informal discussion), if σ^0 is a function of the form $\sigma^0(t, x, \mu, \alpha, \xi) \sim \sigma^0(t, x, \mu, \alpha)\delta(x - \xi)$ then the integration with respect to the spatial part of the random measure W^0 gives

$$\int_{\mathbb{R}^d} \sigma^0(t, X_t^i, \overline{\mu}_t^N, \alpha_t^i, \xi) W^0(d\xi, dt) = \sigma^0(t, X_t^i, \overline{\mu}_t^N) W^0(X_t^i, dt),$$

which says that, at time t, the private state of player i is subject to several sources of random shocks: its own idiosyncratic noise W_t^i, but also, an independent white noise shock picked up at the very location/value of his own private state.

2.1 Asymptotics of the Empirical Distribution $\overline{\mu}_t^N$

The rationale for the MFG approach to the search for approximate Nash equilibriums for large games is based on several limiting arguments, including the analysis of the asymptotic behavior as $N \to \infty$ of the empirical distribution $\overline{\mu}_t^N$ coupling the states dynamics of the individual players. By the symmetry of our model and de Finetti's law of large numbers, this limit should exist if we allow only exchangeable strategy profiles $(\alpha_t^1, \ldots, \alpha_t^N)$. This will be the case if we restrict ourselves to distributed strategy profiles of the form $\alpha_t^j = \alpha(t, X_t^j, \overline{\mu}_t^N)$ for some deterministic (smooth) function $(t, x, \mu) \mapsto \alpha(t, x, \mu) \in A$.

In order to understand this limit, we can use an argument from propagation of chaos theory, as presented in Sznitman's lecture notes [20]. A possible alternative is to analyze the action of $\overline{\mu}_t^N$ on test functions for $t \in [0, T]$, T denoting some time horizon. Fixing a smooth test function ϕ with compact support in $[0, T] \times \mathbb{R}^d$ and using Itô's formula, we compute:

$$\begin{aligned} d\langle \phi(t, \cdot), \frac{1}{N}\sum_{j=1}^N \delta_{X_t^j}\rangle &= \frac{1}{N}\sum_{j=1}^N d\phi(t, X_t^j) \\ &= \frac{1}{N}\sum_{j=1}^N \Big(\partial_t\phi(t, X_t^j)dt + \nabla\phi(t, X_t^j)\cdot dX_t^j + \frac{1}{2}\mathrm{trace}\{\nabla^2\phi(t, X_t^j)d[X^j, X^j]_t\}\Big) \end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{N}\sum_{j=1}^{N} \partial_t \phi(t, X_t^j) dt + \frac{1}{N}\sum_{j=1}^{N} \nabla\phi(t, X_t^j) \cdot \sigma\big(t, X_t^j, \overline{\mu}_t^N, \alpha(t, X_t^j, \overline{\mu}_t^N)\big) dW_t^j \\
&\quad + \frac{1}{N}\sum_{j=1}^{N} \nabla\phi(t, X_t^j) \cdot b\big(t, X_t^j, \overline{\mu}_t^N, \alpha(t, X_t^j, \overline{\mu}_t^N)\big) dt \\
&\quad + \frac{1}{N}\sum_{j=1}^{N} \nabla\phi(t, X_t^j) \cdot \int_{\Xi} \sigma^0\big(t, X_t^j, \overline{\mu}_t^N, \alpha(t, X_t^j, \overline{\mu}_t^N), \xi\big) W^0(d\xi, dt) \\
&\quad + \frac{1}{2N}\sum_{j=1}^{N} \text{trace}\bigg\{\bigg([\sigma\sigma^\dagger]\big(t, X_t^j, \overline{\mu}_t^N, \alpha(t, X_t^j, \overline{\mu}_t^N)\big) \\
&\quad + \int_{\Xi} [\sigma^0\sigma^{0\dagger}]\big(t, X_t^j, \overline{\mu}_t^N, \alpha(t, X_t^j, \overline{\mu}_t^N), \xi\big)\nu(d\xi)\bigg) \nabla^2\phi(t, X_t^j)\bigg\} dt.
\end{aligned}
$$

Our goal is to take the limit as $N \to \infty$ in this expression. Using the definition of the measures $\overline{\mu}_t^N$ we can rewrite the above equality as:

$$
\begin{aligned}
\langle \phi(t, \cdot), \overline{\mu}_t^N \rangle - \langle \phi(0, \cdot), \overline{\mu}_0^N \rangle &= O(N^{-1/2}) + \int_0^t \big\langle \partial_t \phi(s, \cdot), \overline{\mu}_s^N \big\rangle ds \\
&\quad + \int_0^t \big\langle \nabla\phi(s, \cdot) \cdot b\big(s, \cdot, \overline{\mu}_s^N, \alpha(s, \cdot, \overline{\mu}_s^N)\big), \overline{\mu}_s^N \big\rangle ds \\
&\quad + \frac{1}{2}\int_0^t \bigg\langle \text{trace}\bigg\{\bigg([\sigma\sigma^\dagger]\big(s, \cdot, \overline{\mu}_s^N, \alpha(s, \cdot, \overline{\mu}_s^N)\big) \\
&\quad + \int_{\Xi} [\sigma^0\sigma^{0\dagger}]\big(s, \cdot, \overline{\mu}_s^N, \alpha(s, \cdot, \overline{\mu}_s^N), \xi\big)\nu(d\xi)\bigg) \nabla^2\phi(s, \cdot)\bigg\}, \overline{\mu}_s^N \bigg\rangle ds \\
&\quad + \int_0^t \big\langle \nabla\phi(s, \cdot) \cdot \int_{\Xi} \sigma^0\big(s, \cdot, \overline{\mu}_s^N, \alpha(s, \cdot, \overline{\mu}_s^N), \xi\big) W^0(d\xi, ds), \overline{\mu}_s^N \big\rangle,
\end{aligned}
$$

which shows (formally) after integration by parts that, in the limit $N \to \infty$,

$$
\mu_t = \lim_{N\to\infty} \overline{\mu}_t^N
$$

appears as a solution of the Stochastic Partial Differential Equation (SPDE)

$$
\begin{aligned}
d\mu_t = &-\nabla \cdot \big[b\big(t, \cdot, \mu_t, \alpha(t, \cdot, \mu_t)\big)\mu_t \big] dt \\
&- \nabla \cdot \bigg(\int_{\Xi} \sigma^0\big(t, \cdot, \mu_t, \alpha(t, \cdot, \mu_t), \xi\big) W^0(d\xi, dt)\mu_t \bigg) \\
&+ \frac{1}{2} \mathrm{trace}\bigg[\nabla^2 \bigg(\big[\sigma\sigma^\dagger\big]\big(t, \cdot, \mu_t, \alpha(t, \cdot, \mu_t)\big) \\
&+ \int_{\Xi} \big[\sigma^0\sigma^{0\dagger}\big]\big(t, \cdot, \mu_t, \alpha(t, \cdot, \mu_t), \xi\big)\nu(d\xi) \bigg) \mu_t \bigg] dt. \qquad (2)
\end{aligned}
$$

This SPDE reads as a stochastic Kolmogorov equation. It describes the flow of marginal distributions of the solution of a conditional McKean-Vlasov equation, namely:

$$
\begin{aligned}
dX_t = &\, b\big(t, X_t, \mu_t, \alpha(t, X_t, \mu_t)\big)dt + \sigma\big(t, X_t, \mu_t, \alpha(t, X_t, \mu_t)\big)dW_t \\
&+ \int_{\Xi} \sigma^0(t, X_t, \mu_t, \alpha(t, X_t, \mu_t), \xi\big) W^0(d\xi, dt), \qquad (3)
\end{aligned}
$$

subject to the constraint $\mu_t = \mathcal{L}(X_t | \mathcal{F}_t^0)$, where $\mathbb{F}^0 = (\mathcal{F}_t^0)_{t \geq 0}$ is the filtration generated by the spatial white noise measure W^0. Throughout the whole paper, the letter $\mathcal{L}$ refers to the law, so that $\mathcal{L}(X_t | \mathcal{F}_t^0)$ denotes the conditional law of X_t given $\mathcal{F}_t^0$. The connection between (2) and (3) can be checked by expanding $(\langle \phi(t, \cdot), \mu_t \rangle = \mathbb{E}(\phi(X_t)|\mathcal{F}_t^0))_{0 \leq t \leq T}$ by means of Itô's formula.

For the sake of illustration we rewrite this SPDE in a few particular cases which we will revisit later on:

1. If we assume that $\sigma(t, x, \mu, \alpha) \equiv \sigma$ is a constant, that $\sigma^0(t, x, \mu, \alpha) \equiv \sigma^0(t, x)$ is also uncontrolled and that the spatial white noise is actually scalar, namely $W(d\xi, dt) = dW_t^0$ for a scalar Wiener process W^0 independent of the Wiener processes $(W^i)_{i \geq 1}$, then the stochastic differential equations giving the dynamics of the state of the system read

$$
dX_t^i = b(t, X_t^i, \overline{\mu}_t^N, \alpha_t^i)dt + \sigma dW_t^i + \sigma^0(t, X_t^i)dW_t^0, \qquad i = 1, \ldots, N \qquad (4)
$$

and the limit μ_t of the empirical distributions satisfies the equation

$$
\begin{aligned}
d\mu_t = &-\nabla \cdot \big[b\big(t, \cdot, \mu_t, \alpha(t, \cdot, \mu_t)\big)\mu_t \big] dt - \nabla \cdot \big(\sigma^0(t, \cdot)dW_t^0 \mu_t \big) \\
&+ \frac{1}{2} \mathrm{trace}\Big[\nabla^2 \Big(\big[\sigma\sigma^\dagger + \sigma^0\sigma^{0\dagger}\big](t, \cdot) \Big) \mu_t \Big] dt. \qquad (5)
\end{aligned}
$$

Once coupled with the corresponding version (3), rough paths theory can be used to express the dynamics of the path $(X_t)_{t \geq 0}$ conditional on the values of W^0. This would be still another way to express the dynamics of the conditional marginal laws of $(X_t)_{t \geq 0}$ given W^0.

2. Note that, when the ambient noise is not present (i.e. either $\sigma^0 \equiv 0$ or $W^0 \equiv 0$), this SPDE reduces to a deterministic PDE. It is the Kolmogorov equation giving the forward dynamics of the distribution at time t of the nonlinear diffusion process $(X_t)_{t\geq 0}$ (nonlinear in McKean-Vlasov's sense).

2.2 Solution Strategy for Mean Field Games

When players are assigned a cost functional, a natural (and challenging) question is to characterize and identify equilibriums for the population. A typical framework is to assume that the cost to player i, for any $i \in \{1, \dots, N\}$, writes

$$J^i(\alpha^1, \dots, \alpha^N) = \mathbb{E}\Big[\int_0^T f\big(t, X_t^i, \overline{\mu}_t^N, \alpha_t^i\big)dt + g\big(X_T^i, \overline{\mu}_T^N\big)\Big],$$

for some functions $f : [0, T] \times \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \times A \to \mathbb{R}$ and $g : \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}$. Each cost functional J^i depends upon all the controls $((\alpha_t^j)_{0\leq t\leq T})_{j\in\{1,\dots,N\}}$ through the flow of empirical measures $(\overline{\mu}_t^N)_{0\leq t\leq T}$.

In the search for a Nash equilibrium α, one assumes that all the players j but one keep the same strategy profile α, and the remaining player deviates from this strategy in the hope of being better off. If the number of players is large (think $N \to \infty$), one expects that the empirical measure $\overline{\mu}_t^N$ will not be affected much by infinitesimal deviations by one single player, and for all practical purposes, one can assume that the empirical measure $\overline{\mu}_t^N$ is approximately equal to its limit μ_t. So in the case of large symmetric games, the search for approximate Nash equilibriums could be approached through the solution of the optimization problem of one single player (typically the solution of a stochastic control problem instead of a large game) when the empirical measure $\overline{\mu}_t^N$ is replaced by the solution μ_t of the SPDE (2) appearing in this limiting regime, the 'α' plugged in (2) denoting the strategy used by the players at equilibrium.

The implementation of this method can be broken down into three steps for pedagogical reasons:

(i) Given an initial distribution μ_0 on $\mathbb{R}^d$, fix an arbitrary measure valued adapted stochastic process $(\mu_t)_{0\leq t\leq T}$ over the probability space of the random measure W^0. It stands for a possible candidate for being a Nash equilibrium.

(ii) Solve the (standard) stochastic control problem (with random coefficients)

$$\inf_{(\alpha_t)_{0\leq t\leq T}} \mathbb{E}\left[\int_0^T f(t, X_t, \mu_t, \alpha_t)dt + g(X_T, \mu_T)\right] \tag{6}$$

subject to

$$dX_t = b\big(t, X_t, \mu_t, \alpha_t\big)dt + \sigma\big(t, X_t, \mu_t, \alpha_t\big)dW_t + \int_{\Xi} \sigma^0\big(t, X_t, \mu_t, \alpha_t, \xi\big)W^0(d\xi, dt),$$

with $X_0 \sim \mu_0$, over progressively measurable admissible controls.

(iii) If and when an optimal control exists in feedback form $\alpha(t, x, \mu_t)$, plug it in the SPDE (2), and determine the measure valued stochastic process $(\mu_t)_{0\leq t\leq T}$ in step (i) so that the solution of the SPDE (2) for $\alpha(t, x, \mu_t)$ obtained in point (ii) is precisely $(\mu_t)_{0\leq t\leq T}$ we started from.

Clearly, this last item requires the solution of a fixed point problem in an infinite dimensional space, while the second item involves the solution of an optimization problem in a space of stochastic processes. Thanks to the connection between the SPDE (2) and the McKean-Vlasov equation (3), the fixed point item (iii) reduces to the search for a flow of random measures $(\mu_t)_{0\leq t\leq T}$ such that the law of the optimally controlled process (resulting from the solution of the second item) is in fact μ_t, i.e.

$$\forall t \in [0, T], \quad \mu_t = \mathcal{L}(X_t|\mathcal{F}^0_t).$$

In the absence of the ambient random field noise term W^0, the measure valued adapted stochastic process $(\mu_t)_{0\leq t\leq T}$ can be taken as a deterministic function $[0, T] \ni t \mapsto \mu_t \in \mathcal{P}(\mathbb{R}^d)$, and the control problem in item (ii) is a standard Markovian control problem. Moreover, the fixed point item (iii) reduces to the search for a deterministic flow of measures $[0, T] \ni t \mapsto \mu_t \in \mathcal{P}(\mathbb{R}^d)$ such that the optimally controlled process (resulting from the solution of the second item) satisfies $\mathcal{L}(X_t) = \mu_t$ for each t.

2.3 Stochastic HJB Equation

In this subsection, we study the stochastic control problem (ii) when the flow of random measures $\mu = (\mu_t)_{0\leq t\leq T}$ is fixed, and as mentioned earlier, adapted to the filtration $\mathbb{F}^0$ of the common noise. Optimization is performed over sets $\mathbb{A}_t$ of $\mathbb{F}$-progressively measurable A-valued processes $(\alpha_s)_{t\leq s\leq T}$ satisfying

$$\mathbb{E}\int_t^T |\alpha_s|^2 ds < \infty,$$

and we use the notation $\mathbb{A}$ for $\mathbb{A}_0$. For each $(t, x) \in [0, T]\times\mathbb{R}^d$, we let $(X_s^{t,x})_{t\leq s\leq T}$ be the solution of the stochastic differential equation (being granted that it is well-posed)

$$
\begin{aligned}
dX_s &= b(s, X_s, \mu_s, \alpha_s)ds + \sigma(s, X_s, \mu_s, \alpha_s)dW_s \\
&\quad + \int_{\Xi} \sigma^0(s, X_s, \mu_s, \alpha_s, \xi) W^0(d\xi, ds),
\end{aligned}
\tag{7}
$$

with $X_t = x$. With this notation, we define the (conditional) cost

$$
J^{\mu}_{t,x}\big((\alpha_s)_{t \le s \le T}\big) = \mathbb{E}\Big[\int_t^T f(s, X^{t,x}_s, \mu_s, \alpha_s)ds + g(X^{t,x}_T, \mu_T)\Big|\mathcal{F}^0_t\Big]
\tag{8}
$$

and the (conditional) value function

$$
V^{\mu}(t,x) = \operatorname*{ess\,inf}_{(\alpha_s)_{t \le s \le T} \in \mathbb{A}_t} J^{\mu}_{t,x}\big((\alpha_s)_{t \le s \le T}\big).
\tag{9}
$$

We shall drop the superscript and write X_s for $X^{t,x}_s$ when no confusion is possible. Under some regularity assumptions, we can show that, for each $x \in \mathbb{R}^d$, $(V^{\mu}(t,x))_{0 \le t \le T}$ is an $\mathbb{F}^0$-semi-martingale, and deduce by identification of its Itô decomposition, that it solves a form of stochastic Hamilton-Jacobi Bellman (HJB) equation. Because of the special form of the state dynamics (7), we introduce the (random and nonlocal) operator symbol

$$
\begin{aligned}
&L^*\big(t, x, y, z, (z^0(\xi))_{\xi \in \Xi}, \mu_t\big) \\
&\quad = \inf_{\alpha \in A} \Big[b(t, x, \mu_t, \alpha) \cdot y + \frac{1}{2}\mathrm{trace}\big([\sigma\sigma^{\dagger}](t, x, \mu_t, \alpha) \cdot z\big) + f(t, x, \mu_t, \alpha) \\
&\qquad + \frac{1}{2}\mathrm{trace}\Big(\int_{\Xi} [\sigma^0 \sigma^{0\dagger}](t, x, \mu_t, \alpha, \xi) d\nu(\xi) \cdot z\Big) \\
&\qquad + \int_{\Xi} \sigma^0\big(t, x, \mu_t, \alpha, \xi) \cdot z^0(\xi) d\nu(\xi) \Big].
\end{aligned}
\tag{10}
$$

Assuming that the value function is smooth enough, we can use a generalization of the dynamic programming principle to the present set-up of conditional value functions to show that $V^{\mu}(t,x)$ satisfies a form of stochastic HJB equation as given by a parametric family of BSDEs in the sense that:

$$
\begin{aligned}
V^{\mu}(t,x) &= g(x, \mu_T) + \int_t^T L^*\big(s, x, \partial_x V^{\mu}(s,x), \partial^2_x V^{\mu}(s,x), (Z^{\mu}(s,x,\xi))_{\xi \in \Xi}, \mu_s\big) ds \\
&\quad + \int_t^T Z^{\mu}(s, x, \xi) W^0(d\xi, ds).
\end{aligned}
\tag{11}
$$

Noticing that W^0 enjoys the martingale representation theorem (see Chap. 1 in [18]), this result can be seen as part of the folklore of the theory of backward SPDEs (see for example [19] or [17]).

2.4 Towards the Master Equation

The definition of L^* in (10) suggests that the optimal feedback in (8) could be identified as a function $\hat{\alpha}$ of $t, x, \mu_t, V^\mu(t, \cdot)$ and $Z^\mu(t, \cdot, \cdot)$ realizing the infimum appearing in the definition of L^*. Plugging such a choice for α in the SPDE (2), we deduce that the fixed point condition in item (iii) of the definition of an MFG equilibrium could be reformulated in terms of an infinite dimensional FBSDE, the forward component of which being the Kolmogorov SPDE (2) (with the specific choice of α), and the backward component the stochastic HJB equation (11). The forward variable would be $(\mu_t)_{0\le t\le T}$ and the backward one would be $(V^\mu(t, \cdot))_{0\le t\le T}$. Standard FBSDE theory suggests the existence of a *decoupling field* expressing the backward variable in terms of the forward one, in other words that $V^\mu(t, x)$ could be written as $V(t, x, \mu_t)$ for some function V, or equivalently, that $V^\mu(t, \cdot)$ could be written as $V(t, \cdot, \mu_t)$. Using a special form of Itô's change of variable formula proven in the appendix at the end of the paper, these decoupling fields are easily shown, at least when they are smooth, to satisfy PDEs or SPDEs in the case of FBSDEs with random coefficients. The definition of the special notion of smoothness required for this form of Itô formula is recalled in the appendix. This is our hook to Lions' master equation. In order to make this point transparent in the sequel, we strive to provide a better understanding of the mapping $V : [0, T] \times \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}$ and of its dynamics.

3 An Explicitly Solvable Model

This section is devoted to the analysis of an explicitly solvable model. It was introduced and solved in Ref. [7]. There, the players were banks $i \in \{1, \ldots, N\}$, the states X_t^i represented the log-capitalizations of these banks at time t, and it was assumed that each bank controlled its rate of borrowing and lending to a central bank through the drift of X_t^i according to Ornstein-Uhlenbeck dynamics specified below in (12). Here, we ignore the financial interpretation of the model, and we concentrate on some of the mathematical properties of the equilibriums. We reproduce the parts of the solution which are relevant to the present discussion of the master equation. Our interest in this model is the fact that the finite player game can be solved explicitly and the limit $N \to \infty$ of the solution can be controlled. We use it as motivation and testbed for the introduction of the master equation of mean field games with a common noise.

3.1 *Constructions of Exact Nash Equilibria for the N-Player Game*

We assume that the dynamics of the states X_t^i are given by the stochastic differential equations:

$$dX_t^i = \Big[a(m_t^N - X_t^i) + \alpha_t^i\Big]dt + \sigma\Big(\sqrt{1-\rho^2}dW_t^i + \rho dW_t^0\Big), \tag{12}$$

where W_t^i, $i = 0, 1, \dots, N$ are independent scalar Wiener processes, $\sigma > 0$, $a \geq 0$, and m_t^N denotes the sample mean of the X_t^i as defined by $m_t^N = (X_t^1 + \cdots + X_t^N)/N$. So, in the notation introduced in (1), we have

$$b(t, x, \mu, \alpha) = a(m - x) + \alpha, \qquad \text{with} \quad m = \int_{\mathbb{R}} x' d\mu(x'),$$

since the drift of $(X_t^i)_{t\geq 0}$ at time t depends only upon X_t^i itself and the mean m_t^N of the empirical distribution $\overline{\mu}_t^N$ of $X_t = (X_t^1, \dots, X_t^N)$, and

$$\sigma(t, x, \mu, \alpha) = \sigma\sqrt{1-\rho^2}, \qquad \text{and} \qquad \sigma^0(t, x) = \sigma\rho.$$

Player $i \in \{1, \dots, N\}$ controls its state at time t by choosing the control α_t^i in order to minimize

$$J^i(\alpha^1, \dots, \alpha^N) = \mathbb{E}\Bigg[\int_0^T f(t, X_t^i, \overline{\mu}_t^N, \alpha_t^i)dt + g(X_T^i, \overline{\mu}_T^N)\Bigg], \tag{13}$$

where the running and terminal cost functions f and g are given by:

$$\begin{aligned} f(t, x, \mu, \alpha) &= \frac{1}{2}\alpha^2 - q\alpha(m - x) + \frac{\epsilon}{2}(m - x)^2, \\ g(x, \mu) &= \frac{c}{2}(m - x)^2, \end{aligned} \tag{14}$$

for some positive constants q, ϵ and c. As before, m denotes the mean of the measure μ. Clearly, this is a *Linear-Quadratic* (LQ) model and, thus, its solvability should be equivalent to the well-posedness of a matrix Riccati equation. However, given the special structure of the interaction, the Riccati equation is in fact scalar and can be solved explicitly as we are about to demonstrate.

Given an N-tuple $(\hat{\alpha}^i)_{1\leq i\leq N}$ of functions from $[0, T] \times \mathbb{R}$ into $\mathbb{R}$, we define, for each $i \in \{1, \dots, N\}$, the related value function V^i by:

$$V^i(t,x^1,\dots,x^N) = \inf_{(\alpha_s^i)_{t\le s\le T}} \mathbb{E}\Big[\int_t^T f\big(s, X_s^i, \mu_s^N, \alpha_s^i\big)ds + g(X_T^i, \overline{\mu}_T^N)\Big| X_t = x\Big],$$

with the cost functions f and g given in (14), and where the dynamics of $(X_s^1,\dots,X_s^N)_{t\le s\le T}$ are given in (12) with $X_t^j = x^j$ for $j \in \{1,\dots,N\}$ and $\alpha_s^j = \hat{\alpha}^j(s, X_s^j)$ for $j \neq i$. By dynamic programming, the N scalar functions V^i must satisfy the system of HJB equations:

$$\partial_t V^i(t,x) + \inf_{\alpha\in\mathbb{R}}\Big\{\big(a(\overline{x}-x^i)+\alpha\big)\partial_{x^i}V^i(t,x) + \frac{1}{2}\alpha^2 - q\alpha\big(\overline{x}-x^i\big)\Big\} + \frac{\epsilon}{2}(\bar{x}-x^i)^2$$
$$+ \sum_{j\neq i}\big(a(\overline{x}-x^j) + \hat{\alpha}^j(t,x^j)\big)\partial_{x^j}V^j(t,x)$$
$$+ \frac{\sigma^2}{2}\sum_{j=1}^N\sum_{k=1}^N \big(\rho^2 + \delta_{j,k}(1-\rho^2)\big)\,\partial^2_{x^j x^k}V^i(t,x) = 0,$$

for $(t,x)\in[0,T]\times\mathbb{R}^N$, where we use the notation $\overline{x}$ for the mean $\overline{x} = (x^1+\dots+x^N)/N$ and with the terminal condition $V^i(T,x) = (c/2)(\overline{x}-x^i)^2$. The infima in these HJB equations can be computed explicitly:

$$\inf_{\alpha\in\mathbb{R}}\Big\{\big(a(\overline{x}-x^i)+\alpha\big)\partial_{x^i}V^i(t,x) + \frac{1}{2}\alpha^2 - q\alpha\big(\overline{x}-x^i\big)\Big\}$$
$$= a(\overline{x}-x^i)\partial_{x^i}V^i(t,x) - \frac{1}{2}\big[q\big(\overline{x}-x^i\big) - \partial_{x^i}V^i(t,x)\big]^2,$$

the infima being attained for

$$\alpha = q\big(\overline{x}-x^i\big) - \partial_{x^i}V^i(t,x),$$

which suggests to solve the system of N coupled HJB equations:

$$\partial_t V^i + \sum_{j=1}^N\Big[(a+q)\Big(\overline{x}-x^j\Big) - \partial_{x^j}V^j\Big]\partial_{x^j}V^i$$
$$+ \frac{\sigma^2}{2}\sum_{j=1}^N\sum_{k=1}^N\Big(\rho^2+\delta_{j,k}(1-\rho^2)\Big)\partial^2_{x^j x^k}V^i$$
$$+ \frac{1}{2}(\epsilon - q^2)\Big(\overline{x}-x^i\Big)^2 + \frac{1}{2}(\partial_{x^i}V^i)^2 = 0, \qquad i = 1,\dots,N, \tag{15}$$

with the same boundary terminal condition as above. Then, the feedback functions $\hat{\alpha}^i(t,x) = q(\overline{x}-x^i) - \partial_{x^i}V^i(t,x)$ are expected to give the optimal Markovian strategies. Generally speaking, these systems of HJB equations are difficult to solve.

Here, because of the particular forms of the couplings and the terminal conditions, we can solve the system by inspection, checking that a solution can be found in the form

$$V^i(t,x) = \frac{\eta_t}{2}(\overline{x} - x^i)^2 + \chi_t, \tag{16}$$

for some deterministic scalar functions $t \mapsto \eta_t$ and $t \mapsto \chi_t$ satisfying $\eta_T = c$ and $\chi_T = 0$ in order to match the terminal conditions for the V^is. Indeed, the partial derivatives $\partial_{x^j} V^i$ and $\partial_{x^j x^k} V^i$ read

$$\partial_{x^j} V^i(t,x) = \eta_t\big(\frac{1}{N} - \delta_{i,j}\big)\left(\overline{x} - x^i\right), \quad \partial^2_{x^j x^k} V^i(t,x) = \eta_t\big(\frac{1}{N} - \delta_{i,j}\big)(\frac{1}{N} - \delta_{i,k}),$$

and plugging these expressions into (15), and identifying term by term, we see that the system of HJB equations is solved if an only if

$$\begin{cases} \dot{\eta}_t = 2(a+q)\eta_t + \big(1 - \frac{1}{N^2}\big)\eta_t^2 - (\epsilon - q^2), \\ \dot{\chi}_t = -\frac{1}{2}\sigma^2(1-\rho^2)\big(1 - \frac{1}{N}\big)\eta_t, \end{cases} \tag{17}$$

with the terminal conditions $\eta_T = c$ and $\chi_T = 0$. As emphasized earlier, the Riccati equation is scalar and can be solved explicitly. One gets:

$$\eta_t = \frac{-(\epsilon - q^2)\big(e^{(\delta^+ - \delta^-)(T-t)} - 1\big) - c\big(\delta^+ e^{(\delta^+ - \delta^-)(T-t)} - \delta^-\big)}{\big(\delta^- e^{(\delta^+ - \delta^-)(T-t)} - \delta^+\big) - c(1 - 1/N^2)\left(e^{(\delta^+ - \delta^-)(T-t)} - 1\right)}, \tag{18}$$

provided we set:

$$\delta^{\pm} = -(a+q) \pm \sqrt{R}, \qquad \text{with} \qquad R = (a+q)^2 + \left(1 - \frac{1}{N^2}\right)(\epsilon - q^2) > 0. \tag{19}$$

Observe that the denominator in (18) is always negative since $\delta^+ > \delta^-$, so that η_t is well defined for any $t \le T$. The condition $q^2 \le \epsilon$ implies that η_t is positive with $\eta_T = c$. Once η_t is computed, one solves for χ_t (remember that $\chi_T = 0$) and finds:

$$\chi_t = \frac{1}{2}\sigma^2(1-\rho^2)\left(1 - \frac{1}{N}\right)\int_t^T \eta_s \, ds. \tag{20}$$

For the record, we note that the optimal strategies read

$$\hat{\alpha}^i_t = q\big(\overline{X}_t - X^i_t\big) - \partial_{x^i} V^i = \left(q + (1 - \frac{1}{N})\eta_t\right)\big(\overline{X}_t - X^i_t\big), \tag{21}$$

and the optimally controlled dynamics:

$$dX_t^i = \Big(a + q + (1 - \frac{1}{N})\eta_t\Big)\big(\overline{X}_t - X_t^i\big)dt + \sigma\Big(\sqrt{1-\rho^2}dW_t^i + \rho dW_t^0\Big). \quad (22)$$

3.2 The Mean Field Limit

In this subsection, we emphasize the dependence upon the number N of players by writing η_t^N and χ_t^N for the solutions η_t and χ_t of the system (17), and $V^{i,N}(t,x) = (\eta^N/2)(\overline{x} - x^i)^2 + \chi_t^N$ for the value function of player i. Clearly,

$$\lim_{N\to\infty} \eta_t^N = \eta_t^\infty, \qquad \text{and} \qquad \lim_{N\to\infty} \chi_t^N = \chi_t^\infty,$$

where the functions η_t^∞ and χ_t^∞ solve the system:

$$\begin{cases} \dot{\eta}_t^\infty = 2(a+q)\eta_t^\infty + (\eta_t^\infty)^2 - (\epsilon - q^2), \\ \dot{\chi}_t^\infty = -\frac{1}{2}\sigma^2(1-\rho^2)\eta_t^\infty, \end{cases} \quad (23)$$

which is solved as in the case N finite. We find

$$\eta_t^\infty = \frac{-(\epsilon - q^2)\big(e^{(\delta^+ - \delta^-)(T-t)} - 1\big) - c\big(\delta^+ e^{(\delta^+ - \delta^-)(T-t)} - \delta^-\big)}{\big(\delta^- e^{(\delta^+ - \delta^-)(T-t)} - \delta^+\big) - c\left(e^{(\delta^+ - \delta^-)(T-t)} - 1\right)}, \quad (24)$$

and

$$\chi_t^\infty = \frac{1}{2}\sigma^2(1-\rho^2)\int_t^T \eta_s^\infty \, ds. \quad (25)$$

Next we consider the equilibrium behavior of the players' value functions $V^{i,N}$. For the purpose of the present discussion we notice that the value functions $V^{i,N}$ of all the players in the N player game can be written as

$$V^{i,N}\big(t,(x^1,\dots,x^N)\big) = V^N\Big(t, x^i, \frac{1}{N}\sum_{j=1}^N \delta_{x^j}\Big)$$

where the single function V^N is defined as

$$V^N(t,x,\mu) = \frac{\eta_t^N}{2}\Big(x - \int_{\mathbb{R}} x' d\mu(x')\Big)^2 + \chi_t^N, \qquad (t,x,\mu) \in [0,T]\times\mathbb{R}\times\mathcal{P}_1(\mathbb{R}),$$

where $\mathcal{P}_1(\mathbb{R})$ denotes the space of probability measures on $\mathbb{R}$ with a finite first moment. Since the dependence upon the measure is only through the mean, we shall often use the function

$$v^N(t,x,m) = \frac{\eta_t^N}{2}(x-m)^2 + \chi_t^N, \qquad (t,x,m) \in [0,T] \times \mathbb{R} \times \mathbb{R},$$

Notice that, at least for (t,x,m) fixed, we have

$$\lim_{N\to\infty} v^N(t,x,m) = v^\infty(t,x,m)$$

where

$$v^\infty(t,x,m) = \frac{\eta_t^\infty}{2}(x-m)^2 + \chi_t^\infty, \qquad (t,x,m) \in [0,T] \times \mathbb{R} \times \mathbb{R}.$$

Similarly, all the optimal strategies in (21) may be expressed through a single feedback function $\hat{\alpha}^N(t,x,m) = [q + (1-1/N)\eta_t^N](m-x)$ as $\hat{\alpha}_t^i = \hat{\alpha}^N(t, X_t^i, m_t^N)$. Clearly,

$$\lim_{N\to\infty} \hat{\alpha}^N(t,x,m) = \hat{\alpha}^\infty(t,x,m),$$

where $\hat{\alpha}^\infty(t,x,m) = [q+\eta_t](m-x)$.

Repeating the analysis in Sect. 2.1, we find that the limit of the empirical distributions satisfies the following version of (5):

$$\begin{aligned} d\mu_t = &-\partial_x\Big([a(m_t - \cdot\,) + \alpha^\infty(t,\,\cdot\,)]\mu_t\Big)dt \\ &+ \frac{\sigma^2}{2}\partial_{xx}^2\mu_t dt - \sigma\rho\partial_x\mu_t dW_t^0, \quad t \in [0,T], \end{aligned} \tag{26}$$

where $m_t = \int_{\mathbb{R}^d} x d\mu_t(x)$, which is the Kolmogorov equation for the conditional marginal law, given W^0, of the solution of the McKean-Vlasov equation:

$$d\overline{X}_t = \big[a\big(m_t - \overline{X}_t\big) + \alpha^\infty(t,\overline{X}_t)\big]dt + \sigma\left(\rho dW_t^0 + \sqrt{1-\rho^2}dW_t\right), \tag{27}$$

subject to the condition $m_t = \mathbb{E}[\overline{X}_t|\mathcal{F}_t^0]$. Applying the Kolmogorov equation to the test function $\phi(x) = x$, we get

$$dm_t = \left(\int \alpha^\infty(t,x)d\mu_t(x)\right)dt + \sigma\rho dW_t^0. \tag{28}$$

We now write the stochastic HJB equation (11) in the present context. Remember that we assume that the stochastic flow $(\mu_t)_{0\le t\le T}$ is given (as the solution of (26) with some prescribed initial condition $\mu_0 = \mu$), and hence so is $(m_t)_{0\le t\le T}$. Here

$$L^*(t,x,y,z,z^0,m_t) = \inf_{\alpha \in A}\Bigg[[a(m_t - x) + \alpha]y + \frac{\sigma^2}{2}z + \sigma\rho z^0 \\ + \frac{\alpha^2}{2} - q\alpha(m_t - x) + \frac{\epsilon}{2}(m_t - x)^2\Bigg].$$

Since the quantity to minimize is quadratic in α, we need to compute it for $\bar{\alpha} = \bar{\alpha}(t,x,m_t,y)$ with $\bar{\alpha}(t,x,m,y) = q(m-x) - y$. We get:

$$L^*(t,x,y,z,z^0,m_t) = (a+q)(m_t - x)y - \frac{1}{2}y^2 + \frac{\sigma^2}{2}z + \sigma\rho z^0 + \frac{1}{2}(\epsilon - q^2)(m_t - x)^2.$$

Accordingly, the stochastic HJB equation takes the form

$$\begin{aligned} d_t V^\mu(t,x) = \Bigg[& -(a+q)(m_t - x)\partial_x V^\mu(t,x) + \frac{1}{2}[\partial_x V^\mu(t,x)]^2 \\ & - \frac{\sigma^2}{2}\partial_x^2 V^\mu(t,x) - \sigma\rho\partial_x Z^\mu(t,x) - \frac{1}{2}(\epsilon - q^2)(m_t - x)^2\Bigg]dt \\ & - Z^\mu(t,x)dW_t^0, \end{aligned} \tag{29}$$

with the terminal condition $V^\mu(T,x) = (c/2)(m_T - x)^2$.

3.3 Search for a Master Equation

A natural candidate for solving (29) is the random field $(t,x) \mapsto v^\infty(t,x,m_t)$, where as above $(m_t)_{0\le t\le T}$ denotes the means of the solution $(\mu_t)_{0\le t\le t}$ of the Kolmogorov SPDE (26). This can be checked rigorously by using the expression of v^∞ and by expanding $(v^\infty(t,x,m_t))_{0\le t\le T}$ by Itô's formula and taking advantage of (28). As suggested at the end of the previous section, this shows that the stochastic HJB equation admits a solution $V^\mu(t,x)$ that can be expressed as a function of the current value μ_t of the solution of the Kolmogorov SPDE, namely

$$V^\mu(t,x) = v^\infty\Big(t,x,\int_{\mathbb{R}^d} x' d\mu_t(x')\Big).$$

The same argument shows that $(\overline{X}_t)_{0\le t\le T}$ defined in (27) as a solution of a McKean-Vlasov SDE is in fact the optimal trajectory of the control problem considered in the item (ii) of the definition of a MFG, see (6), when the fixed flow of measures is the solution $(\mu_t)_{0\le t\le T}$ of the stochastic PDE (26). Put it differently, $(\mu_t)_{0\le t\le T}$ is a solution of the MFG problem, and the function α^∞ is the associated feedback control, as suggested by the asymptotic analysis performed in the previous paragraph.

A natural question is to characterize the properties of the function v^∞ in an intrinsic way. By definition of the value function (see (9)), we have

$$V^{\mu}(t,\overline{X}_t) = \mathbb{E}\Big[\int_t^T f\big(s,\overline{X}_s,\mu_s,\hat{\alpha}^{\infty}(s,\overline{X}_s)\big)ds + g\big(\overline{X}_T,\mu_T\big)\big|\mathcal{F}_t\Big].$$

Notice that the expectation is here conditional on $\mathcal{F}_t$ whereas in (9), the expression of the value function was conditional on $\mathcal{F}_t^0$. The reason is that, in the above formula, the randomness of the initial condition has to be taken into account.

We deduce

$$dV^{\mu}(t,\overline{X}_t) = -f\big(t,\overline{X}_t,\mu_t,\hat{\alpha}^{\infty}(t,\overline{X}_t)\big)dt + dM_t, \quad t\in[0,T],$$

for some $(\mathcal{F}_t)_{0\le t\le T}$-martingale $(M_t)_{0\le t\le T}$. Recalling that $\bar{\alpha}(t,x,m,y)=q(m-x) - y$, $\partial_x v^{\infty}(t,x,m) = \eta_t^{\infty}(x-m)$, and $\hat{\alpha}^{\infty}(t,x,m) = [q+\eta_t](m-x)$, we deduce that

$$\hat{\alpha}^{\infty}(t,x,m) = \bar{\alpha}\big(t,x,m,\partial_x v^{\infty}(t,x,m)\big),$$

which is the standard relationship in stochastic optimal control for expressing the optimal feedback in terms of the minimizer $\bar{\alpha}$ of the underlying extended Hamiltonian and of the gradient of the value function v^{∞}. We deduce that

$$\begin{aligned} f\big(t,\overline{X}_t,\mu_t,\hat{\alpha}^{\infty}(t,\overline{X}_t)\big) &= -\frac{1}{2}\big(q(m_t-\overline{X}_t) - \partial_x v^{\infty}(t,\overline{X}_t,m_t)\big)\big(q(m_t-\overline{X}_t) \\ &\quad + \partial_x v^{\infty}(t,\overline{X}_t,m_t)\big) + \frac{\epsilon}{2}\big(m_t-\overline{X}_t\big)^2, \end{aligned}$$

so that

$$dV^{\mu}(t,\overline{X}_t) = \Big(-\frac{1}{2}(\epsilon-q^2)(m_t-\overline{X}_t)^2 - \frac{1}{2}\big[\partial_x v^{\infty}(t,\overline{X}_t,m_t)\big]^2\Big)dt + dM_t. \quad (30)$$

We are to compare this Itô expansion with the Itô expansion of $(v^{\infty}(t,\overline{X}_t,m_t))_{0\le t\le T}$. Using the short-hand notation v_t^{∞} for $v^{\infty}(t,\overline{X}_t,m_t)$ and standard Itô's formula, we get:

$$\begin{aligned} dv_t^{\infty} &= \partial_t v_t^{\infty}dt + \partial_x v_t^{\infty}d\overline{X}_t + \partial_m v_t^{\infty}dm_t \\ &\quad + \frac{\sigma^2}{2}\partial_{xx}^2 v_t^{\infty} + \frac{\sigma^2}{2}\rho^2\partial_{mm}^2 v_t^{\infty} + \sigma^2\rho^2\partial_{xm}^2 v_t^{\infty} \\ &= \Big[\partial_t v_t^{\infty} + \partial_x v_t^{\infty}a(m_t-\overline{X}_t) + \partial_x v_t^{\infty}\hat{\alpha}^{\infty}(t,\overline{X}_t) + \partial_m v_t^{\infty}\langle\mu_t,\alpha^{\infty}(t,\cdot)\rangle \qquad (31) \\ &\quad + \frac{\sigma^2}{2}\partial_x^2 v_t^{\infty} + \frac{\sigma^2}{2}\rho^2\partial_m^2 v_t^{\infty} + \sigma^2\rho^2\partial_{xm}^2 v_t^{\infty}\Big]dt \\ &\quad + \sigma\rho[\partial_x v_t^{\infty} + \partial_m v_t^{\infty}]dW_t^0 + \sigma\sqrt{1-\rho^2}\partial_x v_t^{\infty}dW_t. \end{aligned}$$

Identifying the bounded variation terms in (30) and (31), we get:

$$\begin{aligned}
&\partial_t v_t^\infty + \partial_x v_t^\infty a(m_t - \overline{X}_t) + \partial_x v_t^\infty \hat{\alpha}^\infty(t, \overline{X}_t) + \partial_m v_t^\infty \langle \mu_t, \alpha^\infty(t, \cdot) \rangle \\
&\quad + \frac{\sigma^2}{2}\partial_x^2 v_t^\infty + \frac{\sigma^2}{2}\rho^2 \partial_m^2 v_t^\infty + \sigma^2\rho^2\partial_{xm}^2 v_t^\infty \\
&\qquad = -\frac{1}{2}(\epsilon - q^2)(m_t - \overline{X}_t)^2 - \frac{1}{2}\big[\partial_x v_t^\infty\big]^2,
\end{aligned}$$

where $\hat{\alpha}^\infty(t, x, m) = q(m - x) - \partial_x v^\infty(t, x, m)$. Therefore, for a general smooth function $V : (t, x, m) \mapsto V(t, x, m)$, the above relationship with v^∞ replaced by V holds if

$$\begin{aligned}
&\partial_t V(t, x, m) \\
&\quad + (a + q)(m - x)\partial_x V(t, x, m) + \frac{1}{2}(\epsilon - q^2)(m - x)^2 - \frac{1}{2}[\partial_x V(t, x, m)]^2 \\
&\quad + \frac{\sigma^2}{2}\partial_x^2 V(t, x, m) + \frac{\sigma^2}{2}\rho^2\partial_m^2 V(t, x, m) + \sigma^2\rho^2\partial_{xm}^2 V(t, x, m) = 0,
\end{aligned} \tag{32}$$

for all $(t, x, m) \in [0, T] \times \mathbb{R}^d \times \mathbb{R}^d$ provided we have

$$\int \partial_x V(t, x, m) d\mu(x) = 0, \qquad 0 \le t \le T, \tag{33}$$

Equation (33) being used to get rid of the interaction between μ_t and α^∞. Obviously, v^∞ satisfies (33). (Notice that this implies that the stochastic Kolmogorov equation becomes: $dm_t = \rho\sigma dW_t^0$.)

Equation (32) reads as the dynamics for the decoupling field permitting to express the value function V^μ as a function of the current statistical state μ_t of the population. We call it the master equation of the problem.

4 The Master Equation

While we only discussed mean field games so far, it turns out that the concept of master equation applies as well to the control of dynamics of McKean-Vlasov type whose solution also provides approximate equilibriums for large populations of individuals interacting through mean field terms. See Ref. [3] for a detailed analysis. We first outline a procedure common to the two problems. Next we specialize this procedure to the two cases of interest, deriving a master equation in each case. Finally, we highlight the differences to better understand what differentiates these two related and often confused problems.

4.1 General Set-Up

Stated in loose terms, the problem is to minimize the quantity

$$\mathbb{E}\Big[\int_0^T f(s, X_s^\alpha, \mu_s, \alpha_s)ds + g(X_T^\alpha, \mu_T)\Big] \tag{34}$$

over the space of square integrable $\mathbb{F}$-adapted controls $(\alpha_s)_{0\le s\le T}$ under the constraint that

$$\begin{aligned} dX_s^\alpha &= b\big(s, X_s^\alpha, \mu_s, \alpha_s\big)ds + \sigma(s, X_s^\alpha, \mu_s, \alpha_s)dW_s \\ &\quad + \int_\Xi \sigma^0(s, X_s^\alpha, \mu_s, \alpha_s, \xi)W^0(d\xi, ds). \end{aligned} \tag{35}$$

Yet the notion of what we call a minimizer must be specified. Obvious candidates for a precise definition of the minimization problem lead to different solutions. We consider two specifications: *mean field games* on the one hand, and *control of McKean-Vlasov dynamics* on the other.

1. When handling mean-field games, minimization is performed along a frozen flow of measures $(\mu_s = \hat{\mu}_s)_{0\le s\le T}$ describing a statistical equilibrium of the population. Then, the stochastic process $(\hat{X}_s)_{0\le s\le T}$ formed by the optimal paths of the optimal control problem (34) is required to satisfy the matching constraints $\hat{\mu}_s = \mathcal{L}(\hat{X}_s|\mathcal{F}_s^0)$ for $0 \le s \le T$. This is exactly the procedure described in Sect. 2.2.
2. Alternatively, minimization can be performed over the set of all the solutions of (35) subject to the McKean-Vlasov constraint $(\mu_s = \mu_s^\alpha)_{0\le s\le T}$, with $\mu_s^\alpha = \mathcal{L}(X_s^\alpha|\mathcal{F}_s^0)$ for $0 \le s \le T$, in which case the problem consists in minimizing the cost functional (34) over McKean-Vlasov diffusion processes.

As discussed painstakingly in Ref. [6], the two problems have different solutions since, in mean field games, the minimization is performed first and the fitting of the distribution of the optimal paths is performed next, whereas in the control of McKean-Vlasov dynamics, the McKean-Vlasov constraint is imposed first and the minimization is handled next. Still, we show here that both problems can be reformulated in terms of master equations, and we highlight the differences between the two equations resulting from these reformulations.

The main reason for handling both problems within the same framework is because in both cases, we rely on manipulations of a *value function* defined over the *enlarged* state space $\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d)$. For technical reasons, we restrict ourselves to measures in $\mathcal{P}_2(\mathbb{R}^d)$ which denotes the space of square integrable probability measures (i.e. probability measures with a finite second moment). For each $(t, x, \mu) \in [0, T] \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d)$, we would like to define $V(t, x, \mu)$ as the expected future costs:

$$V(t,x,\mu) = \mathbb{E}\bigg[\int_t^T f(s, X_s^{\hat{\alpha}}, \hat{\mu}_s, \hat{\alpha}_s)ds + g(X_T^{\hat{\alpha}}, \hat{\mu}_T)\big|X_t^{\hat{\alpha}} = x\bigg], \tag{36}$$

where $\hat{\alpha}$ minimizes the quantity (34) when we add the constraint $\mu_t = \mu$ and compute the time integral between t and T. In other words:

$$(\hat{\alpha}_s)_{t\le s\le T} = \text{argmin}_{\alpha}\mathbb{E}\bigg[\int_t^T f(s, X_s^{\alpha}, \mu_s, \alpha_s)ds + g(X_T^{\alpha}, \mu_T)\bigg], \tag{37}$$

the rule for computing the infimum being articulated above, either from the mean field game procedure as in 1, or from the optimization over McKean-Vlasov dynamics as explained in 2. In both cases, the flow $(\hat{\mu}_s)_{t\le s\le T}$ appearing in (36) satisfies the fixed point condition $(\hat{\mu}_s = \mathcal{L}(X_s^{\hat{\alpha}}|\mathcal{F}_s^{0,t}))_{t\le s\le T}$, which is true in both cases as $(X_s^{\hat{\alpha}})_{t\le s\le T}$ is an optimal path. Here and in the following $(\mathcal{F}_s^{0,t})_{t\le s\le T}$ is the filtration generated by the future increments of the common noise W^0, in the sense that $\mathcal{F}_s^{0,t} = \sigma\{W_r^0 - W_t^0 : t \le r \le s\}$. Recall that we use the notation W_r^0 for $\{W^0(\Lambda, [0,r)\}_\Lambda$ when Λ varies through the Borel subsets of Ξ. Below, the symbol 'hat' always refers to optimal quantities, and $(X_s^{\hat{\alpha}})_{t\le s\le T}$ is sometimes denoted by $(\hat{X}_s)_{t\le s\le T}$.

Generally speaking, the definition of the (deterministic) function $V(t,x,\mu)$ makes sense whenever the minimizer $(\hat{\alpha}_s)_{t\le s\le T}$ exists and is unique. When handling mean-field games, some additional precaution is needed to guarantee the consistency of the definition. Basically, we also need that, given the initial distribution μ at time t, there exists a unique[1] equilibrium flow of conditional probability measures $(\hat{\mu}_s)_{t\le s\le T}$ satisfying $\hat{\mu}_t = \mu$ and $\hat{\mu}_s = \mathcal{L}(\hat{X}_s|\mathcal{F}_s^{0,t})$ for all $s \in [t,T]$, where $(\hat{X}_s)_{t\le s\le T}$ is the optimal path of the underlying minimization problem (performed under the fixed flow of measures $(\hat{\mu}_s)_{t\le s\le T}$). In that case, the minimizer $(\hat{\alpha}_s)_{t\le s\le T}$ reads as the optimal control of $(\hat{X}_s)_{t\le s\le T}$. In the case of the optimal control of McKean-Vlasov stochastic dynamics, minimization is performed over the set of *conditional McKean-Vlasov diffusion processes* with the prescribed initial distribution μ at time t, in other words, satisfying (35) with $\mathcal{L}(X_t) = \mu$ and $\mu_s = \mu_s^{\alpha} = \mathcal{L}(X_s^{\alpha}|\mathcal{F}_s^{0,t})$ for all $s \in [t,T]$. In that case, the mapping $(t,\mu) \mapsto \int_{\mathbb{R}^d} V(t,x,\mu)d\mu(x)$ appears as the value function of the optimal control problem:

$$\mathbb{E}\big[V(t,\chi,\mu)\big] = \inf_{\alpha}\mathbb{E}\bigg[\int_t^T f\big(s, X_s^{\alpha}, \mathcal{L}(X_s^{\alpha}|\mathcal{F}_s^{0,t}), \alpha_s\big)ds + g\big(X_T^{\alpha}, \mathcal{L}(X_T^{\alpha}|\mathcal{F}_T^{0,t})\big)\bigg], \tag{38}$$

subject to $X_t^{\alpha} = \chi$ where χ is a random variable with distribution μ, i.e. $\chi \sim \mu$.

---

[1] We refer to the Lasry-Lions monotonicity conditions in Ref. [2] for a typical set of assumptions under which uniqueness holds. See also Ref. [5] for a discussion of uniqueness in the presence of a common noise.

Our goal is to characterize the function V as the solution of a partial differential equation (PDE) on the space $[0,T] \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d)$. In the framework of mean-field games, such an equation was touted in several presentations, and called the *master equation*. See for example [2, 10, 16]. We discuss the derivation of this equation below in Sect. 4.4. Using a similar strategy, we also derive a master equation in the case of the optimal control of McKean-Vlasov stochastic dynamics in Sect. 4.5.

4.2 Dynamic Programming Principle

In order to understand better the definition (36), we consider the case in which the minimizer $(\hat{\alpha}_s)_{t \le s \le T}$ has a feedback form, namely $\hat{\alpha}_s$ reads as $\hat{\alpha}(s, X_s^{\hat{\alpha}}, \hat{\mu}_s)$ for some function $\hat{\alpha} : [0,T] \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$. In this case, (36) becomes

$$V(t,x,\mu) = \mathbb{E}\bigg[\int_t^T f\big(s, X_s^{\hat{\alpha}}, \hat{\mu}_s, \hat{\alpha}(s, X_s^{\hat{\alpha}}, \hat{\mu}_s)\big) ds + g(X_T^{\hat{\alpha}}, \hat{\mu}_T) \big| X_t^{\alpha} = x \bigg], \quad (39)$$

where $(X_s^{\hat{\alpha}})_{t \le s \le T}$ is the solution (if well-defined) of (35) with α_s replaced by $\hat{\alpha}(s, X_s^{\hat{\alpha}}, \hat{\mu}_s)$. It is worth recalling that, in that writing, $\hat{\mu}_s$ matches the conditional law $\mathcal{L}(X_s^{\hat{\alpha}} | \mathcal{F}_s^{0,t})$ and is forced to start from $\hat{\mu}_t = \mu$ at time t.

Following the approach used in finite dimension, a natural strategy is then to use (39) as a basis for deriving a dynamic programming principle for V. Quite obviously, a very convenient way to do so consists in requiring the optimal pair $(\hat{X}_s = X_s^{\hat{\alpha}}, \hat{\mu}_s)_{t \le s \le T}$ to be Markov in $\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d)$, in which case we get

$$\begin{aligned} V(t+h, X_{t+h}^{\hat{\alpha}}, \hat{\mu}_{t+h}) = \mathbb{E}\bigg[&\int_{t+h}^T f(s, X_s^{\hat{\alpha}}, \hat{\mu}_s, \hat{\alpha}_s) ds \\ &+ g(X_T^{\hat{\alpha}}, \hat{\mu}_T) \big| \mathcal{F}_{t+h}^{0,t} \vee \sigma\big\{ X_t^{\hat{\alpha}}, (W_s - W_t)_{s \in [t,t+h]} \big\} \bigg]. \end{aligned}$$

Here, the σ-field $\mathcal{F}_{t+h}^{0,t} \vee \sigma\{X_t^{\hat{\alpha}}, (W_s - W_t)_{s \in [t,t+h]}\}$ comprises all the relevant events observed up until time $t+h$.

The rigorous proof of the Markov property for the path $(\hat{X}_s = X_s^{\hat{\alpha}}, \hat{\mu}_s)_{t \le s \le T}$ is left open. Intuitively, it sounds reasonable to expect that the Markov property holds if, for any initial distribution μ, there exists a unique equilibrium $(\hat{\mu}_s)_{t \le s \le T}$ starting from $\hat{\mu}_t = \mu$ at time $t \in [0,T]$. The reason is that, when uniqueness holds, there is no need to investigate the past of the optimal path in order to decide of the future of the dynamics. Such an argument is somehow quite generic in probability theory. In particular, the claim is expected to be true in both cases, whatever the meaning of what an equilibrium is. Of course, this suggests that the following dynamic version of (36)

$$V(t,x,\mu) = \mathbb{E}\Big[\int_t^{t+h} f(s, X_s^{\hat{\alpha}}, \hat{\mu}_s, \hat{\alpha}_s)ds + V\big(t+h, X_{t+h}^{\hat{\alpha}}, \hat{\mu}_{t+h}\big)\big| X_t^{\hat{\alpha}} = x\Big] \quad (40)$$

must be valid. The fact that (40) should be true in both cases is the starting point for our common analysis of the master equation. For instance, as a by-product of (40), we can derive a variational form of the dynamic programming principle:

$$\mathbb{E}\big[V(t,\chi,\mu)\big] = \inf \mathbb{E}\Big[\int_t^{t+h} f(s, X_s^{\alpha}, \mu_s, \alpha_s)ds + V(t+h, X_{t+h}^{\alpha}, \mu_{t+h})\Big], \quad (41)$$

which must be true in both cases as well, provided the random variable χ has distribution μ, i.e. $\chi \sim \mu$, and is $\mathcal{F}_t$-measurable, the minimization being defined as above according to the situation we are considering.

The proof of (41) is as follows. First, we observe from (39) that (41) must be valid when $t + h = T$. Then, (40) implies that the left-hand side is greater than the ride-hand side by choosing $(\hat{\alpha}_s)_{t \le s \le T}$ as a control. To prove the converse inequality, we choose an arbitrary control $(\alpha_s)_{t \le s \le t+h}$ between times t and $t+h$. In the control of McKean-Vlasov dynamics, this means that the random measures $(\mu_s)_{t \le s \le t+h}$ are chosen accordingly, as they depend on $(\alpha_s)_{t \le s \le t+h}$, so that μ_{t+h} is equal to the conditional law of X_{t+h}^{α} at time $t + h$. At time $t + h$, this permits to switch to the optimal strategy starting from $(X_{t+h}^{\alpha}, \mu_{t+h})$. The resulting strategy is of a greater cost than the optimal one. By (39), this cost is exactly given by the right-hand side in (41).

In the framework of mean field games, the argument for proving that the left-hand side is less than the right-hand side in (41) is a bit different. Indeed, in that case, the flow $(\mu_s)_{t \le s \le T}$ is fixed and matches $(\hat{\mu}_s)_{t \le s \le T}$, so that $\hat{\alpha}(s, X_s^{\hat{\alpha}}, \hat{\mu}_s)$ appears as an optimal control for optimizing (34) in the *environment* $(\mu_s = \hat{\mu}_s)_{t \le s \le T}$. So in that case, $V(t,x,\mu)$ is expected to match the optimal conditional cost

$$V(t,x,\mu) = \inf \mathbb{E}\Big[\int_t^{T} f(s, X_s^{\alpha}, \hat{\mu}_s, \alpha_s)ds + g(X_T^{\alpha}, \hat{\mu}_T)\big| X_t^{\alpha} = x\Big], \quad (42)$$

where $(X_s^{\alpha})_{t \le s \le T}$ solves the SDE (35) with $(\mu_s = \hat{\mu}_s)_{t \le s \le T}$ therein. Going back to (41), the choice of an arbitrary control $(\alpha_s)_{t \le s \le t+h}$ between times t and $t+h$ doesn't affect the value of $(\mu_s)_{t \le s \le t+h}$, which remains equal to $(\hat{\mu}_s)_{t \le s \le t+h}$. At time $t + h$, this permits to switch to the optimal strategy starting from X_{t+h}^{α} in the environment $(\hat{\mu}_s)_{t \le s \le T}$. Again, the resulting strategy is of a greater cost than the optimal one and, by (39), this cost is exactly given by the right-hand side in (41).

We emphasize that, when controlling McKean-Vlasov dynamics, (42) fails as in that case, the flow of measures is not frozen during the minimization procedure. In particular, the fact that (42) holds true in mean-field games only suggests that V satisfies a stronger dynamic programming principle in that case:

$$V(t,x,\mu) = \inf \mathbb{E}\Bigg[\int_t^{t+h} f(s, X_s^\alpha, \hat{\mu}_s, \alpha_s)ds + V\big(t+h, X_{t+h}^\alpha, \hat{\mu}_{t+h}\big)\big|X_t^\alpha = x\Bigg]. \tag{43}$$

The reason is the same as above. On the one hand, (40) implies that the left-hand side is greater than the ride-hand side by choosing $(\hat{\alpha}_s)_{t\le s\le T}$ as a control. On the other hand, choosing an arbitrary control $(\hat{\alpha}_s)_{t\le s\le t+h}$ between t and $t+h$ and switching to the optimal control starting from X_{t+h}^α in the environment $(\hat{\mu}_s)_{t\le s\le T}$, the left-hand side must be less than the right-hand side.

4.3 Derivation of the Master Equation

As illustrated earlier (see also the discussion of the second example below), the derivation of the master equation can be based on a suitable chain rule for computing the dynamics of V along paths of the form (35). This requires V to be smooth enough in order to apply an *Itô-like formula*.

In the example considered in the previous section, the dependence of V upon the measure reduces to a dependence upon the mean of the measure, and a standard version of Itô's formula could be used. In general, the measure argument lives in infinite dimension and different tools are needed. The approach advocated by P.L. Lions in his lectures at the *Collège de France* suggests to *lift-up* the mapping V into

$$\tilde{V} : [0,T] \times \mathbb{R}^d \times L^2(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}}; \mathbb{R}^d) \ni (t,x,\tilde{\chi}) \mapsto \tilde{V}(t,x,\tilde{\chi}) = V(t,x,\mathcal{L}(\tilde{\chi})),$$

where $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$ can be viewed as a copy of the space $(\Omega, \mathcal{F}, \mathbb{P})$. The resulting $\tilde{V}$ is defined on the product of $[0,T] \times \mathbb{R}^d$ and a Hilbert space, for which the standard notion of Fréchet differentiability can be used. Demanding V to be smooth in the measure argument is then understood as demanding $\tilde{V}$ to be smooth in the Fréchet sense. In that perspective, expanding $(V(s, X_s^\alpha, \mu_s))_{t\le s\le T}$ is then the same as expanding $(\tilde{V}(s, X_s^\alpha, \tilde{\chi}_s))_{t\le s\le T}$, where the process $(\tilde{\chi}_s)_{t\le s\le T}$ is an Itô process with $(\mu_s)_{t\le s\le T}$ as flow of marginal conditional distributions (conditional on $\mathcal{F}^{0,t}$).

The fact that we require $(\tilde{\chi}_s)_{t\le s\le T}$ to have $(\mu_s)_{t\le s\le T}$ as flow of marginal conditional distributions calls for some precaution in the construction of the lifting. A way to do just this consists in writing $(\Omega, \mathcal{F}, \mathbb{P})$ in the form $(\Omega^0 \times \Omega^1, \mathcal{F}^0 \otimes \mathcal{F}^1, \mathbb{P}^0 \otimes \mathbb{P}^1)$, $(\Omega^0, \mathcal{F}^0, \mathbb{P}^0)$ supporting the common noise W^0, and $(\Omega^1, \mathcal{F}^1, \mathbb{P}^1)$ the idiosyncratic noise W. So an element $\omega \in \Omega$ can be written as $\omega = (\omega^0, \omega^1) \in \Omega^0 \times \Omega^1$.

Considering a copy $(\tilde{\Omega}^1, \tilde{\mathcal{F}}^1, \tilde{\mathbb{P}}^1)$ of the space $(\Omega^1, \mathcal{F}^1, \mathbb{P}^1)$, it then makes sense to consider the process $(\tilde{\chi}_s)_{t \le s \le T}$ as the solution of an equation of the same form of (35), but on the space $(\Omega^0 \times \tilde{\Omega}^1, \mathcal{F}^0 \otimes \tilde{\mathcal{F}}^1, \mathbb{P}^0 \otimes \tilde{\mathbb{P}}^1)$, $(\tilde{\Omega}^1, \tilde{\mathcal{F}}^1, \tilde{\mathbb{P}}^1)$ being endowed with a copy $\tilde{W}$ of W. The realization at some $\omega^0 \in \Omega^0$ of the conditional law of $\tilde{\chi}_s$ given $\mathcal{F}^0$ then reads as the law of the random variable $\tilde{\chi}_s(\omega^0, \cdot) \in L^2(\tilde{\Omega}^1, \tilde{\mathcal{F}}^1, \tilde{\mathbb{P}}^1; \mathbb{R}^d)$. Put in our framework, this makes rigorous the identification of $\mathcal{L}(\tilde{\chi}_s(\omega^0, \cdot))$ with $\mu_s(\omega^0)$.

Generally speaking, we expect that $(\tilde{V}(s, X_s^\alpha, \tilde{\chi}_s) = \tilde{V}(s, X_s^\alpha(\omega^0, \omega^1), \tilde{\chi}_s(\omega^0, \cdot)))_{t \le s \le T}$ can be expanded as

$$d\tilde{V}\big(s, X_s^\alpha, \tilde{\chi}_s\big) = \big[\partial_t \tilde{V}(s, X_s^\alpha, \tilde{\chi}_s) + A_x^\alpha \tilde{V}(s, X_s^\alpha, \tilde{\chi}_s) + A_\mu^\alpha \tilde{V}(s, X_s^\alpha, \tilde{\chi}_s) + A_{x\mu}^\alpha \tilde{V}(s, X_s^\alpha, \tilde{\chi}_s)\big]ds + dM_s, \quad t \le s \le T, \tag{44}$$

with $\tilde{V}(T, x, \tilde{\chi}) = g(x, \mathcal{L}(\tilde{\chi}))$ as terminal condition, where

(i) A_x^α denotes the second-order differential operator associated to the process $(X_s^\alpha)_{t \le s \le T}$. It acts on functions of the state variable $x \in \mathbb{R}^d$ and thus on the variable x in $\tilde{V}(t, x, \tilde{\chi})$ in (44).
(ii) A_μ^α denotes some second-order differential operator associated to the process $(\tilde{\chi}_s)_{t \le s \le T}$. It acts on functions from $L^2(\tilde{\Omega}^1, \tilde{\mathcal{F}}^1, \tilde{\mathbb{P}}^1; \mathbb{R}^d)$ into $\mathbb{R}$ and thus on the variable $\tilde{\chi}$ in $\tilde{V}(t, x, \tilde{\chi})$.
(iii) $A_{x\mu}^\alpha$ denotes some second-order differential operator associated to the cross effect of $(X_s^\alpha)_{t \le s \le T}$ and $(\tilde{\chi}_s)_{t \le s \le T}$, as both feel the same noise W^0. It acts on functions from $\mathbb{R}^d \times L^2(\tilde{\Omega}^1, \tilde{\mathcal{F}}^1, \tilde{\mathbb{P}}^1; \mathbb{R}^d)$ into $\mathbb{R}$ and thus on the variables $(x, \tilde{\chi})$ in $\tilde{V}(t, x, \tilde{\chi})$.
(iv) $(M_s)_{t \le s \le T}$ is a martingale.

A proof of (44) is given in the appendix at the end of the paper. Observe that $A_{x\mu} \equiv 0$ if there is no common noise W^0. Plugging (44) into (41) and letting h tend to 0, we then expect:

$$\partial_t \mathbb{E}\big[\tilde{V}(t, \chi, \tilde{\chi})\big] + \inf_\alpha \mathbb{E}\big[A_x^\alpha \tilde{V}(t, \chi, \tilde{\chi}) + A_\mu^\alpha \tilde{V}(t, \chi, \tilde{\chi}) + A_{x\mu}^\alpha \tilde{V}(t, \chi, \tilde{\chi}) + f(t, \chi, \mu, \alpha)\big] = 0, \tag{45}$$

where χ and $\tilde{\chi}$ random variables defined on $(\Omega^1, \mathcal{F}^1, \mathbb{P}^1)$ and $(\tilde{\Omega}^1, \tilde{\mathcal{F}}^1, \tilde{\mathbb{P}}^1)$ respectively, both being distributed according to μ. If the minimizer has a feedback form, namely if the optimization over random variables α reduces to optimization over random variables of the form $\alpha(t, \chi, \mu)$, α being a function defined on $[0, T] \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d)$, then if we denote by $\hat{\alpha}$ the optimum, the same strategy applied to (40), shows that $\tilde{V}$ satisfies the master equation

$$\partial_t \tilde{V}(t, x, \tilde{\chi}) + A_x^{\hat{\alpha}(t,x,\mu)} \tilde{V}(t, x, \tilde{\chi}) + A_\mu^{\hat{\alpha}(t,x,\mu)} \tilde{V}(t, x, \tilde{\chi}) + A_{x\mu}^{\hat{\alpha}(t,x,\mu)} \tilde{V}(t, x, \tilde{\chi}) + f\big(t, x, \mu, \hat{\alpha}(t, x, \mu)\big) = 0. \tag{46}$$

Of course, the rule for computing the infimum in (45) depends upon the framework. In the case of the optimal control of McKean-Vlasov diffusion processes, $(\tilde{\chi}_s(\omega^0, \tilde{\omega}^1))_{t \le s \le T}$ in (44) is chosen as a copy, denoted by $(\tilde{X}_s^\alpha(\omega^0, \tilde{\omega}^1))_{t \le s \le T}$, of $(X_s^\alpha(\omega^0, \omega^1))_{t \le s \le T}$ on the space $(\Omega^0 \times \tilde{\Omega}^1, \mathcal{F}^0 \otimes \tilde{\mathcal{F}}^1, \mathbb{P}^0 \otimes \tilde{\mathbb{P}}^1)$. In that case, A_μ^α depends on α explicitly. In the framework of mean field games, $(\tilde{\chi}_s(\omega^0, \tilde{\omega}^1))_{t \le s \le T}$ is chosen as a copy of the optimal path $(\hat{X}_s)_{t \le s \le T}$ of the optimization problem (36) under the statistical equilibrium flow initialized at μ at time t. Such a choice for $\tilde{\chi}$ is dictated by the optimization procedure (6), in which the flow of measures is chosen as the flow of measures at equilibrium. Since $\tilde{\chi}$ does not depend on α, neither does A_μ^α. Therefore, $A_\mu = A_\mu^\alpha$ has no role in the computation of the infimum.

For the sake of illustration, we further specialize the form of of the master equation (46) to the simpler case when (35) reduces to

$$dX_s = b(s, X_s, \mu_s, \alpha_s)ds + \sigma(X_s)dW_s + \sigma^0(X_s)dW_s^0.$$

In that case, we know from the results presented in the appendix that

$$\begin{aligned}
A_x^\alpha \tilde{\varphi}(t, x, \tilde{\chi}) &= \langle b\big(t, x, \mathcal{L}(\tilde{\chi}), \alpha\big), \partial_x \tilde{\varphi}(t, x, \tilde{\chi})\rangle \\
&\quad + \frac{1}{2}\mathrm{Trace}\Big[\big[\sigma(x)\big(\sigma(x)\big)^\dagger + \sigma^0(x)\big(\sigma^0(x)\big)^\dagger\big]\partial_x^2 \tilde{\varphi}(t, x, \tilde{\chi})\Big], \\
A_\mu^\alpha \tilde{\varphi}(t, x, \tilde{\chi}) &= b\big(t, \tilde{\chi}, \mathcal{L}(\tilde{\chi}), \tilde{\beta}\big) \cdot D_\mu \tilde{\varphi}(t, x, \tilde{\chi}) \\
&\quad + \frac{1}{2} D_\mu^2 \tilde{\varphi}\big(t, x, \tilde{\chi}\big)\big[\sigma^0(\tilde{\chi}), \sigma^0(\tilde{\chi})\big] + \frac{1}{2} D_\mu^2 \tilde{\varphi}\big(t, x, \tilde{\chi}\big)\big[\sigma(\tilde{\chi})\tilde{G}, \sigma(\tilde{\chi})\tilde{G}\big], \\
A_{x\mu}^\alpha \tilde{\varphi}(t, x, \tilde{\chi}) &= \big\langle \big\{\partial_x D_\mu \tilde{\varphi}\big(t, x, \tilde{\chi}\big) \cdot \sigma^0(\tilde{\chi})\big\}, \sigma^0(x)\big\rangle,
\end{aligned} \tag{47}$$

where $\tilde{G}$ is an $\mathcal{N}(0, 1)$ random variable on the space $(\tilde{\Omega}^1, \tilde{\mathcal{F}}^1, \tilde{\mathbb{P}}^1)$, independent of $\tilde{W}$. The notations D_μ and D_μ^2 refer to Fréchet derivatives of smooth functions on the space $L^2(\tilde{\Omega}^1, \tilde{\mathcal{F}}^1, \tilde{\mathbb{P}}^1; \mathbb{R}^d)$. For a random variable $\tilde{\zeta} \in L^2(\tilde{\Omega}^1, \tilde{\mathcal{F}}^1, \tilde{\mathbb{P}}^1; \mathbb{R}^d)$, the notation $D_\mu \tilde{\varphi}(t, x, \tilde{\chi}) \cdot \tilde{\zeta}$ denotes the action of the differential of $\tilde{\varphi}(t, x, \cdot)$ at point $\tilde{\chi}$ along the direction $\tilde{\zeta}$. Similarly, the notation $D_\mu^2 \tilde{\varphi}(t, x, \tilde{\chi})[\tilde{\zeta}, \tilde{\zeta}]$ denotes the action of the second-order differential of $\tilde{\varphi}(t, x, \cdot)$ at point $\tilde{\chi}$ along the directions $(\tilde{\zeta}, \tilde{\zeta})$. We refer to the appendix for a more detailed account.

Notice that $\tilde{\chi}$ in $A_\mu^\alpha \tilde{\varphi}(t, x, \tilde{\chi})$ denotes the copy of χ, χ standing for the value at time t of the controlled diffusion process $(\chi_s)_{t \le s \le T}$. The control process driving $(\chi_s)_{t \le s \le T}$ is denoted by $(\beta_s)_{t \le s \le T}$. Specifying the values of χ and β according to the framework used for performing the optimization, we derive below the appropriate form of the resulting master equation. Notice also that $A_{x\mu}^\alpha \tilde{\varphi}(t, x, \tilde{\chi})$ does not depend upon α as the coefficients σ^0 and σ do not depend on it.

4.4 The Case of Mean Field Games

In the framework of Mean-Field Games, $(\tilde{\chi}_s)_{t\le s\le T}$ is chosen as a copy of the optimal path $(\hat{X}_s)_{t\le s\le T}$. This says that, in (47), $\tilde{\chi}$ stands for the value at time t of the optimally controlled state from the optimization problem (36) under the statistical equilibrium flow initialized at μ at time t. Therefore, the minimization in (45) reduces to

$$\begin{aligned}&\inf_{\alpha} \mathbb{E}\big[\langle b(t,\chi,\mu,\alpha), \partial_x \tilde{V}(t,\chi,\tilde{\chi})\rangle + f(t,\chi,\mu,\alpha)\big] \\ &\quad = \inf_{\alpha} \mathbb{E}\big[\langle b(t,\chi,\mu,\alpha), \partial_x V(t,\chi,\mu)\rangle + f(t,\chi,\mu,\alpha)\big],\end{aligned} \tag{48}$$

the equality following from the fact that $\partial_x \tilde{V}(t,x,\tilde{\chi})$ is the same as $\partial_x V(t,x,\mu)$ (as the differentiation is performed in the component x).

Assume now that there exists a measurable mapping $\bar{\alpha} : [0,T]\times\mathbb{R}^d\times\mathcal{P}_2(\mathbb{R}^d)\times\mathbb{R}^d \ni (t,x,\mu)\mapsto \bar{\alpha}(t,x,\mu,y)$, providing the argument of the minimization:

$$\bar{\alpha}(t,x,\mu,y) = \arg\inf_{\alpha\in A} H(t,x,\mu,y,\alpha), \tag{49}$$

where the reduced Hamiltonian H is defined as:

$$H(t,x,\mu,y,\alpha) = \langle b(t,x,\mu,\alpha), y\rangle + f(t,x,\mu,\alpha), \tag{50}$$

Then, the minimizer in (48) must be $\alpha = \bar{\alpha}(t,\chi,\mu,\partial_x V(t,\chi,\mu))$, hence showing that $\hat{\alpha}(t,x,\mu) = \bar{\alpha}(t,x,\mu,\partial_x V(t,x,\mu))$ is an optimal feedback. By (46), the master equation reads

$$\begin{aligned}&\partial_t \tilde{V}(t,x,\tilde{\chi}) + \inf_{\alpha} H\big(t,x,\mu,\partial_x \tilde{V}(t,x,\tilde{\chi}),\alpha\big) + \big(A_\mu + A_{x\mu}\big)\tilde{V}(t,x,\tilde{\chi}) \\ &\quad + \frac{1}{2}\mathrm{Trace}\Big[\big[\sigma(x)\big(\sigma(x)\big)^\dagger + \sigma^0(x)\big(\sigma^0(x)\big)^\dagger\big]\partial_x^2 \tilde{V}(t,x,\tilde{\chi})\Big] = 0,\end{aligned} \tag{51}$$

the optimization over α being now performed in the set A (and thus in finite dimension).

By identification of the transport term, this says that the statistical equilibrium of the MFG with μ as initial distribution must be given by the solution of the conditional McKean-Vlasov equation:

$$d\hat{X}_s = b\big(s,\hat{X}_s,\hat{\mu}_s,\bar{\alpha}\big(s,\hat{X}_s,\hat{\mu}_s,\partial_x V(s,\hat{X}_s,\hat{\mu}_s)\big) + \sigma\big(\hat{X}_s\big)dW_s + \sigma^0\big(\hat{X}_s\big)dW_s^0, \tag{52}$$

subject to the constraint $\hat{\mu}_s = \mathcal{L}(\hat{X}_s|\mathcal{F}_s^0)$ for $s\in[t,T]$, with $\hat{X}_t \sim \mu$. We indeed claim

Proposition 4.1 *On the top of the above assumptions and notations, assume that, for all $t\in[0,T]$, $x\in\mathbb{R}^d$ and $\mu\in\mathcal{P}_2(\mathbb{R}^d)$*

$$|\bar{\alpha}(t, x, \mu, y)| \leq C\left[1 + |x| + |y| + \left(\int_{\mathbb{R}^d} |x'|^2 d\mu(x')\right)^{1/2}\right], \tag{53}$$

and that the coefficients b, σ and σ^0 satisfy a similar bound. Assume also that $\tilde{V}$ is a (classical) solution of (51) *satisfying, for all $t \in [0, T]$, $x \in \mathbb{R}^d$ and $\tilde{\chi} \in L^2(\tilde{\Omega}^1, \tilde{\mathcal{F}}^1, \tilde{\mathbb{P}}^1; \mathbb{R}^d)$,*

$$|\partial_x \tilde{V}(t, x, \tilde{\chi})| + \|D_\mu \tilde{V}(t, x, \tilde{\chi})\|_{L^2(\tilde{\Omega}^1)} \leq C\Big(1 + |x| + \|\tilde{\chi}\|_{L^2(\tilde{\Omega}^1)}\Big), \tag{54}$$

and that, for any initial condition $(t, \mu) \in [0, T] \times \mathcal{P}_2(\mathbb{R}^d)$, Eq. (52) *has a unique solution. Then, the flow $(\mathcal{L}(\hat{X}_s|\mathcal{F}_s^0))_{t \leq s \leq T}$ solves the mean field game with (t, μ) as initial condition.*

Proof The proof consists of a verification argument. First, we notice from (53) and (54) that the solution of (52) is square integrable in the sense that its supremum norm over $[0, T]$ is square integrable. Similarly, for any square integrable control α, the supremum of X^α (with $X_t^\alpha \sim \mu$) is square integrable. Next we plug $\hat{\mu}_s = \mathcal{L}(\hat{X}_s|\mathcal{F}_s^0)$ in the right-hand side of (42), replace g by $V(T, \cdot, \cdot)$ and apply the version of Itô's formula proven in the appendix (see Proposition 6.5), using the growth and integrability assumptions to guarantee that the expectation of the martingale part is zero. We conclude that the right-hand side is indeed greater than $V(t, x, \mu)$. Choosing $(\alpha_s = \bar{\alpha}(s, \hat{X}_s, \hat{\mu}_s, \partial_x V(s, \hat{X}_s, \hat{\mu}_s))_{t \leq s \leq T}$, equality holds. This proves that $(\hat{X}_s)_{t \leq s \leq T}$ is a minimization path of the optimization problem driven by its own flow of conditional distributions, which is precisely the definition of an MFG equilibrium. □

Remark 4.2 Proposition 4.1 says that the solution of the master equation (51) contains all the information needed to solve the mean field game problem. It implies that the flow of conditional distributions $(\hat{\mu}_s = \mathcal{L}(\hat{X}_s|\mathcal{F}_s^0))_{t \leq s \leq T}$ solves the SPDE (2), with $\alpha(s, \cdot, \hat{\mu}_s) = \bar{\alpha}(s, x, \hat{\mu}_s, \partial_x V(s, x, \hat{\mu}_s))$.

Remark 4.3 Notice that $(Y_s = \partial_x V(s, \hat{X}_s, \hat{\mu}_s))_{t \leq s \leq T}$ may be reinterpreted as the adjoint process in the stochastic Pontryagin principle derived for mean field games in Ref. [4] (at least when there is no common noise W^0). Furthermore, the function $(t, x, \mu) \mapsto \partial_x V(t, x, \mu)$ appears as the decoupling field of the McKean-Vlasov FBSDE derived from the stochastic Pontryagin principle. It plays the same role as the gradient of the value function in standard optimal control theory. See Sect. 4.6.

4.5 The Case of the Control of McKean-Vlasov Dynamics

When handling the control of McKean-Vlasov dynamics, $(\tilde{\chi}_s)_{t \leq s \leq T}$ is chosen as a copy of $(X_s^\alpha)_{t \leq s \leq T}$. So if $\tilde{\alpha}$ denotes a copy of α, the minimization in (45) takes the form

$$
\begin{aligned}
&\inf_{\alpha} \mathbb{E}\big[\langle b(t,\chi,\mu,\alpha), \partial_x \tilde{V}(t,\chi,\tilde{\chi})\rangle + b(t,\tilde{\chi},\mu,\tilde{\alpha}) \cdot D_\mu \tilde{V}(t,\chi,\tilde{\chi}) + f(t,\chi,\mu,\alpha)\big] \\
&\quad = \inf_{\alpha} \mathbb{E}^1\Big[\langle b(t,\chi,\mu,\alpha), \partial_x V(t,\chi,\mu)\rangle + \tilde{\mathbb{E}}^1\big[\langle b(t,\tilde{\chi},\mu,\tilde{\alpha}), \partial_\mu V(t,\chi,\mu)(\tilde{\chi})\rangle\big] \\
&\qquad\qquad + f(t,\chi,\mu,\alpha)\Big],
\end{aligned}
$$

where the function $\partial_\mu V(t,x,\mu)(\cdot)$ *represents* the Fréchet derivative $D_\mu \tilde{V}(t,x,\tilde{\chi})$, that is $D_\mu \tilde{V}(t,x,\tilde{\chi}) = \partial_\mu V(t,x,\mu)(\tilde{\chi})$. See the appendix at the end of the paper for details on the definitions and the properties of these differentials. By Fubini's theorem, the minimization can be reformulated as

$$
\inf_{\alpha} \mathbb{E}^1\Big[\big\langle b(t,\chi,\mu,\alpha), \partial_x V(t,\chi,\mu) + \tilde{\mathbb{E}}^1\big[\partial_\mu V(t,\tilde{\chi},\mu)(\chi)\big]\big\rangle + f(t,\chi,\mu,\alpha)\Big]. \tag{55}
$$

The strategy is then the same as in the previous subsection. Assume indeed that there exists a measurable mapping $\bar{\alpha} : [0,T] \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \times \mathbb{R}^d \ni (t,x,\mu,y) \mapsto \bar{\alpha}(t,x,\mu,y)$ minimizing the reduced Hamiltonian as in (49), then the minimizer in (55) must be

$$
\begin{aligned}
\hat{\alpha} &= \bar{\alpha}\big(t,\chi,\mu,\partial_x V(t,\chi,\mu) + \tilde{\mathbb{E}}^1[\partial_\mu V(t,\tilde{\chi},\mu)(\chi)]\big) \\
&= \bar{\alpha}\bigg(t,\chi,\mu,\partial_x V(t,\chi,\mu) + \int_{\mathbb{R}^d} \partial_\mu V(t,x',\mu)(\chi) d\mu(x')\bigg),
\end{aligned}
$$

showing that $\hat{\alpha}(t,x,\mu) = \bar{\alpha}(t,x,\mu,\partial_x V(t,x,\mu) + \int_{\mathbb{R}^d} \partial_\mu V(t,x',\mu)(x) d\mu(x'))$ is an optimal feedback. By (46), this permits to make explicit the form of the master equation:

$$
\begin{aligned}
&\partial_t \tilde{V}(t,x,\tilde{\chi}) + \langle b\big(t,x,\mu,\hat{\alpha}(t,x,\mu)\big), \partial_x \tilde{V}(t,x,\tilde{\chi})\rangle \\
&\quad + b\big(t,\tilde{\chi},\mu,\hat{\alpha}(t,\tilde{\chi},\mu)\big) \cdot D_\mu \tilde{V}(t,x,\tilde{\chi}) \\
&\quad + \frac{1}{2}\mathrm{Trace}\Big[\big[\sigma(x)\big(\sigma(x)\big)^\dagger + \sigma^0(x)\big(\sigma^0(x)\big)^\dagger\big]\partial_x^2 \tilde{V}(t,x,\tilde{\chi})\Big] \\
&\quad + \frac{1}{2}\Big[D_\mu^2 \tilde{V}\big(t,x,\tilde{\chi}\big)\big[\sigma^0(\tilde{\chi}),\sigma^0(\tilde{\chi})\big] + D_\mu^2 \tilde{V}\big(t,x,\tilde{\chi}\big)\big[\sigma(\tilde{\chi})\tilde{G},\sigma(\tilde{\chi})\tilde{G}\big]\Big] \\
&\quad + \langle\big\{\partial_x D_\mu \tilde{V}\big(t,x,\tilde{\chi}\big) \cdot \sigma^0(\tilde{\chi})\big\}, \sigma^0(x)\rangle + f\big(t,x,\mu,\hat{\alpha}(t,x,\mu)\big) = 0.
\end{aligned} \tag{56}
$$

Moreover, the optimal path solving the optimal control of McKean-Vlasov dynamics is given by:

$$
\begin{aligned}
&d\hat{X}_s \\
&= b\bigg[s,\hat{X}_s,\hat{\mu}_s,\bar{\alpha}\bigg(s,\hat{X}_s,\hat{\mu}_s,\partial_x V(s,\hat{X}_s,\hat{\mu}_s) + \int_{\mathbb{R}^d} \partial_\mu V(s,x',\hat{\mu}_s)(\hat{X}_s) d\hat{\mu}_s(x')\bigg)\bigg]ds \\
&\quad + \sigma\big(\hat{X}_s\big)dW_s + \sigma^0\big(\hat{X}_s\big)dW_s^0,
\end{aligned} \tag{57}
$$

subject to the constraint $\hat{\mu}_s = \mathcal{L}(\hat{X}_s|\mathcal{F}^0_s)$ for $s \in [t, T]$, with $\hat{X}_t \sim \mu$. We indeed claim

Proposition 4.4 *On the top of the assumptions and notations introduced above, assume that $\bar{\alpha}$, b, σ and σ^0 satisfy (53), that $\tilde{V}$ is a classical solution of (56) satisfying, for all $t \in [0, T]$, $x \in \mathbb{R}^d$ and $\tilde{\chi} \in L^2(\tilde{\Omega}^1, \tilde{\mathcal{F}}^1, \tilde{\mathbb{P}}^1; \mathbb{R}^d)$,*

$$|\partial_x \tilde{V}(t, x, \tilde{\chi})| + \|D_\mu \tilde{V}(t, x, \tilde{\chi})\|_{2,\tilde{\Omega}^1} \leq C\Big(1 + |x| + \|\tilde{\chi}\|_{L^2(\tilde{\Omega}^1)}\Big), \tag{58}$$

and that, for any initial condition $(t, \mu) \in [0, T] \times \mathcal{P}_2(\mathbb{R}^d)$, Eq. (57) has a unique solution. Then, the flow $(\mathcal{L}(\hat{X}_s|\mathcal{F}^0_s))_{t \leq s \leq T}$ solves the minimization problem (34) over controlled McKean-Vlasov dynamics.

Proof The proof consists again of a verification argument. As for mean field games, we notice from (53) and (58) that the supremum over $[0, T]$ of the solution of (57) is square integrable and that, for any square integrable control α, the supremum of X^α (with $X^\alpha_t \sim \mu$) is also square integrable. Next, we replace g by $V(T, \cdot, \cdot)$ in (38), and apply the version of Itô's formula proven in the appendix (see Proposition 6.5), the integrability condition (58) ensuring that the expectation of the martingale part is zero. Using the same Fubini argument as in (55), we deduce that the right-hand side is indeed greater than $\mathbb{E}[V(t, \chi, \mu)]$. Choosing $\alpha_s = \hat{\alpha}(s, \hat{X}_s, \hat{\mu}_s)$, with $\hat{\alpha}(t, x, \mu) = \bar{\alpha}(t, x, \mu, \partial_x V(t, x, \mu) + \int_{\mathbb{R}^d} \partial_\mu V(t, x', \mu)(x) d\mu(x'))$, equality must hold. □

Remark 4.5 Notice that the combination of the terms in $\hat{\alpha}$ in (56) does not read as an infimum, namely:

$$\begin{aligned}
&\langle b\big(t, x, \mu, \hat{\alpha}(t, x, \mu)\big), \partial_x \tilde{V}(t, x, \tilde{\chi})\rangle + b\big(t, \tilde{\chi}, \mu, \hat{\alpha}(t, \tilde{\chi}, \mu)\big) \cdot D_\mu \tilde{V}(t, x, \tilde{\chi}) \\
&\quad + f\big(t, x, \mu, \hat{\alpha}(t, x, \mu)\big) \neq \inf_\alpha \big[\langle b(t, x, \mu, \alpha), \partial_x \tilde{V}(t, x, \tilde{\chi})\rangle + b(t, x, \mu, \tilde{\alpha}\big) \\
&\quad \times D_\mu \tilde{V}(t, x, \tilde{\chi}) + f(t, x, \mu, \alpha)\big],
\end{aligned}$$

which shows that equation (56) cannot be put in a variational form of the same type as equation (51), the minimization in (51) being performed over $\alpha \in A$. The reason is that the minimization in (55) is performed over random variables, and not over finite dimensional variables, the functional to minimize being written as the integrated version of the one which is above.

Actually, the variational structure has to be read in (45). Under the assumption of Proposition 4.4, the map $[0, T] \times L^2(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^d) \ni (t, \chi) \mapsto \mathbb{E}[\tilde{V}(t, \chi, \tilde{\chi})]$ can be shown to satisfy (45) by taking expectation in (56), provided that the time derivative and the expectation can be interchanged.

Remark 4.6 The flow of conditional distributions $(\hat{\mu}_s = \mathcal{L}(\hat{X}_s|\mathcal{F}^0_s))_{t \leq s \leq T}$ solves an SPDE, of the same form as (2). The precise formulation of that SPDE is left to the reader.

Remark 4.7 Notice that $(\partial_x V(s, \hat{X}_s, \hat{\mu}_s) + \int_{\mathbb{R}^d} \partial_\mu V(s, x, \hat{\mu}_s)(\hat{X}_s) d\hat{\mu}_s(x))_{t \le s \le T}$ may be reinterpreted as the adjoint process in the stochastic Pontryagin principle derived for the control of McKean-Vlasov dynamics in Ref. [3] (at least when there is no common noise W^0). In particular, the function $(t, x, \mu) \mapsto \partial_x V(s, x, \mu) + \int_{\mathbb{R}^d} \partial_\mu V(s, x', \mu)(x) d\mu(x')$ reads as the decoupling field of the McKean-Vlasov FBSDE deriving from the stochastic Pontryagin principle for the control of McKean-Vlasov dynamics. It is interesting to notice that the fact that the formula contains two different terms is a perfect reflection of the backward propagation of the terminal condition of the FBSDE. Indeed, as seen in Ref. [3], this terminal condition has two terms corresponding to the partial derivatives of the terminal cost function g with respect to the state variable x and the distribution μ. See Sect. 4.6.

4.6 Viscosity Solutions

In the previous paragraph, we used the master equation within the context of a verification argument to identify optimal paths of the underlying optimal control problem, and we alluded to the connection with purely probabilistic methods derived from the Pontryagin stochastic maximum principle which works as follows: under suitable conditions, optimal paths are identified with the forward component of a McKean-Vlasov FBSDE. In that framework, our discussion permits to identify the gradient of the function V with the decoupling field of the FBSDE. This FBSDE has the form:

$$
\begin{aligned}
dX_s &= b(s, X_s, \mu_s, \bar{\alpha}(s, X_s, \mu_s, Y_s))ds + \sigma^0(X_s)dW_s^0 + \sigma(X_s)dW_s, \\
dY_s &= -\Psi\big(s, X_s, \nu_s, Y_s, \bar{\alpha}(s, X_s, \mu_s, Y_s)\big)ds + Z_s^0 dW_s^0 + Z_s dW_s, \\
Y_T &= \phi(X_T, \mu_T)
\end{aligned}
\tag{59}
$$

for some functions $(t, x, \nu, y, \alpha) \mapsto \Psi(t, x, \nu, y, \alpha)$ and $(x, \mu) \mapsto \phi(x, \mu)$, the McKean-Vlasov nature of the FBSDE being due to the constraints $\mu_s = \mathcal{L}(X_s|\mathcal{F}_s^0)$ and $\nu_s = \mathcal{L}((X_s, Y_s)|\mathcal{F}_s^0)$. The function $\bar{\alpha}$ is given by (49).

In the mean field game case, the stochastic Pontryagin principle takes the form

$$
\Psi(t, x, \nu, y, \alpha) = \partial_x H\big(t, x, \mu, y, \alpha\big), \quad \phi(x, \mu) = \partial_x g(x, \mu), \tag{60}
$$

where μ denotes the first marginal of ν, and

$$
\begin{aligned}
\Psi(t, x, \nu, y, \alpha) &= \partial_x H\big(t, x, \mu, y, \alpha\big) \\
&\quad + \int_{\mathbb{R}^d \times \mathbb{R}^d} \big(\partial_\mu H\big(t, x', \mu, y', \alpha'\big)(x)\big)_{|\alpha' = \hat{\alpha}(t, x', \mu, y')} \nu(dx', dy'), \\
\phi(x, \mu) &= \partial_x g(x, \mu) + \int_{\mathbb{R}^d} \partial_\mu g(x', \mu)(x) \mu(dx')
\end{aligned}
\tag{61}
$$

in the case of the control of McKean-Vlasov dynamics.

One may wonder if a converse to the strategy discussed previously is possible: how could we reconstruct a solution of the master equation from a purely probabilistic approach? Put it differently, given the solution of the McKean-Vlasov FBSDE characterizing the optimal path *via* the Pontryagin stochastic maximum principle, is it possible to reconstruct V and to prove that it satisfies a PDE or SPDE which we could identify to the *master equation*?

In the forthcoming paper [8], the authors investigate the differentiability of the flow of a McKean-Vlasov FBSDE and reconstruct , in some cases, V as a classical solution of the master equation.

A more direct approach consists in checking that V is a viscosity solution of the master equation. This direct approach was used in Ref. [2] for non-stochastic games. In all cases the fundamental argument relies on a suitable form of the dynamic programming principle. This was our motivation for the discussion in Sect. 4.2. Still we must remember that Sect. 4.2 remains mostly at the heuristic level, and that a complete proof of the dynamic programming principle in this context would require more work. This is where the stochastic maximum principle may help. If uniqueness of the optimal paths and of the equilibrium are known (see for instance [4] and [3]), then the definition of V in (36) makes sense. In this case, not only do we have the explicit form of the optimal paths, but the dynamic programming principle is expected to hold.

We refrain from going into the gory details in this review paper. Instead, we take the dynamic programming principle for granted. The question is then to derive the master equation solved by V in the viscosity sense, from the three possible versions (43), (40) and (41). In the present context, since differentiability with respect to one of the variables is done through a lifting of the functions, we will be using the following definition of viscosity solutions.

Definition 4.8 We say that V is a super-solution (resp. sub-solution) in the sense of viscosity of the master equation if whenever $(t, x, \mu) \in [0, T] \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d)$ and the function $[0, T] \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \ni (s, y, \nu) \mapsto \varphi(s, y, \nu)$ is continuously differentiable, once in the time variable s, and twice in the variables y and ν, satisfies $V(t, x, \mu) = \varphi(t, x, \mu)$ and $V(s, y, \nu) \geq \varphi(s, y, \nu)$ (resp. $V(s, y, \nu) \leq \varphi(s, y, \nu)$) for all (s, y, ν) then we have (45) and/or (46), with $\tilde{V}$ replaced by $\tilde{\varphi}$ and $= 0$ replaced by ≤ 0 (respectively by ≥ 0). Notice that the signs are reversed since the equation is set backward.

The reason why we say *and/or* might look rather strange. This will be explained below, the problem being actually more subtle than it seems at first.

Following the approach used in standard stochastic optimal control problems, the proof could consist in applying Itô's formula to $\tilde{\varphi}(s, X_s^{\hat{\alpha}}, \hat{\mu}_s)_{t \leq s \leq t+h}$. In fact, there is no difficulty in proving the *viscosity inequality* (46) by means of (40). Still, this result is rather useless as the optimizer $\hat{\alpha}$ is expected to depend upon the gradient of $\tilde{V}$ and much more, as $\hat{\alpha}$ reads as $\bar{\alpha}$ applied to the gradient of $\tilde{V}$. The question is thus to decide whether it makes sense to replace the gradient of $\tilde{V}$ in $\bar{\alpha}$ by the gradient of $\tilde{\varphi}$. To answer the question, we must distinguish the two problems:

1. In the framework of mean field games, the answer is yes. The reason is that, when V is smooth, the inequality $V \geq \varphi$ in the neighborhood of (t, x, μ) implies $\partial_x V(t, x, \mu) = \partial_x \varphi(t, x, \mu)$. This says that we expect $\tilde{\varphi}$ to satisfy (51) with $= 0$ replaced by ≤ 0. Actually, this can be checked rigorously by means of the stronger version (43) of the dynamic programming principle, following the discussion in Ref. [9].

2. Unfortunately, this is false when handling the control of McKean-Vlasov dynamics. Indeed, the gradient of V is then understood as $\partial_x V(t, x, \mu) + \int_{\mathbb{R}^d} \partial_\mu V(t, x', \mu)(x) d\mu(x')$, which is 'non-local' in the sense that it involves values of $V(t, x', \mu)$ for x' far away from x. In particular, there is no way one can replace $\partial_x V(t, x, \mu) + \int_{\mathbb{R}^d} \partial_\mu V(t, x', \mu)(x) d\mu(x')$ by $\partial_x \varphi(t, x, \mu) + \int_{\mathbb{R}^d} \partial_\mu \varphi(t, x', \mu)(x) d\mu(x')$ on the single basis of the comparison of φ and V. This implies that, in the optimal control of McKean-Vlasov dynamics, viscosity solutions must be discussed in the framework of (45). Obviously, this requires adapting the notion of viscosity solution as only the function $(t, \mu) \mapsto \int_{\mathbb{R}^d} V(t, x, \mu) d\mu(x)$ matters in the dynamic programming principle (41). Comparison is then done with test functions of the form $(t, \mu) \mapsto \int_{\mathbb{R}^d} \phi(t, x, \mu) d\mu(x)$ (or simply $\phi(t, \mu)$). The derivation of an inequality in (45) is then achieved by a new application of Itô's formula.

4.7 Comparison of the Two Master Equations

We repeatedly reminded the reader that the function V obtained in the case of mean field games (whether or not there is a common noise) *is not a value function* in the usual sense of optimal control. Indeed, solving a mean field game problem is finding a fixed point more than solving an optimization problem. For this reason, the master equation should not read (and should not be interpreted) as a Hamilton-Jacobi-Bellman equation. Indeed, even though the first terms in Eq. (51) are of Hamiltonian type, the extra term $\mathcal{A}_\mu$ (specifically the first order term in $\mathcal{A}_\mu$) shows that this equation is not an HJB equation. On the other hand, the previous subsection shows that the master equation for the control of McKean-Vlasov dynamics, which comes from an optimization problem, can be viewed as an HJB equation when put in the form (45). In that case, the solution reads as the value function $(t, \mu) \mapsto \int_{\mathbb{R}^d} V(t, x, \mu) d\mu(x)$ of the corresponding optimization problem.

In the case of mean-field games, the master equation (51) matches the one given in Ref. [1]. Another type of differential calculus is used in Ref. [1] for handling the infinite dimensional component, but the master equation is indeed the same. The reason is that the master equation has the same interpretation: the solution $V(t, x, \mu)$ is also defined as the value function of the game when, at time t, the population is initialized with the distribution μ and the representative player with the state x.

This identification can be checked by connecting the two types of differential calculus. Roughly speaking, it holds $\partial_v[\partial_m V(t,x,m)(v)] = \partial_\mu V(t,x,\mu)(v)$, where $V(t,x,m)$ in the left-hand side refers to the concepts used in Ref. [1] and $V(t,x,\mu)$ in the right-hand side refers to the concepts we use in this paper. Formally, m is intended to be the density of μ. The notation $\partial_m V(t,x,m)(v)$ refers to the differential calculus used in Ref. [1] for differentiating a functional of m in the Gâteaux sense. The gradient $\partial_m V(t,x,m)$ has to be understood as a function $\mathbb{R}^d \ni v \mapsto \partial_m V(t,x,m)(v)$. The notation $\partial_\mu V(t,x,\mu)(v)$ is explained in detail in Sect. 6. Essentially, this function $\mathbb{R}^d \ni v \mapsto \partial_\mu V(t,x,\mu)(v)$ is such that $D_\mu \tilde{V}(t,x,\tilde{\chi}) = \partial_\mu V(t,x,\mu)(\tilde{\chi})$ whenever χ has law μ. The connection can be easily checked when $V(t,x,m) = \int_{\mathbb{R}^d} \varphi(t,x,v)m(v)dv$ or equivalently $V(t,x,\mu) = \int_{\mathbb{R}^d} \varphi(t,x,v)d\mu(v)$ for a test function φ. Then, $\partial_m V(t,x,m)(v) = \varphi(t,x,v)$ whereas, as shown in [2], $\partial_\mu V(t,x,\mu)(v) = \partial_v \varphi(t,x,v)$. The relationship $\partial_v[\partial_m V(t,x,m)(v)] = \partial_\mu V(t,x,\mu)(v)$ between the first-order derivatives can be extended to the second-order derivatives. With the same notations, it indeed holds that $\partial_v \partial_{v'}[\partial_m^2 V(t,x,m)(v,v')] = \partial_\mu^2 V(t,x,\mu)(v)(v')$, where the function $\mathbb{R}^d \times \mathbb{R}^d \ni (v,v') \mapsto \partial_m^2 V(t,x,m)(v,v') \in \mathbb{R}^{d\times d}$ represents the second-order derivatives of V with respect to m according to the concept used in Ref. [1] and $\partial_\mu^2 V(t,x,\mu)(v)(v')$ is obtained by differentiating $\partial_\mu V(t,x,\mu)(v)$ with respect to μ according to the rules detailed in Sect. 6, see in particular Remark 6.4. The identity between the second-order derivatives can be easily checked when V has the form $V(t,x,m) = \int_{\mathbb{R}^d\times\mathbb{R}^d} \varphi(t,x,v,v')m(v)m(v')dvdv'$ or equivalently $V(t,x,\mu) = \int_{\mathbb{R}^d\times\mathbb{R}^d} \varphi(t,x,v,v')d\mu(v)d\mu(v')$, in which case $\partial_m^2 V(t,x,m)(v,v') = 2\varphi(t,x,v,v')$ and $\partial_\mu^2 V(t,x,m)(v)(v') = 2\partial_v\partial_{v'}\varphi(t,x,v,v')$.

In the case of the control of McKean-Vlasov SDEs, the master equation (56) does not match the one given in Ref. [1]. Therein, the master equation for the control of McKean-Vlasov SDEs has a different interpretation. Indeed, its solution appears as the derivative in the sense used in [1], of the value function of the HJB equation (45). Intuitively, the derivative of the value function of Eq. (45) is expected to be the decoupling field of an infinite dimensional forward-backward system, in the spirit of the approach based on the Pontryagin stochastic maximum principle. The fact that it is the decoupling field of an infinite dimensional forward-backward system is the main reason why it is called the solution of the master equation in Ref. [1], in full analogy with what happens in the framework of mean-field games.

5 A Second Example: A Simple Growth Model

The following growth model was introduced and studied in Ref. [11]. We review its main features by recasting it in the framework of the present discussion of the master equation of mean field games with common noise. In fact the common noise W^0 is the only noise of the model since $\sigma \equiv 0$ and the idiosyncratic noises do not appear.

5.1 Background

As it is the case in many economic models, the problem in Ref. [11] is set for an infinite time horizon ($T = \infty$) with a positive discount rate $r > 0$. As we just said, $\sigma \equiv 0$. Moreover, the common noise is a one dimensional Wiener process $(W_t^0)_{t\geq 0}$. As before, we denote by $\mathbb{F}^0 = (\mathcal{F}_t^0)_{t\geq 0}$ its filtration. We also assume that its volatility is linear, that is $\sigma^0(x) = \sigma x$ for some positive constant σ, and that each player controls the drift of its state so that $b(t, x, \mu, \alpha) = \alpha$. In other words, the dynamics of the state of player i read:

$$dX_t^i = \alpha_t^i dt + \sigma X_t^i dW_t^0. \tag{62}$$

We shall restrict ourselves to Markovian controls of the form $\alpha_t^i = \alpha(t, X_t^i)$ for a deterministic function $(t, x) \mapsto \alpha(t, x)$, which will be assumed non-negative and Lipschitz in the variable x. Under these conditions, for any player, say player 1, $X_t^1 \geq 0$ at all times $t > 0$ if $X_0^1 \geq 0$ and for any two players, say players 1 and 2, the homeomorphism property of Lipschitz SDEs implies that $X_t^1 \leq X_t^2$ at all times $t > 0$ if $X_0^1 \leq X_0^2$.

Note that in the particular case

$$\alpha(t, x) = \gamma x \tag{63}$$

for some $\gamma > 0$, then

$$X_t^2 = X_t^1 + (X_0^2 - X_0^1)e^{(\gamma - \sigma^2/2)t + \sigma W_t^0}. \tag{64}$$

We assume that $k > 0$ is a fixed parameter and we introduce a special notation for the family of scaled Pareto distributions with decay parameter k. For any real number $q \geq 1$, we denote by $\mu^{(q)}$ the Pareto distribution:

$$\mu^{(q)}(dx) = k\frac{q^k}{x^{k+1}}\mathbf{1}_{[q,\infty)}(x)dx. \tag{65}$$

Notice that $X \sim \mu^{(1)}$ is equivalent to $qX \sim \mu^{(q)}$. We shall use the notation μ_t for the conditional distribution of the state X_t of a generic player at time $t \geq 0$ conditioned by the knowledge of the past up to time t as given by $\mathcal{F}_t^0$. Under the prescription (63), we claim that, if $\mu_0 = \mu^{(1)}$, then $\mu_t = \mu^{(q_t)}$ where $q_t = e^{(\gamma - \sigma^2/2)t + \sigma W_t^0}$. In other words, conditioned on the history of the common noise, the distribution of the states of the players remains Pareto with parameter k if it started that way, and the left-hand point of the distribution q_t can be understood as a sufficient statistic characterizing the distribution μ_t. This remark is an immediate consequence of formula (64) applied to $X_t^1 = q_t$, in which case $q_0 = 1$, and $X_t^2 = X_t$, implying that $X_t = X_0 q_t$. So if $X_0 \sim \mu^{(1)}$, then $\mu_t \sim \mu^{(q_t)}$. In particular, we have an explicit solution of the conditional Kolmogorov equation in the case of the particular linear feedback controls.

5.2 Optimization Problem

We now introduce the cost functions and define the optimization problem. We first assume that the problem is set for a finite horizon T. For the sake of convenience, we skip the stage of the N player game for N finite, and discuss directly the limiting MFG problem in order to avoid dealing with the fact that empirical measures do not have densities. The shape of the terminal cost g will be specified later on. Using the same notation as in Ref. [11], we define the running cost function f by

$$f(x,\mu,\alpha) = c\frac{x^a}{[(d\mu/dx)(x)]^b} - \frac{E}{p}\frac{\alpha^p}{[\mu([x,\infty))]^b},$$

for some positive constants a, b, c, E and $p > 1$ whose economic meanings are discussed in Ref. [11]. We use the convention that the density is the density of the absolutely continuous part of the Lebesgue's decomposition of the measure μ, and that in the above sum, the first term is set to 0 when this density is not defined or is itself 0. The extended Hamiltonian of the system (see (49)) reads

$$H(x,y,\mu,\alpha) = \alpha y + c\frac{x^a}{[(d\mu/dx)(x)]^b} - \frac{E}{p}\frac{\alpha^p}{[\mu([x,\infty))]^b}$$

and the value $\bar{\alpha}$ of α minimizing H is given by (for $y \geq 0$):

$$\bar{\alpha} = \bar{\alpha}(x,\mu,y) = \left(\frac{y}{E}\big[\mu([x,\infty))\big]^b\right)^{1/(p-1)} \tag{66}$$

so that:

$$\begin{aligned} H(x,y,\mu,\bar{\alpha}) &= \left(\frac{y}{E}\big[\mu([x,\infty))\big]^b\right)^{1/(p-1)} y + c\frac{x^a}{[(d\mu/dx)(x)]^b} \\ &\quad - \frac{E}{p}\frac{\left((y/E)[\mu([x,\infty))]^b\right)^{p/(p-1)}}{[\mu([x,\infty))]^b} \\ &= \frac{p-1}{p}E^{-1/(p-1)}y^{p/(p-1)}\big[\mu([x,\infty))\big]^{b/(p-1)} + c\frac{x^a}{[(d\mu/dx)(x)]^b}. \end{aligned}$$

In the particular case of linear controls (63), using the explicit formula (65) for the density of $\mu^{(q)}$ and the fact that

$$\mu^{(q)}([x,\infty)) = 1 \wedge \frac{q^k}{x^k},$$

we get

$$f\big(x, \mu^{(q)}, \alpha\big) = c\frac{x^a}{(kq^k/x^{k+1})^b}\mathbf{1}_{\{x\ge q\}} - \frac{E}{p}\frac{\alpha^p}{1 \wedge (q^{kb}/x^{kb})}$$
$$= \frac{c}{k^b q^{kb}}x^{a+b(k+1)}\mathbf{1}_{\{x\ge q\}} - \frac{E}{pq^{kb}}\alpha^p\big(x^{kb} \vee q^{kb}\big),$$

and

$$\bar{\alpha}(x, \mu, y) = \left[\frac{y}{E}\Big(\frac{q^{kb}}{x^{kb}} \wedge 1\Big)\right]^{1/(p-1)}, \tag{67}$$

so that

$$H(x, y, \mu^{(q)}, \bar{\alpha}) = \frac{p-1}{p}E^{-1/(p-1)}y^{p/(p-1)}\Big(\frac{q^{kb/(p-1)}}{x^{kb/(p-1)}} \wedge 1\Big) + c\frac{x^{a+(k+1)b}}{k^b q^{kb}}\mathbf{1}_{\{x\ge q\}}.$$

5.3 *Search for an Equilibrium*

Assuming that the initial distribution of the values of the state is given by the Pareto distribution $\mu^{(1)}$, we now restrict ourselves in searching for equilibriums with Pareto distributions, which means that the description of the equilibrium flow of measures $(\hat{\mu}_t)_{0\le t\le T}$ can be reduced to the description of the flow of corresponding Pareto parameters $(\hat{q}_t)_{0\le t\le T}$. Introducing the letter V for denoting the solution of the master equation, we know from (51) and Proposition 4.1 that the optimal feedback control must read

$$\hat{\alpha}(t, x) = \bar{\alpha}\big(x, \hat{\mu}_t, \partial_x V(t, x, \hat{\mu}_t)\big) = \left[\frac{\partial_x V(t, x, \hat{\mu}_t)}{E}\Big(\frac{\hat{q}_t^{kb}}{x^{kb}} \wedge 1\Big)\right]^{1/(p-1)}.$$

In order to guarantee that the equilibrium flow of measures is of Pareto type, it must satisfy the condition:

$$\gamma x = \left(\frac{\partial_x V(t, x, \hat{\mu}_t)}{E}\frac{\hat{q}_t^{kb}}{x^{kb}}\right)^{1/(p-1)}, \quad x \ge \hat{q}_t. \tag{68}$$

for some $\gamma > 0$. There is no need for checking the condition for $x < \hat{q}_t$ as the path driven by the Pareto distribution is then always greater than or equal to $(\hat{q}_t)_{t\ge 0}$.

Since we focus on equilibriums of Pareto type, we compute the function V at distributions of Pareto type only. It then makes sense to *parameterize* the problem and to seek for V in the factorized form:

$$\mathcal{V}(t, x, q) = V(t, x, \mu^{(q)}),$$

for some function $\mathcal{V} : (t, x, q) \in [0, T] \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$. Then, the relationship (68) takes the form:

$$\gamma x = \left(\frac{\partial_x \mathcal{V}(t, x, q)}{E} \frac{q^{kb}}{x^{kb}} \right)^{1/(p-1)}, \quad x \geq q.$$

The point is then to write the equation satisfied by $\mathcal{V}$, namely the equivalent of (51) but satisfied by $\mathcal{V}$ instead of V. First, we observe that, in (51), $\sigma(x) \equiv 0$. Obviously, the difficult point is to rewrite A_μ and $A_{x\mu}$ as differential operators acting on the variables q and (x, q) respectively.

A natural solution is to redo the computations used for deriving (51) by replacing Itô's formula for the measures $(\hat{\mu}_t)_{0 \leq t \leq T}$ by Itô's formula for $(\hat{q}_t)_{0 \leq t \leq T}$, taking benefit that $(\hat{q}_t)_{0 \leq t \leq T}$ solves the SDE

$$d\hat{q}_t = \gamma \hat{q}_t dt + \sigma \hat{q}_t dW_t, \tag{69}$$

which is a consequence of (63) and (64). Then the term $A_\mu \tilde{V}$ in (51), which reads as the Itô expansion of V along $(\hat{\mu}_t)_{0 \leq t \leq T}$, turns into the second-order differential operator associated to the SDE satisfied by $\hat{q}_t$, namely

$$A_q \mathcal{V}(t, x, q) = \gamma q \partial_q \mathcal{V}(t, x, q) + \frac{1}{2} \sigma^2 q^2 \partial_q^2 \mathcal{V}(t, x, q).$$

Similarly, the term $A_{x\mu} \tilde{V}$ in (51), which reads as the bracket of the components in $\mathbb{R}^d$ and in $\mathcal{P}_2(\mathbb{R}^d)$ in the Itô expansion, turns into the second-order differential operator associated to bracket of the SDEs satisfied by $(X_t)_{0 \leq t \leq T}$ in (62) and by $(\hat{q}_t)_{0 \leq t \leq T}$, namely

$$A_{xq} \mathcal{V}(t, x, q) = \sigma^2 x q \partial_{xq}^2 \mathcal{V}(t, x, q).$$

Rewriting (51), we get

$$\begin{aligned}
&\partial_t \mathcal{V}(t, x, q) + \frac{p-1}{p} E^{-1/(p-1)} \big(\partial_x \mathcal{V}(t, x, q) \big)^{p/(p-1)} \Big(\frac{q^{kb/(p-1)}}{x^{kb/(p-1)}} \wedge 1 \Big) \\
&\quad + c \frac{x^{a+(k+1)b}}{k^b q^{kb}} \mathbf{1}_{\{x \geq q\}} + \gamma q \partial_q \mathcal{V}(t, x, q) \\
&\quad + \frac{1}{2} \sigma^2 \big[x^2 \partial_x^2 \mathcal{V}(t, x, q) + q^2 \partial_q^2 \mathcal{V}(t, x, q) + 2 x q \partial_{xv}^2 \mathcal{V}(t, x, q) \big] = 0.
\end{aligned} \tag{70}$$

Now we look for a constant $B > 0$ such that

$$\mathcal{V}(t, x, q) = \mathcal{V}(x, q) = B \frac{x^{p+bk}}{q^{bk}}, \tag{71}$$

solves the parameterized master equation (70) on the set $\{x \geq q\}$. Under the additional condition that $a + b = p$, B must be the solution of the equation

$$\frac{p-1}{p} E^{-1/(p-1)} \big(B(p+bk)\big)^{p/(p-1)} + \frac{c}{k^b} - \gamma Bbk + \frac{\sigma^2}{2} Bp(p-1) = 0.$$

The condition (68) reads

$$\gamma = \Big(\frac{B(p+bk)}{E}\Big)^{1/(p-1)},$$

so that the above equation for B becomes

$$(p+bk)^{1/(p-1)} E^{-1/(p-1)} \big(p - 1 - \frac{bk}{p}\big) B^{p/(p-1)} + \frac{\sigma^2}{2} p(p-1)B + \frac{c}{k^b} = 0.$$

which always admits a solution if $p(p-1) < bk$. The fact that (70) is satisfied for $x \geq q$ is enough to prove that

$$\bigg(\mathcal{V}(\hat{X}_t, \hat{q}_t) + \int_0^t f\big(\hat{X}_s, \hat{\mu}_s, \gamma \hat{X}_s\big) ds\bigg)_{0 \leq t \leq T}, \quad \text{with } \hat{\mu}_s = \mu^{(\hat{q}_s)} \text{ for } \quad s \in [0, T],$$

is a martingale, whenever

$$d\hat{X}_t = \gamma \hat{X}_t dt + \sigma \hat{X}_t dW_t^0, \quad t \in [0, T],$$

with $\hat{X}_0 \sim \mu^{\hat{q}_0}$, and $(\hat{q}_t)_{0 \leq t \leq T}$ also solves (69). The reason is that $\hat{X}_t > \hat{q}_t$ for any $t \in [0, T]$ (equality $\hat{X}_t = \hat{q}_t$ holds along scenarios for which $\hat{X}_0 = \hat{q}_0$, which are of zero probability).

The martingale property is a part of the verification Proposition 4.1 for proving the optimality of $(\hat{X}_t)_{0 \leq t \leq T}$ when $(\hat{\mu}_t)_{0 \leq t \leq T}$ is the flow of conditional measures, but this is not sufficient. We must evaluate $\mathcal{V}$ along a pair $(X_t, \hat{q}_t)_{0 \leq t \leq T}$, $(X_t)_{0 \leq t \leq T}$ denoting a general controlled process satisfying (62). Unfortunately, things then become more difficult as X_t might not be larger than $\hat{q}_t$. In other words, we are facing the fact that $\mathcal{V}$ satisfies the PDE (70) on the set $\{x \geq q\}$ only. In order to circumvent this problem, a strategy consists in replacing $\mathcal{V}$ by

$$\mathcal{V}(x, q) = Bx^p \Big(\frac{x^{bk}}{q^{bk}} \wedge 1\Big),$$

for the same constant B as above. Obviously, the PDE (70) is not satisfied when $x < q$, but $\mathcal{V}$ defines a subsolution on the set $\{0 \leq x < q\}$, as (70) holds but with $= 0$ replaced by ≥ 0. Heuristically, this should show that

$$\left(\mathcal{V}(X_t, \hat{q}_t) + \int_0^t f\big(X_s, \hat{\mu}_s, \alpha_s\big) ds\right)_{0 \le t \le T} \tag{72}$$

is a submartingale when $(X_t)_{0 \le t \le T}$ is an arbitrary controlled process driven by the control $(\alpha_t)_{0 \le t \le T}$. Still, the justification requires some precaution as the function $\mathcal{V}$ is not $\mathcal{C}^2$ (which is the standard framework to apply Itô's expansion), its first-order derivatives being discontinuous on the diagonal $\{x = q\}$. The argument for justifying the Itô expansion is a bit technical so that we just give a sketchy proof of it. Basically, we can write $\mathcal{V}(X_t, \hat{q}_t) = B(X_t)^p [\varphi(X_t/\hat{q}_t)]^{bk}$, with $\varphi(r) = \min(1, r)$. The key point is that $(X_t/\hat{q}_t)_{0 \le t \le T}$ is always a bounded variation process, so that the expansion of $(\phi(X_t/\hat{q}_t))_{0 \le t \le T}$, for some function ϕ, only requires to control ϕ' and not ϕ''. Then, we can regularize φ by a sequence $(\varphi_n)_{n \ge 1}$ such that $(\varphi_n)'(r) = 0$, for $r \le 1 - 1/n$, $(\varphi_n)'(r) = 1$, for $r \ge 1$ and $(\varphi_n)'(r) \in [0, 1]$ for $r \in [1 - 1/n, 1]$. The fact that $(\varphi_n)'(r)$ is uniformly bounded in n permits to expand $(B(X_t)^p [\varphi_n(X_t/\hat{q}_t)]^{bk})_{0 \le t \le T}$ and then to pass to the limit.

The submartingale property shows that

$$\int_{\mathbb{R}^d} \mathcal{V}(x, \hat{q}_0) d\mu^{\hat{q}_0}(x) \le \inf_{(\alpha_t)_{0 \le t \le T}} \left[\int_0^T f(X_t, \hat{q}_t, \alpha_t) dt + \mathcal{V}(X_T, \hat{q}_T)\right], \tag{73}$$

which, together with the martingale property along $(\hat{X}_t)_{0 \le t \le T}$, shows that equality holds and that the Pareto distributions $(\hat{\mu}_t)_{0 \le t \le T}$ form a MFG equilibrium, provided g is chosen as $\mathcal{V}$. This constraint on the choice of g can be circumvented by choosing $T = \infty$, as done in Ref. [11], in which case f must be replaced by $e^{-rt} f$ for some discount rate $r > 0$.

The analysis in the case $T = \infty$ can be done in the following way. In the proof of the martingale and submartingale properties, $\mathcal{V}$ must replaced by $e^{-rt}\mathcal{V}$. Plugging $e^{-rt}\mathcal{V}$ and $e^{-rt} f$ in (70) instead of $\mathcal{V}$ and f, we understand that $\mathcal{V}$ must now satisfy (70) but with an additional $-r\mathcal{V}$ in the left-hand side. Then, we can repeat the previous argument in order to identify the value of B in (71). Finally, if r is large enough, $\mathbb{E}[e^{-rT}\mathcal{V}(\hat{X}_T, \hat{q}_T)]$ tends to 0 as T tends to the infinity in the martingale property (72). Similarly, if we restrict ourselves to a class of feedback controls with a suitable growth, $\mathbb{E}[e^{-rT}\mathcal{V}(X_T, \hat{q}_T)]$ tends to 0 in (73), which permits to conclude.

5.4 Control of McKean-Vlasov Equations

A similar framework could be used for considering the control of McKean-Vlasov equations. The analog of the strategy exposed in the previous paragraph would consist in limiting the optimization procedure to controlled processes in (62) driven by controls $(\alpha_t)_{0 \le t \le T}$ of the form $(\alpha_t = \gamma_t X_t)_{0 \le t \le T}$ for some deterministic $(\gamma_t)_{0 \le t \le T}$. Using an obvious extension of (64), this would force the conditional marginal distributions of $(X_t)_{0\le t\le T}$ to be Pareto distributed. Exactly as above, this would transform the problem into a finite dimensional problem. Precisely, this would transform the problem into a finite dimensional optimal control problem. In that perspective, the corresponding master equation could be reformulated as an HJB equation in finite dimension. In comparison with, we emphasize, once again, that the master equation (70) for the mean field game is not a HJB equation.

6 Appendix: A Generalized Form of Itô's Formula

Our derivation of the master equation requires the use of a form of Itô formula in a space of probability measures. This subsection is devoted to the proof of such a formula.

6.1 Notion of Differentiability

In Sect. 4, we alluded to a specific notion of differentiability for functions of probability measures. The choice of this notion is dictated by the fact that (1) the probability measures we are dealing with appear as laws of random variables; (2) in trying to differentiate functions of measures, the infinitesimal variations which we consider are naturally expressed as infinitesimal variations in the linear space of those random variables. The relevance of this notion of differentiability was argued by P.L. Lions in his lectures at the *Collège de France* [16]. The notes [2] offer a readable account, and [3] provides several properties involving empirical measures. It is based on the *lifting* of functions $\mathcal{P}_2(\mathbb{R}^d) \ni \mu \mapsto H(\mu)$ into functions $\tilde{H}$ defined on the Hilbert space $L^2(\tilde{\Omega};\mathbb{R}^d)$ over some probability space $(\tilde{\Omega},\tilde{\mathcal{F}},\tilde{\mathbb{P}})$ by setting $\tilde{H}(\tilde{X}) = H(\mathcal{L}(\tilde{X}))$, for $\tilde{X} \in L^2(\tilde{\Omega};\mathbb{R}^d)$, $\tilde{\Omega}$ being a Polish space and $\tilde{\mathbb{P}}$ an atomless measure.

Then, a function H is said to be differentiable at $\mu_0 \in \mathcal{P}_2(\mathbb{R}^d)$ if there exists a random variable $\tilde{X}_0$ with law μ_0, in other words satisfying $\mathcal{L}(\tilde{X}_0) = \mu_0$, such that the lifted function $\tilde{H}$ is Fréchet differentiable at $\tilde{X}_0$. Whenever this is the case, the Fréchet derivative of $\tilde{H}$ at $\tilde{X}_0$ can be viewed as an element of $L^2(\tilde{\Omega};\mathbb{R}^d)$ by identifying $L^2(\tilde{\Omega};\mathbb{R}^d)$ and its dual. It turns out that its distribution depends only upon the law μ_0 and not upon the particular random variable $\tilde{X}_0$ having distribution μ_0. See Sect. 6 in Ref. [2] for details. This Fréchet derivative $[D\tilde{H}](\tilde{X}_0)$ is called the representation of the derivative of H at μ_0 along the variable $\tilde{X}_0$. It is shown in Ref. [2] that, as a random variable, it is of the form $\tilde{h}(\tilde{X}_0)$ for some deterministic measurable function $\tilde{h} : \mathbb{R}^d \to \mathbb{R}^d$, which is uniquely defined μ_0-almost everywhere on $\mathbb{R}^d$. The equivalence class of $\tilde{h}$ in $L^2(\mathbb{R}^d,\mu_0)$ being uniquely defined, it can be denoted by $\partial_\mu H(\mu_0)$ (or $\partial H(\mu_0)$ when no confusion is possible). It is then natural to call $\partial_\mu H(\mu_0)$ the derivative of H at μ_0 and to identify it with a function $\partial_\mu H(\mu_0)(\,\cdot\,)$: $\mathbb{R}^d \ni v \mapsto \partial_\mu H(\mu_0)(v) \in \mathbb{R}^d$.

This procedure permits to express $[D\tilde{H}](\tilde{X}_0)$ as a function of any random variable $\tilde{X}_0$ with distribution μ_0, irrespective of where this random variable is defined.

Remark 6.1 Since it is customary to identify a Hilbert space to its dual, we will identify $L^2(\tilde{\Omega})$ with its dual, and in so doing, any derivative $D\tilde{H}(\tilde{X})$ will be viewed as an element of $L^2(\tilde{\Omega})$. In this way, the derivative in the direction $\tilde{Y}$ will be given by the inner product $[D\tilde{H}(\tilde{X})] \cdot \tilde{Y}$. Accordingly, the second Frechet derivative $D^2\tilde{H}(\tilde{X})$ which should be a linear operator from $L^2(\tilde{\Omega})$ into itself because of the identification with its dual, will be viewed as a bilinear form on $L^2(\tilde{\Omega})$. In particular, we shall use the notation $D^2\tilde{H}(\tilde{X})[\tilde{Y}, \tilde{Z}]$ for $\big([D^2\tilde{H}(\tilde{X})](\tilde{Y})\big) \cdot \tilde{Z}$.

Remark 6.2 The following result (see [3] for a proof) gives, though under stronger regularity assumptions on the Fréchet derivatives, a convenient way to handle this notion of differentiation with respect to probability distributions. If the function $\tilde{H}$ is Fréchet differentiable and if its Fréchet derivative is uniformly Lipschitz (i.e. there exists a constant $c > 0$ such that $\|D\tilde{H}(\tilde{X}) - D\tilde{H}(\tilde{X}')\| \le c|\tilde{X} - \tilde{X}'|$ for all $\tilde{X}, \tilde{X}'$ in $L^2(\tilde{\Omega})$), then there exists a function $\partial_\mu H$

$$\mathcal{P}_2(\mathbb{R}^d) \times \mathbb{R}^d \ni (\mu, v) \mapsto \partial_\mu H(\mu)(v)$$

such as $|\partial_\mu H(\mu)(v) - \partial_\mu H(\mu)(v')| \le c|v - v'|$ for all $v, v' \in \mathbb{R}^d$ and $\mu \in \mathcal{P}_2(\mathbb{R}^d)$, and for every $\mu \in \mathcal{P}_2(\mathbb{R}^d)$, $\partial_\mu H(\mu)(\tilde{X}) = D\tilde{H}(\tilde{X})$ almost surely if $\mu = \mathcal{L}(\tilde{X})$.

A.2 Itô's Formula Along a Flow of Conditional Measures

In the derivation of the master equation, the value function is expanded along a flow of conditional measures. As already explained in Sect. 4.3, this requires a suitable construction of the lifting.

Throughout this section, we assume that $(\Omega, \mathcal{F}, \mathbb{P})$ is of the form $(\Omega^0 \times \Omega^1, \mathcal{F}^0 \otimes \mathcal{F}^1, \mathbb{P}^0 \otimes \mathbb{P}^1)$, $(\Omega^0, \mathcal{F}^0, \mathbb{P}^0)$ supporting the common noise W^0, and $(\Omega^1, \mathcal{F}^1, \mathbb{P}^1)$ the idiosyncratic noise W. So an element $\omega \in \Omega$ can be written as $\omega = (\omega^0, \omega^1) \in \Omega^0 \times \Omega^1$, and functionals $H(\mu(\omega^0))$ of a random probability measure $\mu(\omega^0) \in \mathcal{P}_2(\mathbb{R}^d)$ with $\omega^0 \in \Omega^0$, can be lifted into $\tilde{H}(\tilde{X}(\omega^0, \cdot)) = H(\mathcal{L}(\tilde{X}(\omega^0, \cdot)))$, where $\tilde{X}(\omega^0, \cdot)$ is an element of $L^2(\tilde{\Omega}^1, \tilde{\mathcal{F}}^1, \mathbb{P}^1; \mathbb{R}^d)$ with $\mu(\omega^0)$ as distribution, $(\tilde{\Omega}^1, \tilde{\mathcal{F}}^1, \tilde{\mathbb{P}}^1)$ being Polish and atomless. Put it differently, the random variable $\tilde{X}$ is defined on $(\tilde{\Omega} = \Omega^0 \times \tilde{\Omega}^1, \tilde{\mathcal{F}} = \mathcal{F}^0 \otimes \tilde{\mathcal{F}}^1, \tilde{\mathbb{P}} = \mathbb{P}^0 \otimes \tilde{\mathbb{P}}^1)$.

The objective is then to expand $(\tilde{H}(\tilde{\chi}_t(\omega^0, \cdot)))_{0 \le t \le T}$, where $(\tilde{\chi}_t)_{0 \le t \le T}$ is the copy so constructed, of an Itô process on $(\Omega, \mathcal{F}, \mathbb{P})$ of the form:

$$\chi_t = \chi_0 + \int_0^t \beta_s ds + \int_0^t \int_\Xi \varsigma^0_{s,\xi} W^0(d\xi, ds) + \int_0^t \varsigma_s dW_s,$$

for $t \in [0, T]$, assuming that the processes $(\beta_t)_{0\le t\le T}$, $(\varsigma_t)_{0\le t\le T}$ and $(\varsigma^0_{t,\xi})_{0\le t\le T,\xi\in\Xi}$ are progressively measurable with respect to the filtration generated by W and W^0 and square integrable, in the sense that

$$\mathbb{E}\int_0^T\biggl(|\beta_t|^2+|\varsigma_t|^2+\int_\Xi|\varsigma^0_{t,\xi}|^2d\nu(\xi)\biggr)dt<+\infty. \tag{74}$$

Denoting by $(\tilde W_t)_{0\le t\le T}$, $(\tilde\beta_t)_{0\le t\le T}$, $(\tilde\varsigma_t)_{0\le t\le T}$ and $(\tilde\varsigma^0_{t,\xi})_{0\le t\le T,\xi\in\Xi}$ the copies of $(W_t)_{0\le t\le T}$, $(\beta_t)_{0\le t\le T}$, $(\varsigma_t)_{0\le t\le T}$ and $(\varsigma^0_{t,\xi})_{0\le t\le T,\xi\in\Xi}$, we then have

$$\tilde\chi_t=\tilde\chi_0+\int_0^t\tilde\beta_s ds+\int_0^t\int_\Xi\tilde\varsigma^0_{s,\xi}W^0(d\xi,ds)+\int_0^t\tilde\varsigma_s d\tilde W_s,$$

for $t\in[0,T]$. In this framework, we emphasize that it makes sense to look at $\tilde H(\tilde\chi_t(\omega^0,\cdot))$, for $t\in[0,T]$, since

$$\mathbb{E}^0\tilde{\mathbb{E}}^1\bigl[\sup_{0\le t\le T}|\tilde\chi_t|^2\bigr]=\mathbb{E}^0\mathbb{E}^1\bigl[\sup_{0\le t\le T}|\chi_t|^2\bigr]<+\infty,$$

where $\mathbb{E}^0$, $\mathbb{E}^1$ and $\tilde{\mathbb{E}}^1$ are the expectations associated to $\mathbb{P}^0$, $\mathbb{P}^1$ and $\tilde{\mathbb{P}}^1$ respectively.

In order to simplify notations, we let $\check\chi_t(\omega^0)=\tilde\chi_t(\omega^0,\cdot)$ for $t\in[0,T]$, so that $(\check\chi_t)_{0\le t\le T}$ is $L^2(\tilde\Omega^1,\tilde{\mathcal{F}}^1,\tilde{\mathbb{P}}^1;\mathbb{R}^d)$-valued, $\mathbb{P}^0$ almost surely. Similarly, we let $\check\beta_t(\omega^0)=\tilde\beta_t(\omega^0,\cdot)$, $\check\varsigma_t(\omega^0)=\tilde\varsigma_t(\omega^0,\cdot)$ $\check\varsigma_{t,\xi}(\omega^0)=\tilde\varsigma_{t,\xi}(\omega^0,\cdot)$, for $t\in[0,T]$ and $\xi\in\Xi$. We then claim

Proposition 6.3 *On the top of the assumption and notation introduced right above, assume that $\tilde H$ is twice continuously Fréchet differentiable. Then, we have $\mathbb{P}^0$ almost surely, for all $t\in[0,T]$,*

$$\begin{aligned}\tilde H\bigl(\check\chi_t\bigr)&=\tilde H\bigl(\check\chi_0\bigr)+\int_0^t D\tilde H\bigl(\check\chi_s\bigr)\cdot\check\beta_s ds+\int_0^t\int_\Xi D\tilde H\bigl(\check\chi_s\bigr)\cdot\check\varsigma^0_{s,\xi}\,W^0(d\xi,ds)\\&\quad+\frac12\int_0^t\biggl(D^2\tilde H(\check\chi_s)\bigl[\check\varsigma_s\tilde G,\check\varsigma_s\tilde G\bigr]+\int_\Xi D^2\tilde H\bigl(\check\chi_s\bigr)\bigl[\check\varsigma^0_{s,\xi},\check\varsigma^0_{s,\xi}\bigr]d\nu(\xi)\biggr)ds.\end{aligned} \tag{75}$$

where $\tilde G$ is an $\mathcal{N}(0,1)$-distributed random variable on $(\tilde\Omega^1,\tilde{\mathcal{F}}^1,\tilde{\mathbb{P}}^1)$, independent of $(\tilde W_t)_{t\ge0}$.

Remark 6.4 Following Remark 6.2, one can specialize Itô's formula to a situation with smoother derivatives. See Ref. [8] for a more detailed account. Indeed, if one assumes that

1. the function H is C^1 in the sense given above and its first derivative is Lipschitz;
2. for each fixed $v \in \mathbb{R}^d$, the function $\mu \mapsto \partial_\mu H(\mu)(v)$ is differentiable with Lipschitz derivative, and consequently, there exists a function

$$(\mu, v', v) \mapsto \partial_\mu^2 H(\mu)(v)(v') \in \mathbb{R}^{d\times d}$$

which is Lipschitz in v' uniformly with respect to v and μ and such that $\partial_\mu^2 H(\mu)(v)(\tilde{X})$ gives the Fréchet derivative of $\mu \mapsto \partial_\mu H(\mu)(v)$ for every $v \in \mathbb{R}^d$ as long as $\mathcal{L}(\tilde{X}) = \mu$;
3. for each fixed $\mu \in \mathcal{P}_2(\mathbb{R}^d)$, the function $v \mapsto \partial_\mu H(\mu)(v)$ is differentiable with Lipschitz derivative, and consequently, there exists a bounded function $(v, \mu) \mapsto \partial_v \partial_\mu H(\mu)(v) \in \mathbb{R}^{d\times d}$ giving the value of its derivative;
4. the functions $(\mu, v', v) \mapsto \partial_\mu^2 H(\mu)(v)(v')$ and $(\mu, v) \mapsto \partial_v \partial_\mu H(\mu)(v)$ are continuous (the space $\mathcal{P}_2(\mathbb{R}^d)$ being endowed with the 2-Wasserstein distance).

Then, the second order term appearing in Itô's formula can be expressed as the sum of two explicit operators whose interpretations are more natural. Indeed, the second Fréchet derivative $D^2\tilde{H}(\tilde{X})$ can be written as the linear operator $\tilde{Y} \mapsto A\tilde{Y}$ on $L^2(\tilde{\Omega}^1, \tilde{\mathcal{F}}^1, \mathbb{P}^1; \mathbb{R}^d)$ defined by

$$[A\tilde{Y}](\tilde{\omega}^1) = \int_{\tilde{\Omega}^{1,\prime}} \partial_\mu^2 H\big(\mathcal{L}(\tilde{X})\big)\big(\tilde{X}(\tilde{\omega}^1)\big)\big(\tilde{X}'(\omega')\big)\tilde{Y}'(\omega')\, d\tilde{\mathbb{P}}^{1,\prime}(\omega') \\ + \partial_v\partial_\mu H\big(\mathcal{L}(\tilde{X})\big)\big(\tilde{X}(\tilde{\omega}^1)\big)\tilde{Y}(\tilde{\omega}^1),$$

where $(\tilde{\Omega}^{1,\prime}, \tilde{\mathcal{F}}^{1,\prime}, \tilde{\mathbb{P}}^{1,\prime})$ is another Polish and atomless probability space endowed with a copy $(\tilde{X}', \tilde{Y}')$ of $(\tilde{X}, \tilde{Y})$.

In particular, when $\tilde{Y}$ is replaced by $\tilde{Y} \times \tilde{G}$, with $\tilde{G} \sim \mathcal{N}(0, 1)$ and independent of $(\tilde{X}, \tilde{Y})$, the integral over $\tilde{\Omega}^{1,\prime}$ in the right-hand side vanishes. We then obtain

$$D^2\tilde{H}(\tilde{X})\big[\tilde{Y}, \tilde{Y}\big] = \tilde{\mathbb{E}}^1\tilde{\mathbb{E}}^{1,\prime}\big\{\mathrm{trace}\big[\partial_\mu^2 H\big(\mathcal{L}(\tilde{X})\big)(\tilde{X})(\tilde{X}')\tilde{Y}\big(\tilde{Y}'\big)^\top\big]\big\} \\ + \tilde{\mathbb{E}}^1\big\{\mathrm{trace}\big[\partial_v\partial_\mu H\big(\mathcal{L}(\tilde{X})\big)(\tilde{X})\tilde{Y}\tilde{Y}^\top\big]\big\},$$
$$D^2\tilde{H}(\tilde{X})\big[\tilde{Y}\tilde{G}, \tilde{Y}\tilde{G}\big] = \tilde{\mathbb{E}}^1\big\{\mathrm{trace}\big[\partial_v\partial_\mu H\big(\mathcal{L}(\tilde{X})\big)(\tilde{X})\tilde{Y}\tilde{Y}^\top\big]\big\}.$$

The derivation of the master equation actually requires a more general result than Proposition 6.3. Indeed one needs to expand $(\tilde{H}(X_t, \check{\chi}_t))_{0\le t\le T}$ for a function $\tilde{H}$ of $(x, \tilde{X}) \in \mathbb{R}^d \times L^2(\tilde{\Omega}^1, \tilde{\mathcal{F}}^1, \tilde{\mathbb{P}}^1; \mathbb{R}^d)$. As before, $(\check{\chi}_t)_{0\le t\le T}$ is understood as $(\tilde{\chi}_t(\omega^0, \cdot))_{0\le t\le T}$. The process $(X_t)_{0\le t\le T}$ is assumed to be another Itô process, defined on the original space $(\Omega, \mathcal{F}, \mathbb{P}) = (\Omega^0 \times \Omega^1, \mathcal{F}^0 \otimes \mathcal{F}^1, \mathbb{P}^0 \otimes \mathbb{P}^1)$, with dynamics of the form

$$X_t = X_0 + \int_0^t b_s ds + \int_0^t \int_\Xi \sigma^0_{s,\xi} W^0(d\xi, ds) + \int_0^t \sigma_s dW_s,$$

for $t \in [0, T]$, the processes $(b_t)_{0\le t\le T}$, $(\sigma_t)_{0\le t\le T}$ and $(\sigma^0_{t,\xi})_{0\le t\le T,\xi\in\Xi}$ being progressively-measurable with respect to the filtration generated by W and W^0, and square integrable as in (74). Under these conditions, the result of Proposition 6.3 can be extended to:

Proposition 6.5 *On the top of the above assumptions and notations, assume that $\tilde{H}$ is twice continuously Fréchet differentiable on $\mathbb{R}^d \times L^2(\tilde{\Omega}^1, \tilde{\mathcal{F}}^1, \tilde{\mathbb{P}}^1; \mathbb{R}^d)$. Then, we have $\mathbb{P}$ almost surely, for all $t \in [0, T]$,*

$$\begin{aligned}
\tilde{H}\big(X_t, \check{\chi}_t\big) &= \tilde{H}\big(X_0, \check{\chi}_0\big) \\
&+ \int_0^t \Big(\langle \partial_x \tilde{H}\big(X_s, \check{\chi}_s\big), b_s\rangle + D_\mu \tilde{H}\big(X_s, \check{\chi}_s\big) \cdot \check{\beta}_s\Big) ds + \int_0^t \big[\partial_x \tilde{H}\big(X_s, \check{\chi}_s\big)\big]^\dagger \sigma_s dW_s \\
&+ \int_0^t \int_\Xi \Big(\big[\partial_x \tilde{H}\big(X_s, \check{\chi}_s\big)\big]^\dagger \sigma^0_{s,\xi} + D_\mu \tilde{H}\big(X_s, \check{\chi}_s\big) \cdot \check{\varsigma}^0_{s,\xi}\Big)\, W^0(d\xi, ds) \\
&+ \frac{1}{2} \int_0^t \int_\Xi \Big(\mathrm{trace}\big[\partial_x^2 \tilde{H}\big(X_s, \check{\chi}_s\big) \sigma^0_{s,\xi} (\sigma^0_{s,\xi})^\dagger\big] + D^2_\mu \tilde{H}\big(X_s, \check{\chi}_s\big)\big[\check{\varsigma}^0_{s,\xi}, \check{\varsigma}^0_{s,\xi}\big]\Big) d\nu(\xi) ds \\
&+ \frac{1}{2} \int_0^t \Big(\mathrm{trace}\big[\partial_x^2 \tilde{H}\big(X_s, \check{\chi}_s\big) \sigma_s (\sigma_s)^\dagger\big] + D^2_\mu \tilde{H}\big(X_s, \check{\chi}_s\big)\big[\check{\varsigma}_s \tilde{G}, \check{\varsigma}_s \tilde{G}\big]\Big) ds \\
&+ \int_0^t \int_\Xi \big\langle \partial_x D_\mu \tilde{H}\big(X_s, \check{\chi}_s\big) \cdot \check{\varsigma}^0_{s,\xi}\, ,\ \sigma^0_{s,\xi}\big\rangle d\nu(\xi) ds.
\end{aligned}$$

where $\tilde{G}$ is an $\mathcal{N}(0, 1)$-distributed random variable on $(\tilde{\Omega}^1, \tilde{\mathcal{F}}^1, \tilde{\mathbb{P}}^1)$, independent of $(\tilde{W}_t)_{t\ge 0}$. The partial derivatives in the infinite dimensional component are denoted with the index 'μ'. In that framework, the term $\langle \partial_x D_\mu \tilde{H}(X_s, \check{\chi}_s) \cdot \check{\varsigma}^0_{s,\xi}, \sigma^0_{s,\xi}\rangle$ reads

$$\sum_{i=1}^d \{\partial_{x_i} D_\mu \tilde{H}(X_s, \check{\chi}_s) \cdot \check{\varsigma}^0_{s,\xi}\} \big(\sigma^0_{s,\xi}\big)_i.$$

A.3 Proof of Itô's Formula

We only provide the proof of Proposition 6.3 as the proof of Proposition 6.5 is similar.

By a standard continuity argument, it is sufficient to prove that Eq. (75) holds for any $t \in [0, T]$ $\mathbb{P}^0$-almost surely. In particular, we can choose $t = T$. Moreover, by a standard approximation argument, it is sufficient to consider the case of simple processes $(\beta_t)_{0 \le t \le T}$, $(\varsigma_t)_{0 \le t \le T}$ and $(\varsigma^0_{t,\xi})_{0 \le t \le T,\xi}$ of the form

$$\beta_t = \sum_{i=0}^{M-1} \beta_i \mathbf{1}_{[\tau_i,\tau_{i+1})}(t), \quad \varsigma_t = \sum_{i=0}^{M-1} \varsigma_i \mathbf{1}_{[\tau_i,\tau_{i+1})}(t), \quad \varsigma^0_{t,\xi} = \sum_{i=0}^{M-1} \sum_{j=1}^{N} \varsigma^0_{i,j} \mathbf{1}_{[\tau_i,\tau_{i+1})}(t) \mathbf{1}_{A_j}(\xi),$$

where $M, N \ge 1$, $0 = \tau_0 < \tau_1 < \cdots < \tau_M = T$, $(A_j)_{1 \le j \le N}$ are piecewise disjoint Borel subsets of Ξ and $(\beta^i, \varsigma^i, \varsigma^0_{i,j})_{1 \le j \le N}$ are bounded $\mathcal{F}_{\tau_i}$-measurable random variables.

The strategy is taken from Ref. [8] and consists in splitting $\tilde{H}(\check{\chi}_T) - \tilde{H}(\check{\chi}_0)$ into

$$\tilde{H}(\check{\chi}_T) - \tilde{H}(\check{\chi}_0) = \sum_{k=0}^{K-1} \big(\tilde{H}(\check{\chi}_{t_{k+1}}) - \tilde{H}(\check{\chi}_{t_k})\big),$$

where $0 = t_0 < \cdots < t_K = T$ is a subdivision of $[0, T]$ of step h such that, for any $k \in \{0, \ldots, K-1\}$, there exists some $i \in \{0, \ldots, M-1\}$ such that $[t_k, t_{k+1}) \subset [\tau_i, \tau_{i+1})$. We then start with approximating a general increment $\tilde{H}(\check{\chi}_{t_{k+1}}) - \tilde{H}(\check{\chi}_{t_k})$, omitting to specify the dependence upon ω^0. By Taylor's formula, we know that we can find some $\delta \in [0, 1]$ such that

$$\begin{aligned}
&\tilde{H}(\check{\chi}_{t_{k+1}}) - \tilde{H}(\check{\chi}_{t_k}) \\
&= D\tilde{H}(\check{\chi}_{t_k}) \cdot (\check{\chi}_{t_{k+1}} - \check{\chi}_{t_k}) \\
&\quad + \frac{1}{2} D^2 \tilde{H}\big(\check{\chi}_{t_k} + \delta(\check{\chi}_{t_{k+1}} - \check{\chi}_{t_k})\big)\big(\check{\chi}_{t_{k+1}} - \check{\chi}_{t_k}, \check{\chi}_{t_{k+1}} - \check{\chi}_{t_k}\big) \\
&= D\tilde{H}(\check{\chi}_{t_k}) \cdot (\check{\chi}_{t_{k+1}} - \check{\chi}_{t_k}) + \frac{1}{2} D^2 \tilde{H}(\check{\chi}_{t_k})\big(\check{\chi}_{t_{k+1}} - \check{\chi}_{t_k}, \check{\chi}_{t_{k+1}} - \check{\chi}_{t_k}\big) \\
&\quad + \frac{1}{2}\big[D^2 \tilde{H}\big(\check{\chi}_{t_k} + \delta(\check{\chi}_{t_{k+1}} - \check{\chi}_{t_k})\big) - D^2 \tilde{H}\big(\check{\chi}_{t_k}\big)\big]\big(\check{\chi}_{t_{k+1}} - \check{\chi}_{t_k}, \check{\chi}_{t_{k+1}} - \check{\chi}_{t_k}\big).
\end{aligned} \tag{76}$$

By Kolmogorov continuity theorem, we know that, $\mathbb{P}^0$ almost surely, the mapping $[0, T] \ni t \mapsto \tilde{\chi}_t \in L^2(\tilde{\Omega}^1, \tilde{\mathcal{F}}^1, \tilde{\mathbb{P}}^1; \mathbb{R}^d)$ is continuous. Therefore, $\mathbb{P}^0$ almost surely, the mapping $(s, t, \delta) \mapsto D^2 \tilde{H}(\check{\chi}_t + \delta(\check{\chi}_s - \check{\chi}_t))$ is continuous from $[0, T]^2 \times [0, 1]$ to the space of bounded operators from $L^2(\tilde{\Omega}^1, \tilde{\mathcal{F}}^1, \tilde{\mathbb{P}}^1; \mathbb{R}^d)$ into itself, which proves that, $\mathbb{P}^0$ almost surely,

$$\lim_{h \searrow 0} \sup_{s,t \in [0,T], |t-s| \le h} \sup_{\delta \in [0,1]} \big\|\!\big|D^2 \tilde{H}\big(\check{\chi}_t + \delta(\check{\chi}_s - \check{\chi}_t)\big) - D^2 \tilde{H}\big(\check{\chi}_t\big)\big\|\!\big|_{2,\tilde{\Omega}^1} = 0,$$

$|||\cdot|||_{2,\tilde{\Omega}^1}$ denoting the operator norm on the space of bounded operators on $L^2(\tilde{\Omega}^1, \tilde{\mathcal{F}}^1, \tilde{\mathbb{P}}^1; \mathbb{R}^d)$. Now,

$$\begin{aligned}
&\Big|\sum_{k=0}^{K-1}\big[D^2\tilde{H}\big(\check{\chi}_{t_k}+\delta(\check{\chi}_{t_{k+1}}-\check{\chi}_{t_k})\big)-D^2\tilde{H}\big(\check{\chi}_{t_k}\big)\big]\big(\check{\chi}_{t_{k+1}}-\check{\chi}_{t_k},\check{\chi}_{t_{k+1}}-\check{\chi}_{t_k}\big)\Big| \\
&\quad\le \sup_{s,t\in[0,T],|t-s|\le h}\ \sup_{\delta\in[0,1]} |||D^2\tilde{H}\big(\check{\chi}_t+\delta(\check{\chi}_s-\check{\chi}_t)\big) \\
&\qquad - D^2\tilde{H}\big(\check{\chi}_t\big)|||_{2,\tilde{\Omega}^1}\sum_{k=0}^{K-1}\|\check{\chi}_{t_{k+1}}-\check{\chi}_{t_k}\|^2_{L^2(\tilde{\Omega})}.
\end{aligned}$$

Since

$$\mathbb{E}^0\Big[\sum_{k=0}^{K-1}\|\check{\chi}_{t_{k+1}}-\check{\chi}_{t_k}\|^2_{L^2(\tilde{\Omega})}\Big]\le C\sum_{k=0}^{K-1}\big(t_{k+1}-t_k\big)\le CT,$$

we deduce that

$$\Big|\sum_{k=0}^{K-1}\big[D^2\tilde{H}\big(\check{\chi}_{t_k}+\delta(\check{\chi}_{t_{k+1}}-\check{\chi}_{t_k})\big)-D^2\tilde{H}\big(\check{\chi}_{t_k}\big)\big]\cdot\big(\check{\chi}_{t_{k+1}}-\check{\chi}_{t_k},\check{\chi}_{t_{k+1}}-\check{\chi}_{t_k}\big)\Big|\to 0 \quad (77)$$

in $\mathbb{P}^0$ probability as h tends to 0. We now compute the various terms appearing in (76). We write

$$\begin{aligned}
D\tilde{H}(\check{\chi}_{t_k})\cdot(\check{\chi}_{t_{k+1}}-\check{\chi}_{t_k}) &= D\tilde{H}(\check{\chi}_{t_k})\cdot\int_{t_k}^{t_{k+1}}\tilde{\beta}_s(\omega^0,\cdot)ds \\
&\quad + D\tilde{H}(\check{\chi}_{t_k})\cdot\Big[\Big(\int_{t_k}^{t_{k+1}}\int_{\Xi}\varsigma^0_{s,\xi}W^0(d\xi,ds)\Big)(\omega^0,\cdot)\Big] \\
&\quad + D\tilde{H}(\check{\chi}_{t_k})\cdot\Big[\Big(\int_{t_k}^{t_{k+1}}\tilde{\varsigma}_s d\tilde{W}_s\Big)\Big](\omega^0,\cdot).
\end{aligned}$$

Assume that, for some $0\le i\le M-1$, $\tau_i\le t_k<t_{k+1}\le\tau_{i+1}$. Then,

$$D\tilde{H}(\check{\chi}_{t_k})\cdot\int_{t_k}^{t_{k+1}}\tilde{\beta}_s(\omega^0,\cdot)ds=\big(t_{k+1}-t_k\big)D\tilde{H}(\check{\chi}_{t_k})\cdot\tilde{\beta}_{t_k}(\omega^0,\cdot). \quad (78)$$

Note that the right-hand side is well-defined as β_{t_k} is bounded. Similarly, we notice that

$$D\tilde{H}(\check{\chi}_{t_k})\cdot\left[\left(\int_{t_k}^{t_{k+1}}\tilde{\varsigma}_s d\tilde{W}_s\right)(\omega^0,\cdot)\right]=\big(t_{k+1}-t_k\big)D\tilde{H}(\check{\chi}_{t_k})\cdot\big[\tilde{\varsigma}_{t_k}(\omega^0,\cdot)\big(\tilde{W}_{t_{k+1}}-\tilde{W}_{t_k}\big)\big].$$

Now, using the specific form of $D\tilde{H}$, $D\tilde{H}(\check{\chi}_{t_k}(\omega^0)) = (\tilde{\omega}^1 \mapsto \partial_\mu H(\mathcal{L}(\check{\chi}_{t_k}(\omega^0)))$ $(\tilde{\chi}_{t_k}(\omega^0,\tilde{\omega}^1))$ appears to be a $\tilde{\mathcal{F}}_{t_k}$-measurable random variable, and as such, it is orthogonal to $\tilde{\varsigma}_{t_k}(\omega^0,\cdot)(\tilde{W}_{t_{k+1}}-\tilde{W}_{t_k})$, which shows that

$$D\tilde{H}(\check{\chi}_{t_k})\cdot\left[\left(\int_{t_k}^{t_{k+1}}\tilde{\varsigma}_s d\tilde{W}_s\right)(\omega^0,\cdot)\right]=0. \tag{79}$$

Finally,

$$\begin{aligned}
&D\tilde{H}(\check{\chi}_{t_k})\cdot\left[\left(\int_{t_k}^{t_{k+1}}\int_{\Xi}\tilde{\varsigma}^0_{s,\xi}W^0(d\xi,ds)\right)(\omega^0,\cdot)\right]\\
&\quad = D\tilde{H}(\check{\chi}_{t_k})\cdot\left[\sum_{j=1}^{N}\tilde{\varsigma}^0_{i,j}(\omega^0,\cdot)W^0\big(A_j\times[t_k,t_{k+1})\big)(\omega^0)\right].
\end{aligned}$$

Now, $W^0\big(A_j\times[t_k,t_{k+1})\big)(\omega^0)$ behaves as a constant in the linear form above. Therefore,

$$\begin{aligned}
&D\tilde{H}(\check{\chi}_{t_k})\cdot\left[\left(\int_{t_k}^{t_{k+1}}\int_{\Xi}\tilde{\varsigma}^0_{s,\xi}W^0(d\xi,ds)\right)(\omega^0,\cdot)\right]\\
&\quad =\sum_{j=1}^{N}D\tilde{H}(\check{\chi}_{t_k})\cdot\tilde{\varsigma}^0_{i,j}(\omega^0,\cdot)W^0\big(A_j\times[t_k,t_{k+1})\big)(\omega^0)\\
&\quad =\left[\int_{t_k}^{t_{k+1}}\int_{\Xi}\big\{D\tilde{H}(\check{\chi}_{t_k})\cdot\tilde{\varsigma}^0_{s,\xi}(\omega^0,\cdot)\big\}W^0(d\xi,ds)\right](\omega^0).
\end{aligned}\tag{80}$$

Therefore, in analogy with (77), we deduce from (78), (79) and (80) that

$$\sum_{k=0}^{K-1}D\tilde{H}(\check{\chi}_{t_k})\cdot(\check{\chi}_{t_{k+1}}-\check{\chi}_{t_k})\to\int_0^T D\tilde{H}(\tilde{X}_s)\cdot\check{\beta}_s ds+\int_0^T\int_{\Xi}\big\{D\tilde{H}(\check{\chi}_s)\cdot\tilde{\varsigma}^0_{s,\xi}\big\}W^0(d\xi,ds),$$

in $\mathbb{P}^0$ probability as h tends to 0.

We now reproduce this analysis for the second order derivatives. We need to compute:

$$\begin{aligned}\Gamma_k := D^2\tilde{H}(\check{\chi}_{t_k})\Big[&\tilde{\beta}_{t_k}(\omega^0,\cdot)\big(t_{k+1}-t_k\big)+\tilde{\varsigma}_{t_k}(\omega^0,\cdot)\big(\tilde{W}_{t_{k+1}}-\tilde{W}_{t_k}\big)\\&+\sum_{j=1}^N \tilde{\varsigma}^0_{i,j}(\omega^0,\cdot)W^0\big([t_k,t_{k+1})\times A_j\big)(\omega^0),\\&\tilde{\beta}_{t_k}(\omega^1,\cdot)\big(t_{k+1}-t_k\big)+\tilde{\varsigma}_{t_k}(\omega^0,\cdot)\big(\tilde{W}_{t_{k+1}}-\tilde{W}_{t_k}\big)\\&+\sum_{j=1}^N \tilde{\varsigma}^0_{i,j}(\omega^0,\cdot)W^0\big([t_k,t_{k+1})\times A_j\big)(\omega^0)\Big].\end{aligned}$$

Clearly, the drift has very low influence on the value of Γ_k. Precisely, for investigating the limit (in $\mathbb{P}^0$ probability) of $\sum_{k=0}^{K-1}\Gamma_k$, we can focus on the 'reduced' version of Γ_k:

$$\begin{aligned}\Gamma_k := D^2\tilde{H}(\check{\chi}_{t_k})\Big[&\tilde{\varsigma}_{t_k}(\omega^0,\cdot)\big(\tilde{W}_{t_{k+1}}-\tilde{W}_{t_k}\big)+\sum_{j=1}^N \varsigma^0_{i,j}(\omega^0,\cdot)W^0\big([t_k,t_{k+1})\times A_j\big)(\omega^0),\\&\tilde{\varsigma}_{t_k}(\omega^0,\cdot)\big(\tilde{W}_{t_{k+1}}-\tilde{W}_{t_k}\big)+\sum_{j=1}^N \varsigma^0_{i,j}(\omega^0,\cdot)W^0\big([t,t+h]\times A_j\big)(\omega^0)\Big].\end{aligned}$$

We first notice that

$$D^2\tilde{H}(\check{\chi}_{t_k})\big[\tilde{\varsigma}_{t_k}(\omega^0,\cdot)\big(\tilde{W}_{t_{k+1}}-\tilde{W}_{t_k}\big),\tilde{\varsigma}^0_{i,j}(\omega^0,\cdot)W^0\big([t_k,t_{k+1})\times A_j\big)(\omega^0)\big]=0$$

(and the same for the symmetric term), the reason being that

$$\begin{aligned}&D^2\tilde{H}(\check{\chi}_{t_k})\big[\tilde{\varsigma}_{t_k}(\omega^0,\cdot)\big(\tilde{W}_{t_{k+1}}-\tilde{W}_{t_k}\big),\tilde{\varsigma}^0_{i,j}(\omega^0,\cdot)W^0\big([t_k,t_{k+1})\times A_j\big)(\omega^0)\big]\\&=\lim_{\epsilon\to 0}\epsilon^{-1}\big[D\tilde{H}\big(\check{\chi}_{t_k}+\epsilon\tilde{\varsigma}^0_{i,j}(\omega^0,\cdot)W^0\big([t_k,t_{k+1})\times A_j\big)(\omega^0)\big)\\&\qquad-D\tilde{H}(\check{\chi}_{t_k})\big]\cdot\big[\tilde{\varsigma}_{t_k}(\omega^0,\cdot)\big(\tilde{W}_{t_{k+1}}-\tilde{W}_{t_k}\big)\big],\end{aligned}$$

which is zero by the independence argument used in (79). Following the proof of (80),

$$\begin{aligned}&D^2\tilde{H}(\check{\chi}_{t_k})\Big[\sum_{j=1}^N \tilde{\varsigma}^0_{i,j}(\omega^0,\cdot)W^0\big([t_k,t_{k+1})\times A_j\big)(\omega^0),\\&\qquad\sum_{j=1}^N \tilde{\varsigma}^0_{i,j}(\omega^0,\cdot)W^0\big([t_k,t_{k+1})\times A_j\big)(\omega^0)\Big]\\&=\sum_{j,j'=1}^N D^2\tilde{H}(\check{\chi}_{t_k})\big[\tilde{\varsigma}^0_{i,j}(\omega^0,\cdot),\tilde{\varsigma}^0_{i,j'}(\omega^0,\cdot)\big]\\&\qquad\times\ W^0\big([t_k,t_{k+1})\times A_j\big)(\omega^0)W^0\big([t_k,t_{k+1})\times A_{j'}\big)(\omega^0).\end{aligned}$$

The second line reads as a the bracket of a discrete stochastic integral. Letting $\check{\varsigma}^0_{i,j}(\omega^0) = \tilde{\varsigma}^0_{i,j}(\omega^0, \cdot)$, it is quite standard to check

$$\sum_{k=0}^{K-1} \sum_{j,j'=1}^{N} D^2 \tilde{H}(\check{\chi}_{t_k})\big[\check{\varsigma}^0_{i,j}, \check{\varsigma}^0_{i,j'}\big] W^0\big([t_k, t_{k+1}) \times A_j\big) W^0\big([t_k, t_{k+1}) \times A_{j'}\big)$$
$$- \sum_{k=0}^{K-1} \sum_{j=1}^{N} D^2 \tilde{H}(\check{\chi}_{t_k})\big[\check{\varsigma}^0_{i,j}, \check{\varsigma}^0_{i,j}\big]\big(t_{k+1} - t_k\big)\nu(A_j) \to 0$$

in $\mathbb{P}^0$ probability as h tends to 0. Noticing that

$$\sum_{k=0}^{K-1} \sum_{j=1}^{N} D^2 \tilde{H}(\check{\chi}_{t_k})\big[\check{\varsigma}^0_{i,j}, \check{\varsigma}^0_{i,j}\big]\big(t_{k+1} - t_k\big)\nu(A_j)$$
$$= \sum_{k=0}^{K-1} \int_{t_k}^{t_{k+1}} \int_{\Xi} D^2 \tilde{H}(\check{\chi}_{t_k})\big[\check{\varsigma}^0_{s,\xi}, \check{\varsigma}^0_{s,\xi}\big] d\nu(\xi) ds,$$

we deduce that

$$\sum_{k=0}^{K-1} \sum_{j,j'=1}^{N} D^2 \tilde{H}(\check{\chi}_{t_k})\big[\check{\varsigma}^0_{i,j}, \check{\varsigma}^0_{i,j'}\big] W^0\big([t_k, t_{k+1}) \times A_j\big) W^0\big([t_k, t_{k+1}) \times A_{j'}\big)$$
$$- \int_0^T \int_{\Xi} D^2 \tilde{H}(\check{\chi}_s)\big[\check{\varsigma}^0_{s,\xi}, \check{\varsigma}^0_{s,\xi}\big] d\nu(\xi) ds \to 0$$

in $\mathbb{P}^0$ probability as h tends to 0. It remains to compute

$$D^2 \tilde{H}(\check{\chi}_{t_k})\big[\tilde{\varsigma}_{t_k}(\omega^0, \cdot)\big(\tilde{W}_{t_{k+1}} - \tilde{W}_{t_k}\big), \tilde{\varsigma}_{t_k}(\omega^0, \cdot)\big(\tilde{W}_{t_{k+1}} - \tilde{W}_{t_k}\big)\big].$$

Recall that this is the limit

$$\lim_{\varepsilon \to 0} \frac{1}{\varepsilon^2}\big[\tilde{H}\big(\tilde{\chi}_{t_k}(\omega^0, \cdot) + \varepsilon \tilde{\varsigma}_{t_k}(\omega^0, \cdot)(\tilde{W}_{t_{k+1}} - \tilde{W}_{t_k})\big)$$
$$+ \tilde{H}\big(\tilde{\chi}_{t_k}(\omega^0, \cdot) - \varepsilon \tilde{\varsigma}_{t_k}(\omega^0, \cdot)(\tilde{W}_{t_{k+1}} - \tilde{W}_{t_k})\big) - 2\tilde{H}\big(\tilde{\chi}_{t_k}(\omega^0, \cdot)\big)\big],$$

which is the same as

$$\lim_{\varepsilon \to 0} \frac{1}{\varepsilon^2}\big[\tilde{H}\big(\tilde{\chi}_{t_k}(\omega^0, \cdot) + \varepsilon \tilde{\varsigma}_{t_k}(\omega^0, \cdot)\sqrt{t_{k+1} - t_k}\,\tilde{G}\big) - \tilde{H}\big(\tilde{\chi}_{t_k}(\omega^0, \cdot)\big)\big],$$

where $\tilde{G}$ is independent of $(\tilde{W}_t)_{0 \le t \le T}$, and $\mathcal{N}(0, 1)$ distributed. Therefore,

$$D^2\tilde{H}(\check{\chi}_{t_k})\big[\check{\varsigma}_{t_k}\big(\tilde{W}_{t_{k+1}}-\tilde{W}_{t_k}\big),\check{\varsigma}_{t_k}\big(\tilde{W}_{t_{k+1}}-\tilde{W}_{t_k}\big)\big]=\big(t_{k+1}-t_k\big)D^2\tilde{H}(\check{\chi}_{t_k})\big[\check{\varsigma}_{t_k}\tilde{G},\check{\sigma}_{t_k}\tilde{G}\big],$$

which is enough to prove that

$$\sum_{k=0}^{K-1}D^2\tilde{H}(\check{\chi}_{t_k})\big[\check{\varsigma}_{t_k}\big(\tilde{W}_{t_{k+1}}-\tilde{W}_{t_k}\big),\check{\varsigma}_{t_k}\big(\tilde{W}_{t_{k+1}}-\tilde{W}_{t_k}\big)\big]\to\int_0^T D^2\tilde{H}(\check{\chi}_s)\big[\check{\varsigma}_s\tilde{G},\check{\varsigma}_s\tilde{G}\big]ds$$

in $\mathbb{P}^0$ probability as h tends to 0.

References

1. A. Bensoussan, J. Frehse, P. Yam, The master equation in mean-field theory. Technical report. http://arxiv.org/abs/1404.4150
2. P. Cardaliaguet, Notes on mean field games. Notes from P.L. Lions' lectures at the Collège de France https://www.ceremade.dauphine.fr/cardalia/MFG100629.pdf (2012)
3. R. Carmona, F. Delarue, Forward-backward stochastic differential equations and controlled McKean Vlasov dynamics. in Annals of Probability To appear
4. R. Carmona, F. Delarue, Probabilistic analysis of mean field games. SIAM J. Control Optim. **51**, 2705–2734 (2013)
5. R. Carmona, F. Delarue, D. Lacker, Mean field games with a common noise. Technical report. http://arxiv.org/abs/1407.6181
6. R. Carmona, F. Delarue, A. Lachapelle, Control of McKean-Vlasov versus mean field games. Math. Financ. Econ. **7**, 131–166 (2013)
7. R. Carmona, J.P. Fouque, A. Sun, Mean field games and systemic risk. To appear in Communications in Mathematical Sciences
8. J.F. Chassagneux, D. Crisan, F. Delarue, McKean-Vlasov FBSDEs and related master equation. Work in progress
9. W. Fleming, M. Soner, *Controlled Markov Processes and Viscosity Solutions* (Springer, New York, 2010)
10. D.A. Gomes, J. Saude, Mean field games models—a brief survey. Technical report (2013)
11. O. Guéant, J.M. Lasry, P.L. Lions, Paris Princeton Lectures in Mathematical Finance IV, in *Mean Field Games and Applications*, Lecture Notes in Mathematics, ed. by R. Carmona, et al. (Springer, Berlin, 2010)
12. M. Huang, P.E. Caines, R.P. Malhamé, Large population stochastic dynamic games: closed-loop McKean-Vlasov systems and the Nash certainty equivalence principle. Commun. Inf. Syst. **6**, 221–252 (2006)
13. J.M. Lasry, P.L. Lions, Jeux à champ moyen I. Le cas stationnaire. Comptes Rendus de l'Académie des Sciences de Paris, ser. A **343**(9), 619–625 (2006)
14. J.M. Lasry, P.L. Lions, Jeux à champ moyen II. Horizon fini et contrôle optimal. Comptes Rendus de l'Académie des Sciences de Paris, ser. A **343**(10), 679–684 (2006)
15. J.M. Lasry, P.L. Lions, Mean field games. Jpn. J. Math. **2**(1), 229–260 (2007)
16. P.L. Lions, Théorie des jeux à champs moyen et applications. Technical report, 2007–2008
17. J. Ma, H. Yin, J. Zhang, On non-Markovian forward-backward SDEs and backward stochastic PDEs. Stoch. Process. Appl. **122**, 3980–4004 (2012)
18. D. Nualart, *The Malliavin Calculus and Related Topics*, Probability and its Applications (Springer, New York, 1995)

19. S. Peng, Stochastic Hamilton Jacobi Bellman equations. SIAM J. Control Optim. **30**, 284–304 (1992)
20. A.S. Sznitman, Topics in propagation of chaos, in D.L. Burkholder et al., Ecole de Probabilités de Saint Flour, XIX-1989. Lecture Notes in Mathematics, vol. 1464 (Springer, Heidelberg, 1989), pp. 165–251

The Filtering Equations Revisited

Thomas Cass, Martin Clark and Dan Crisan

Abstract The problem of nonlinear filtering has engendered a surprising number of mathematical techniques for its treatment. A notable example is the change-of–probability-measure method introduced by Kallianpur and Striebel to derive the filtering equations and the Bayes-like formula that bears their names. More recent work, however, has generally preferred other methods. In this paper, we reconsider the change-of-measure approach to the derivation of the filtering equations and show that many of the technical conditions present in previous work can be relaxed. The filtering equations are established for general Markov signal processes that can be described by a martingale-problem formulation. Two specific applications are treated.

Keywords Measure valued processes · Non-linear filtering · Kallianpur-Striebel formula · Change of probability measure method · Kazamaki criterion

1 Introduction

The aim of nonlinear filtering is to estimate an evolving dynamical system, customarily modelled by a stochastic process and called the signal process. The signal process cannot be measured directly, but only via a related process, termed the observation process. The filtering problem consists in computing the conditional distribution of the signal at the current time given the observation data accumulated up to that time. In order to describe the contribution of the paper, we start with a few historical comments on the subject.

T. Cass · D. Crisan (✉)
Department of Mathematics, Imperial College London, London, UK
e-mail: d.crisan@imperial.ac.uk

M. Clark
Department of Electrical and Electronic Engineering, Imperial College London, London, UK

D. Crisan et al. (eds.), *Stochastic Analysis and Applications 2014*,
Springer Proceedings in Mathematics & Statistics 100,
DOI 10.1007/978-3-319-11292-3_5

The development of the modern theory of nonlinear filtering started in the sixties with the publications of Stratonovich [35, 36], Kushner [14, 15] and Shiryaev [33] for diffusions and Wonham for pure-jump Markov processes [38]; these introduced the basic form of the class of stochastic differential equations for the conditional distributions of partially observed Markov processes, which are now known generically as the filtering equation. This class of equations has inspired authors to introduce a rich variety of mathematical techniques to justify their structure, together with that of their un-normalized form, the Zakai (or Duncan-Mortensen-Zakai) equation, [9, 23, 41], and to establish the existence, uniqueness and regularity of their solutions. A description of much of the work on this equation and its generalizations can be found in [13] for papers before 1980, in [16, 17] for papers before 2000 and in [2, 6, 39] for more recent work.

For instance, Fujisaki et al. [10] exploited a stochastic-integral representation theorem in order to enable them to express conditional distributions as functionals of an "innovations" martingale (a concept introduced in the Gaussian case by Kailath [20]). Krylov et al. [18, 19, 24], Chap. 6 in [6] and other authors developed a general theory of stochastic partial differential equations that led to a direct 'PDE' approach to the filtering equations, but there are many other approaches For example, see the work of Grigelionis and Mikulevicius on filtering for signal and observation processes with jumps [4, Chap. 4] and that of Kurtz and Nappo on the filtered martingale problem [4, Chap. 5].

In parallel with the above developments, Snyder [34], Brémaud [3] and van Schuppen [28] have initiated the study of the filtering problem for observations of counting process type. A large number of papers have been written on this class of filtering problems. Some of the early contributors to this topic include Boel, Davis, Segal, Varaiya, Willems and Wong, see [7, 29–32, 37]. Also, Grigelionis [11] looked at filtering problems with common jumps of the unobserved state process and of the observations. For further developments in this directions see [4, Chap. 10].

A probabilistic approach, initially considered formally by Bucy [4], but developed in detail by Kallianpur and Striebel [21, 22], made use of a functional form of Bayes formula for processes, now known as the Kallianpur-Striebel formula. This technique, which is based on a change of probability measure that makes, at each time, the future observation process independent of past processes, is effective for filtering problems in which the observation process is of the "signal plus white noise" variety, where the signal is independent of the noise process, but less so for the "correlated case"; that is, for problems in which observed and unobserved components are coupled via a common noise process. For this reason, among probabilistic methods, the "innovations" approach is often preferred to the "change of measure" method. The awkwardness in the application of the latter results from the fact that an exponential local martingale, constructed via Girsanov's theorem as a process of potential densities, has to be verified as a true martingale, and this is generally requires ad hoc techniques peculiar to the particular filtering problem being considered.

In this paper we re-visit the change-of-measure method and show that it can be used to derive the filtering equations for a broad class of Markov processes with coupled observed and unobserved components. This class includes diffusions with jumps obeying only mild linear growth conditions on their characteristic coefficients. Propositions are also presented that serve to test whether the filtering equations are derivable by the change-of-measure method for a particular filtering problem.

2 The Filtering Framework

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space together with a filtration $(\mathcal{F}_t)_{t\geq 0}$ which satisfies the usual conditions.[1] On $(\Omega, \mathcal{F}, \mathbb{P})$ we consider an $\mathcal{F}_t$-adapted process $\overline{X}$ with càdlàg paths. The process $\overline{X}$ consists in a pair of processes X and Y, $\overline{X} = (X, Y)$. The process X is called the *signal* process and is assumed to take values in a complete separable metric space $\mathbb{S}$ (the state space). The process Y is assumed to take values in $\mathbb{R}^m$ and is called the *observation* process.

Let $\mathcal{B}(\mathbb{S} \times \mathbb{R}^m)$ be the associated product Borel σ-algebra on $\mathbb{S} \times \mathbb{R}^m$ and $b\mathcal{B}(\mathbb{S} \times \mathbb{R}^m)$ be the space of bounded $\mathcal{B}(\mathbb{S} \times \mathbb{R}^m)$-measurable functions. Let $A\colon b\mathcal{B}(\mathbb{S} \times \mathbb{R}^m) \to b\mathcal{B}(\mathbb{S} \times \mathbb{R}^m)$ and write $\mathcal{D}(A) \subseteq b\mathcal{B}(\mathbb{S} \times \mathbb{R}^m)$ for the domain of A. We assume that $\mathbf{1} \in \mathcal{D}(A)$ and $A\mathbf{1} = 0$. In the following we will assume that the distribution of X_0 is $\pi_0 \in \mathcal{P}(\mathbb{S})$ and that $Y_0 = 0$. Since $Y_0 = 0$, the initial distribution of X, is identical with the conditional distribution of X_0 given Y_0 and we use the same notation for both. Further we will assume that $\overline{X}$ is a solution of the martingale problem for $(A, \pi_0 \times \delta_0)$. In other words, we assume that the process $M^\varphi = \{M_t^\varphi,\ t \geq 0\}$ defined as

$$M_t^\varphi = \varphi(\overline{X}_t) - \varphi(\overline{X}_0) - \int_0^t A\varphi(\overline{X}_s)\mathrm{d}s, \quad t \geq 0, \tag{1}$$

is an $\mathcal{F}_t$-adapted martingale for any $\varphi \in \mathcal{D}(A)$. In addition, let $h = (h_i)_{i=1}^m : \mathbb{S} \to \mathbb{R}^m$ be a measurable function such that

$$P\left(\int_0^t \left|h^i(\overline{X}_s)\right|^2 ds < \infty\right) = 1. \tag{2}$$

for all $t \geq 0$. Let W be a standard $\mathcal{F}_t$-adapted m-dimensional Brownian motion defined on $(\Omega, \mathcal{F}, \mathbb{P})$. We will assume that Y satisfies the following evolution equation

[1] The probability space $(\Omega, \mathcal{F}, \mathbb{P})$ together with the filtration $(\mathcal{F}_t)_{t\geq 0}$ satisfies the usual conditions provided: a. $\mathcal{F}$ is complete i.e. $A \subset B$, $B \in \mathcal{F}$ and $\mathbb{P}(B) = 0$ implies that $A \in \mathcal{F}$ and $\mathbb{P}(A) = 0$, b. The filtration $\mathcal{F}_t$ is right continuous i.e. $\mathcal{F}_t = \mathcal{F}_{t+}$. c. $\mathcal{F}_0$ (and consequently all $\mathcal{F}_t$ for $t \geq 0$) contains all the $\mathbb{P}$-null sets.

$$Y_t = \int_0^t h(\overline{X}_s)\,\mathrm{d}s + W_t. \tag{3}$$

To complete the description we need to identify the covariation process between $M^\varphi = \{M_t^\varphi,\ t \geq 0\}$ and W. For this we introduce m operators $B^i : b\mathcal{B}(\mathbb{S} \times \mathbb{R}^m) \to b\mathcal{B}(\mathbb{S} \times \mathbb{R}^m)$, $i = 1, \ldots, m$ with $\mathcal{D}(A) \subseteq \mathcal{D}(B^i) \subseteq b\mathcal{B}(\mathbb{S} \times \mathbb{R}^m)$. We assume that $\mathbf{1} \in \mathcal{D}(A)$ and $A\mathbf{1} = 0$. We will assume that,

$$\left\langle M^\varphi, W^i \right\rangle_t = \int_0^t B^i\varphi\left(\overline{X}_s\right) ds + \int_0^t \frac{\partial\varphi}{\partial y_i}\left(\overline{X}_s\right) ds, \tag{4}$$

for any $t \geq 0$ and for test functions φ both in the domain of A and with bounded partial derivatives in the y direction. In particular, for functions that are constant in the second component, then we have

$$\left\langle M^\varphi, W \right\rangle_t = \int_0^t B^i\varphi\left(X_s, Y_s\right) ds. \tag{5}$$

Let $\{\mathcal{Y}_t,\ t \geq 0\}$ be the usual augmentation of the filtration associated with the process Y, viz

$$\mathcal{Y}_t = \bigcap_{\varepsilon>0} \sigma(Y_s,\ s \in [0, t+\varepsilon]) \vee \mathcal{N}, \quad \mathcal{Y} = \bigvee_{t\in\mathbb{R}_+} \mathcal{Y}_t. \tag{6}$$

where $\mathcal{N}$ is the class of all $\mathbb{P}$-null sets. Note that Y_t is $\mathcal{F}_t$-adapted, hence $\mathcal{Y}_t \subset \mathcal{F}_t$. In the following we will assume that $\mathcal{Y}_t$ is a right continuous filtration.

Definition 1 The filtering problem consists in determining the conditional distribution π_t of the signal X at time t given the information accumulated from observing Y in the interval $[0, t]$; that is, for $\varphi \in b\mathcal{B}(\mathbb{S})$, computing

$$\pi_t(\varphi) = \mathbb{E}[\varphi(X_t) \mid \mathcal{Y}_t]. \tag{7}$$

There exists a suitable regularisation of the process $\pi = \{\pi_t,\ t \geq 0\}$, so that π_t is an optional $\mathcal{Y}_t$-adapted probability measure-valued process for which (7) holds almost surely. [2] In addition, since $\mathcal{Y}_t$ is right-continuous, it follows that π has a càdlàg version (see Corollary 2.26 in [2]). In the following, we take π to be this version.

In the following we deduce the evolution equation for π. A new measure is constructed under which Y becomes a Brownian motion and π has a representation in terms of an associated unnormalised version ρ. This ρ is then shown to satisfy a linear evolution equation which leads to the evolution equation for π by an application of Itô's formula.

---

[2] See Theorem 2.1 in [2].

2.1 Preliminary Results

Definition 2 We define $H^1(\mathbb{P})$ to be the set of càdlàg real-valued $\mathcal{F}_t$-martingales $M = \{M_t,\ t \geq 0\}$ such that the associated process $M^* = \{M_t^*,\ t \geq 0\}$ defined as $M_t^* := \sup_{0\leq s\leq t} |M_s|$ for $t \geq 0$ is a submartingale. In particular, $\mathbb{E}[M_t] < \infty$. for any $t \geq 0$.

Remark 3 $H^1(\mathbb{P})$ together with the distance function

$$d(M, N) := \sum_{n=1}^{\infty} \frac{1}{2^n} \min\left(\mathbb{E}\left[(M - N)_n^*\right], 1\right)$$

is a Fréchet space with translation invariant metric. Suppose $(W_t)_{t\geq 0}$ is an $\mathbb{R}^d$-valued Brownian motion and $H = (H^i)_{i=1}^d$ is an $\mathcal{F}_t$-adapted measurable $\mathbb{R}^d$-valued process such that

$$P\left(\int_0^t |H_s|^2\, ds < \infty\right) = 1. \tag{8}$$

Define $Z = (Z_t)_{t\geq 0}$ to be the exponential local martingale[3]

$$Z_t = \exp\left(\int_0^t H_s^\top dW_s - \frac{1}{2}\int_0^t |H_s|^2\, ds\right),$$

where $\int_0^t H_s^\top dW_s := \sum_{i=1}^d \int_0^t H_s^i dW_s^i$.

Lemma 4 (The $Z \log Z$ lemma) *For any $t \geq 0$ we have*

$$\sup_{\tau\in\mathcal{T}_t} \mathbb{E}\left[Z_\tau \log Z_\tau\right] = \frac{1}{2}\mathbb{E}\left[\int_0^t Z_s |H_s|^2\, ds\right] \in [0, \infty], \tag{9}$$

where $\mathcal{T}_t$ is the set of $(\mathcal{F}_t)$ -stopping times bounded by t. If furthermore the terms in (9) *are finite, then they are both equal to $\mathbb{E}\left[Z_t \log Z_t\right]$. We also have*

$$\mathbb{E}\left[Z_t^*\right] \leq \frac{e+1}{e-1} + \frac{e}{2(e-1)}\mathbb{E}\left[\int_0^t Z_s |H_s|^2\, ds\right] \in [0, \infty]. \tag{10}$$

As an immediate consequence of this lemma we have

[3] Here and later if $a = (a_i)_{i=1}^d \in \mathbb{R}^d$, then $|a|^2 = \sum_{i=1}^d a_i^2$. Hence, for example, in the expression for Z from $\int_0^t |H_s|^2\, ds = \sum_{i=1}^d \int_0^t \left(H_s^i\right)^2 ds$.

Corollary 5 *If the terms in* (9) *are finite,then* $(Z_t)_{t\geq 0}$ *is a genuine martingale, uniformly integrable over any finite interval* $[0,t]$*, that belongs to* $H^1(\mathbb{P})$.

Remark 6 The first part of this corollary—that Z is a martingale if the terms in (9) are finite—is not new. At the time of going to press J. Ruf brought to the authors' attention that it is a consequence of the either of two more general results: see Theorem 1 and Corollary 5 in [27]. The additional generality of these results is in fact unnecessary for us. Since the governing considerations of our presentation are those of economy and self-sufficiency, we include a short proof of our result below.

Proof Let $L_t := Z_t \log Z_t$ for $t \geq 0$. If we assume that $\sup_{\tau\in\mathcal{T}_t}\mathbb{E}[L_\tau]$ is finite, then for all $K \geq e$

$$\sup_{\tau\in\mathcal{T}_t}\mathbb{E}\left[|Z_\tau|\,1_{\{|Z_\tau|\geq K\}}\right] \leq \sup_{\tau\in\mathcal{T}_t}\frac{\mathbb{E}\left[|Z_\tau|\log Z_\tau 1_{\{|Z_\tau|\geq K\}}\right]}{\log K} \leq \frac{1}{\log K}\left(\sup_{\tau\in\mathcal{T}_t}\mathbb{E}[L_\tau]+e^{-1}\right)$$

the right hand side of which tends to zero as $K\to\infty$. Hence the family random variables

$$\{Z_\tau : \tau\in\mathcal{T}_t\}$$

is uniformly integrable. Z is thus a martingale over $[0,t]$ and L, by Jensen's inequality, is a submartingale. Using $P(0<Z_t<\infty,\ \text{for all } t<\infty)=1$ we have from Itô's formula that

$$L_t = \underbrace{\int_0^t (1+\log Z_s)\,Z_s H_s^\top dW_s}_{:=M_t} + \underbrace{\frac{1}{2}\int_0^t Z_s\,|H_s|^2\,ds}_{:=A_t}.$$

M is a local martingale; hence the stopped process $M^{\sigma_n}_{\cdot} := M_{\cdot\wedge\sigma_n}$ is a martingale for some localising sequence $0\leq\sigma_n\leq\sigma_{n+1}\uparrow\infty$ as $n\to\infty$. For any $\tau\in\mathcal{T}_t$ we obtain

$$\mathbb{E}\left[L_\tau^{\sigma_n}\right]=\mathbb{E}\left[A_\tau^{\sigma_n}\right]\leq\mathbb{E}[L_\tau]\leq\mathbb{E}[L_t].$$

Then, using Fatou's lemma[4] and the monotone convergence theorem, we have

$$\mathbb{E}[L_\tau]\leq\lim\inf_{n\to\infty}\mathbb{E}\left[L_\tau^{\sigma_n}\right]=\lim\inf_{n\to\infty}\mathbb{E}\left[A_\tau^{\sigma_n}\right]=\mathbb{E}[A_\tau]\leq\mathbb{E}[L_\tau]\leq\mathbb{E}[L_t].$$

Finally taking the supremum over $\tau\in\mathcal{T}_t$ yields

$$\mathbb{E}[L_t]\leq\sup_{\tau\in\mathcal{T}_t}\mathbb{E}[L_\tau]\leq\sup_{\tau\in\mathcal{T}_t}\mathbb{E}[A_\tau]\leq\mathbb{E}[A_t]\leq\sup_{\tau\in\mathcal{T}_t}\mathbb{E}[L_\tau]\leq\mathbb{E}[L_t].$$

and the equality (9) holds in this case. If instead we know that $\mathbb{E}[A_t]<\infty$, then by defining the sequence of stopping times $(\tau_n)_{n=1}^{\infty}$, $0\leq\tau_n\leq\tau_{n+1}$ by

4 Which we may do since L is bounded from below by $-e^{-1}$.

$$\tau_n = \inf \left\{ t \geq 0 : |Z_t| = \frac{1}{n} \text{ or } |Z_t| = n \right\}$$

we have

$$\mathbb{E}\left[M_{t\wedge\tau_n}^2\right] = \mathbb{E}\left[\int_0^{t\wedge\tau_n} (1+\log Z_s)^2 Z_s^2 |H_s|^2 \, ds\right] \leq 2n^2 (1+\log n)^2 \, \mathbb{E}[A_t] < \infty.$$

From this we deduce that the stopped process $M_\cdot^{\tau_n} := M_{\cdot\wedge\tau_n}$ is a square-integrable martingale over $[0, t]$. Combining this with the fact that $A_{t\wedge\tau_n} \leq A_t$ yields

$$\mathbb{E}\left[L_{\tau\wedge\tau_n}\right] = \mathbb{E}\left[A_{\tau\wedge\tau_n}\right] \leq \mathbb{E}[A_t]$$

for any $\tau \in \mathcal{T}_t$. We notice that $\tau_n \uparrow \infty$, and hence $Z_{t\wedge\tau_n} \to Z_t$ a.s. as $n \to \infty$. Then applying Fatou's lemma and taking the supremum over all $\tau \in \mathcal{T}_t$ then gives that $\sup_{\tau\in\mathcal{T}_t} \mathbb{E}[L_\tau] \leq \mathbb{E}[A_t] < \infty$. The equality

$$\mathbb{E}[L_t] = \mathbb{E}[A_t] = \sup_{\tau\in\mathcal{T}_t} \mathbb{E}[L_\tau] \in [0, \infty)$$

then follows from the first part of the proof. It is clear from the argument that A_t is not integrable if and only if $\sup_{\tau\in\mathcal{T}_t} \mathbb{E}[L_\tau] = \infty$.

Turning attention to (10), we observe that the stopped process L^{τ_n} is a bounded submartingale, with a bounded martingale part given by M^{τ_n}. Hence, by a modification of a standard maximal inequality (see p. 52 in [25]), we deduce that

$$\begin{aligned}
\mathbb{E}\left[\left(Z^{\tau_n}\right)_t^*\right] &\leq \frac{e+1}{e-1} + \frac{e}{e-1}\mathbb{E}\left[L_{t\wedge\tau_n}\right] \\
&\leq \frac{e+1}{e-1} + \frac{e}{e-1}\mathbb{E}\left[A_{t\wedge\tau_n}\right].
\end{aligned}$$

The proof is finished by an application of the monotone convergence theorem. □

Remark 7 (A comparison with Kazamaki's criterion) The criterion of finite transformed average energy:

$$\mathbb{E}\left[\int_0^t Z_s |H_s|^2 \, ds\right] < \infty, \tag{11}$$

turns out to be a criterion for Z to be a martingale that is independent of Kazamaki's criterion—and therefore of Novikov's criterion—in the sense that one is sometimes applicable when the other is not.

We give two examples to illustrate this. First, we can make use of a simple example introduced in Revuz and Yor [25] (p. 366, Exercise 2.10.40) in which Kazamaki's criterion fails. Let W be a scalar Brownian motion with $W_0 = 0$

and set $H_t = \alpha W_t$ for some $\alpha > 0$. Recall that Kazamaki's criterion is that $\exp\left(\frac{1}{2}\int_0^{\cdot} H_s^T dW_s\right)$ should be a submartinagle. But, as Revuz and Yor point out, $Z_\cdot = \exp\left(\alpha \int_0^{\cdot} W_s dW_s - \frac{\alpha^2}{2}\int_0^{\cdot} W_s^2 ds\right)$ is a true martingale on $[0, \infty)$ for all α, but $\exp\left(\frac{\alpha}{2}\int_0^t W_s dW_s\right)$ ceases to be a submartingale for $t \geq \alpha^{-1}$. However, under the transformed probability measure $\tilde{\mathbb{P}}$, defined on the σ-ring $\cup_{t\geq 0}\mathcal{F}_t$ by

$$\left.\frac{d\tilde{\mathbb{P}}}{d\mathbb{P}}\right|_{\mathcal{F}_t} = Z_t,$$

W is turned into a Gaussian semimartingale satisfying

$$W_t = \int_0^t \alpha W_s ds + B_t$$

for some $\left(\{\mathcal{F}_t\}_{t\geq 0}, \tilde{\mathbb{P}}\right)$ Brownian motion B. But W can also be expressed as

$$W_t = \int_0^t e^{\alpha(t-s)} dB_s$$

and then it is straightforward to show that for all $t \geq 0$

$$\mathbb{E}\left[\int_0^t Z_s H_s^2 ds\right] = \tilde{\mathbb{E}}\left[\alpha^2 \int_0^t W_s^2 ds\right] = \frac{1}{4}\left(e^{2\alpha t} - 2\alpha t - 1\right).$$

Hence the transformed average energy condition is applicable in this case.

To give an example in the other direction, we construct a stopping time $S < 1$ a.s., a continuous local martingale X on $[0, 1]$ with quadratic variation

$$\langle X \rangle_\cdot = \int_0^{S\wedge\cdot} \frac{dr}{(1-r)^2}$$

such that $e^{\frac{1}{2}X_\cdot}$ is a submartingale on $[0, 1]$ and the transformed average energy satisfies

$$\mathbb{E}\left[\int_0^1 \frac{\zeta_r}{(1-r)^2} dr\right] = \mathbb{E}\left[\int_0^S \frac{\zeta_r}{(1-r)^2} dr\right] = \infty,$$

where ζ is the exponential local martingale $\zeta_t = e^{X_t - \frac{1}{2}\langle X \rangle_t}$. For this example, Kazamaki's criterion implies that ζ is a martingale on the closed interval $[0, 1]$, while the average energy condition fails to do so for $t = 1$.

Suppose W is an $\{\mathcal{F}_t\}$-adapted Brownian motion, null at zero, on a filtered probability space $\left(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, \tilde{\mathbb{P}}\right)$ and N is an $\mathcal{F}_0$−measurable integer-valued random variable, independent of W, with distribution under $\mathbb{P}$ given by $\mathbb{P}(N = n) = 1/(n(n+1))$ for $n \in \mathbb{N}$. Introduce a sequence of stopping times

$$T_n := \inf \{t \geq 0 : W_t = -1 \text{ or } W_t = n\},$$

with the convention that $T_n = \infty$ if this set is empty.

For each n, $\tilde{\mathbb{P}}(T_n < \infty) = 1$ and $W_{T_n \wedge \cdot}$ is a zero-mean bounded martingale with $\tilde{\mathbb{P}}\left(W_{T_n} = -1\right) = n/(n+1)$ and $\tilde{\mathbb{P}}\left(W_{T_n} = n\right) = 1/(n+1)$. Furthermore, by Jensen's inequality, $\exp\left(-\frac{1}{2} W_{T_n \wedge \cdot}\right)$ is a positive submartingale which is bounded uniformly in n and t by $e^{1/2}$. We now let $T = T_N$. The process $e^{-\frac{1}{2} W_{T \wedge \cdot}}$ is also a bounded submartingale since for all stopping times $R < S$ and for all n

$$\begin{aligned} \mathbb{E}_{\tilde{\mathbb{P}}}\left[e^{-\frac{1}{2} W_{T \wedge R}}; N = n\right] &= \mathbb{E}_{\tilde{\mathbb{P}}}\left[e^{-\frac{1}{2} W_{T_n \wedge R}}; N = n\right] \leq \mathbb{E}_{\tilde{\mathbb{P}}}\left[e^{-\frac{1}{2} W_{T_n \wedge S}}; N = n\right] \\ &= \mathbb{E}_{\tilde{\mathbb{P}}}\left[e^{-\frac{1}{2} W_{T \wedge S}}; N = n\right]. \end{aligned}$$

Now the strictly positive local martingale $\tilde{Z}_\cdot := e^{-W_{T \wedge \cdot} - \frac{1}{2} T \wedge \cdot}$ is bounded and hence is a uniformly integrable martingale of the form $\tilde{Z}_t = \mathbb{E}_{\tilde{\mathbb{P}}}\left[\tilde{Z}_T \middle| \mathcal{F}_t\right]$. Let $\mathbb{P}$ be the probability measure which is equivalent to $\tilde{\mathbb{P}}$ defined by $d\mathbb{P} = \tilde{Z}_T d\tilde{\mathbb{P}}$. Define on $[0, \infty]$ the process Y :

$$\begin{aligned} Y_t &= W_{T \wedge t} + T \wedge t \text{ for } t \in [0, \infty), \text{ and} \\ Y_\infty &= (W_T + T) \, 1_{\{T < \infty\}}. \end{aligned}$$

Girsanov's Theorem tells us that Y is a local martingale under $\mathbb{P}$. Set $Z_t = \left(\tilde{Z}_t\right)^{-1}$. Then $Z_\cdot = e^{Y_\cdot - T \wedge \cdot}$ on $[0, \infty)$. We need to show that $e^{\frac{1}{2} Y_\cdot}$ is a submartingale under $\mathbb{P}$. But this follows from the fact that for any finite stopping times $R < S$,

$$\begin{aligned} \mathbb{E}_{\mathbb{P}}\left[e^{\frac{1}{2} Y_R}\right] &= \mathbb{E}_{\tilde{\mathbb{P}}}\left[\tilde{Z}_R e^{\frac{1}{2} Y_R}\right] = \mathbb{E}_{\tilde{\mathbb{P}}}\left[e^{-\frac{1}{2} W_R}\right] \\ &\leq \mathbb{E}_{\tilde{\mathbb{P}}}\left[e^{-\frac{1}{2} W_S}\right] = \mathbb{E}_{\mathbb{P}}\left[e^{\frac{1}{2} Y_S}\right], \end{aligned}$$

where we have used the fact that $e^{-\frac{1}{2} W_\cdot}$ is a submartingale under $\tilde{\mathbb{P}}$. So Kazamaki's criterion allows us to construct a probability measure $\overline{\mathbb{P}}$ such that, for all stopping times S, $d\overline{\mathbb{P}} = Z_S d\mathbb{P}$ on $\mathcal{F}_S \cap \{S < \infty\}$. Since $Z_S = Z_{S \wedge T} = \tilde{Z}^{-1}_{S \wedge T}$, and $\mathbb{P}(T < \infty) =$

$\tilde{\mathbb{P}}(T < \infty) = 1$ the measures $\overline{\mathbb{P}}$ and $\tilde{\mathbb{P}}$ coincide on $\mathcal{F}_T$. Now the quadratic variation $\langle Y \rangle_\cdot = T \wedge \cdot$, and the integral in the relevant transformed average energy condition is

$$\begin{aligned}\mathbb{E}_{\mathbb{P}}\left[\int_0^T Z_s ds\right] &= \mathbb{E}_{\mathbb{P}}\left[T Z_T\right] = \mathbb{E}_{\tilde{\mathbb{P}}}\left[T\right] \\ &= \mathbb{E}_{\tilde{\mathbb{P}}}\left[W_T^2\right] \\ &= \tilde{\mathbb{P}}\left(W_T = -1\right) + \mathbb{E}_{\tilde{\mathbb{P}}}\left[W_T^2; W_T \geq 1\right] \\ &= \tilde{\mathbb{P}}\left(W_T = -1\right) + \sum_{n=1}^{\infty} \frac{n}{(n+1)^2} \\ &= \infty.\end{aligned}$$

We now turn to the construction of X and ζ. Let $\sigma : [0, 1] \to [0, \infty]$ be the time-change $\sigma(t) = t(1-t)^{-1}$. Let $X_t = Y_{\sigma(t)}$ and $\zeta_t = Z_{\sigma(t)}$. Then $X, e^{\frac{1}{2}X}$ and ζ inherit, respectively, the local martingale, the submartingale and the uniformly integrable martingale properties of $Y, e^{\frac{1}{2}Y}$ and Z, though with respect to the filtration $\left\{\mathcal{F}_{\sigma(t)}\right\}_{0 \leq t < 1}$. Set $S = T(1+T)^{-1}$; that is, $\sigma(S) = T$. Then the quadratic variation

$$\langle X \rangle_t = T \wedge \sigma(t) = \frac{S \wedge t}{1 - S \wedge t} = \int_0^{S \wedge t} \frac{dr}{(1-r)^2}.$$

Furthermore

$$\mathbb{E}_{\mathbb{P}}\left[\int_0^S \frac{\zeta_r}{(1-r)^2} dr\right] = \mathbb{E}_{\mathbb{P}}\left[\int_0^T Z_s ds\right] = \infty.$$

This completes the justification of the properties of the example.

Remark 8 We record four observations:

1. The proof does not require the a priori assumption that $\mathbb{E}\left[\int_0^t |H_s|^2 ds\right] < \infty$. However observe that

$$\mathbb{E}\left[\int_0^t Z_s |H_s|^2 ds\right] = \mathbb{E}\left[\int_0^t \mathbb{E}\left[Z_t|\mathcal{F}_s\right] |H_s|^2 ds\right] = \mathbb{E}\left[Z_t \int_0^t |H_s|^2 ds\right].$$

2. If the Brownian motion W is independent of H then using the sequence of stopping times $(\tau_n)_{n=1}^{\infty}$, $0 \leq \tau_n \leq \tau_{n+1}$, defined by

$$\tau_n = \inf\left\{t \geq 0 : |H_t| \geq n\right\},$$

we get that

$$\mathbb{E}\left[Z_{t\wedge\tau_n}|H\right] = \mathbb{E}\left[\exp\left(\int_0^{t\wedge\tau_n} H_s^\top dW_s - \frac{1}{2}\int_0^{t\wedge\tau_n} |H_s|^2\, ds\right)\middle| H\right]$$

$$= \exp\left(-\frac{1}{2}\int_0^{t\wedge\tau_n} |H_s|^2\, ds\right)\mathbb{E}\left[\exp\left(\int_0^{t\wedge\tau_n} H_s^\top dW_s\right)\middle| H\right] = 1.$$

In particular, the stopped process Z^{τ_n} is a martingale. Moreover

$$\mathbb{E}\left[\int_0^{t\wedge\tau_n} Z_s\,|H_s|^2\, ds\middle| H\right] = \int_0^t \mathbb{E}\left[Z_{s\wedge\tau_n}\middle| H\right]\left|H_{s\wedge\tau_n}\right|^2 ds = \int_0^{t\wedge\tau_n} |H_s|^2\, ds.$$

Hence, by an application of the monotone convergence theorem

$$\mathbb{E}\left[\int_0^t Z_s\,|H_s|^2\, ds\right] = \mathbb{E}\left[\int_0^t |H_s|^2\, ds\right].$$

By the same argument one can prove directly that Z is a martingale under the weaker condition (8). This result is contained in Lemma 11.3.1 of [13].

3. Assume that $\mathbb{E}[A_t] < \infty$ for all $t \geq 0$, then $(Z-1)$ is a zero-mean martingale and $\mathbb{E}\left[(Z-1)_t^*\right] < 1 + \mathbb{E}\left[Z_t^*\right] < \infty$. Since $\langle Z-1\rangle_t = \int_0^t Z_s^2\,|H_s|^2\, ds$ the Burkholder-Davis-Gundy inequalities give

$$\mathbb{E}\left[\left(\int_0^t Z_s^2\,|H_s|^2\, ds\right)^{1/2}\right] < \infty$$

for all $t \geq 0$.

4. The finiteness of the transformed average energy does not imply that the average energy itself is finite. The following example illustrates this. Let $W = (W_t)_{0\leq t\leq 1}$ be a one-dimensional $(\mathcal{F}_t)_{0\leq t\leq 1}$ -adapted Brownian motion with $W_0 = 0$, and suppose that $\mathcal{F}_0$ carries a uniform [0, 1] random variable which is independent of W. Then we will prove there exists an $(\mathcal{F}_t)$-optional process $H = (H_t)_{0\leq t\leq 1}$ such that the local martingale Z given by

$$Z_t = \exp\left(\int_0^t H_s dW_s - \frac{1}{2}\int_0^t H_s^2 ds\right)$$

is a martingale on [0, 1] for which

$$\mathbb{E}\left[\int_0^1 Z_s H_s^2 ds\right] < \infty \text{ and } \mathbb{E}\left[\int_0^1 H_s^2 ds\right] = \infty.$$

To construct Z we will make use of the Gaussian martingale $B_t = \int_0^t \frac{1}{1-s} dW_s$ defined on [0, 1). We notice that $(1-t)\,B_t$ is a Brownian bridge on [0, 1) and the related process $V_t := \exp\left[B_t - \frac{t}{2(1-t)}\right]$ is just the martingale of densities on $(\mathcal{F}_t)_{0\le t<1}$ that turns W into a Brownian bridge, cf. [25]. But the property we exploit is the existence of a Brownian motion $\overline{B}$ on $[0,\infty)$ such that $B_t = \overline{B}_{\sigma(t)}$ wherein $\sigma(t) := t(1-t)^{-1}$. Let

$$X_t = \int_0^t \frac{V_s ds}{(1-s)^2}$$

be defined on [0, 1] and introduce the sequence of stopping times

$$T_n = \inf\left\{t \ge 0 : X_t = \frac{n(1-t)}{n(1-t)+t}\right\}.$$

Since $X.$ is non-negative and increasing with $X_0 = 0$ and the function $t \mapsto \frac{n(1-t)}{n(1-t)+t}$ is strictly decreasing to 0, each T_n is strictly less than one. Furthermore the sequence $(T_n)_{n=1}^{\infty}$ increases to a limit $T_\infty \le 1$. We need to prove that $\mathbb{P}(T_\infty = 1) > 0$. Using the fact that

$$\lim_{n\to\infty} \frac{n(1-t)}{n(1-t)+t} = 1 \text{ for all } t < 1,$$

it follows that $\mathbb{P}(T_\infty = 1) = \mathbb{P}(X_1 < 1)$. However,

$$\begin{aligned} X_1 &= \int_0^1 \frac{1}{(1-t)^2} \exp\left[B_t - \frac{t}{2(1-t)}\right] \mathrm{d}t \\ &= \int_0^\infty \exp\left(\overline{B}_t - \frac{1}{2}s\right) \mathrm{d}s \end{aligned}$$

and it is a result of Dufresne [8] (see also Yor [40], p. 15) that this latter integral is distributed as twice the inverse of a standard exponential random variable Y. In particular $\mathbb{P}(X_1 < 1) = \mathbb{P}(Y > 2) = e^{-2}$, from which it follows that $\mathbb{P}(T_\infty = 1) > 0$ and, that, $\mathbb{E}\left[\frac{T_\infty}{1-T_\infty}\right] = \infty$. The monotone convergence theorem implies that the sequence

$$m(n) := \mathbb{E}\left[\frac{T_n}{1-T_n}\right] \uparrow \infty \text{ as } n \to \infty.$$

Let U be the uniform $[0, 1]$ random variable on $\mathcal{F}_0$ referred to earlier. We can construct, as a measurable function of U, an integer random variable N satisfying

$$\mathbb{E}[m(N)] = \infty.$$

If T denotes the stopping time T_N then $T < 1$, but also

$$\mathbb{E}\left[\frac{T}{1-T}\right] = \mathbb{E}[m(N)] = \infty.$$

Finally we take $Z_t := M_{t\wedge T}$ on $[0, 1]$ and define H to be the corresponding integrand

$$H_t = \begin{cases} (1-t)^{-1} & \text{on } [0, T) \\ 0 & \text{on } [T, 1] \end{cases},$$

whereupon we have

$$\mathbb{E}\left[\int_0^1 Z_s H_s^2 ds\right] = \mathbb{E}[X_T] = \mathbb{E}\left[\frac{N(1-T)}{N(1-T)+T}\right] < 1, \text{ but}$$
$$\mathbb{E}\left[\int_0^1 H_s^2 ds\right] = \mathbb{E}\left[\int_0^T \frac{1}{(1-t)^2} dt\right] = \mathbb{E}\left[\frac{T}{1-T}\right] = \infty$$

as required.

Remark 9 For any $K > 0$, it is possible to decompose the local martingale M as

$$M = M^{sq,K} + M^{d,K},$$

where $M^{sq,K}$ is a locally square-integrable martingale with jumps bounded by a constant K and $M^{d,K}$ is a purely discontinuous local martingale with locally integrable total variation, with jumps greater than K, in such a manner that the quadratic variation process $\left[M^{sq,K}, M^{d,K}\right]$ is identically equal to 0. In what follows we will discard the dependence on the constant K in the notation for $M^{sq,K}$ and $M^{d,K}$. The first part of the statement is essentially Proposition I.4.17 in [12] while the second part follows from Theorem I.4.18 of the same reference, i.e., from the classical decomposition of the local martingale M^{sq} into its continuous and purely discontinuous parts

$$M^{sq} = M^{sq,c} + M^{sq,d}.$$

We have that

$$\left[M^{sq}, M^{d}\right] = \left[M^{sq,c}, M^{d}\right] + \left[M^{sq,d}, M^{d}\right] = 0.$$

Here $\left[M^{sq,c}, M^{d}\right]$ is null since it is the quadratic variation between a continuous and a purely discontinuous martingale and $\left[M^{sq,d}, M^{d}\right]$ is null since it is the quadratic variation of two purely discontinuous martingales with no jumps occurring at the same time.

For the following proposition, we introduce a positive $\mathcal{F}_t$ -adapted càdlàg semimartingale of the form

$$U_t = U_0 + \int_0^t a_s ds + M_t,$$

where a is a measurable $\mathcal{F}_t$-adapted process and M is a local $\mathcal{F}_t$-martingale null at zero.[5] We also assume that $E\left[U_0\right] < \infty$ and additionally that the quadratic variation processes $\left\langle W^i, M\right\rangle$ $i = 1, \ldots, m$ are absolutely continuous. In particular, there exists a measurable m-dimensional $\mathcal{F}_t$-adapted process $N = (N^i)_{i=1}^m$ such that

$$\left\langle W^i, M\right\rangle_t = \int_0^t N_s^i ds, \qquad t \geq 0, \qquad i = 1, \ldots, m.$$

Moreover we will assume that there exists a positive constant c such that

$$\max\left(|a_t|, |N_t|^2\right) \leq c \max\left(U_t, U_{t-}\right), \qquad t \geq 0. \tag{12}$$

Proposition 10 *Assume that the $\mathcal{F}_t$-adapted measurable process $H = (H^i)_{i=1}^d$ satisfies the inequality*

$$|H_t|^2 \leq c \max\left(U_t, U_{t-}\right) \quad t \geq 0. \tag{13}$$

Then the functions $t \to \mathbb{E}\left[Z_t |H_t|^2\right]$, $t \to \mathbb{E}\left[|H_t|^2\right]$ are locally bounded. In particular Lemma 4 allows us to deduce that the process Z is a $H^1(\mathbb{P})$ martingale.

Proof Let $(T_n)_{n>0}$ be a localizing sequence of stopping times such that the stopped process $\left(M^{sq}_{T_n\wedge\cdot}\right)$ is a square integrable martingale and the process $\left(M^{d}_{T_n\wedge\cdot}\right)$ is a martingale with integrable total variation $Var\left(M^d\right)_{T_n\wedge\cdot}$. Now introduce the localizing sequence $(S_n)_{n>0}$ where

---

[5] We will use the notation $[\cdot,\cdot]$ to denote the quadratic variation process of two local martingales. In addition, we will use the notation $\langle\cdot,\cdot\rangle$ to denote the predictable quadratic variation process of two locally square integrable martingales. The two processes coincide if one of the martingales is continuous. For further details see, for example, Chap. 4 of [26].

$$S_n = \inf \left\{ t \geq 0 \middle| \max \left\{ Z_t, \int_0^t |a_s|\, ds, U_{t-} \right\} \geq n \right\} \wedge T_n.$$

Note that the left continuity of the processes listed in the inner brackets implies that these processes, when stopped at S_n are bounded by n. Consider now the evolution equation for ZU, that is

$$Z_t U_t = U_0 + \int_0^t Z_s \left(a_s + H_s^\top N_s \right) ds + \int_0^t Z_s \left(H_s^\top dW_s + dM_s^{sq} + dM_s^d \right). \quad (14)$$

It follows that the expected value of $Z_t U_t$ is controlled by the sum of the expected values of the six terms on the right hand side of (14). The stochastic integral terms in (14), when stopped at S_n become genuine martingales. They can be controlled as follows:

$$\begin{aligned} \mathbb{E}\left[\left(\int_0^{t\wedge S_n} Z_s U_{s-} H_s^\top dW_s \right)^2 \right] &= \mathbb{E}\left[\int_0^{t\wedge S_n} Z_s^2 U_{s-}^2 \, |H_s|^2 \, ds \right] \\ &\leq cn^4 \mathbb{E}\left[\int_0^{t\wedge S_n} \max\left(U_t, U_{t-}\right) ds \right] \leq cn^5 t. \end{aligned}$$

Here we have used the fact that, for all $t \geq 0$, $\int_0^t P\left(U_s \neq U_{s-}\right) ds = 0$. We also have that

$$\mathbb{E}\left[\left(\int_0^{t\wedge S_n} Z_s dM_s^{sq} \right)^2 \right] = \mathbb{E}\left[\int_0^{t\wedge S_n} Z_s^2 d\left\langle M^{sq} \right\rangle_s \right] \leq n^2 \mathbb{E}\left[\left\langle M^{sq} \right\rangle_{t\wedge S_n} \right] < \infty$$

$$\mathbb{E}\left[\left| \int_0^{t\wedge S_n} Z_s dM_s^d \right| \right] \leq n\mathbb{E}\left[Var\left(M^d\right)_{S_n \wedge t} \right] < \infty.$$

By taking the expectation of both sides in (14) stopped at $t \wedge S_n$, we deduce that

$$\begin{aligned} \mathbb{E}\left[Z_t U_t 1_{\{t \leq S_n\}} \right] &\leq \mathbb{E}\left[Z_{t\wedge S_n} U_{t\wedge S_n} \right] \\ &= \mathbb{E}\left[U_0\right] + \mathbb{E}\left[\int_0^{t\wedge S_n} Z_s \left(a_s + H_s^\top N_s \right) ds \right] \end{aligned}$$

$$\leq \mathbb{E}[U_0] + 2c\mathbb{E}\left[\int_0^t Z_s \max(U_s, U_{s-}) 1_{\{s\leq S_n\}} ds\right]$$

$$\leq \mathbb{E}[U_0] + 2c\int_0^t \mathbb{E}\left[Z_s U_s 1_{\{s\leq S_n\}}\right] ds \leq e^{2ct}\mathbb{E}[U_0] < \infty.$$

Note that the last inequality follows from Gronwall's lemma. Since $\lim_{n\to\infty} S_n = \infty$, we can then deduce by the monotone convergence theorem that, for all $t > 0$,

$$\sup_{s\in[0,t]} \mathbb{E}[Z_s U_s] \leq e^{2ct}\mathbb{E}[U_0]. \tag{15}$$

The local boundedness of $t \to \mathbb{E}\left[Z_t |H_t|^2\right]$ follows from (13) and (15). Similarly we show that for all $t > 0$,

$$\sup_{s\in[0,t]} \mathbb{E}[U_s] < \infty.$$

by using the above argument with $H = 0$ for all $t \geq 0$ (and therefore $Z_t = 1$). This in turn implies the local boundedness of the functions $t \to \mathbb{E}\left[|H_t|^2\right]$. □

3 Two Particular Cases

3.1 The Signal is a Jump-Diffusion Process

We continue to assume that the observation process follows (3), and suppose that $X_t = (X_t^i)_{i=1}^d$, for all $t \geq 0$, is a càdlàg solution of a d-dimensional stochastic differential equation. This is driven by a triplet (V, W, L) comprising a p-dimensional Brownian motion $V = (V^j)_{j=1}^p$, the m-dimensional Brownian motion $W = (W^j)_{j=1}^m$ driving the observation process Y, and an $\mathbb{R}^r$-valued Lévy process $L = (L^j)_{j=1}^r$ with no centred Gaussian component and with Lévy measure F such that $F(\{0\}) = 0$. viz.

$$X_t^i = X_0^i + \int_0^t f^i(X_{s-})\,\mathrm{d}s + \sum_{j=1}^p \int_0^t \sigma^{ij}(X_{s-})\,\mathrm{d}V_s^j + \sum_{k=1}^m \int_0^t \overline{\sigma}^{ik}(X_{s-})\,\mathrm{d}W_s^k + \sum_{l=1}^r \int_0^t \tilde{\sigma}^{il}(X_{s-})\,\mathrm{d}L_s^l, \tag{16}$$

for $i = 1, \ldots, d$. We write $f = (f^i)_{i=1}^d : \mathbb{R}^d \to \mathbb{R}^d$, $\sigma = (\sigma^{ij})_{i=1,\ldots,d,j=1,\ldots,p} : \mathbb{R}^d \to \mathbb{R}^{d\times p}$, $\overline{\sigma} = (\overline{\sigma}^{ij})_{i=1,\ldots,d,j=1,\ldots,m} : \mathbb{R}^d \to \mathbb{R}^{d\times m}$ and $\tilde{\sigma} = (\tilde{\sigma}^{ij})_{i=1,\ldots,d,j=1,\ldots,r} : \mathbb{R}^d \to \mathbb{R}^{d\times r}$.

We recall that a function $g : E \to F$ between two normed spaces $(E, ||\cdot||_E)$ and $(F, ||\cdot||_F)$ has *at most linear growth* if there exists $K < \infty$ such that

$$||g(e)||_F \leq K(1 + ||e||_E)$$

for all $e \in E$. We record the assumptions to be made on the coefficients in the Eq. (16).

Condition 11 We assume f, σ, $\overline{\sigma}$ and $\tilde{\sigma}$ are Borel and have at most linear growth.

We will use μ to denote the Poisson random measure associated with L, i.e. for every $t \geq 0$ and $A \in \mathcal{B}(\mathbb{R}^r \setminus \{0\})$ the random measure $\mu(t, \cdot)$ defined by

$$\mu(t, A) := \sum_{0\leq s\leq t} 1_A(\Delta L_s).$$

We let $\nu(t, \cdot) := F(\cdot)\, t = \mathbb{E}[\mu(1, \cdot)]\, t$, where $F(\cdot)$ is the Lévy measure of L, and denote the compensated measure by $\tilde{\mu}(t, A) = \mu(t, A) - \nu(t, A)$. L then has a Lévy-Ito decomposition of the form

$$L_t = at + \int_{0<|\rho|<1} \rho\tilde{\mu}(t, \mathrm{d}\rho) + \int_{|\rho|\geq 1} \rho\mu(t, \mathrm{d}\rho). \tag{17}$$

Condition 12 Let $L = (L_t)_{t\geq 0}$ be a Lévy process with Lévy measure F. We assume the square integrability condition

$$\int_{|\rho|\geq 1} \rho^2 F(\mathrm{d}\rho) < \infty.$$

Remark 13 Whenever this condition is in force we have that

$$\int_{|\rho|\geq 1} \rho F(\mathrm{d}\rho) < \infty \quad \text{for every} \quad t \geq 0, \tag{18}$$

and hence the Lévy-Ito decomposition (17) may be rewritten as

$$L_t = bt + \int_{\mathbb{R}^r\setminus\{0\}} \rho\tilde{\mu}(t, \mathrm{d}\rho),$$

where $b := a - \int_{|\rho|\geq 1} \rho F(\mathrm{d}\rho)$.

We continue to assume the dynamics for the observation process described in (3), and we now assume that (18) holds. We can restate this example in the language of Sect. 2.1 by noticing that the process $\overline{X} = (X, Y)$ is a solution to a martingale problem, with generator A now given by

$$\begin{aligned} A\phi\left(\overline{x}\right) &= A\phi\left(x, y\right) \\ &= \mathcal{L}\phi\left(x, y\right) + \int\limits_{\mathbb{R}^r\setminus\{0\}} \Big[\phi\left(x + \tilde{\sigma}(x)\eta, y\right) - \phi\left(x, y\right) \\ &\qquad - \sum_{i=1}^{d}\sum_{l=1}^{r} \frac{\partial\phi\left(x, y\right)}{\partial x_i}\tilde{\sigma}^{il}(x)\eta^l\Big] F\left(d\eta\right) \end{aligned}$$

where

$$\begin{aligned} \mathcal{L} &= \sum_{i=1}^{d}\tilde{f}^i(x)\frac{\partial}{\partial x_i} + \sum_{k=1}^{m} h^k(x, y)\frac{\partial}{\partial y_k} \\ &\quad + \frac{1}{2}\sum_{i,j=1}^{d}\left(a^{ij}\left(x\right) + \overline{a}^{ij}\left(x\right)\right)\frac{\partial^2}{\partial x_i\partial x_j} + \frac{1}{2}\sum_{k=1}^{m}\frac{\partial^2}{\partial x_k^2}, \end{aligned}$$

with $\tilde{f}^i(x) := f^i(x) + b^i$, and $a = (a^{ij})_{i,j=1,\ldots,d} : \mathbb{R}^d \to \mathbb{R}^{d\times d}$, $\overline{a} = (\overline{a}^{ij})_{i,j=1,\ldots,d} : \mathbb{R}^d \to \mathbb{R}^{d\times d}$ are the matrix-valued function defined respectively as

$$a^{ij} = \frac{1}{2}\sum_{k=1}^{p}\sigma^{ik}\sigma^{jk} = \frac{1}{2}\left(\sigma\sigma^\top\right)^{ij} \text{ and } a^{ij} = \frac{1}{2}\sum_{k=1}^{m}\sigma^{ik}\sigma^{jk} = \frac{1}{2}\left(\overline{\sigma\sigma}^\top\right)^{ij}$$

for all $i, j = 1, \ldots, d$.

To ensure the filtering equations described in Sect. 6 can be applied to this example, we wish to establish that the functions $\mathbb{E}\left[Z_{\cdot}\left|h\left(X_{\cdot}\right)\right|^2\right]$ and $\mathbb{E}\left[\left|h\left(X_{\cdot}\right)\right|^2\right]$ are locally bounded.

Corollary 14 *Assume the coefficients in* (16) *satisfy Condition* 11 *and that* $\overline{\sigma}$ *is uniformly bounded. Let* $X_t = (X_t^i)_{i=1}^d$ *denote a* d*-dimensional jump-diffusion process which solves* (16) *for all* $t \geq 0$. *Suppose the driving Lévy process* L *has a Lévy measure* F *which satisfies* $F\left(\{0\}\right) = 0$ *and has no Gaussian part. Assume Condition* 12 *and further suppose that* X_0, V, W *and* L *are independent with* $\mathbb{E}\left[|X_0|^2\right] < \infty$. *Let* $h : \mathbb{R}^d \to \mathbb{R}^m$ *be any Borel measurable function for which there exists* $K > 0$ *such that for all* $x \in \mathbb{R}^d$

$$\left|h\left(x\right)\right| \leq K\left(1 + |x|\right),$$

and let $Z = (Z_t)_{t\geq 0}$ *be the positive local martingale which solves* $Z_t = 1 + \int_0^t Z_s h\left(X_s\right)^T dW_s$. *Then* $\mathbb{E}\left[Z_{\cdot}\left|h\left(X_{\cdot}\right)\right|^2\right]$ *and* $\mathbb{E}\left[\left|h\left(X_{\cdot}\right)\right|^2\right]$ *are locally bounded.*

Proof By exploiting Remark 13 we can rewrite the SDE governing X as

$$dX_t = \tilde{f}(X_{t-})\,\mathrm{d}t + \sigma(X_{t-})\,\mathrm{d}V_t + \overline{\sigma}(X_{t-})\,\mathrm{d}W_t + \int\limits_{\mathbb{R}^r\setminus\{0\}} \tilde{\sigma}(X_{t-})\rho\,\tilde{\mu}\,(\mathrm{d}t,\mathrm{d}\rho)\,,$$

where $\tilde{f}(x) = f(x) + b$ (b is as given in Remark 13) is clearly still locally Lipschitz. In order to apply the local boundedness lemma we need to find a suitable process U and the component processes in its decomposition. To this end we let

$$U_t = 1 + |X_t|^2\,.$$

and use Itô's formula to obtain

$$U_t = 1 + |X_0|^2 + 2\int\limits_0^t X_{s-}^T \mathrm{d}X_s + [X,X]_t\,,$$

where the quadratic variation $[X,X]$ may be computed as

$$\begin{aligned}
[X,X]_t &= \int\limits_0^t \mathrm{tr}\left[\sigma\,(X_{s-})^T\,\sigma(X_{s-}) + \overline{\sigma}\,(X_{s-})^T\,\overline{\sigma}\,(X_{s-})\right]\mathrm{d}s \\
&\quad + \int\limits_0^t \int\limits_{\mathbb{R}^r\setminus\{0\}} \mathrm{tr}\left[\tilde{\sigma}(X_{s-})\rho\rho^T\tilde{\sigma}(X_{s-})^T\right]\mu\,(\mathrm{d}s,\mathrm{d}\rho) \\
&= \int\limits_0^t \mathrm{tr}\left[\sigma\,(X_{s-})^T\,\sigma(X_{s-}) + \overline{\sigma}\,(X_{s-})^T\,\overline{\sigma}\,(X_{s-})\right]\mathrm{d}s \\
&\quad + \sum_{0\le s\le t} \mathrm{tr}\left[\tilde{\sigma}(X_{s-})\Delta L_s \Delta L_s^T \tilde{\sigma}(X_{s-})^T\right].
\end{aligned}$$

Hence we may write U as

$$U_t = U_0 + \int\limits_0^t a_s \mathrm{d}s + M_t,$$

where

$$U_0 = 1 + |X_0|^2$$
$$a_t = 2X_{t-}^T \tilde{f}(X_{t-}) + \mathrm{tr}\left[\sigma(X_{t-})^T \sigma(X_{t-}) + \overline{\sigma}(X_{t-})^T \overline{\sigma}(X_{t-})\right]$$
$$+ \int_0^t \int_{\mathbb{R}^r \setminus \{0\}} \mathrm{tr}\left[\tilde{\sigma}(X_{s-})\rho\rho^T \tilde{\sigma}(X_{s-})^T\right] F(\mathrm{d}\rho)\,\mathrm{d}s$$

and M is the local martingale

$$M_t = \int_0^t 2X_{s-}^T \left[\sigma(X_{s-})\,\mathrm{d}V_s + \overline{\sigma}(X_{s-})\,\mathrm{d}W_s\right]$$
$$+ \int_0^t \int_{\mathbb{R}^r \setminus \{0\}} \mathrm{tr}\left[\tilde{\sigma}(X_{s-})\rho\rho^T \tilde{\sigma}(X_{s-})^T\right] \tilde{\mu}(\mathrm{d}s, \mathrm{d}\rho).$$

Condition 11 on $\tilde{f}, \sigma, \overline{\sigma}$ and $\tilde{\sigma}$ ensures the existence of $C > 0$ such that

$$a_t \leq C\left(U_{t-} \vee U_t\right),$$

moreover the boundedness of $\overline{\sigma}$ gives rise to the estimate

$$\left|\langle W, M\rangle_t'\right| = |\overline{\sigma}(X_{t-})X_{t-}| \leq K\,|X_{t-}| \leq KU_{t-}^{1/2}.$$

The result then follows from Proposition 10. □

Remark 15 We may adapt this example to the case where X be an $\{\mathcal{F}_t\}$-adapted Markov process with values in a finite state space I.

3.2 The Change-Detection Filtering Problem

The following is a simple example with real-world applications which fits within the above framework. The effect we try to capture is a sudden change in the parameters of the model which describes the (stochastic) evolution of the observed process. The following illustrates how such an effect might be incorporated into the framework presented previously.

We assume that Y is the real-valued process with dynamics

$$Y_t = \int_0^t \left(b_0 + B1_{[T,\infty)}(s)\right) Y_s ds + W_t,$$

where $W = \{W_t,\ t \geq 0\}$ is a standard Brownian motion, b_0 a constant and B and T independent random variables, which are also independent of W. We also assume that $T \geq 0$ and that $\mathbb{E}\left[e^{\lambda B^2}\right] < \infty$ for all $\lambda \in \mathbb{R}$. The process $X_t = \left(X_t^1, X_t^2\right)$ is then defined by

$$X_t^1 = B \text{ and } X_t^2 = I_{[T,\infty)}(t), \quad t \geq 0,$$

whereupon the process $\overline{X}_t = \left(X_t^1, X_t^2, Y_t\right)$ is adapted to the filtration

$$\{\mathcal{F}_t\}_{t\geq 0} := \left\{\sigma\left(B, I_{[T,\infty)}(s), W_s : s \leq t\right) \vee \mathcal{N}\right\}_{t\geq 0},$$

where $\mathcal{N}$ is the class of null sets of the completed σ -field $\mathcal{F}_\infty = \overline{\sigma}(B, T, W_s, s < \infty)$. We introduce the uniquely defined càdlàg $(\mathcal{B}(\mathbb{R}) \times \mathcal{F}_t)$-optional processes

$$(t, b, \omega) \mapsto H_t^b(\omega) = \left(b_0 + b1_{[T(\omega),\infty)}(t)\right) Y_t^b(\omega)$$

$$(t, b, \omega) \mapsto Y_t^b(\omega) = \int_0^t H_s^b(\omega)\, ds + W_t(\omega),$$

and set $Z_t^b := \exp\left[-\int_0^t H_s^b dW_s - \frac{1}{2}\int_0^t (H_s)^2\, ds\right]$. Notice that B is $\mathcal{F}_0$-measurable, and hence the continuous process $\left(Z_t^B\right)_{t\geq 0}$ is an $\{\mathcal{F}_t\}$ -adapted exponential local martingale. Again, as in the previous example, we need to show that the functions $\mathbb{E}\left[Z_\cdot^B\left(H_\cdot^B\right)^2\right]$ and $\mathbb{E}\left[\left(H_\cdot^B\right)^2\right]$ are locally bounded. To do this, fix $b \in \mathbb{R}$ and take the terms U_t and c in Proposition 10 to be

$$U_t = U_t^b := 1 + \left(Y_t^b\right)^2 \text{ and } c = c(b) := 4 + (b_0 + b)^2.$$

Then we may verify that the conditions of Proposition 10 are satisfied. It is immediate from its proof that the conclusion of Proposition 10 can be strengthened to give the estimate

$$\max\left\{\mathbb{E}\left[Z_t^b\left(H_t^b\right)^2\right], \mathbb{E}\left[\left(H_t^b\right)^2\right]\right\} \leq e^{c(b)t}\mathbb{E}\left[U_0^b\right] = e^{c(b)t}.$$

Consequently

$$\mathbb{E}\left[Z_t^B\left(H_t^B\right)^2\right] = \mathbb{E}\left[\mathbb{E}\left[Z_t^b\left(H_t^b\right)^2\right]\Big|_{b=B}\right] \leq \mathbb{E}\left[e^{c(B)t}\right]$$

and similarly

$$\mathbb{E}\left[\left(H_t^B\right)^2\right] \leq \mathbb{E}\left[e^{c(B)t}\right].$$

These inequalities, together with the moment condition on B, give the required result.

4 The Change of Probability Measure Method

We now have all the ingredients required for introducing a probability measure with respect to which the process Y becomes a Brownian motion. We return to the set-up of Sect. 2. Define $Z = (Z_t)_{t\geq 0}$ to be the exponential local martingale

$$Z_t = \exp\left(-\int_0^t h\left(\overline{X}_s\right)^\top dW_s - \frac{1}{2}\int_0^t \left|h\left(\overline{X}_s\right)\right|^2 ds\right).$$

The change of probability measure method consists in modifying the probability measure on Ω by means of Girsanov's theorem. As we require Z to be a martingale in order to construct the change of measure, Lemma 4 suggests the following as a suitable condition to impose upon h,

$$\mathbb{E}\left[\int_0^t Z_s \left\|h(\overline{X}_s)\right\|^2 \mathrm{d}s\right] < \infty, \quad \forall t > 0. \tag{19}$$

Let us assume that (19) holds. Then, by Lemma 4, Z is a true martingale. Let $\tilde{\mathbb{P}}$ be the probability measure defined on the field $\bigcup_{0\leq t<\infty} \mathcal{F}_t$ that is specified by its Radon–Nikodym derivative Z_t on each $\mathcal{F}_t$ with respect to the corresponding trace of $\mathbb{P}$; that is, for each $t \geq 0$:

$$\left.\frac{\mathrm{d}\tilde{\mathbb{P}}}{\mathrm{d}\mathbb{P}}\right|_{\mathcal{F}_t} = Z_t.$$

$\tilde{\mathbb{P}}$ restricted to each $\mathcal{F}_t$ is equivalent to $\mathbb{P}$ since Z_t is a positive random variable.[6]

Let $\tilde{Z} = \{\tilde{Z}_t,\ t \geq 0\}$ be the process defined as $\tilde{Z}_t = Z_t^{-1}$ for $t \geq 0$. Under $\tilde{\mathbb{P}}$, $\tilde{Z}_t$ satisfies the following stochastic differential equation,

$$\mathrm{d}\tilde{Z}_t = \sum_{i=1}^m \tilde{Z}_t h^i(X_t)\,\mathrm{d}Y_t^i \tag{20}$$

[6] Note that we have not defined $\tilde{\mathbb{P}}$ on $\mathcal{F}_\infty$, where $\mathcal{F}_\infty = \bigvee_{t=0}^\infty \mathcal{F}_t = \sigma\left(\bigcup_{0\leq t<\infty} \mathcal{F}_t\right)$.

and since $\tilde{Z}_0 = 1$,

$$\tilde{Z}_t = \exp\left(\sum_{i=1}^{m}\int_0^t h^i(X_s)\,\mathrm{d}Y_s^i - \frac{1}{2}\sum_{i=1}^{m}\int_0^t h^i(X_s)^2\,\mathrm{d}s\right), \tag{21}$$

then $\tilde{\mathbb{E}}[\tilde{Z}_t] = \mathbb{E}[\tilde{Z}_t Z_t] = 1$. So $\tilde{Z}$ is an $\mathcal{F}_t$-adapted martingale under $\tilde{\mathbb{P}}$ and

$$\left.\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\tilde{\mathbb{P}}}\right|_{\mathcal{F}_t} = \tilde{Z}_t \quad \text{for } t \geq 0.$$

$\mathbb{P}$ and $\tilde{\mathbb{P}}$ are therefore equivalent on each $\mathcal{F}_t$ for $t \geq 0$.

Proposition 16 *If Condition (19) is satisfied, then under $\tilde{\mathbb{P}}$ the observation process Y is a Brownian motion. Let $\varphi \in \mathcal{D}(A)$ have bounded derivatives in the y -direction, and let $\tilde{M}^{\varphi}$ denote the semimartingale*

$$\tilde{M}_t^{\varphi} := M_t^{\varphi} + \int_0^t \sum_{i=1}^{m}\left(h^i B^i\varphi + \frac{\partial\varphi}{\partial y_i}\right)\left(\overline{X}_t\right)\mathrm{d}s.$$

Then the stochastic integral $\int_0^{\cdot} \tilde{Z}_s d\tilde{M}_s^{\varphi}$ is a zero-mean martingale under $\tilde{\mathbb{P}}$.

Proof Lemma 4, together with Condition 19, ensures that Z is a martingale (under $\mathbb{P}$) and that $\tilde{\mathbb{P}}$ is a probability measure on each $\mathcal{F}_t$. That Y becomes a Brownian motion under $\tilde{\mathbb{P}}$ is an immediate consequence of Girsanov's theorem. For brevity, let β denote the process defined by

$$\beta_t := \sum_{i=1}^{m}\left(h^i B^i\varphi + \frac{\partial\varphi}{\partial y_i}\right)\left(\overline{X}_t\right);$$

then $\tilde{M}_t^{\varphi}$ can be expressed as $M_t^{\varphi} + \int_0^t \beta_s \mathrm{d}s$. It also follows from (4) and the definition of $\tilde{Z}$ that $\left\langle M^{\varphi}, \tilde{Z}\right\rangle_t = \int_0^t \tilde{Z}_s\beta_s \mathrm{d}s$. But by Itô's integration-by-parts formula

$$\begin{aligned}
\tilde{Z}_t M_t^{\varphi} &= \int_0^t M_s^{\varphi}\mathrm{d}\tilde{Z}_s + \int_0^t \tilde{Z}_s \mathrm{d}M_s^{\varphi} + \left\langle M^{\varphi}, \tilde{Z}\right\rangle_t \\
&= \int_0^t M_s^{\varphi}\mathrm{d}\tilde{Z}_s + \int_0^t \tilde{Z}_s\left(\mathrm{d}M_s^{\varphi} + \beta_s \mathrm{d}s\right) \\
&= \int_0^t M_s^{\varphi}\mathrm{d}\tilde{Z}_s + \int_0^t \tilde{Z}_s \mathrm{d}\tilde{M}_s^{\varphi}.
\end{aligned} \tag{22}$$

However M^φ being a martingale under $\tilde{\mathbb{P}}$ implies that $\tilde{Z}M^\varphi$ is a martingale under $\tilde{\mathbb{P}}$, and the first integral on the right-hand side is a martingale under $\tilde{\mathbb{P}}$ because M^φ is bounded on finite intervals and $\tilde{Z}$ itself is a martingale. The conclusion of the proposition follows. □

Remark 17 Since $\mathbb{P}$ and $\tilde{\mathbb{P}}$ are absolutely continuous with respect to each other, they have the same class of null sets $\mathcal{N}$ and therefore the (augmented) observation filtration is the same both under $\mathbb{P}$ and $\tilde{\mathbb{P}}$. Since Y is a Brownian motion under $\tilde{\mathbb{P}}$ it follows that the filtration $\{\mathcal{Y}_t,\ t \geq 0\}$ is right-continuous both under $\mathbb{P}$ and $\tilde{\mathbb{P}}$. To put it differently, $\{\mathcal{Y}_t,\ t \geq 0\}$ satisfies the usual conditions both under $\mathbb{P}$ and under $\tilde{\mathbb{P}}$.

The following proposition is a consequence of the Brownian motion property of the process Y under $\tilde{\mathbb{P}}$.

Proposition 18 *Let U be an integrable $\mathcal{F}_t$-measurable random variable. Then we have*

$$\tilde{\mathbb{E}}[U \mid \mathcal{Y}_t] = \tilde{\mathbb{E}}[U \mid \mathcal{Y}]. \tag{23}$$

Proof Let us denote by

$$\mathcal{Y}'_t = \sigma(Y_{t+u} - Y_t;\ u \geq 0);$$

then $\mathcal{Y} = \sigma(\mathcal{Y}_t, \mathcal{Y}'_t)$. Under the probability measure $\tilde{\mathbb{P}}$ the σ-algebra $\mathcal{Y}'_t \subset \mathcal{Y}$ is independent of $\mathcal{F}_t$ because Y is an $\mathcal{F}_t$-adapted Brownian motion. Hence since U is $\mathcal{F}_t$-adapted using the property (f) of conditional expectation

$$\tilde{\mathbb{E}}[U \mid \mathcal{Y}_t] = \tilde{\mathbb{E}}[U \mid \sigma(\mathcal{Y}_t, \mathcal{Y}'_t)] = \tilde{\mathbb{E}}[U \mid \mathcal{Y}].$$ □

5 Unnormalised Conditional Distribution

In this section we first prove the Kallianpur–Striebel formula and use this to define the unnormalized conditional distribution process. The notation $\tilde{\mathbb{P}}(\mathbb{P})$-a.s. below means that the result holds both $\tilde{\mathbb{P}}$-a.s. and $\mathbb{P}$-a.s. We only need to show that it holds true in the first sense since $\tilde{\mathbb{P}}$ and $\mathbb{P}$ are equivalent probability measures.

Proposition 19 (Kallianpur–Striebel) *Assume that Condition* (19) *holds. For every $\varphi \in b\mathcal{B}(\mathbb{S})$, for fixed $t \in [0, \infty)$,*

$$\pi_t(\varphi) = \frac{\tilde{\mathbb{E}}[\tilde{Z}_t \varphi(X_t) \mid \mathcal{Y}]}{\tilde{\mathbb{E}}[\tilde{Z}_t \mid \mathcal{Y}]} \qquad \tilde{\mathbb{P}}(\mathbb{P})\text{-a.s.} \tag{24}$$

Proof It is clear from the definition that $\tilde{Z}_t > 0$ $\tilde{\mathbb{P}}(\mathbb{P})$-a.s. as a consequence of which $\tilde{\mathbb{E}}[\tilde{Z}_t \mid \mathcal{Y}] > 0$ $\mathbb{P}$-a.s. and the right-hand side of (24) is well defined. It suffices to show that

$$\pi_t(\varphi)\tilde{\mathbb{E}}[\tilde{Z}_t \mid \mathcal{Y}_t] = \tilde{\mathbb{E}}[\tilde{Z}_t\varphi(X_t) \mid \mathcal{Y}_t] \qquad \tilde{\mathbb{P}}\text{-a.s.}$$

As both the left- and right-hand sides of this equation are $\mathcal{Y}_t$-measurable, this is equivalent to showing that for any bounded $\mathcal{Y}_t$-measurable random variable b,

$$\tilde{\mathbb{E}}[\pi_t(\varphi)\tilde{\mathbb{E}}[\tilde{Z}_t \mid \mathcal{Y}_t]b] = \tilde{\mathbb{E}}[\tilde{\mathbb{E}}[\tilde{Z}_t\varphi(X_t) \mid \mathcal{Y}_t]b].$$

A consequence of the definition of the process π_t is that $\pi_t\varphi = \mathbb{E}[\varphi(X_t) \mid \mathcal{Y}_t]$ $\tilde{\mathbb{P}}$-a.s., so from the definition of Kolmogorov conditional expectation

$$\mathbb{E}\left[\pi_t(\varphi)b\right] = \mathbb{E}\left[\varphi(X_t)b\right].$$

Writing this under the measure $\tilde{\mathbb{P}}$,

$$\tilde{\mathbb{E}}\left[\pi_t(\varphi)b\tilde{Z}_t\right] = \tilde{\mathbb{E}}\left[\varphi(X_t)b\tilde{Z}_t\right].$$

Since the function b is $\mathcal{Y}_t$-measurable, by the tower property of the conditional expectation,

$$\tilde{\mathbb{E}}\left[\pi_t(\varphi)\tilde{\mathbb{E}}[\tilde{Z}_t \mid \mathcal{Y}_t]b\right] = \tilde{\mathbb{E}}\left[\tilde{\mathbb{E}}[\varphi(X_t)\tilde{Z}_t \mid \mathcal{Y}_t]b\right]$$

which proves that the result holds $\tilde{\mathbb{P}}$-a.s. □

Let $\zeta = \{\zeta_t,\ t \geq 0\}$ be the process defined by

$$\zeta_t = \tilde{\mathbb{E}}[\tilde{Z}_t \mid \mathcal{Y}_t], \tag{25}$$

then as $\tilde{Z}_t$ is an $\mathcal{F}_t$-martingale under $\tilde{\mathbb{P}}$ and $\mathcal{Y}_s \subseteq \mathcal{F}_s$, it follows that for $0 \leq s < t$,

$$\tilde{\mathbb{E}}[\zeta_t \mid \mathcal{Y}_s] = \tilde{\mathbb{E}}[\tilde{Z}_t|\mathcal{Y}_s] = \tilde{\mathbb{E}}\left[\tilde{\mathbb{E}}[\tilde{Z}_t \mid \mathcal{F}_s] \mid \mathcal{Y}_s\right] = \tilde{\mathbb{E}}[\tilde{Z}_s \mid \mathcal{Y}_s] = \zeta_s.$$

Therefore by Doob's regularization theorem (see Rogers and Williams [26, Theorem II.67.7]) since the filtration $\mathcal{Y}_t$ satisfies the usual conditions we can choose a càdlàg version of ζ_t which is a $\mathcal{Y}_t$-martingale. In what follows, assume that $\{\zeta_t, t \geq 0\}$ has been chosen to be such a version. $\mathcal{Y}_t$-optional projection of $\tilde{Z}_t$ with respect to the probability measure $\tilde{\mathbb{P}}$. Given such a ζ, Proposition 19 suggests the following definition.

Definition 20 Define the *unnormalised conditional distribution* of X to be the measure-valued process $\rho = \{\rho_t,\ t \geq 0\}$ given by $\rho_t = \zeta_t\pi_t$ for any $t \geq 0$.

Lemma 21 *The process* $\{\rho_t,\ t \geq 0\}$ *is càdlàg and* $\mathcal{Y}_t$*-adapted. Furthermore, for any* $t \geq 0$*,*

$$\rho_t(\varphi) = \tilde{\mathbb{E}}\left[\tilde{Z}_t\varphi(X_t) \mid \mathcal{Y}_t\right] \qquad \tilde{\mathbb{P}}(\mathbb{P})\text{-}a.s. \tag{26}$$

Proof Both $\pi_t(\varphi)$ and ζ_t are $\mathcal{Y}_t$-adapted. By construction $\{\zeta_t,\ t \geq 0\}$ is also càdlàg. We know that $\{\pi_t,\ t \geq 0\}$ is càdlàg and $\mathcal{Y}_t$-adapted; therefore the process $\{\rho_t, t \geq 0\}$ is also càdlàg and $\mathcal{Y}_t$-adapted.

For the second part, from Propositions 18 and 19 it follows that

$$\pi_t(\varphi)\tilde{\mathbb{E}}[\tilde{Z}_t \mid \mathcal{Y}_t] = \tilde{\mathbb{E}}[\tilde{Z}_t\varphi(X_t) \mid \mathcal{Y}_t] \qquad \tilde{\mathbb{P}}\text{-a.s.},$$

From (25), $\tilde{\mathbb{E}}[\tilde{Z}_t \mid \mathcal{Y}_t] = \zeta_t$ a.s. from which the result follows. □

Corollary 22 *Assume that Condition* (19) *holds. For every* $\varphi \in B(\mathbb{S})$*,*

$$\pi_t(\varphi) = \frac{\rho_t(\varphi)}{\rho_t(\mathbf{1})} \qquad \forall t \in [0, \infty) \qquad \tilde{\mathbb{P}}(\mathbb{P})\ \text{-}a.s. \tag{27}$$

Proof It is clear from Definition 20 that $\zeta_t = \rho_t(\mathbf{1})$. The result then follows immediately. □

The Kallianpur–Striebel formula explains the usage of the term *unnormalised* in the definition of ρ_t as the denominator $\rho_t(\mathbf{1})$ can be viewed as the normalising factor.

Lemma 23 i. *Let* $\{u_t,\ t \geq 0\}$ *be an* $\mathcal{F}_t$*-progressively measurable process such that for all* $t \geq 0$*, we have*

$$\tilde{\mathbb{E}}\left[\left(\int_0^t u_s^2\,\mathrm{d}s\right)^{1/2}\right] < \infty; \tag{28}$$

then, for all $t \geq 0$*, and* $j = 1, \ldots, m$*, we have*

$$\tilde{\mathbb{E}}\left[\int_0^t u_s\,\mathrm{d}Y_s^j \,\middle|\, \mathcal{Y}\right] = \int_0^t \tilde{\mathbb{E}}[u_s \mid \mathcal{Y}]\,\mathrm{d}Y_s^j. \tag{29}$$

ii. *Let* $\tilde{M}^\varphi$ *be as defined in Proposition* 16*. Then for all* $t \geq 0$

$$\tilde{\mathbb{E}}\left[\int_0^t \tilde{Z}_s\,\mathrm{d}\tilde{M}_s^\varphi \,\middle|\, \mathcal{Y}\right] = \sum_{j=1}^m \int_0^t \tilde{\mathbb{E}}\left[\left(B^j\varphi + \frac{\partial\varphi}{\partial y_j}\right)(X_s)\,\tilde{Z}_s \,\middle|\, \mathcal{Y}\right]\mathrm{d}Y_s^j, \tag{30}$$

Proof i. To deduce the results we introduce the set of uniformly bounded test random variables

$$S_t = \left\{ \varepsilon_t = \exp\left(i \int_0^t r_s^\top \, \mathrm{d}Y_s + \frac{1}{2} \int_0^t \|r_s\|^2 \, \mathrm{d}s \right) : r \in L^\infty \left([0,t], \mathbb{R}^m \right) \right\}. \quad (31)$$

Then S_t is a total set. That is, if $a \in L^1(\Omega, \mathcal{Y}_t, \tilde{\mathbb{P}})$ and $\tilde{\mathbb{E}}[a\varepsilon_t] = 0$, for all $\varepsilon_t \in S_t$, then $a = 0$ $\tilde{\mathbb{P}}$-a.s. For a proof of this result see, for example, Lemma B.39 p. 355 in Bain and Crisan [2]. In addition, if $\varepsilon_t \in S_t$, then

$$\varepsilon_t = 1 + \int_0^t i\varepsilon_s r_s^\top \, \mathrm{d}Y_s.$$

From Condition (28) it follows, by Burkholder-Davis-Gundy's inequalities that both processes $t \to \int_0^t u_s \, \mathrm{d}Y_s^j$ and $t \to \int_0^t \tilde{\mathbb{E}}\left[u_s \mid \mathcal{Y} \right] \mathrm{d}Y_s^j$ belong to $H^1(\tilde{\mathbb{P}})$. In particular they are zero-mean martingales. We observe the following sequence of identities

$$\begin{aligned}
\tilde{\mathbb{E}}\left[\varepsilon_t \tilde{\mathbb{E}}\left[\int_0^t u_s \, \mathrm{d}Y_s^j \,\middle|\, \mathcal{Y} \right] \right] &= \tilde{\mathbb{E}}\left[\varepsilon_t \int_0^t u_s \, \mathrm{d}Y_s^j \right] \\
&= \tilde{\mathbb{E}}\left[\int_0^t u_s \, \mathrm{d}Y_s^j \right] + \tilde{\mathbb{E}}\left[\int_0^t i\varepsilon_s r_s^j u_s \, \mathrm{d}s \right] \\
&= \tilde{\mathbb{E}}\left[\tilde{\mathbb{E}}\left[\int_0^t i\varepsilon_s r_s^j u_s \, \mathrm{d}s \,\middle|\, \mathcal{Y} \right] \right] \\
&= \tilde{\mathbb{E}}\left[\int_0^t i\varepsilon_s r_s^j \, \tilde{\mathbb{E}}[u_s \mid \mathcal{Y}] \, \mathrm{d}s \right] \\
&= \tilde{\mathbb{E}}\left[\varepsilon_t \int_0^t \tilde{\mathbb{E}}[u_s \mid \mathcal{Y}] \, \mathrm{d}Y_s^j \right],
\end{aligned}$$

which completes the proof of (29).

ii. From Proposition 16 we know that $\int_0^\cdot \tilde{Z}_s \, \mathrm{d}\tilde{M}_s^\varphi$ is a zero-mean martingale under $\tilde{\mathbb{P}}$. It is therefore integrable and its conditional expectation is well defined. Notice that

$$\left\langle \tilde{M}^\varphi, Y^j \right\rangle_t = \left\langle M^\varphi, W^j \right\rangle_t = \int_0^t \left(B^j \varphi + \frac{\partial \varphi}{\partial y_j} \right) \left(\overline{X}_s \right) \, \mathrm{d}s$$

The rest of the proof of (30) is similar to that of (29). Once again we choose ε_t from the set S_t and in this case we obtain the following sequence of identities.

$$
\begin{aligned}
\tilde{\mathbb{E}}\left[\varepsilon_t \tilde{\mathbb{E}}\left[\int_0^t \tilde{Z}_s \,\mathrm{d}\tilde{M}_s^{\varphi} \,\middle|\, \mathcal{Y}\right]\right] &= \tilde{\mathbb{E}}\left[\varepsilon_t \int_0^t \tilde{Z}_s \,\mathrm{d}\tilde{M}_s^{\varphi}\right] \\
&= \tilde{\mathbb{E}}\left[\int_0^t \tilde{Z}_s \,\mathrm{d}\tilde{M}_s^{\varphi}\right] + \sum_{j=1}^m \tilde{\mathbb{E}}\left\langle \int_0^{\cdot} i\varepsilon_s r_s^j \,\mathrm{d}Y_s^j, \int_0^{\cdot} \tilde{Z}_s \,\mathrm{d}\tilde{M}_s^{\varphi} \right\rangle_t \\
&= \tilde{\mathbb{E}}\left[\int_0^t \tilde{Z}_s \,\mathrm{d}\tilde{M}_s^{\varphi}\right] + \sum_{j=1}^m \tilde{\mathbb{E}} \int_0^t i\varepsilon_s r_s^j \tilde{Z}_s \,\mathrm{d}\left\langle \tilde{M}^{\varphi}, Y^j \right\rangle_s \\
&= \sum_{j=1}^m \tilde{\mathbb{E}} \int_0^t i\varepsilon_s r_s^j \tilde{Z}_s \left(B^j \varphi + \frac{\partial \varphi}{\partial y_j}\right)(\overline{X}_s)\,\mathrm{d}s \\
&= \sum_{j=1}^m \tilde{\mathbb{E}}\left[\varepsilon_t \int_0^t \tilde{\mathbb{E}}\left[\left(B^j \varphi + \frac{\partial \varphi}{\partial y_j}\right)(\overline{X}_s)\,\tilde{Z}_s \,\middle|\, \mathcal{Y}\right] \mathrm{d}Y_s^j\right].
\end{aligned}
$$

As the identities hold for an arbitrary choice of $\varepsilon_t \in S_t$, the proof of (30) is complete. □

6 The Filtering Equations

To simplify the analysis, we will impose onto $\tilde{Z}$ a similar Condition to (19). More precisely, we will assume that,

$$
\tilde{\mathbb{E}}\left[\int_0^t \tilde{Z}_s \left\|h(\overline{X}_s)\right\|^2 \mathrm{d}s\right] < \infty, \quad \forall t > 0. \tag{32}
$$

Reverting back to $\mathbb{P}$, Condition (32) is equivalent to

$$
\mathbb{E}\left[\int_0^t \left\|h(\overline{X}_s)\right\|^2 \mathrm{d}s\right] < \infty, \quad \forall t > 0. \tag{33}
$$

From Corollary 5, it follows that $\tilde{Z}$ is an $H^1(\tilde{\mathbb{P}})$-martingale. Then $\left(\tilde{Z}_\cdot - 1\right)$ is a zero-mean martingale and $\mathbb{E}\left[\left(\tilde{Z}_\cdot - 1\right)_t^*\right] < 1 + \mathbb{E}\left[\tilde{Z}_t^*\right] < \infty$. Since $\left\langle \tilde{Z}_\cdot - 1\right\rangle_t = \int_0^t \tilde{Z}_s^2 \left|h(\overline{X}_s)\right|^2 ds$ the Burkholder-Davis-Gundy inequalities give

$$\mathbb{E}\left[\left(\int_0^t \tilde{Z}_s^2 \left|h(\overline{X}_s)\right|^2 ds\right)^{1/2}\right] < \infty \tag{34}$$

for all $t \geq 0$ and hence, for any $\varphi \in b\mathcal{B}(\mathbb{S} \times \mathbb{R}^m)$, the processes

$$t \to \int_0^t \varphi(\overline{X}_t)\tilde{Z}_t h(\overline{X}_s)^\top \mathrm{d}Y_s$$

$$t \to \int_0^t \tilde{\mathbb{E}}[\varphi(\overline{X}_t)\tilde{Z}_t h(\overline{X}_s)^\top \mid \mathcal{Y}_t]\, \mathrm{d}Y_s$$

are zero-mean $H^1(\tilde{\mathbb{P}})$ martingales. In the following, for any function $\varphi \in b\mathcal{B}(\mathbb{S} \times \mathbb{R}^m)$ such that $\varphi \in \mathcal{D}(A)$ and that has bounded partial derivatives in the y direction we will denote by $D_i\varphi, j = 1, \ldots, m$ the functions

$$D_j\varphi = h^j\left(\varphi + B^j\varphi + \frac{\partial\varphi}{\partial y_j}\right) \qquad j = 1, \ldots, m.$$

Theorem 24 *If Conditions (19) and (32) are satisfied then,*

$$\tilde{\mathbb{E}}[\tilde{Z}_t\varphi(\overline{X}_t) \mid \mathcal{Y}] = \pi_0(\varphi) + \int_0^t \tilde{\mathbb{E}}[\tilde{Z}_s A\varphi(\overline{X}_s) \mid \mathcal{Y}]\, \mathrm{d}s + \sum_{j=1}^m \tilde{\mathbb{E}}[\tilde{Z}_s D_j\varphi(\overline{X}_s) \mid \mathcal{Y}]\mathrm{d}Y_s^j \tag{35}$$

for any $\varphi \in b\mathcal{B}(\mathbb{S} \times \mathbb{R}^m)$ be a function such that $\varphi, \varphi^2 \in \mathcal{D}(A)$ and that has bounded partial derivatives in the y direction. In particular the process ρ_t satisfies the following evolution equation

$$\rho_t(\varphi) = \rho_0(\varphi) + \int_0^t \rho_s\,(A\varphi)\, \mathrm{d}s + \int_0^t \rho_s((h^\top + B^\top)\varphi)\, \mathrm{d}Y_s, \quad \tilde{\mathbb{P}}\text{-}a.s. \quad \forall t \geq 0 \tag{36}$$

for any function $\varphi \in b\mathcal{B}(\mathbb{S})$ be a function such that $\varphi \in \mathcal{D}(A)$.

Proof Using Itô's formula and integration-by-parts, we find

$$\begin{aligned}
\mathrm{d}\left(\tilde{Z}_t\varphi(\overline{X}_t)\right) &= \tilde{Z}_t A\varphi(\overline{X}_t)\,\mathrm{d}t + \tilde{Z}_t \mathrm{d}M_t^{\varphi} + \varphi(\overline{X}_t)\tilde{Z}_t h^{\top}(\overline{X}_t)\,\mathrm{d}Y_t + \sum_{j=1}^{m} \tilde{Z}_t h^i(\overline{X}_t)\left\langle M^{\varphi}, Y^i\right\rangle_t \\
&= \tilde{Z}_t\left[A\varphi(\overline{X}_t) + \sum_{j=1}^{m} h^i(\overline{X}_t)\left(B^i\varphi\left(\overline{X}_t\right) + \frac{\partial\varphi}{\partial y_i}\left(\overline{X}_t\right)\right)\right]\mathrm{d}t \\
&\quad + \tilde{Z}_t\mathrm{d}M_t^{\varphi} + \varphi(\overline{X}_t)\tilde{Z}_t h^{\top}(\overline{X}_t)\,\mathrm{d}Y_t \\
&= \tilde{Z}_t A\varphi(\overline{X}_t)\mathrm{d}t + \tilde{Z}_t\mathrm{d}\tilde{M}_t^{\varphi} + \varphi(\overline{X}_t)\tilde{Z}_t h^{\top}(\overline{X}_t)\,\mathrm{d}Y_t.
\end{aligned} \tag{37}$$

We next take the conditional expectation with respect to $\mathcal{Y}$ and obtain

$$\begin{aligned}
\tilde{\mathbb{E}}[\tilde{Z}_t\varphi(\overline{X}_t) \mid \mathcal{Y}] &= \tilde{\mathbb{E}}[\tilde{Z}_0\varphi(\overline{X}_t) \mid \mathcal{Y}] + \int_0^t \tilde{\mathbb{E}}[\tilde{Z}_t A\varphi(\overline{X}_t) \mid \mathcal{Y}]\,\mathrm{d}s \\
&\quad + \tilde{\mathbb{E}}\left[\int_0^t \tilde{Z}_s\mathrm{d}\tilde{M}_s^{\varphi} \mid \mathcal{Y}\right] + \tilde{\mathbb{E}}\left[\int_0^t \varphi(\overline{X}_s)\tilde{Z}_s h^{\top}(\overline{X}_s)\,\mathrm{d}Y_s \mid \mathcal{Y}\right],
\end{aligned} \tag{38}$$

where we have used Fubini's theorem (the conditional version) to get the second term on the right hand side of (38). Observe that, since $\tilde{Z}$ is an $H^1(\tilde{\mathbb{P}})$-martingale, we have

$$\tilde{\mathbb{E}}\left[\left(\int_0^t \tilde{Z}_s^2\,\mathrm{d}s\right)^{1/2}\right] \leq \sqrt{t}\tilde{\mathbb{E}}\left[\tilde{Z}_s^*\right] < \infty.$$

Also from (34) we get that

$$\tilde{\mathbb{E}}\left[\left(\int_0^t \left(\varphi(\overline{X}_s)\tilde{Z}_s h^j(\overline{X}_s)\right)^2\,\mathrm{d}s\right)^{1/2}\right] \leq ||\varphi||\,\mathbb{E}\left[\left(\int_0^t \tilde{Z}_s^2\left|h(\overline{X}_s)\right|^2 ds\right)^{1/2}\right] < \infty.$$

In other words Condition (28) is satisfied for $u = \varphi\tilde{Z}h^j$. The identity (35) then follows from (38) by applying (29) and (30). Identity (36) follows immediately after observing that the terms containing the partial derivatives in the y direction $\frac{\partial\varphi}{\partial y_i}$ are zero since the function no longer depends on y. □

Theorem 25 *If Conditions* (19) *and* (32) *are satisfied then the conditional distribution of the signal* π_t *satisfies the following evolution equation*

$$\pi_t(\varphi) = \pi_0(\varphi) + \int_0^t \pi_s(A\varphi)\,\mathrm{d}s + \int_0^t \left(\pi_s(\varphi h^\top) - \pi_s(h^\top)\pi_s(\varphi) + \pi_t(B^\top\varphi)\right)(\mathrm{d}Y_s - \pi_s(h)\,\mathrm{d}s), \quad (39)$$

for any $\varphi \in \mathcal{D}(A)$.

Proof Since $A\mathbf{1} = 0$, it follows from (1) that $M^{\mathbf{1}} \equiv 0$, which together with (4) implies that

$$\int_0^t B^i\mathbf{1}\left(\overline{X}_s\right) ds = 0,$$

for any $t \geq 0$ and $i = 1, \ldots, m$, so

$$\sum_{j=1}^m \int_0^t \rho_s\left(h^j B^j \mathbf{1}\right) \mathrm{d}s = 0.$$

Hence, from (36), one obtains that $\rho_t(\mathbf{1})$ satisfies the following equation

$$\rho_t(\mathbf{1}) = 1 + \int_0^t \rho_s(h^\top)\,\mathrm{d}Y_s.$$

Let $(U_n)_{n>0}$ be the sequence of stopping times

$$U_n = \inf\left\{t \geq 0 \,\middle|\, \rho_t(\mathbf{1}) \leq \frac{1}{n}\right\}.$$

Then

$$\rho_t^{U_n}(\mathbf{1}) = \rho_{t\wedge U_n}(\mathbf{1}) = 1 + \int_0^{t\wedge U_n} \rho_s(h^\top)\,\mathrm{d}Y_s,$$

We apply Itô's formula to the stopped process $t \to \rho_{t\wedge U_n}(\mathbf{1})$ and the function $x \mapsto \frac{1}{x}$ to obtain that

$$\frac{1}{\rho_t^{U_n}(\mathbf{1})} = 1 - \int_0^{t\wedge U_n} \frac{\rho_s(h^\top)}{\rho_s(\mathbf{1})^2}\mathrm{d}Y_s + \int_0^{t\wedge U_n} \frac{\rho_s(h^\top)\rho_s(h)}{\rho_s(\mathbf{1})^{\,3}}\,\mathrm{d}s \quad (40)$$

By using (stochastic) integration by parts, (40), the equation for $\rho_t(\varphi)$ and the Kallianpur–Striebel formula, we obtain

$$\frac{\rho_t^{U_n}(\varphi)}{\rho_t^{U_n}(\mathbf{1})} = \pi_0(\varphi) + \int_0^{t\wedge U_n} \pi_s(A\varphi)\,\mathrm{d}s + \int_0^{t\wedge U_n} \pi_s((h^\top + B^\top)\varphi)\,\mathrm{d}Y_s - \int_0^{t\wedge U_n} \pi_s(\varphi)\pi_s(h^\top)\mathrm{d}Y_s$$
$$+ \int_0^{t\wedge U_n} \pi_s(\varphi)\pi_s(h^\top)\pi_s(h)\,\mathrm{d}s - \int_0^{t\wedge U_n} \pi_s((h^\top + B^\top)\varphi)\,\pi_s(h)\mathrm{d}s$$

As $\lim_{n\to\infty} U_n = \infty$ almost surely, we obtain the result by taking the limit as n tends to infinity. □

Remark 26 The jump-diffusion example and the change detection model discussed in Sect. 3 both satisfy Conditions (19) and (33). Therefore the two previous theorems can be applied to these two cases.

Acknowledgments The authors are grateful to J. Ruf for setting us straight on the provenance of Corollary 5 and on other points in the paper.

References

1. D. Applebaum, *Lévy Processes and Sochastic Calculus Cambridge*, vol. 93, Studies in Advanced Mathematics (Cambridge University Press, Cambridge, 2004)
2. A. Bain, D. Crisan, *Fundamentals of Stochastic Filtering*, vol. 60, Stochastic Modelling and Applied Probability (Springer, Berlin, 2008)
3. P. M. Brémaud, A martingale approach to point processes, Ph.D. thesis, University of California, Berkeley (1972)
4. R.S. Bucy, Nonlinear filtering. IEEE Trans. Autom. Control **10**, 198 (1965)
5. J.M.C. Clark, D. Crisan, On a robust version of the integral representation formula of nonlinear filtering. Probab. Theory Relat. Fields **133**(1), 43–56 (2005)
6. D. Crisan, B. Rozovsky, *The Oxford Handbook of Nonlinear Filtering* (Oxford University Press, Oxford, 2011)
7. M.H.A. Davis, Detection Theory of Poisson Processes, Publication 73/8, Department of Computing and Control (Imperial College of Science and Technology, London, 1973)
8. D. Dufresne, The distribution of a perpetuity, with applications to risk theory and pension funding. Scand. Actuar. J. **1990**(1–2), 39–79 (1990)
9. T.E. Duncan, Likelihood functions for stochastic signals in white noise. Inf. Control **16**, 303–310 (1970)
10. M. Fujisaki, G. Kallianpur, H. Kunita, Stochastic differential equations for the non linear filtering problem. Osaka J. Math. **9**, 19–40 (1972)
11. B.I. Grigelionis, On stochastic equations for nonlinear filtering problem of stochastic processes. Lietuvos Matematikos Rinkinys **12**, 37–51 (1972)
12. J. Jacod, A.N. Shiryaev, *Limit Theorems for Stochastic Processes, 288* (Springer, Berlin, 1987)
13. G. Kallianpur, *Stochastic Filtering Theory, Applications of Mathematics, 13* (Springer, Berlin, 1980)
14. H. Kushner, On the differential equations satisfied by conditional densities of Markov processes, with applications. SIAM J. Control **2**, 106–119 (1964)

15. H.J. Kushner, Dynamical equations for optimal nonlinear filtering. J. Differ. Equ. **3**, 179–190 (1967)
16. R.S. Liptser, A.N. Shiryaev, Statistics of Random Processes. II. Applications, Translated from the 1974 Russian original by A.B. Aries. Second, revised and expanded edition. Applications of Mathematics (New York), 6. Stochastic Modelling and Applied Probability. Springer, Berlin (2001)
17. R.S. Liptser, A.N. Shiryaev, Statistics of Random Processes. I. General theory. Translated from the 1974 Russian original by A.B. Aries. Second, revised and expanded edition. Applications of Mathematics (New York), 5. Stochastic Modelling and Applied Probability. Springer, Berlin (2001)
18. N.V. Krylov, B.L. Rozovsky, The Cauchy problem for linear stochastic partial differential equations. Izv. Akad. Nauk SSSR Ser. Mat. **41**(6), 1329–1347 (1977)
19. N.V. Krylov, B.L. Rozovsky, Conditional distributions of diffusion processes. Izv. Akad. Nauk. SSSR Ser. Mat. **42**(2), 356–378 (1978)
20. T. Kailath, An innovations approach to least-squares estimation. I. Linear filtering in additive white noise. IEEE Trans. Autom. Control **13**, 646–655 (1968)
21. G. Kallianpur, C. Striebel, Estimation of stochastic systems: arbitrary system process with additive noise observation errors. Ann. Math. Stat. **39**, 785–801 (1968)
22. G. Kallianpur, C. Striebel, Stochastic differential equations occurring in the estimation of continuous parameter stochastic processes. Teor. Veroyatn. Primen. **14**(4), 597–622 (1969)
23. R.E. Mortensen, Stochastic optimal control with noisy observations. Int. J. Control **1**(4), 455–464 (1966)
24. E. Pardoux, Equations aux dérivées partielles stochastiques non linéarires monotones. Ph.D. thesis, Univ Paris XI, Orsay (1975)
25. D. Revuz, M. Yor, *Continuous Martingales and Brownian motion*, vol. 293, 3rd edn. (Springer, Berlin, 1999)
26. L.C.G. Rogers, D. Williams, *Diffusions, Markov Processes, and Martingales. Vol. 1. Reprint of the second*, Cambridge Mathematical Library (Cambridge University Press, Cambridge, 1994). 2000
27. J. Ruf, A new proof for the conditions of Novikov and Kazamaki. Stoch. Process. Appl. **123**, 404–421 (2013)
28. J.H. van Schuppen, Estimation theory for continuous time processes, a martingale approach, Ph.D. thesis, University of California, Berkeley (1973)
29. J.H. Van Schuppen, Filtering, prediction and smoothing for counting process observations, a martingale approach. SIAM J. Appl. Math. **32**(3), 552–570 (1977)
30. J.H. Van Schuppen, E. Wong, Transformation of local martingales under a change of law. Ann. Prob. **2**, 879–888 (1974)
31. R. Boel, P. Varaiya, E. Wong, Martingales on jump processes. II. Applications. SIAM J. Control **13**(5), 1022–1061 (1975)
32. A. Segall, A martingale approach to modeling, estimation and detection of jump processes, Ph.D. thesis, Stanford University, Stanford, California (1973)
33. A.N. Shiryaev, On stochastic equations in the theory of conditional Markov process. Teor. Verojatnost. i Primenen. **11**, 200–206 (1966)
34. D.L. Snyder, IEEE Trans. Inf. Theory, Filtering and detection for doubly stochastic Poisson processes. **18**, 91–102 (1972)
35. R.L. Stratonovich, Application of the theory of Markov processes for optimum filtration of signals. Radio Eng. Electron. Phys **1**, 119 (1960)
36. R.L. Stratonovich, Conditional Markov processes. Theory Prob. Appl. **5**(2), 156–178 (1960)
37. J.C. Willems, J.H. van Schuppen, Stochastic systems and the problem of state space realization. Geometrical methods for the study of linear systems (Proc. NATO Adv. Study Inst., Harvard Univ., Cambridge, Mass., 1979), pp. 283–313, NATO Adv. Study Inst. Ser., Ser. C: Math. Phys. Sci., 62, Reidel, Dordrecht-Boston, Mass. (1980)
38. W.M. Wonham, Some applications of stochastic differential equations to optimal nonlinear filtering. J. Soc. Ind. Appl. Math. Ser. A Control **2**, 347–369 (1965)

39. J. Xiong, *An Introduction to Stochastic Filtering Theory*, Oxford Graduate Texts in Mathematics, 18 (Oxford University Press, Oxford, 2008)
40. M. Yor, *Exponential Functionals of Brownian Motion and Related Processes*, 2nd edn. (Springer, Berlin, 1991)
41. M. Zakai, On the optimal filtering of diffusion processes. Z. Wahrs. und Verw. Gebiete **11**, 230–243 (1969)

On the Stochastic Least Action Principle for the Navier-Stokes Equation

Ana Bella Cruzeiro and Remi Lassalle

Abstract In this paper we extend the class of stochastic processes allowed to represent solutions of the Navier-Stokes equation on the two dimensional torus to certain non-Markovian processes which we call admissible. More precisely, using the variations of Ref. [3], we provide a criterion for the associated mean velocity field to solve this equation. Due to the fluctuations of the shift a new term of pressure appears which is of purely stochastic origin. We provide an alternative formulation of this least action principle by means of transformations of measure. Within this approach the action is a function of the law of the processes, while the variations are induced by some translations on the space of the divergence free vector fields. Due to the renormalization in the definition of the cylindrical Brownian motion, our action is only related to the relative entropy by an inequality. However we show that, if we cut the high frequency modes, this new approach provides a least action principle for the Navier-Stokes equation based on the relative entropy.

Keywords Navier-Stokes · Entropy

1 Introduction

Let (W_t) be a suitably renormalized Brownian motion on the space of vector fields on the two dimensional torus $\mathbb{T}^2$ with a well chosen Sobolev regularity. In the case where (u_t) is a deterministic vector field, it was shown that equations of the form

$$dg_t = (\circ dW_t + u_t dt)(g_t);\ g_t = e \tag{1.1}$$

A.B. Cruzeiro (✉)
GFMUL and Department Matemática Instituto Superior Técnico, Universidade de Lisboa, Av. Rovisco Pais, 1049-001 Lisbon, Portugal
e-mail: abcruz@math.ist.utl.pt

R. Lassalle
GFMUL Instituto para a Investigação Interdisciplinar da Universidade de Lisboa, Av. Prof. Gama Pinto, 2, 1649-003 Lisbon, Portugal
e-mail: lassalle@telecom-paristech.fr

D. Crisan et al. (eds.), *Stochastic Analysis and Applications 2014*,
Springer Proceedings in Mathematics & Statistics 100,
DOI 10.1007/978-3-319-11292-3_6

could model the Navier-Stokes flows (see for instance the review article [2] and references within). More precisely it was shown that (u_t) solves the Navier-Stokes equation if and only if a certain associated action is stationary. Subsequently, models of the form

$$dg_t = (\circ dW_t + \dot{v}_t(\omega)dt)(g_t);\ g_t = e \tag{1.2}$$

where considered in Ref. [1], together with a notion of generalized stochastic flows with fixed marginals. In these latter models, the shift $\dot{v}_t(\omega)$ is allowed to be random: the drift changes from one realization of the noise to another which seems to fit accurately with the microscopic models of the Navier-Stokes equation one encounters in physics. In particular such processes are not necessarily Markovian.

In the case of (1.2) there is no reason why we should hope $\dot{v}(\omega)$ to solve the Navier-Stokes equation for any ω *a.s.*, and we should focus on the mean velocity field

$$u : (t, x) \in [0, 1] \times \mathbb{T}^2 \to u(t, x) = E_\eta[\dot{v}_t(x)] \in T_x\mathbb{T}^2$$

where η is the underlying probability on the canonical path space, and where $T_x\mathbb{T}^2$ is the tangent space at x.

We extend here the criterion of Ref. [2] from equations with the form (1.1) to equations of type (1.2) for a wide class of stochastic drifts. Namely we focus on drifts v associated with a probability η with finite entropy with respect to the law μ of the renormalized Brownian motion on the corresponding path space. We exhibit a class of such drifts (they will be called admissible) whose mean velocity field solves the Navier-Stokes equation if and only if the associated action, which will be noted $\mathcal{S}(\eta|\mu)$, is critical. We then prove that this notion naturally extends the variational principle of Ref. [2]. One of the aspects of this model is to allow that the fluctuations of the drift itself may contribute to the pressure. Then we provide an alternative formulation to the least action principle by means of transformation of measure. However in this case, due to the renormalization involved in the definition of the cylindrical Brownian motion, our action for a process with law η is only related to the corresponding relative entropy

$$\mathcal{H}(\eta|\mu) := E_\eta\left[\ln \frac{d\eta}{d\mu}\right]$$

by an inequality. Nevertheless, by introducing a cut-off, the action $\mathcal{S}(\nu|\eta)$ becomes proportional to the relative entropy, and by cutting the high modes, we provide a least action principle to the Navier-Stokes equation by means of the relative entropy.

The structure of this paper is the following. In Sect. 2 we introduce the general framework as well as the main notations of the paper. In Sect. 3 we provide a characterization of solutions of the Navier-Stokes equation as critical flows of the action. In Sect. 4 this criterion is proved to extend those of Refs. [2, 3]. In Sect. 5 we introduce a cut-off in order to transform variations of the action in variations of the entropy. (Sect. 6).

2 Preliminaries and Notations

2.1 A Basis of Vector Fields on the Two Dimensional Torus

Let $M := \mathbb{T}^2$ be the set of pairs (θ_1, θ_2) of real numbers modulo 2π, and denote $m_{\mathbb{T}} = \frac{\lambda^L \otimes \lambda^L}{4\pi^2}$ where λ^L is the Lebesgue measure on $[0, 2\pi]$. Integration will often be noted dx instead of $m_{\mathbb{T}}(dx)$. A basis of the tangent space T_xM at $x = (\theta_1, \theta_2) \in M$ is given by $(\partial_i|_x) := (\frac{\partial}{\partial\theta_i}|_{x=(\theta_1,\theta_2)})$. We define a scalar product $\langle ., .\rangle_{T_xM}$ on each T_xM by setting $\langle \partial_i|_x, \partial_j|_x\rangle_{T_xM} = \delta^{i,j}$ where $\delta^{i,j} = 1$ if $i = j$ and 0 if $i \neq j$. When there is no ambiguity, we will sometimes note $X.Y$ instead of $\langle X, Y\rangle_{T_xM}$ for $X, Y \in T_xM$. If $\mathcal{X}(M)$ consists of sections of TM, $\mathcal{X}(M) = \{X : M \to T(M)\}$, and considering its L^2 equivalence class, we set

$$\mathcal{G} = \left\{ X \in \mathcal{X}(M)|\ \operatorname{div}(X) = 0 \text{ and } \int_M |X(x)|^2_{T_xM} dx < \infty \right\}$$

which is a separable Hilbert space with the product

$$\langle X, Y\rangle_{\mathcal{G}} := \int_M \langle X(x), Y(x)\rangle_{T_xM} dx$$

An Hilbertian basis of $\mathcal{G}$ is given by a subset $(e_\alpha)_{\alpha=1}^{\infty}$, whose definition is the following. Let $k : \alpha \in \mathbb{N}/\{0\} \to k(\alpha) := (k_1(\alpha), k_2(\alpha)) \in (\mathbb{Z} \times \mathbb{Z})/\{(0, 0)\}$ be a bijection such that $|k(\alpha)| := \sqrt{k_1^2(\alpha) + k_2(\alpha)^2} \uparrow \infty$; we set

$$e_\alpha(x) := \sum_j a^{\alpha,j}(x)\partial_j|_x$$

where

$$a^{\alpha,i}(x) := \begin{cases} 1 & \text{if } (\alpha, i) \in (1, 1) \cup (2, 2) \\ 0 & \text{if } (\alpha, i) \in (2, 1) \cup (1, 2) \\ \sqrt{2}\frac{k_2(m)}{|k(m)|}\cos(k(m).x) & \text{if } (\alpha, i) = (2m + 2, 1), m \geq 1 \\ -\sqrt{2}\frac{k_1(m)}{|k(m)|}\cos(k(m).x) & \text{if } (\alpha, i) = (2m + 2, 2), m \geq 1 \\ \sqrt{2}\frac{k_2(m)}{|k(m)|}\sin(k(m).x) & \text{if } (\alpha, i) = (2m + 1, 1), m \geq 1 \\ -\sqrt{2}\frac{k_1(m)}{|k(m)|}\sin(k(m).x) & \text{if } (\alpha, i) = (2m + 1, 2), m \geq 1 \end{cases}$$

and where, for $k = (k_1, k_2) \in \mathbb{Z} \times \mathbb{Z}$ and $x = (\theta_1, \theta_2) \in M$, $k.x := k_1\theta_1 + k_2\theta_2$. Any $X \in \mathcal{G}$ can be written

$$X(x) = \sum_j X^j(x)\partial_j|_x$$

where $X_j(x) = \sum_\alpha \langle X, e_\alpha\rangle_{\mathcal{G}} a^{\alpha,j}(x)$. Let $Y(x) := \sum_j Y_j(x)\partial_j|_x$ be another vector field: it is straightforward to check that we also have

$$\langle X, Y\rangle_{\mathcal{G}} = \int_M \sum_j X^j(x)Y^j(x)dx$$

We recall the following formulae

$$\mathrm{div}(X) := \sum_j \partial_j X^j,$$

$$\Delta X := \sum_i (\sum_j \partial^2_{j,j} X_i)\partial_i|_x$$

and

$$(X.\nabla)Y := \sum_j (\sum_i X_i(\partial_i Y_j))\partial_j|_x$$

2.2 *The Group of the Volume Preserving Homeomorphisms*

Let G be the group of the homeomorphisms of M which leaves $m_{\mathbb{T}}$ invariant

$$G := \{\phi : M \to M, \textit{homeomorphisms}, \phi_\star m_{\mathbb{T}} = m_{\mathbb{T}}\}$$

We note e the identity on G and $\phi.\psi$ the group operation of $\phi, \psi \in G$ (given by the composition of the two maps). We recall [6] that the subset of G consisting of maps which are, together with their inverses, in the Sobolev class H^r, for $r > 2$ is a Hilbert manifold and a topological group. It is not, strictly speaking, a Lie algebra since left translation is not smooth. T_eG is given by the set of the vector fields $X : x \in M \to X_x \in T_xM$ such that $\mathrm{div}(X) = 0$. Let $X \in T_eG$, and let

$$c : t \in \mathbb{R} \to c_t \in G; c_0 = e$$

be a smooth curve on G to which X is tangent. We recall that, by setting $\widehat{c} : (t, x) \in \mathbb{R} \times M \to c_t(x) \in M$, the value of X at $x \in M$ is given by

$$X(x) = \partial_t \widehat{c}(t, x)|_{t=0} \in T_xM$$

Furthermore X can be uniquely extended to a right invariant vector field $\widehat{X}$ on G by setting

$$\widehat{X} : \phi \in G \to \widehat{X}_\phi \in T_\phi G$$

where $\widehat{X}_\phi$ is given by

$$\widehat{X}_\phi : x \in M \to \widehat{X}_\phi(x) := X(\phi(x)) \in T_{\phi(x)}M$$

i.e. $\widehat{X}_\phi$ is tangent to the curve $c^\phi : t \in \mathbb{R} \to c_t.\phi \in G$. In particular for any smooth f on M and $x \in M$ denote f^x the map $\phi \in G \to f^x(\phi) := f(\phi(x)) \in \mathbb{R}$. Then f^x is smooth on G and we have

$$(\widehat{X}f^x)(\phi) := \widehat{X}_\phi f^x = \partial_t f(c_t.\phi(x))|_{t=0} = \partial_t f(\widehat{c}(t, \phi(x)))|_{t=0} = X(\phi(x))f := (Xf)(\phi(x))$$

In the sequel we will simply write X instead of $\widehat{X}$ since it will be clear from the context whether we consider X as an element of the tangent space, or as a right-invariant vector field on G. In order to kill the noise in the higher modes and to control the integrability of the derivatives, we introduce the following Sobolev spaces $(\mathcal{G}_\lambda)_{\lambda>1}$ and the associated abstract Wiener spaces $(W, H_\lambda, \mu_\lambda)$.

2.3 Sobolev Vector Fields

To any positive real number $\lambda > 1$ we associate a sequence $(\lambda_\alpha)_{\alpha\in\mathbb{N}}$ defined by

$$\lambda_\alpha = \frac{|k([\frac{\alpha-1}{2}])|^{2\lambda}}{K(\lambda)}$$

where $[.]$ is the floor function and where $K(\lambda)$ is chosen so that

$$\sum_\alpha \frac{a^{\alpha,i}(x)}{\sqrt{\lambda_\alpha}} \frac{a^{\alpha,j}(x)}{\sqrt{\lambda_\alpha}} = \delta^{i,j}$$

Such a $K(\lambda)$ exists from standard results on Riemann series since $\lambda > 1$, and we have $K(\lambda) \uparrow \infty$ as $\lambda \downarrow 1$. For $\lambda > 1$, let S_λ be the positive, definite, trace class operator defined by

$$S_\lambda x := \sum_i \frac{1}{\lambda_i} \langle x, e_i \rangle_{\mathcal{G}} e_i$$

and let

$$\mathcal{G}_\lambda := \sqrt{S_\lambda}(\mathcal{G})$$

which is an Hilbert space for the scalar product $\langle ., . \rangle_{\mathcal{G}_\lambda}$ characterized by

$$\left\langle \sqrt{S_\lambda} x, \sqrt{S_\lambda} y \right\rangle_{\mathcal{G}_\lambda} = \langle x, y \rangle_{\mathcal{G}}.$$

A natural Hilbertian basis of $\mathcal{G}_\lambda$ is given by $(H_\alpha^\lambda)_{\alpha=1}^\infty$ where

$$H_\alpha^\lambda := \frac{e_\alpha}{\sqrt{\lambda_\alpha}} \tag{2.3}$$

We set

$$A_{\alpha,j}^\lambda(x) = \frac{a_{\alpha,j}}{\sqrt{\lambda_\alpha}}$$

so that

$$\sum_\alpha A_{\alpha,i}^\lambda(x) A_{\alpha,j}^\lambda(x) = \delta^{i,j} \tag{2.4}$$

and

$$H_\alpha^\lambda(x) = \sum_j A_{\alpha,j}^\lambda(x) \partial_j|_x$$

Since $\sqrt{S_\lambda}$ is Hilbert-Schmidt, it is well known that $|.|_{\mathcal{G}}$ is a measurable semi-norm on the Hilbert space $\mathcal{G}_\lambda$ (see [9]). In particular $(\mathcal{G}_\lambda, \mathcal{G})$ is an abstract Wiener space [9, 12], which allows to regard the cylindrical Brownian motion below as a Brownian sheet (note that we could have defined a Wiener measure directly on the Wiener space $(\mathcal{G}_\lambda, \mathcal{G})$, but we won't use this in the sequel since we are interested in the path space).

2.4 Associated Wiener Spaces

The space

$$H_\lambda := \left\{ h : [0, 1] \to \mathcal{G}_\lambda : h := \int_0^{\cdot} \dot{h}_s ds, \int_0^1 |\dot{h}_s|^2_{\mathcal{G}_\lambda} ds < \infty \right\}$$

is an Hilbert space whose product will be noted $\langle ., . \rangle_\lambda$. On the other hand the space

$$W := C_0([0, 1], \mathcal{G})$$

is a separable Banach space for the uniform convergence norm. We denote by i_λ the injection of H_λ in W. Since for $\lambda > 1$ $|.|_{\mathcal{G}}$ is a measurable semi-norm on $\mathcal{G}_\lambda$, it is a classical result on Wiener spaces that $(i_\lambda, W, H_\lambda)$ is also an abstract Wiener space. If μ_λ is the standard Wiener measure on W for the A.W.S. $(W, H_\lambda, i_\lambda)$, we recall that under this probability the coordinate process $t \to W_t(\omega) = \omega(t) \in \mathcal{G}$ is an abstract Brownian motion with respect to its own filtration $(\mathcal{F}_t)$ (see for instance [10, 12]). From the Itô Nisio theorem, we have μ_λ−a.s.

$$W_t = \sum_\alpha W_t^\alpha H_\alpha^\lambda$$

with $W_t^\alpha := \widehat{\delta} H_\alpha(W_t)$, and where $\left\{\widehat{\delta}(X), X \in \mathcal{G}_\lambda\right\}$ is the isonormal Gaussian process on $\mathcal{G}_\lambda$. We recall that under μ_λ, $\left\{\widehat{\delta}(X)(W_s), X \in \mathcal{G}_\lambda, s \in [0,1]\right\}$ is a Gaussian process with covariance

$$E_{\mu_\lambda}[\widehat{\delta}(X)(W_s)\widehat{\delta}(Y)(W_t)] = (s \wedge t)\langle X, Y\rangle_{\mathcal{G}_\lambda}$$

so that $(W_\cdot^\alpha)$ is a family of real valued independent Brownian motions under μ_λ. Under μ_λ, the coordinate process $t \to W_t$ is called the cylindrical Brownian motion. The difference with respect to the case where the state space is finite dimensional is that it is a renormalized sum of independent Brownian motions, the renormalization appearing in (2.3). For a measure $\eta \ll \mu_\lambda$ and a $u \in L_a^0(\eta, H_\lambda)$, the stochastic integral $\delta^W u := \int_0^1 \dot{u}_s dW_s$ is well defined as an abstract stochastic integral [10, 12]. Let η be a probability which is absolutely continuous with respect to μ_λ. Then there is a unique $v \in L_a^0(\eta, H_\lambda)$ such that $\eta - a.s.$

$$\frac{d\eta}{d\mu_\lambda} := \exp\left(\delta^W v - \frac{|v|_\lambda^2}{2}\right) \tag{2.5}$$

Moreover $W^\eta := I_W - v$ is a $(\mathcal{F}_t)$-Brownian motion on $(W, \mathcal{F}, \eta)$. We call v the **velocity field** associated to η. The famous formula of [7] (which in fact holds in a more general framework: [10, 12]) reads

$$2\mathcal{H}(\eta|\mu_\lambda) = E_\eta\left[\int_0^1 |\dot{v}_t|_{\mathcal{G}_\lambda}^2 dt\right] \tag{2.6}$$

where

$$\mathcal{H}(\eta|\mu_\lambda) := E_\eta\left[\ln \frac{d\eta}{d\mu_\lambda}\right]$$

is the relative entropy of η with respect to μ_λ. Note that since $\mathcal{G}_\lambda \subset \mathcal{G} \subset T_eG$ it makes sense to consider $(Xf)(\phi)$ for $\phi \in G$, for f smooth on G and for $X \in \mathcal{G}_\lambda$ or $X \in \mathcal{G}$.

3 Navier-Stokes Flows with Stochastic Drifts

Henceforth and until the end of Sect. 5 we assume that the renormalization sequence is fixed for a $\lambda \geq 2$, and we drop the indices λ of the notations except for $\mathcal{G}_\lambda$.

3.1 Constraints on the Kinematics: Regular and Admissible Flows

Definition 1 A probability η which is absolutely continuous with respect to μ with finite entropy ($\mathcal{H}(\eta|\mu) < \infty$) is called a **regular flow** if $u \in C^1([0,1] \times M)$ and dt- $a.s.$ $\partial_t u \in \mathcal{G}$, where $u(t,x) := E_\eta[\dot{v}_t(x)]$, and where $v := \int_0^{\cdot} \dot{v}_s ds$ is the **velocity field** of η (see (2.5)). We call u the **mean velocity field** of η. Moreover we say that a regular flow is **admissible** if there is a $C^1([0,1] \times M)$ mapping $p^\star : [0,1] \times M \to \mathbb{R}$ such that

$$Cov(\dot{v}_t(x)) = p^\star(t,x) I_d$$

i.e. for $i, j \in \mathbb{N} \cap [1, d]$

$$E_\eta \left[\left(\dot{v}_t^i(x) - u_t^i(x) \right) \left(\dot{v}_t^j(x) - u_t^j(x) \right) \right] = p^\star(x,t) \delta^{i,j} \tag{3.7}$$

where $(\dot{v}_t^j(x))$ denotes the jth (random) component of $(\dot{v}_t^j)$ at x i.e. it is such that $\dot{v}_t(x) = \sum_j \dot{v}_t^j(x) \partial_j|_x$, and where $u_t^j(x) := E_\eta[\dot{v}_t^j(x)]$.

3.2 Constraints on the Dynamics: Critical Flows

Definition 2 Let η be a regular flow whose velocity field is denoted by v^η (see (2.5)). For any $k \in C^1([0,1]; \mathcal{G})$ we set

$$L_k \mathcal{S}(\eta|\mu) := E_\eta \left[\int_0^1 \left(\int_M < \dot{v}_t^\eta(x), \partial_t k + (\dot{v}_t^\eta . \nabla) k + \frac{\Delta k}{2} >_{T_x M} dx \right) dt \right]$$

The probability η is said to be **critical** if and only if for any $k \in C_0^1([0,1], \mathcal{G})$

$$L_k \mathcal{S}(\eta|\mu) = 0$$

where

$$C_0^1([0,1], \mathcal{G}) := \left\{ k \in C^1([0,1]; \mathcal{G}) : k(0,.) = k(1,.) = 0 \right\}$$

The dynamic of the mean velocity field of a critical flow is given by the following theorem

Theorem 1 *Let η be a regular flow with a velocity field v and a mean velocity field $u \in \mathcal{G}_\lambda$. Then η is critical (Definition 2) if and only if there is a function $\widehat{p}(t,x)$ such that*

$$\partial_t u + E_\eta[(\dot{v}_t(x).\nabla)\dot{v}_t(x)] = \frac{\Delta u}{2} - \nabla\widehat{p}(t,x) \tag{3.8}$$

In other words, let

$$\beta(t,x) := E_\eta[((\dot{v}_t(x) - u_t(x)).\nabla)(\dot{v}_t(x) - u_t(x))] \tag{3.9}$$

Then u solves, in the weak L^2 sense, the following equation :

$$\partial_t u + (u_t.\nabla)u = \frac{\Delta u}{2} - \nabla\widehat{p} - \beta \tag{3.10}$$

Proof For any $k \in {C_0}^1([0,1];\mathcal{G})$ we have $k(0,.) = k(1,.) = 0$, so that an integration by parts yields

$$L_k\mathcal{S}(\eta|\mu) = -\int_M\int_0^1 \left(\partial_t u + E_\eta[(\dot{v}_t.\nabla)\dot{v}_t] - \frac{\Delta u}{2}\right)(t,x).k(t,x)dxdt \tag{3.11}$$

from which we obtain (3.8). Since

$$\begin{aligned}
\beta(t,x) &:= E_\eta\left[[\dot{v}_t(x) - u_t(x)].\nabla)[\dot{v}_t(x) - u_t(x)]\right] \\
&= E_\eta\left[(\dot{v}_t(x).\nabla)\dot{v}_t(x)\right] + (u_t(x).\nabla)u_t(x) - E_\eta\left[(\dot{v}_t(x).\nabla)u_t(x)\right] - E_\eta\left[(u(x).\nabla)\dot{v}_t(x)\right] \\
&= E_\eta\left[(\dot{v}_t(x).\nabla)\dot{v}_t(x)\right] + (u_t(x).\nabla)u_t(x) - (E_\eta\left[\dot{v}_t(x)\right].\nabla)u_t(x) - (u(x).\nabla)E_\eta\left[\dot{v}_t(x)\right] \\
&= E_\eta\left[(\dot{v}_t(x).\nabla)\dot{v}_t(x)\right] - (u_t(x).\nabla)u_t(x)
\end{aligned}$$

we obtain (3.10) from (3.8). □

3.3 Navier-Stokes Flows

Definition 3 A regular flow η (see Definition 1) is a Navier-Stokes flow if its mean velocity field u solves the Navier-Stokes equation, i.e. if and only if there is a function $p : [0,1] \times M \to \mathbb{R}$ which is such that u solves, in the weak L^2 sense, the Navier-Stokes equation

$$\partial_t u + u.\nabla u = \frac{\Delta u}{2} - \nabla p$$

we have:

Corollary 1 *An admissible flow is a Navier-Stokes flow if and only if it is critical.*

Proof Let η be an admissible flow. We recall that by definition there exists a mapping $p^\star$ such that

$$Cov(\dot{v}_t(x)) = p^\star(x,t)I_d \tag{3.12}$$

where $v := \int_0^\cdot \dot{v}_s ds$ is the velocity field of η (see (2.5)). We also recall that

$$u(t,x) := E_\eta[\dot{v}_t(x)]$$

The idea is to apply Theorem 1 and to set

$$p := p^\star + \widehat{p}$$

We have

$$\beta^i(t,x) = \sum_j \partial_j Cov(\dot{v}_t(x))^{i,j}$$

Indeed (repeated indices are summed over) we have

$$\begin{aligned}
\beta^i(t,x) &= E_\eta\left[\left(\dot{v}_t^j(x) - u_t^j(x)\right)\partial_j\left(\dot{v}_t^i(x) - u_t^i(x)\right)\right] \\
&= \partial_j E_\eta\left[\left(\dot{v}_t^i(x) - u_t^i(x)\right)\left(\dot{v}_t^j(x) - u_t^j(x)\right)\right] - E_\eta\left[\left(\dot{v}_t^i(x) - u_t^i(x)\right)\partial_j\left(\dot{v}_t^j(x) - u_t^j(x)\right)\right] \\
&= \partial_j E_\eta\left[\left(\dot{v}_t^i(x) - u_t^i(x)\right)\left(\dot{v}_t^j(x) - u_t^j(x)\right)\right] - E_\eta\left[\left(\dot{v}_t^i(x) - u_t^i(x)\right)\operatorname{div}\left(\dot{v}_t(x) - u_t(x)\right)\right] \\
&= \partial_j E_\eta\left[\left(\dot{v}_t^i(x) - u_t^i(x)\right)\left(\dot{v}_t^j(x) - u_t^j(x)\right)\right] \\
&= \partial_j Cov(\dot{v}_t(x))^{i,j}
\end{aligned}$$

Assumption (3.12) then yields $\beta^i(t,x) = \partial_i p^\star$ i.e.

$$\beta = \nabla p^\star \tag{3.13}$$

□

Remark 1 Note that by this proof, for critical flows, $p^\star$ appears as a part of the pressure which is originated from the stochastic model. Specifically it expresses the fluctuations of the drift itself. Indeed by (3.13) and (3.9) for an admissible flow η we have

$$\nabla p^\star(t,x) = E_\eta[(\dot{v}_t(x).\nabla)\dot{v}_t(x)] - (u_t(x).\nabla)u_t(x) \tag{3.14}$$

where $p^\star$ is the function associated to the admissible flow η by formula (3.7).

4 Interpretation of Critical Flows by Means of the Stochastic Exponential

In this section we prove that the quantities $L_k\mathcal{S}(\eta|\mu)$ defined in Definition 2 can still be interpreted in terms of certain variations along deterministic paths which extend those of Ref. [3].

4.1 The Stochastic Exponential

Let $C_G = C_e([0,1], G)$ be the space of continuous paths starting from e and with values in G. The coordinate function $(t,\gamma) \in [0,1] \times C_G \to \gamma_t(\omega)$ generates a filtration $\left(\mathcal{F}_t^G\right)$ and we denote $\mathcal{F}^G := \mathcal{F}_1^G$.

Proposition 1 *The equation*

$$dX_t = \circ dB_t;\; X_0 = e \tag{4.15}$$

has a continuous strong solution on the space $\left(W, \mathcal{F}_{\cdot}^W, \mu\right)$ with the canonical Brownian $t \to W_t \in \mathcal{G}$. We note g this solution. By this we mean that for μ−a.s. $g \in C_G$ and, for any smooth f on G,

$$f(g_t) = f(e) + \sum_\alpha \int_0^t (H_\alpha f)(g_t) \circ dW_t$$

where $\circ$ denotes the Stratonovich integral.

Proof See [11]. □

Girsanov theorem on (W, H, μ) implies the following:

Proposition 2 *Let η be a probability which is absolutely continuous with respect to μ whose velocity field is noted v, and set $\widetilde{W} := I_W - v$. Then $(g, \widetilde{W})$ is a solution of*

$$dX_t = (\circ dB_t + \dot{v}_t dt); X_0 = e \tag{4.16}$$

on $(W, \mathcal{F}_{\cdot}, \eta)$.

Proof We have

$$\widetilde{W}_s = \sum_\alpha \widehat{\delta}(H_\alpha)(W_s)H_\alpha - \sum_\alpha \langle v, H_\alpha\rangle_\lambda H_\alpha = \sum_\alpha \widehat{\delta}(H_\alpha)(\widetilde{W}_s)H_\alpha$$

Since $\widetilde{W}_\star\eta = \mu$, $\widetilde{W}_{\cdot}^\alpha := \widehat{\delta}(H_\alpha)(\widetilde{W}_{\cdot})$ are independent Brownian motions on (W, H, η), by Itô's formula we have, $\eta - a.s.$,

$$f(g_t) = f(e) + \int_0^1 \sum_\alpha (H_\alpha f)(g_t) \circ d\widetilde{W}_t^\alpha + \sum_\alpha \int_0^1 (H_\alpha f)(g_s)\langle \dot{v}_s, H_\alpha \rangle_{\mathcal{G}_\lambda} ds$$

i.e.

$$f(g_t) = f(e) + \int_0^1 (H_\alpha f)(g_s) \circ d\widetilde{W}_t^\alpha + \int_0^1 (\dot{v}_t(\omega) f)(g_s) ds \qquad \square$$

Proposition 3 *Let η be a probability absolutely continuous with respect to μ, $v := \int_0^\cdot \dot{v}_s ds$ the associated velocity field, $\widetilde{W} = I_W - v$ and $\widetilde{W}_\cdot^\alpha = \widehat{\delta}(H_\alpha)(\widetilde{W}_\cdot)$. For any smooth function f on $[0,1] \times M$ we have $\eta-a.s.$*

$$f(t, g_t(x)) = f(0,x) + \int_0^t \left(\frac{\Delta}{2} f + (\dot{v}_\sigma . \nabla) f + \partial_\sigma f)(\sigma, g_\sigma(x) \right) d\sigma + \int_0^t \sum_\alpha (H_\alpha f)(\sigma, g_\sigma(x)) d\widetilde{W}_\sigma^\alpha \tag{4.17}$$

and $\eta-a.s.$

$$\lim_{\delta \to 0} E_\eta \left[\frac{f(t+\delta, g_{t+\delta}(x)) - f(t, g_t(x))}{\delta} \middle| \mathcal{F}_t \right] = \left(\partial_t f + (\dot{v}_t(\omega).\nabla) f + \frac{\Delta f}{2} \right)(t, g_t(x)) \tag{4.18}$$

Proof Let $x \in M, f \in C^\infty(M)$. The main part of the proof will be to prove that

$$\sum_\alpha (H_\alpha^2 f^x)(\phi) = (\Delta f)(\phi(x)) \tag{4.19}$$

To see this recall that $f^x : \phi \in G \to f(\phi(x)) \in \mathbb{R}$. We have

$$(H_\alpha f^x)(\phi) := H_\alpha(\phi) f^x = H_\alpha(\phi(x)) f = (H_\alpha f)(\phi(x)) = (H_\alpha f)^x(\phi) \tag{4.20}$$

so that by iterating (4.20) we obtain

$$\sum_\alpha (H_\alpha^2 f^x)(\phi) = \sum_\alpha (H_\alpha^2 f)(\phi(x)) \tag{4.21}$$

On the other hand

$$\sum_\alpha (H_\alpha^2 f)(\phi(x)) = (\Delta f)(\phi(x)) \tag{4.22}$$

Indeed by using the fact that for any α the vector field H^α is divergence free together with (2.4) we obtain

$$\sum_\alpha H_\alpha^2 f = \sum_{\alpha, j} A^{\alpha, j} \partial_j (H_\alpha f)$$

$$
\begin{aligned}
&= \sum_{\alpha,i,j} A^{\alpha,j} A^{\alpha,i} (\partial_j \partial_i f) + A^{\alpha,j} (\partial_j A^{\alpha,i})(\partial_i f) \\
&= \sum_i (\partial^2_{i,i} f) + \sum_{\alpha,i,j} A^{\alpha,j} (\partial_j A^{\alpha,i})(\partial_i f) \\
&= \Delta f + \sum_{\alpha,i,j} A^{\alpha,j} (\partial_j A^{\alpha,i})(\partial_i f) \\
&= \Delta f + \sum_{\alpha,i,j} \partial_j (A^{\alpha,j} A^{\alpha,i})(\partial_i f) - \sum_{\alpha,i,j} (\partial_j A^{\alpha,j}) A^{\alpha,i} (\partial_i f) \\
&= \Delta f + \sum_{i,j} \partial_j (\sum_\alpha A^{\alpha,j} A^{\alpha,i})(\partial_i f) - \sum_{\alpha,i} (\mathrm{div}(H^\alpha) A^{\alpha,i} (\partial_i f) \\
&= \Delta f
\end{aligned}
$$

Finally by putting together (4.21) and (4.22) we get (4.19) which yields

$$
\begin{aligned}
f(t, g_t(x)) &= f^x(t, g_t) \\
&= f^x(s, g_s) + \int_s^t (H_\alpha f^x)(g_\sigma) \circ d\widetilde{W}^\alpha_\sigma + \int_s^t (\partial_\sigma f^x + \dot{v}_\sigma f^x)(g_\sigma) d\sigma \\
&= f(s, g_s(x)) + \int_s^t \left(\frac{\Delta}{2} f + (\dot{v}_\sigma . \nabla) f + \partial_\sigma f \right) (\sigma, g_\sigma(x)) d\sigma \\
&\quad + \int_s^t \sum_\alpha (H_\alpha f_\sigma)(g_\sigma(x)) d\widetilde{W}^\alpha_\sigma
\end{aligned}
$$

On the other hand by the Girsanov theorem, $(\widetilde{W}_t)$ is a $(\mathcal{F}_t)$-Brownian motion on (W, η) so that (4.18) follows from (4.17). □

4.2 Perturbations of the Energy Along Deterministic Paths

For $k \in C^0([0, 1], \mathcal{G}_\lambda)$, $k := \int_0^\cdot \dot{k}_s ds$, we define $e(k)$ to be the solution of the ordinary differential equation on G

$$
d(e_t(k)) = (\dot{k}_t dt)(e_t(k)); \; e_0 = e
$$

i.e. for any smooth $F : G \to \mathbb{R}$,

$$
F(e_t(k)) = F(e) + \int_0^t (\dot{k}_s F)(e_s(k)) ds. \tag{4.23}
$$

Note that $e_{.}(0_H) = e$ i.e. the exponential of the function which is constant and equal to 0_H is constant and equal to e. We denote by $(e^i_t(k))$ the ith component of $(e_t(k))$ in the canonical chart.

Proposition 4 *If η is a probability of finite entropy with respect to μ, for any $k \in C^1_0([0,1], \mathcal{G}_\lambda)$ we have*

$$L_k \mathcal{S}(\eta|\mu) = \frac{d}{d\epsilon} E_\eta \left[\int_0^1 \left(\int_M \frac{|D^\eta e_t(\epsilon k).g_t(x)|^2_{T_{g_t(x)}M}}{2} dx \right) dt \right] |_{\epsilon=0} \tag{4.24}$$

where $L_k \mathcal{S}(\eta|\mu)$ has been defined in Definition 2 and where $D^\eta e_t(\epsilon k).g_t(x)$ is defined a.e. by

$$D^\eta e_t(\epsilon k).g_t(x) := \sum_i \lim_{\delta \to 0} E_\eta \left[\frac{e^i_{t+\delta}(\epsilon k).g_{t+\delta}(x) - e^i_t(\epsilon k).g_t(x)}{\delta} \middle| \mathcal{F}_t \right] \partial_i|_{g_{t(x)}} \tag{4.25}$$

Proof By (4.18) of Proposition 3 we first obtain

$$D^\eta e_t(\epsilon k).g_t(x) := \sum_i \left(\partial_t e^i_t(\epsilon k) + (\dot{v}_t(\omega).\nabla) e^i_t(\epsilon k) + \frac{\Delta e^i_t(\epsilon k)}{2} \right) (g_t(x)) \partial_i|_{g_t(x)} \tag{4.26}$$

On the other hand let $x \in M$ and denote by f a smooth function on M. Considering $F := f^x$ in (4.23) we have

$$f(e_t(\epsilon k)(x)) = f(x) + \epsilon \int_0^t (\dot{k}_s f)(e_s(\epsilon k)(x)) ds$$

Since $e_{.}(0_H)(x) = e(x) = x$, we get :

$$\frac{d}{d\epsilon}|_{\epsilon=0} f(e_t(\epsilon k)(x)) = \int_0^t (\dot{k}_s f)(x) ds = (k_t f)(x)$$

so that

$$\frac{d}{d\epsilon}|_{\epsilon=0} e_t(\epsilon k)(x) = k_t(x) \tag{4.27}$$

By (4.26) and (4.27) we obtain

$$\frac{d}{d\epsilon} D^\eta e_t(\epsilon k).g_t(x)|_{\epsilon=0} = \left(\partial_t k_t + \dot{v}_t.\nabla k_t + \frac{\Delta k_t}{2} \right) (g_t(x)) \tag{4.28}$$

For convenience of notations we denote by A the right hand term of (4.24). By first differentiating the product, then by applying (4.26) at $\epsilon = 0$, then by applying (4.28), and finally by using that g_t preserves the measure we obtain

$$A = E_\eta \left[\int_0^1 \left(\int_M \langle D^\eta g_t(x), \frac{d}{d\epsilon} D^\eta e_t(\epsilon k).g_t(x)|_{\epsilon=0} \rangle_{T_{g_t(x)}M} dx \right) dt \right]$$

$$= E_\eta \left[\int_0^1 \left(\int_M \langle \dot{v}_t(g_t(x)), \frac{d}{d\epsilon} D^\eta e_t(\epsilon k).g_t(x)|_{\epsilon=0} \rangle_{T_{g_t(x)}M} dx \right) dt \right]$$

$$= E_\eta \left[\int_0^1 \left(\int_M \langle \dot{v}_t(g_t(x)), \left(\partial_t k_t + \dot{v}_t.\nabla k_t + \frac{\Delta k_t}{2} \right) (g_t(x)) \rangle_{T_{g_t(x)}M} dx \right) dt \right]$$

$$= E_\eta \left[\int_0^1 \left(\int_M \langle \dot{v}_t(x), \partial_t k_t(x) + \dot{v}_t.\nabla k_t(x) + \frac{\Delta k_t}{2}(x) \rangle_{T_x M} dx \right) dt \right]$$

which proves (4.24). □

5 Variations of the Energy Along Translations

Let η be a probability which is absolutely continuous with respect to μ (as mentioned in the beginning of Sect. 3 we work with a fixed $\lambda \geq 2$) and with velocity field v^η. The stochastic action of η is defined by

$$\mathcal{S}(\eta|\mu) := E_\eta \left[\int_0^1 \frac{|\dot{v}_s^\eta|^2_{\mathcal{G}}}{2} ds \right] \tag{5.29}$$

The motivation for this definition is that, by taking $\epsilon = 0$ in (4.26) and using the fact that g_t preserves the measure, we also have

$$\mathcal{S}(\eta|\mu) = E_\eta \left[\int_0^1 \left(\int_M \frac{|D^\eta g_s(x)|^2_{T_{g_s(x)}M}}{2} dx \right) ds \right]$$

with the notations of Proposition 4. By (2.6), $\mathcal{G}_\lambda \subset \mathcal{G}$ implies that whenever the entropy is finite we have

$$\mathcal{S}(\eta|\mu) < \infty$$

as well. More accurately, by a classical result on abstract Wiener spaces together with (2.6), there exists a $c > 0$ such that for any $\eta \ll \mu$

$$\mathcal{S}(\eta|\mu) \leq c\mathcal{H}(\eta|\mu)$$

In this section we introduce another kind of variations for the functional $\mathcal{S}(\eta|\mu)$, namely we study its variations along translations, These variations are generally different from those introduced above; however, when restricted to admissible flows, they are the same. We also investigate similar variations for the relative entropy. Proposition 5 computes the values of the variations of these quantities along deterministic translations.

Proposition 5 *Let η be a probability absolutely continuous with respect to μ with velocity field v^η and mean velocity $u_s(x) := E_\eta[\dot{v}_s^\eta(x)]$. If $\mathcal{S}(\eta|\mu) < \infty$ we have,*

$$\frac{d}{d\epsilon}\mathcal{S}(\tau_{\epsilon h}\eta|\mu)|_{\epsilon=0} = \int_0^1 \langle u_s, \dot{h}_s \rangle_{\mathcal{G}} ds \tag{5.30}$$

and if $\mathcal{H}(\eta|\mu) < \infty$ we have

$$\frac{d}{d\epsilon}\mathcal{H}(\tau_{\epsilon h}\eta|\mu)|_{\epsilon=0} = \int_0^1 \langle u_s, \dot{h}_s \rangle_{\mathcal{G}_\lambda} ds \tag{5.31}$$

where $\tau_h\eta$ is the image measure of η by the mapping τ_h defined by

$$\tau_h : \omega \in W \to \omega + h \in W$$

Proof A straightforward application of the Cameron-Martin theorem shows that for any $h := \int_0^{\cdot} \dot{h}_s ds \in H$, the velocity field field $v^{\tau_h\eta}$ of $\tau_h\eta$ is given by

$$v^{\tau_h\eta} = \tau_h \circ v^\eta \circ \tau_{-h} = v^\eta \circ \tau_{-h} + h \tag{5.32}$$

Hence by (5.29) we have

$$\mathcal{S}(\tau_h\eta|\mu) = E_\eta\left[\int_0^1 \frac{|\dot{v}_s^\eta + \dot{h}_s|_{\mathcal{G}}^2}{2} ds\right]$$

which yields (5.30). Similarly (5.31) follows by (2.6) and (5.32). □

Let

$$C_0^n([0,1], \mathcal{G}_{\lambda+2}) := \left\{k \in C^n([0,1], \mathcal{G}_{\lambda+2}) : k(0,.) = k(1,.) = 0\right\} \tag{5.33}$$

and let Π be the Helmoltz projection on divergence free vector fields. We set

$$\mathcal{K}_0^\eta := \left\{ h := \int_0^{\cdot} \dot{h}_s(\omega)ds \,\middle|\, \exists k \in C_0^n([0,1], \mathcal{G}_{\lambda+2}), ds - a.s., \dot{h}_s = \partial_s k_s + \Pi((u_s.\nabla)k_s) + \frac{\Delta k_s}{2} \right\} \tag{5.34}$$

so that it makes sense to say that any $h \in \mathcal{K}_0^\eta$ is associated to a $k \in C_0^n([0,1], \mathcal{G}_{\lambda+2})$. For n sufficiently large we have $\mathcal{K}^\eta \subset H$.

Proposition 6 *Let η be a smooth flow whose mean velocity field is given by u. Then u solves the Navier-Stokes equation if and only if for any $h \in \mathcal{K}_0^\eta$*

$$\frac{d}{d\epsilon}\mathcal{S}(\tau_{\epsilon h}\eta|\mu)|_{\epsilon=0} = 0$$

Proof By Proposition 5, and by definition of Π, for any h (which is associated to k) we have

$$\frac{d}{d\epsilon}\mathcal{S}(\tau_{\epsilon h}\eta|\mu)|_{\epsilon=0} = \int_M \int_0^1 \left(\partial_s k + \Pi((u.\nabla)k) + \frac{\Delta k}{2}\right)(s,x).u(s,x)dxds \tag{5.35}$$

$$= \int_M \int_0^1 \left(\partial_s k + (u.\nabla)k + \frac{\Delta k}{2}\right)(s,x).u(s,x)dxds \tag{5.36}$$

and, since $k(0,.) = k(1,.) = 0$, the result directly follows from an integrating by parts. □

We now relate these variations to the ones of Sect. 4. Namely we prove that, for admissible flows, these variations of measure by quasi-invariant transformations yield exactly the same variations as the exponential variations of Sect. 4.

Proposition 7 *Let η be an admissible flow. Then, for any $h \in \mathcal{K}_0^\eta$ (see (5.34)) associated with a $k \in C_0^n([0,1], \mathcal{G}_{\lambda+2})$ (see (5.33)) we have*

$$\frac{d}{d\epsilon}\mathcal{S}(\tau_{\epsilon h}\eta|\mu)|_{\epsilon=0} = L_k\mathcal{S}(\eta|\mu)$$

Proof Let u be the mean velocity field of η. Since η is admissible we have, by (3.14)

$$\langle u_t, (u_t.\nabla)k_t\rangle_{\mathcal{G}} = -\langle (u_t.\nabla)u_t, k_t\rangle_{\mathcal{G}} = -E_\eta[\langle (\dot{v}_t.\nabla)\dot{v}_t, k_t\rangle_{\mathcal{G}}] = E_\eta[\langle \dot{v}_t, (\dot{v}_t.\nabla)k_t\rangle_{\mathcal{G}}]$$

Hence, using (5.36),

$$\frac{d}{d\epsilon}\mathcal{S}(\tau_{\epsilon h}\eta|\mu)|_{\epsilon=0} = E_\eta\left[\int_0^1 \left(\int_M \langle \dot{v}_t^\eta(x), \partial_t k + (\dot{v}_t^\eta.\nabla)k + \frac{\Delta k}{2}\rangle_{T_xM} dx\right) dt\right]$$

which is exactly the definition of $L_k\mathcal{S}(\eta|\mu)$ (Definition 2). □

6 Generalized Flows with a Cut-off

In Sect. 5 we have seen that in the infinite dimensional case, the relative entropy was generally not proportional to the action $\mathcal{S}(\cdot|\mu)$. The reason is that the renormalization procedure gives a different weight to the different modes: hard modes have a weaker weight in the energy than in the relative entropy. However if instead of renormalizing we introduce a cutoff, and rescale the noise accordingly, $\mathcal{S}(\cdot|\mu)$ becomes proportional to the relative entropy $\mathcal{H}(\cdot|\mu)$. Within this framework, we investigate the existence of generalized flows with a given marginal.

6.1 General Framework for a Cut-off at Scale n

We recall that (e_α) denotes the Hilbertian basis of $\mathcal{G}$ of Sect. 2. By induction we define $(I_l)_{l=1}^\infty$ by $I_1 = 1$ and

$$I_{l+1} = \min\left(\{m \geq I_l : |k(m)| > |k(I_l)|\}\right)$$

For $N \in \mathbb{N}, N > 1$ we set

$$n := 2I_N$$

We define $\mathcal{G}^n = Vect(e_1, \ldots, e_n) \subset \mathcal{G}$ and recall that we work under the hypothesis

$$e_\alpha(x) = \sum_j a^{\alpha,j}(x)\partial_j|_x$$

The cut-off has been chosen so that $\exists S(N)$ such that

$$\sum_{\alpha=1}^n a^{\alpha,i}(x)a^{\alpha,j}(x) = \mathcal{S}(N)\delta^{i,j}$$

where $\mathcal{S}(N) \uparrow \infty$. We note

$$H_n := \left\{ h : [0,1] \to \mathcal{G}^n, h := \int_0^{\cdot} \dot{h}_s ds, \int_0^1 |\dot{h}_s|^2_{\mathcal{G}} ds < \infty \right\}$$

and $\langle ., . \rangle_{H_n}$ the associated scalar product. We set $W_n := C([0,1], \mathcal{G}^n)$ endowed with the norm of uniform convergence, and μ_n the Wiener measure on (W_n, H_n) with a parameter

$$\sigma(N) := \frac{2\nu}{\mathcal{S}(N)}$$

$t \to W_t$ is the coordinate process. Define g^n to be the solution of

$$dg^n_t := (\circ dW_t)(g^n_t);\ g^n_0 = e$$

on the Wiener space (W_n, H_n, μ_n) i.e., satisfying, for every smooth f,

$$f(g^n_t) = f(e) + \int_0^t \sum_{\alpha=1}^n (e_\alpha f)(g^n_s) \circ dW^\alpha_s$$

where $W^\alpha := \langle W_t, e_\alpha \rangle_{G_n}$. We are now working with the Wiener measure with parameter $\sigma(N)$. Still by the Girsanov theorem, for any $\eta \ll \mu_n$ there is a unique $v \in L^0(\eta, H_n)$ such that

$$\frac{d\eta}{d\mu} = \exp\left(\delta^W v - \frac{\sigma(N)|v|^2_{H_n}}{2} \right)$$

and $\widetilde{W} := I_W - \sigma(N)v$ is a Brownian motion with parameter $\sigma(N)$ under η. We call v the velocity field of η. Furthermore, Föllmer's formula (c.f. [8]) then reads

$$\mathcal{H}(\eta|\mu_n) = \sigma(N) E_\eta \left[\frac{|v|^2_{H_n}}{2} \right]$$

Hence $(g, \widetilde{W})$ is a solution to

$$dg^n_t := \circ(dW^\nu_t + \sigma(N)\dot{v}_t dt))(g^n_t);\ g^n_0 = e$$

on the probability space (W_n, η) for the filtration generated by the coordinate process $t \to W_t$, i.e., for every smooth f,

$$f(g_t^n) = f(e) + \int_0^t \sum_{\alpha=1}^n (e_\alpha f)(g_s) \circ d\widetilde{W}_s^\alpha + \sigma(N) \int_0^t \sum_{\alpha=1}^n (e_\alpha f)(g_s) \langle \dot{v}_s, e_\alpha \rangle ds$$

Within this framework, by an admissible flow we mean a probability η of finite entropy with respect to η satisfying the same conditions as in Definition 1 with μ_n (resp. $\mathcal{G}_n$) instead of μ (resp. of $\mathcal{G}$).

6.2 Variations of the Action

We now define the action for the cutoff $n \in \mathbb{N}$ by

$$\mathcal{S}(\eta|\mu_n) := E_\eta \left[\int_0^1 \frac{|D_s^\eta g_s^n|_{\mathcal{G}}^2}{2} ds \right] = E_\eta \left[\int_0^1 \frac{|\sigma(N)\dot{v}_s|_{\mathcal{G}}^2}{2} ds \right] = \sigma(N)^2 E_\eta \left[\int_0^1 \frac{|\dot{v}_s|_{\mathcal{G}}^2}{2} ds \right]$$

Therefore

$$\mathcal{S}(\eta|\mu_n) = \sigma(N)\mathcal{H}(\eta|\mu_n) \tag{6.37}$$

Similarly to Proposition 7 we note

$$\mathcal{K}_0^\eta(n) := \left\{ h \in H_n : \exists k \in C_0^1([0,1], \mathcal{G}^n), ds - a.s., \dot{h}_s = \partial_s k + \pi_n \Pi((\sigma(N)u_s.\nabla)k) + \nu \Delta k \right\}$$

where π_n is the orthogonal projection $\pi_n : \mathcal{G} \to \mathcal{G}_n$ and we say that a $h \in \mathcal{K}_0^\eta(n)$ is associated to a $k \in C_0^1([0,1], \mathcal{G}^n)$.

Proposition 8 *For any smooth flow η*

$$u^n(t,x) := \sigma(N) E_\eta[\dot{v}_t(x)]$$

solves the Navier-Stokes equation if and only if for any $h \in \mathcal{K}_0^\eta(n)$ we have

$$\frac{d}{d\epsilon} \mathcal{H}(\tau_{\epsilon h}\eta|\mu_n)]|_{\epsilon=0} = 0$$

for any h associated with a $k \in C_0^1([0,1], \mathcal{G}^n)$. Moreover whenever η is an admissible flow, and $h \in \mathcal{K}_0^\eta(n)$ is associated to $k \in C_0^1([0,1], \mathcal{G}^n)$ we have

$$\frac{d}{d\epsilon} \mathcal{H}(\tau_{\epsilon h}\eta|\mu_n)]|_{\epsilon=0} = \frac{d}{d\epsilon} E_\eta \left[\int_0^1 \left(\int_M \frac{|D^\eta e_t(\epsilon k).g_t^n(x)|_{T_{g_t(x)}M}^2}{2} dx \right) dt \right] |_{\epsilon=0}$$

where the notations are those of Sect. 4.

Proof The first part of the proof is the same as in Proposition 6. We now prove the second part of the claim which is similar to Proposition 7. As in the first subsection we have

$$\sum_{\alpha=1}^{n} e_\alpha^2 f = \mathcal{S}(N)\Delta f$$

Therefore by setting

$$A := \lim_{\epsilon\to 0}\left(\frac{E_\eta\left[\int_0^1\left(\int_M |D^\eta e_t(\epsilon k).g_t^n(x)|^2_{T_{g_t(x)}M}dx\right)dt\right] - E_\eta\left[\int_0^1\left(\int_M |D^\eta g_t^n(x)|^2_{T_{g_t(x)}M}dx\right)dt\right]}{2\epsilon}\right)$$

and using the fact g_t preserves the measure we get

$$A = E_\eta\left[\int_0^1 \langle \dot{v}_t, \partial_t k + \sigma(N)\dot{v}_t.\nabla k + \nu\Delta k\rangle_{\mathcal{G}} dt\right]$$

If η is assumed to be admissible, then similarly to the proof of Proposition 7 we obtain

$$A = \frac{d}{d\epsilon}\mathcal{H}(\tau_{\epsilon h}\eta|\mu_n)]|_{\epsilon=0}$$

□

Concerning existence of Lagrangian Navier-Stokes flows with a cut-off they have been shown to exist in Ref. [4] for deterministic L^2 drifts. Examples of random solutions of Navier-Stokes equations were constructed in Ref. [5] but we did not prove existence of the corresponding flows.

Acknowledgments We thank the anonymous referee for a careful reading of the manuscript and valuable remarks. We acknowledge support from the FCT project PTDC/MAT/120354/2010.

References

1. A. Antoniouk, A.B. Cruzeiro, M. Arnaudon, Generalized Navier-Stokes flows and applications to incompressible viscous fluids. Bull. Sci. Math. **138**(4), 565–584 (2014)
2. M. Arnaudon, A.B. Cruzeiro, Stochastic Lagrangian flows and the Navier-Stokes equation, preprint, http://gfm.cii.fc.ul.pt/people/abcruzeiro/abc-marnaudon-2012-11-08.pdf (2012)
3. A.B. Cruzeiro, F. Cipriano, Navier-Stokes equation and diffusions on the group of homeomorphisms of the torus. Commun. Math. Phys. **275**, 255–269 (2007)
4. A.B. Cruzeiro, F. Cipriano, Variational principle for diffusions on the diffeomorhism group with the H_2 metric, in *Mathematical analysis of random phenomena*, World. Sci. Publ., Hackensack (2007)
5. A.B. Cruzeiro, Zh Qian, Backward stochastic differential equations associated with the vorticity equations. J. Funct. Anal. **267**(3), 660–677 (2014)
6. D.J. Ebin, J. Marsden, Groups of diffeomorphisms and the motion of an incompressible fluid. Ann. Math. **2**(92), 102–163 (1970)

7. H. Föllmer, *Time Reversal on Wiener Space*, Lecture Notes in Mathematics (Springer, Berlin, 1986)
8. H. Föllmer, *Random Fields and Diffusion Processes, Saint Flour XV-XVII* (Springer, New York, 1988)
9. H. Kuo, *Gaussian Measures in Banach Spaces*, Lecture Notes in Mathematics (Springer, New York, 1975), p. 463
10. R. Lassalle, Invertibility of adapted perturbations of the identity on abstract Wiener space. J. Funct. Anal. **262**(6), 2734–2776 (2012)
11. P. Malliavin, *Stochastic Analysis* (Springer, New York, 1997)
12. A.S. Üstünel, M. Zakai, *Transformation of Measure on Wiener Space* (Springer, New York, 1999)

KMT Theory Applied to Approximations of SDE

Alexander Davie

Abstract The dyadic method of Komlós, Major and Tusnády is a powerful way of constructing simultaneous normal approximations to a sequence of partial sums of i.i.d. random variables. We use a version of this KMT method to obtain order 1 approximation in a Vaserstein metric to solutions of vector SDEs under a mild non-degeneracy condition using an easily implemented numerical scheme.

Keywords SDE · Numerical scheme · Vaserstein metric

1 Introduction

The pathwise simulation of solutions of vector stochastic differential equations is challenging because, using standard methods, to obtain approximations to order greater than $\frac{1}{2}$ requires simulation of iterated integrals of the Brownian path, which is difficult. One approach is to seek approximations in a Vaserstein metric, meaning that there is a coupling between the approximate and exact solutions with respect to which the error is of the desired order. Reference [2] describes an easily generated scheme, based on the standard order 1 Milstein scheme, which is order 1 in a Vaserstein metric, provided the SDE has a nondegenerate diffusion term. Here we describe a modified version of the scheme from Ref. [2] which gives order 1 under a weaker nondegeneracy condition. The proof uses a construction of a coupling based on the KMT method.

Section 2 reviews the basics of SDE approximation and states the main result. Section 3 briefly reviews the KMT theorem and presents some required material from coupling and optimal transport theory. The rest of the paper is devoted to the proof of the theorem and a relevant example.

Some other work on SDE approximation using coupling is described in the final chapter of volume 2 of Ref. [6]. We also mention [1] which obtains an order $\frac{2}{3} - \epsilon$ bound in a Vaserstein metric for the Euler method in one dimension.

A. Davie (✉)
School of Mathematics, University of Edinburgh, Edinburgh, Scotland
e-mail: A.Davie@ed.ac.uk

D. Crisan et al. (eds.), *Stochastic Analysis and Applications 2014*,
Springer Proceedings in Mathematics & Statistics 100,
DOI 10.1007/978-3-319-11292-3_7

The author is grateful to the referee for several suggestions which improved the presentation of the paper.

2 Approximation of SDEs

Here we briefly review the Milstein scheme and formulate our new version.

Consider an Itô SDE

$$dx_i(t) = \sum_{k=1}^{d} b_{ik}(t, x(t))dW_k(t), \qquad x_i(0) = x_i^{(0)}, \qquad i = 1, \ldots, q \tag{1}$$

on an interval $[0, T]$, for a q-dimensional vector $x(t)$, with a d-dimensional driving Brownian path $W(t)$. If the coefficients $b_{ik}(t, x)$ satisfy a global Lipschitz condition

$$|a_i(t, x) - a_i(t, y)| \leq C|x - y|, \qquad |b_{ik}(t, x) - b_{ik}(t, y)| \leq C|x - y| \tag{2}$$

for all $x, y \in \mathbb{R}^q$, $t \in [0, T]$ and all i, k, where C is a constant, and if a_i and b_i are continuous in t for each x, then (1) has a unique solution $x(t)$ which is a process adapted to the filtration induced by the Brownian motion. This solution satisfies satisfies $\mathbb{E}|x(t)|^p < \infty$ for each $p \in [1, \infty)$ and $t \in [0, T]$.

The standard approach to the strong or pathwise approximation of the solution of (1), as described for example in Ref. [4], is to divide $[0, T]$ into a finite number N of subintervals, which we shall usually assume to be of equal length $h = T/N$, and to approximate the equation on each subinterval using a stochastic Taylor expansion. Such expansions are described in detail in Chap. 5 of [4]. The simplest such approximation, using only the linear term in the expansion, gives the Euler (also known as Euler-Maruyama) scheme

$$x_i^{(j+1)} = x_i^{(j)} + \sum_{k=1}^{d} b_{ik}(t_j, x^{(j)})V_k^{(j)} \tag{3}$$

while adding the quadratic terms gives the Milstein scheme

$$x_i^{(j+1)} = x_i^{(j)} + \sum_{k=1}^{d} b_{ik}(t_j, x^{(j)})V_k^{(j)} + \sum_{k,l=1}^{d} \rho_{ikl}(t_j, x^{(j)})I_{kl}^{(j)} \tag{4}$$

where $V_k^{(j)} = W_k((j+1)h) - W_k(jh)$, $I_{kl}^{(j)} = \int_{jh}^{(j+1)h} \{W_k(t) - W_k(jh)\}dW_l(t)$ and $\rho_{ikl}(t, x) = \sum_{m=1}^{q} b_{mk}(t, x) \frac{\partial b_{il}}{\partial x_m}(t, x)$.

Assuming (2) the Euler scheme has order $\frac{1}{2}$, in the sense that $\mathbb{E}(\max_{j=1}^{N} |x^{(j)} - x(jh)|^2) = O(h)$ and under a stronger smoothness condition on the b_{ik} the

Milstein scheme has order 1, indeed

$$\mathbb{E}(\max_{j=1}^{N} |x^{(j)} - x(jh)|^2) = O(h^2) \tag{5}$$

(see Kloeden and Platen [2], Sect. 10.3). These L^2 bounds can be extended to L^p for any $p \geq 1$.

The Euler scheme is straightforward to implement, as the only random variables one has to generate are the normally-distributed $V_k^{(j)}$, but for Milstein one has also to generate the 'area integrals' $I_{kl}^{(j)}$ which is non-trivial if $d \geq 2$. Order $\frac{1}{2}$ is the best one can do in general when the only random variables generated are the $V_k^{(j)}$.

We remark here that we can write $I_{kl}^{(j)} = \frac{1}{2}(V_k^{(j)} V_l^{(j)} - h\delta_{kl}) + \zeta_k^{(j)} V_l^{(j)} - \zeta_l^{(j)} V_k^{(j)} + K_{kl}^{(j)}$ with random variables $\zeta_k^{(j)}$, $K_{kl}^{(j)}$ for $1 \leq k, l \leq d$ all having zero mean, variance $\frac{h}{12}$, satisfying $K_{kl}^{(j)} = -K_{lk}^{(j)}$, and such that the $d(d+1)/2$ random variables consisting of $\zeta_k^{(j)} : 1 \leq k \leq d$ and $K_{kl}^{(j)} : 1 \leq k < l \leq d$ are mutually uncorrelated (though not independent).

Motivated by this remark we consider the following modification of Milstein, which requires the generation of normal random variables only.

$$\tilde{x}_i^{(j+1)} = \tilde{x}_i^{(j)} + \sum_{k=1}^{d} b_{ik}(t_j, \tilde{x}^{(j)}) \tilde{V}_k^{(j)} + \sum_{k,l=1}^{d} \rho_{ikl}(t_j, \tilde{x}^{(j)}) J_{kl}^{(j)} \tag{6}$$

where again the $\tilde{V}_k^{(j)}$ are independent $N(0, h)$ and $J_{kl}^{(j)} = \frac{1}{2}(\tilde{V}_k^{(j)} \tilde{V}_l^{(j)} - h\delta_{kl}) + z_k^{(j)} \tilde{V}_l^{(j)} - z_l^{(j)} \tilde{V}_k^{(j)} + \lambda_{kl}^{(j)}$, where the $z_k^{(j)}$ for $1 \leq k \leq d$ and $\lambda_{kl}^{(j)}$ for $1 \leq k < l \leq d$ are independent $N(0, \frac{h}{12})$, and then we set $\lambda_{lk}^{(j)} = -\lambda_{kl}^{(j)}$ for $k < l$ and $\lambda_{kk}^{(j)} = 0$.

Our main result is that, under suitable regularity conditions and a fairly mild nondegeneracy condition, the scheme (6) has order 1 under a suitable coupling. To formulate the nondegeneracy condition, we define for each $(t, x) \in [0, T] \times \mathbb{R}^d$ a linear mapping $L_{t,x} : \mathbb{R}^d \oplus S_d \to \mathbb{R}^q$ by $L_{t,x}(r, s)_i = \sum_{k=1}^{d} b_{ik}(t, x) r_k + \sum_{k,l=1}^{d} \rho_{ikl}(t, x) s_{kl}$ for $r \in \mathbb{R}^d$ and $s \in S_d$, where S_d is the space of skew-symmetric $d \times d$ matrices. We will require that L_x be surjective for each $(t, x) \in [0, T] \times \mathbb{R}^d$; this is equivalent to requiring that for each (t, x) the vectors $b^k(x)$ and $[b^k, b^l](x)$, for $1 \leq k, l \leq d$, will span $\mathbb{R}^q$ (here b^k is the vector whose ith component is b_{ik}, and $[b^k, b^l]$ denotes the Lie bracket, regarding b^k and b^l as vector fields on $\mathbb{R}^q$ for each t). This can be thought of as a strengthened Hörmander condition. In fact we need a version with some uniformity in (t, x), which we state precisely in the main theorem:

Theorem 1 *Suppose that the first and second derivatives of b_{ik} are bounded on $[0, T] \times \mathbb{R}^q$, and that there constants $\delta > 0$ and $K > 0$ such that for each $(t, x) \in [0, T] \times \mathbb{R}^q$ the image under $L_{t,x}$ of the unit ball in $\mathbb{R}^r \oplus S_d$ contains the ball of radius $\delta(1 + |x|)^{-K}$ in $\mathbb{R}^q$.*

Then there is a constant $C > 0$ such that if $N \in \mathbb{N}$ is given we can find independent $N(0, \frac{h}{12})$ random variables $z_k^{(j)}$ for $1 \le k \le d$ and $\lambda_{kl}^{(j)}$ for $1 \le k < l \le d$ and $0 \le j < N$, defined on the same probability space as the Brownian path $W(t)$, such that, if $\tilde{x}^{(j)}$ is as given by the scheme (6), we have $\mathbb{E}|\tilde{x}^{(j)} - x(jh)|^2 \le Ch^2$ for $j = 1, \ldots, N$.

A similar result is proved in Ref. [2] for the scheme

$$x_i^{(j+1)} = x_i^{(j)} + \sum_{k=1}^{d} b_{ik}(t_j, x^{(j)})V_k^{(j)} + \frac{1}{2}\sum_{k,l=1}^{d} \rho_{ikl}(t_j, x^{(j)})(V_k^{(j)}V_l^{(j)} - h\delta_{kl})$$

under a stronger nondegeneracy condition that the matrix (b_{ik}) has rank q.

The proof of Theorem 1 occupies much of the remainder of the paper. We note here some properties of the joint characteristic function χ of the random variables (z_k), (λ_{kl}), regarded as a function on $\mathbb{R}^{d(d+1)/2}$. An explicit expression for χ can be found in Ref. [9]. What we require are the following (taking the case $h = 1$, from which the general case can be deduced by scaling): χ extends to be analytic on a 'strip' $\{x + iy : x, y \in \mathbb{R}^{d(d+1)/2}$ in $\mathbb{C}^{d(d+1)/2}$ and $|y| < \delta\}$ for some $\delta > 0$, and $|\chi(x + iy)| < C(1 + |x|)^{-1}$ on this strip.

3 Coupling and KMT Theory

If we have two probability spaces $(\mathcal{X}, \mathcal{F}, \mathbb{P})$ and $(\mathcal{Y}, \mathcal{G}, \mathbb{Q})$ then a *coupling* between $\mathbb{P}$ and $\mathbb{Q}$ is a measure on $\mathcal{X} \times \mathcal{Y}$ which has $\mathbb{P}$ and $\mathbb{Q}$ as its marginal distributions. Theorem 1 asserts the existence of a coupling between the probability space of the Brownian path and that of the random variables used in the approximation (6). We collect here some results on couplings which we shall need.

First we mention the Vaserstein metrics on probability measures on $\mathbb{R}^n$. If $\mathbb{P}_1$ and $\mathbb{P}_2$ are such measures, we define $\mathbb{W}_p(\mathbb{P}_1, \mathbb{P}_2)$ to be the infimum of $\mathbb{E}|X - Y|^p$, taken over all couplings between $\mathbb{P}_1$ and $\mathbb{P}_2$ where X and Y have distributions $\mathbb{P}_1$ and $\mathbb{P}_2$ respectively. For $p \ge 1$ one can then show that $\mathbb{W}_p$ is a metric on the set of all probability measures $\mathbb{P}$ on $\mathbb{R}^n$ having finite pth moment (i.e. satisfying $\int_{\mathbb{R}^n} |x|^p d\mathbb{P}(x) < \infty$). $\mathbb{W}_p$ is known as the p-Vaserstein metric after Ref. [7] (Note: we use the transliteration 'Vaserstein' from the Cyrillic as that is the one used by Vaserstein himself; 'Wasserstein' is also used in the literature).

We also note the elementary result (see e.g. Proposition 7.10 in [8]) that

$$\mathbb{W}_p(\mu, \nu) \le 2^{(p-1)/p} \left\{ \int |x|^p d|\mu - \nu|(x) \right\}^{1/p} \tag{7}$$

for any two probability measures μ, ν on $\mathbb{R}^n$ and for any $p \ge 1$.

This is quite a good bound if $p = 1$ but is less good for $p > 1$; we shall however use it for bounding some small remainder terms.

The KMT theorem [5] is a form of simultaneous Central Limit Theorem using coupling. A variant of this result (modified from the original to be closer to the type of result we will use) states that if $\mathbb{P}$ is a suitably well-behaved probability distribution on $\mathbb{R}$, with zero mean, variance 1 and zero 3rd moment, then there is a constant $C > 0$ such that the following holds: if $n \in \mathbb{N}$ and $X_1, \ldots, X_n$ are independent with distribution $\mathbb{P}$, and if $Y_1, \ldots, Y_n$ are independent $N(0, 1)$, then there is a coupling between the random vectors $(X_1, \ldots, X_n)$ and $(Y_1, \ldots, Y_n)$ such that

$$\mathbb{E}\left\{\sum_{i=1}^{k}(X_i - Y_i)\right\}^2 \leq C$$

for $k = 1, \ldots, n$.

There are various generalisations in the literature. Einmahl [3] extended the result to vector random variables and Zaitsev [10] further extended it to non-identical distributions which are uniformly non-degenerate. What we require is a variant of this latter result where the distributions are themselves random. It is not clear that this can be easily deduced from results in the literature so we prefer to give a self-contained argument in the context we need. This argument will use the lemma and corollary below, on polynomial perturbations of normal distributions. We denote by ϕ the standard normal $N(0, I)$ distribution on $\mathbb{R}^q$.

Lemma 2 *Let X be an $\mathbb{R}^q$-valued random variable with $N(0, I)$ distribution, let $p : \mathbb{R}^q \to \mathbb{R}^q$ be a polynomial function of degree 3, and define $\rho : \mathbb{R}^q \to \mathbb{R}^q$ by $\rho(x) = x + p(x)$. Let $\mathbb{P}$ be the probability distribution of $\rho(X)$ and let ν be the signed measure on $\mathbb{R}^q$ with density $\phi(y)(1+y.p(y)-\nabla.p(y))$. Then for any $M \geq 1$ we have a bound*

$$\int_{\mathbb{R}^q}(1 + |y|)^M d|\mathbb{P} - \nu|(y) \leq C\epsilon^2 \tag{8}$$

where C is a positive constant depending only on q and M, and ϵ is an upper bound for the absolute values of the coefficients of p.

Proof We use C_1, C_2 etc. to denote positive constants which depend only on q and M. First we can find $C_1 \geq 1$ such that

$$\max(|\rho(x) - x|,\ \|D\rho(x) - I\|) \leq C_1\epsilon(1 + |x|)^3 \tag{9}$$

and

$$\max(|r(x)|,\ \|Dr(x)\|) \leq C_1\epsilon^2(1 + |x|)^9 \tag{10}$$

for all $x \in \mathbb{R}^q$, where $r(x) = p(x) - p(x + p(x))$. Then let $R = (2C_1\epsilon)^{-1/6} - 1$ and let $B_R = \{x \in \mathbb{R}^q : |x| < R\}$ (which will of course be empty if $R \leq 0$, which can happen if ϵ is not very small).

Now define a measure μ as the image under ρ of the restriction to B_R of the $N(0, I)$ distribution on $\mathbb{R}^q$. We also define $\tilde{\nu} = \nu|\rho(B_R)$. Then we have

$$\int_{\mathbb{R}^q} (1 + |y|)^M d|\mathbb{P} - \nu|(y) \leq \Omega_1 + \Omega_2 + \Omega_3$$

where $\Omega_1 = \int_{\mathbb{R}^q} (1 + |y|)^M d|\mu - \tilde{\nu}|(y)$, $\Omega_2 = \int_{\mathbb{R}^q} (1 + |y|)^M d(\mathbb{P} - \mu)(y)$ and $\Omega_3 = \int_{\mathbb{R}^q} (1 + |y|)^M d|\nu - \tilde{\nu}|(y)$.

We first bound Ω_1. To this end we note that, by the definition of R, for $x \in B_R$ the RHS of (9) is bounded by $\frac{1}{2}|\epsilon|^{1/2}$. It then follows from (9) that for $x \in B_R$ we have $\|D\rho(x) - I\| \leq \frac{1}{2}$ and so ρ is bijective on B_R. Then the density f of $\tilde{\nu}$ on $\rho(B_R)$ is given by $f(y) = \det D\rho^{-1}(y)\phi(\rho^{-1}(y))$ and so we have

$$\Omega_1 = \int_{\rho(B_R)} (1 + |y|)^M |\det D\rho^{-1}(y)\phi(\rho^{-1}(y)) - (1 + y.p(y) - \nabla.p(y))\phi(y)|dy$$

To bound the RHS, we fix $x \in B_R$ and set $y = \rho(x)$, noting that $|x - y| \leq \min(1, |y|^{-1})$ by (9). Noting that $x = y - p(y) + r(x)$ and using the bound (10) we readily find that

$$|\phi(x) - (1 + y.p(y) - \nabla.p(y))\phi(y)| \leq C_2\epsilon^2(1 + |y|^{C_3})\phi(y)$$

and

$$|\det D\rho^{-1}(y) - (1 - \nabla.p(y))| \leq C_2\epsilon^2(1 + |y|^{C_3}).$$

From this we easily deduce that $\Omega_1 \leq C_4\epsilon^2$.

Similar bounds for Ω_2 and Ω_3 are also easily proved, using the exponential decay of ϕ, and the result follows. □

We require a corollary of this lemma, for which we first introduce some notation.

Let P denote the space of all real-valued polynomials on $\mathbb{R}^q$, and P^q the space of $\mathbb{R}^q$-valued functions $p = (p_1, \ldots, p_q)$ on $\mathbb{R}^q$ such that each p_i is a polynomial. Let P_0 denote the subspace of $S \in P$ such that $\int_{\mathbb{R}^q} S(y)\phi(y)dy = 0$. We can characterise P_0 as follows. Let $\mathcal{L} : P^q \to P$ be the linear mapping defined by $\mathcal{L}p(x) = \nabla.p(x) - x.p(x)$. Then $\nabla.(\phi p)(x) = \mathcal{L}p(x)\phi(x)$ and it follows from the divergence theorem that $\mathcal{L}p \in P_0$ for every $p \in P^q$. In the converse direction, we note that if $u \in P$ has degree $n \geq 1$ then $\mathcal{L}\nabla u = -nu + r$ where $r \in P$ has degree less than n. If this u is in P_0 then we have $r \in P_0$ and by induction on n we can deduce that u is in the range of $\mathcal{L}$. So P_0 is precisely the range of $\mathcal{L}$.

Corollary 3 *Let $g \in P$ have degree 4, let μ be the measure with density ϕ (i.e. the standard normal probability measure) and let λ be a probability measure on $\mathbb{R}^q$ such that*

$$\int_{\mathbb{R}^q} (1+|y|)^M d|(1+g)\mu - \lambda|(y) \le \alpha \tag{11}$$

Then $\mathbb{W}_M(\mu, \lambda) \le C(\epsilon + (\epsilon^2 + \alpha)^{1/M})$, where C is a positive constant depending only on q and M, and ϵ is an upper bound for the absolute values of the coefficients of g.

Proof Let $\beta = \int g d\mu$. Then (11) gives $|\beta| \le \alpha$, and $g - \beta \in P_0$. So by replacing g by $g - \beta$ we can assume $g \in P_0$.

Then as described above we can find $p \in P^q$ with $\mathcal{L}p = g$, and from the construction of p it is clear that p has degree 3 and its coefficients are bounded by $C_1\epsilon$. Let X be an $N(0, I)$ random variable and let $Y = X + p(X)$. By Lemma 2 we have

$$\int_{\mathbb{R}^q} (1+|y|)^M d|g\mu - \nu|(y) \le C_2\epsilon^2$$

and so

$$\int_{\mathbb{R}^q} (1+|y|)^M d|\nu - \lambda|(y) \le C_2\epsilon^2 + \alpha$$

Hence by (7), $\mathbb{W}_M(\nu, \lambda) \le C_3(\epsilon^2 + \alpha)^{1/M}$. Finally $\mathbb{E}|Y - X|^M = \mathbb{E}|p(X)|^M \le C_4\epsilon^M$ so $\mathbb{W}_M(\mu, \nu) \le C_4^{1/M}\epsilon$ and the result follows by the triangle inequality. □

4 First Reduction

For the Milstein approximation $(x_i^{(j)})$ given above, we know that $\mathbb{E}|x^{(j)} - x(jh)|^2 \le Ch^2$ holds under the assumptions of the theorem. So to prove the theorem it suffices to obtain a bound $\mathbb{E}|\tilde{x}^{(j)} - x^{(j)}|^2 \le Ch^2$. We will construct a coupling between the set of random variables $V_k^{(j)}$, $I_{kl}^{(j)}$ used for Milstein and the set of random variables used by (6), such that this bound holds.

We first split each of the random variables $V_k^{(j)}$ as the sum of two parts: $V_k^{(j)} = Q_k^{(j)} + R_k^{(j)}$ where $Q_k^{(j)} \sim N(0, h - h^2)$ and $R_k^{(j)} \sim N(0, h^2)$ are independent. (See the remarks following the proof of Theorem 1 for discussion of this splitting). Now let $(u_i^{(j)})$ be the modified Euler approximation defined by the recurrence relation

$$u_i^{(j+1)} = u_i^{(j)} + \sum_{k=1}^{d} b_{ik}(u^{(j)}) Q_k^{(j)} \tag{12}$$

with $u^{(0)} = x^{(0)}$. Then define the $q \times q$ matrix $A^{(j)}$ by $A^{(j)}_{il} = \sum_{k=1}^{d} \frac{\partial b_{ik}}{\partial x_l}(u^{(j)})Q^{(j)}_k$, and a modified matrix by $\hat{A}^{(j)} = A^{(j)}$ if $\|A^{(j)}\| \le \frac{1}{2}$, and $\hat{A}^{(j)} = 0$ otherwise. We also define a modified version of $I^{(j)}_{kl}$ by replacing V by Q, namely $\overline{I}^{(j)}_{kl} = \frac{1}{2}(Q^{(j)}_k Q^{(j)}_l - (h-h^2)\delta_{kl}) + \zeta^{(j)}_k Q^{(j)}_l - \zeta^{(j)}_l Q^{(j)}_k + K^{(j)}_{kl}$. Then define $\alpha^{(j)}$ by the recurrence relation

$$\alpha^{(j+1)}_i = \{(I + \hat{A}^{(j)})\alpha^{(j)}\}_i + \sum_{k=1}^{d} b_{ik}(u^{(j)})R^{(j)}_k + \sum_{k,l=1}^{d} \rho_{ikl}(u^{(j)})\overline{I}^{(j)}_{kl} \tag{13}$$

with $\alpha^{(0)} = 0$. Next define $\beta^{(j)} = x^{(j)} - u^{(j)} - \alpha^{(j)} \in \mathbb{R}^q$ and note that then $\beta^{(0)} = 0$. We have $\beta^{(j+1)}_i - \beta^{(j)} = x^{(j+1)}_i - x^{(j)}_i - (u^{(j+1)}_i - u^{(j)}_i) - (\alpha^{(j+1)}_i - \alpha^{(j)}_i)$ and using (4), (12) and (13) we find after some rearrangement that

$$\begin{aligned}
\beta^{(j+1)}_i - \beta^{(j)}_i = &\sum_{k,l=1}^{d} \frac{\partial b_{ik}}{\partial x_l}(u^{(j)})\beta^{(j)}_l Q^{(j)}_k \\
&+ \sum_{k=1}^{d} \left\{ b_{ik}(x^{(j)}) - b_{ik}(u^{(j)}) - \sum_{l=1}^{q} \frac{\partial b_{ik}}{\partial x_l}(u^{(j)})(x^{(j)}_l - u^{(j)}_l) \right\} Q^{(j)}_k \\
&+ \sum_{k=1}^{d} \{b_{ik}(x^{(j)}) - b_{ik}(u^{(j)})\} R^{(j)}_k + \sum_{k,l=1}^{d} (\rho_{ikl}(x^{(j)}) - \rho_{ikl}(u^{(j)})) I_{kl} \\
&+ \{(A^{(j)} - \hat{A}^{(j)})\alpha^{(j)}\}_i + \sum_{k,l=1}^{d} \rho_{ikl}(u^{(j)})(I^{(j)}_{kl} - \overline{I}^{(j)}_{kl})
\end{aligned} \tag{14}$$

We now bound the RHS of (14). First note that, conditional on the random variables $Q^{(i)}, R^{(i)}, \zeta^{(i)}, K^{(i)}$ for $i < j$, each of the 6 terms on the RHS has expectation 0. Also the first term has variance bounded by $C_1\mathbb{E}|\beta^{(j)}|^2 h$. Next, we see that the scheme (12) has order $\frac{1}{2}$, being an Euler scheme with the random term scaled by $\sqrt{1-h} = 1 + O(h)$, so that $\mathbb{E}|x^{(j)} - u^{(j)}|^2 \le C_2 h$. Then we see that each of the other 3 terms on the RHS has variance bounded by $C_3 h^3$. Then we conclude from (14) that $\mathbb{E}|\beta^{(j+1)}|^2 \le (1 + C_1 h)\mathbb{E}|\beta^{(j)}|^2 + C_3 h^3$ and hence that

$$\mathbb{E}|x^{(j)} - u^{(j)} - \alpha^{(j)}|^2 = \mathbb{E}|\beta^{(j)}|^2 \le C_4 h^2 \tag{15}$$

for $j = 1, \ldots, N$.

We can do a similar analysis for $(x^{(j)}_i)$ as defined by (6) using random variables $\tilde{V}^{(j)}_k$, $z^{(j)}_k$ and $\lambda^{(j)}_{kl}$ as above. We again write $\tilde{V}^{(j)}_k = \tilde{Q}^{(j)}_k + \tilde{R}^{(j)}_k$ with $\tilde{Q}^{(j)}_k \sim N(0, h-h^2)$ and $\tilde{R}^{(j)}_k \sim N(0, h^2)$. Our intention is to construct a coupling between the two sets of random variables so that they all all defined on the same probability space, on which we can compare the two approximations. Our coupling will satisfy $\tilde{Q}^{(j)} = Q^{(j)}$, so we will assume this from now on.

Then, using the same $\hat{A}^{(j)}$ and $u^{(j)}$ as above, we define $\tilde{\alpha}^{(j)}$ by the recurrence relation

$$\tilde{\alpha}_i^{(j+1)} = \{(I + \hat{A}^{(j)})\tilde{\alpha}^{(j)}\}_i + \sum_{k=1}^{d} b_{ik}(u^{(j)})\tilde{R}_k^{(j)} + \sum_{k,l=1}^{d} \rho_{ikl}(u^{(j)})\overline{J}_{kl}^{(j)} \tag{16}$$

with $\tilde{\alpha}^{(0)} = 0$, where $\overline{J}_{kl}^{(j)} = \frac{1}{2}(Q_k^{(j)} Q_l^{(j)} - (h - h^2)\delta_{kl}) + z_k^{(j)} Q_l^{(j)} - z_l^{(j)} Q_k^{(j)} + \lambda_{kl}^{(j)}$ and just as before we obtain a bound

$$\mathbb{E}|\tilde{x}^{(j)} - u^{(j)} - \tilde{\alpha}^{(j)}|^2 = \mathbb{E}|\tilde{\beta}^{(j)}|^2 \le C_5 h^2 \tag{17}$$

From the bounds (15) and (17) we see that to prove the theorem it is enough to obtain a bound

$$\mathbb{E}|\alpha^{(j)} - \tilde{\alpha}^{(j)}|^2 \le C h^2 \tag{18}$$

We prove this in the next section.

As preparation we note some properties of the process $(u^{(j)})$. We let $\mathcal{G}$ denote the σ-algebra generated by $Q^{(0)}, \ldots, Q^{(N-1)}$, so that the $u^{(j)}$ and $\hat{A}^{(j)}$ are $\mathcal{G}$-measurable. As $u(j)$ is an Euler approximation to (1), with the random term scaled by $\sqrt{1-h}$, standard bounds apply and we have $\mathbb{E}|u^{(j)}|^p \le C(p)$ for any $p \ge 1$. We also define $B^{(r)} = (I + \hat{A}^{(1)})^{-1} \cdots (I + \hat{A}^{(r)})^{-1}$ and we readily obtain $\mathbb{E}\|B^{(r)}\|^p \le C(p)$ and $\mathbb{E}\|(B^{(r)})^{-1}\|^p \le C(p)$.

5 Proof of Theorem

Throughout we use C to denote a constant which may depend on the SDE but is independent of N; each occurrence may be different.

With $B^{(r)}$ as defined above we set

$$\gamma^{(r)} = B^{(r)} \left\{ \sum_{k=1}^{d} R_k^{(r)} b_k(u^{(r)}) + \sum_{k,l=1}^{d} \sigma_{kl}(u^{(r)}) Q_l^{(r)} \zeta_k^{(r)} + \sum_{1 \le k < l \le d} \sigma_{kl}(u^{(r)}) K_{kl}^{(r)} \right\},$$

$$\tilde{\gamma}^{(r)} = B^{(r)} \left\{ \sum_{k=1}^{d} \tilde{R}_k^{(r)} b_k(u^{(r)}) + \sum_{k,l=1}^{d} \sigma_{kl}(u^{(r)}) Q_l^{(r)} z_k^{(r)} + \sum_{1 \le k < l \le d} \sigma_{kl}(u^{(r)}) \lambda_{kl}^{(r)} \right\},$$

where $\sigma_{kl}(x)$ is the vector in $\mathbb{R}^q$ whose ith component is $\rho_{ikl}(x) - \rho_{ilk}(x)$, and we see that

$$\alpha^{(j)} - \tilde{\alpha}^{(j)} = (B^{(j)})^{-1} \sum_{r=0}^{j-1} (\gamma^{(r)} - \tilde{\gamma}^{(r)}) \tag{19}$$

It is convenient to reformulate the above expressions for $\gamma^{(r)}$ and $\tilde{\gamma}^{(r)}$ using random variables scaled to have variance 1. We let $m = d(d+3)/2$ and define random vectors $X^{(r)} = (X_1^{(r)}, \dots, X_m^{(r)})$ by $X_k^{(r)} = h^{-1} R_k^{(r)}$ for $k = 1, \dots, d$; $X_k^{(r)} = (12/h)^{1/2} \zeta_{k-d}^{(r)}$ for $k = d+1, \dots, 2d$; $X_{(k+1)(d-k/2)+l} = 12^{1/2} h^{-1} K_{kl}^{(r)}$. Then (conditional on $\mathcal{G}$), $X^{(r)}$ has mean 0 and covariance matrix I. We can then write $h^{-1}\gamma^{(r)} = G_r X^{(r)}$ where G_r is a $q \times m$ matrix defined in terms of $B^{(r)}, b_k(u^{(r)}), \sigma_{kl}(u^{(r)}), Q^{(r)}$. In the same way we have $h^{-1}\tilde{\gamma}^{(r)} = G_r \tilde{X}^{(r)}$ where $\tilde{X}^{(r)}$ is $N(0, I)$.

We have inequalities

$$\|G_r\| \le \|B^{(r)}\| \left(\sum_{k=1}^{d} |b_k^{(r)}(u^{(r)}| + \sum_{k,l=1}^{d} |\sigma_{kl}(u^{(r)}|(h^{-1/2}|Q_l^{(r)}| + 1) \right)$$

and $G_r G_r^t \ge B^{(r)} F(u^{(r)}) B^{(r)t}$ where $F(x) = \sum_{k=1}^{d} b_k(x) b_k(x)^t + \frac{1}{12} \sum_{k<l} \sigma_{kl}(x) \sigma_{kl}(x)^t$. We note that the nondegeneracy hypothesis in the theorem implies that $\|(F(x)F(x)^t)^{-1}\| \le C(1+|x|)^{-2K}$. From these bounds and those at the end of the last section we deduce that $\mathbb{E}\|G_r\|^p \le C(p)$ and $\mathbb{E}\|(G_r G_r^t)^{-1}\|^p \le C(p)$ for all $p \ge 1$. We remark that, conditional on $\mathcal{G}$, $\gamma^{(r)}$ and $\tilde{\gamma}^{(r)}$ have the same covariance matrix $h^2 G_r G_r^t$.

From now on we assume for convenience that N is a power of 2, $N = 2^\kappa$ (this can always be arranged by extending the SDE to the interval $[0, 2^\kappa h]$ where κ is the smallest integer such that $2^\kappa \ge N$).

Let $E_0 = \{0, 1, \dots, 2^\kappa - 1\}$. We call a subset E of E_0 dyadic if it is of the form $E = \{m2^n, m2^n + 1, \dots, (m+1)2^n - 1\}$ for some $n \in \{0, 1, \dots, \kappa\}$ and $m \in \{0, 1, \dots, 2^{\kappa-n} - 1\}$. We see then that, for each n, the dyadic sets of size 2^n form a partition of E_0, and each dyadic set of size 2^{n+1} is the union of two dyadic sets of size 2^n. For each dyadic set E of size 2^n we then define $\gamma_E = \sum_{r \in E} \gamma^{(r)}$, $\tilde{\gamma}_E = \sum_{r \in E} \tilde{\gamma}^{(r)}$ and $H_E = 2^{-n} \sum_{r \in E} G_r G_r^t$. Note that since, conditional on $\mathcal{G}$, the random variables $\gamma^{(0)}, \dots, \gamma^{(N-1)}$ are independent, H_E is the (conditional) covariance matrix of $Y_E := 2^{-n/2} h^{-1} \gamma_E$. The same applies to $\tilde{\gamma}_E$.

The idea is to construct couplings between $\tilde{\gamma}_E$ and γ_E recursively, starting with E_0 and proceeding by successive bisection. For this purpose we use the following lemma, which is a version of the Central Limit Theorem saying that the density of γ_E is close to the (Gaussian) density of $\tilde{\gamma}_E$.

Lemma 4 *Let E be a dyadic set of size 2^n, and let f_E be the density function of Y_E, conditional on $\mathcal{G}$. Fix η with $0 < \eta < \frac{1}{12}$. Then, provided $\|G_r\| < 2^{\eta n}$ and $\|(G_r G_r^t)^{-1}\| < 2^{2\eta n}$ for each $r \in E$, we have for $|v| < 2^{\eta n}$ that*

$$\left|\frac{f_E}{\tilde{f}_E}(v) - 1 - p_E(v)\right| < C2^{(16\eta-2)n}$$

where $\tilde{f}_E(v) = (2\pi)^{-q/2}(\det H_E)^{-1/2}\exp(-\frac{1}{2}v^t H_E^{-1} v)$ is the density of $\tilde{Y}_E$ and $p_E(v)$ is a 4th degree polynomial whose coefficients are bounded by $C2^{(4\eta-1)n}$.

Proof Note first that the bounds on G_r imply $\|H_E\| \le 2^{2n\eta}$ and $\|H_E^{-1}\| \le 2^{2n\eta}$.

Let ψ be the characteristic function of the random variable $X^{(r)}$ (which is independent of r). ψ is real-valued and even on $\mathbb{R}^m$, and extends to a complex-analytic function on a 'strip' $\{x + iy : x, y \in \mathbb{R}^m, |y| < a\}$ for some $a > 0$. In a neighbourhood of 0 in $\mathbb{C}$, $\log\psi$ has a convergent expansion $\log\psi(z) = -\frac{1}{2}|z|^2 + c_4(z) + c_6(z) + \cdots$ where $c_k(z)$ is a homogeneous polynomial of degree k, and $|c_k(z)| \le (C|z|)^k$ for even $k \ge 4$. Then $\psi(z) = \exp -\frac{z^t z}{2} + \chi(z)$ where $\chi(z) = c_4(z) + c_6(z) + \cdots$. From this it follows that there exists $\delta > 0$ such that

$$\text{if } x, y \in \mathbb{R} \text{ with } 2|y| \le |x| < \delta \text{ then } |\psi(x+iy)| \le e^{-|x|^2/6} \tag{20}$$

Then using the decay of ψ as $x \to \infty$ and the fact that $|\psi(x)| < 1$ for $x \in \mathbb{R}$ with $x \ne 0$, we can find γ with $0 < \gamma < 1$ and $\delta' > 0$ so that

$$\text{if } x, y \in \mathbb{R} \text{ with } |x| \ge \delta \text{ and } |y| \le \delta' \text{ then } |\psi(x+iy)| \le \min(\gamma, C|x|^{-1}) \tag{21}$$

Now let Ψ be the characteristic function of Y_E; then $\Psi(u) = \prod_{r\in E}\psi(2^{-n/2}G_r^t u)$ and $f_E(v) = (2\pi)^{-q/2}\int_{\mathbb{R}^q} e^{-iu^t v}\Psi(u)du$, which by translating the subspace of integration in $\mathbb{C}^q$ by $-iH^{-1}v$ we can write as

$$f_E(v) = (2\pi)^{-q} e^{-v^t H_E^{-1} v} \int_{\mathbb{R}^q} e^{-iu^t v}\Psi(u - iH_E^{-1}v)du \tag{22}$$

If $|u| \ge 2^{4\eta n+1}$ we can write $\Psi(u - iH_E^{-1}v) = \prod_{r\in E}\psi(2^{-n/2}G_r^t u - i2^{-n/2}G_r^t H_E^{-1}v)$. Now using (20) and (21) we see that each term in the product is bounded by either $\min(\gamma, (C2^{n(\eta+1/2)}|u|^{-1})$ or $\exp(-2^{-(1+2\eta)n}|u|^2/6)$, and we deduce that $|\Psi(u - iH_E^{-1}v)| \le \min(\gamma, (C2^{n(\eta+1/2)}|u|^{-1})^{2^n} + \exp(-2^{-2\eta n}|u|^2/6)$ for $|u| \ge 2^{4\eta n+1}$. It then follows easily that

$$\int_{|u|\geq 2^{4\eta n+1}} |\Psi(u - iH_E^{-1}v)|du \leq C\left\{2^{nm}\gamma^{2^n} + \exp(-2^{6\eta n-1})\right\} \tag{23}$$

To get a bound for $|u| \leq 2^{4\eta n+1}$ we write $w = u - iH_E^{-1}v$ and note that $e^{-iu^t v}$ $\Psi(w) = \exp(\frac{1}{2}v^t H_E^{-1}v - \frac{1}{2}u^t H_E u + \Lambda(w)$ where $\Lambda(w) = \sum_{r\in E}\chi(2^{-n/2}G_r^t w) = \sum_{k=2}^{\infty} S_{2k}(w)$ where $S_{2k}(w) = 2^{-kn}\sum_{r\in E} c_{2k}(G_r^t w)$. We see that S_{2k} is a homogeneous polynomial of degree $2k$ and satisfies $|S_{2k}(w)| \leq C2^{(1-k+2k\eta)n}|w|^{2k}$. We find that $|e^{\Lambda(w)} - 1 - S_4(w)| \leq C2^{(8\eta-2)n}|w|^6$ and hence that

$$e^{-\frac{1}{2}v^t H_E^{-1}v}\int_{|u|\leq 2^{4\eta n+1}} |e^{-iu^t v}\Psi(u - iH_E^{-1}v) - (1 + S_4(u - iH_E^{-1}v))e^{-u^2 H_E u}|du$$
$$\leq C2^{(16\eta-2)n} \tag{24}$$

We also have $\int_{|u|\geq 2^{4\eta n+1}} |1 + S_4(u - iH_E^{-1}v)|e^{-\frac{1}{2}u^t H_E u}du \leq Ce^{-2^{\eta n}}$ and combining these bounds the lemma follows, with $p_E(v) = \int_{\mathbb{R}^q} S_4(u - iH_E^{-1}v)e^{-\frac{1}{2}u^2 H_E u}du$ which is a polynomial of degree 4 whose coefficients are bounded by $C2^{(4\eta-1)n}$. $\square$

Initial step. We start the construction by finding a coupling between $\tilde{Y}_{E_0}$ and Y_{E_0}. Let $\mathcal{E}_0$ be the event that condition (27) below holds with $E = E_0$. Then provided $\mathcal{E}_0$ holds, Lemma 4 gives $|f_{E_0}(y)/\tilde{f}_{E_0}(y) - 1 - p_{E_0}(y)| < C2^{(16-2\eta)n}$ for $|y| \leq 2^{n\eta/3}$. To apply Corollary 3 we write $y = H_{E_0}^{1/2}u$ and $g(u) = (\det H_{E_0})^{1/2} f_{E_0}(H_{E_0}^{1/2}U)$, and deduce that

$$\int_{\mathcal{A}} (1 + |u|)^3 \left|\left\{1 + p_{E_0}\left(H_{E_0}^{1/2}u\right)\right\}\phi(u) - g(u)\right| du < C2^{(16\eta-2)n} \tag{25}$$

where $\mathcal{A} = \{u \in \mathbb{R}^q : |H_{E_0}^{1/2}u| < 2^{n\eta/3}\}$. One can easily see that the integral over $\mathcal{A}^c$ is bounded by $C2^{-2n}$ so that (25) holds with the integral over $\mathbb{R}^q$. And the polynomial $p_{E_0}(H^{1/2}u)$ has coefficients bounded by $C2^{(4\eta-1)n}$ so from Corollary 3 we have $\mathbb{W}_3(g, \phi) \leq C2^{(16\eta-2)n/3}$. Then $\mathbb{W}_3(f_{E_0}, \tilde{f}_{E_0}) \leq \|H_{E_0}^{1/2}\|\mathbb{W}_3(g, \phi) \leq C2^{(17\eta-2)n/3}$. So can can find a coupling between Y_{E_0} and Y_E so that

$$\mathbb{E}|Y_{E_0} - \tilde{Y}_{E_0}|^3 \leq C2^{(17\eta-2)n} \tag{26}$$

Recursive step. Let E be a dyadic set of size 2^n where $n \geq 1$. We can write E in a unique way as the union of two disjoint dyadic sets F and G of size 2^{n-1} and note that $Y_F + Y_G = 2^{1/2}Y_E$ and $\tilde{Y}_F + \tilde{Y}_G = 2^{1/2}\tilde{Y}_E$. We suppose a coupling between $\tilde{Y}_E$ and Y_E has been defined, conditional on $\mathcal{G}$. In other words, for each choice of $Q^{(0)}, \ldots, Q^{(N-1)}$, we have a joint distribution of Y_E and $\tilde{Y}_E$ with the correct conditional marginal distributions. We wish to extend this coupling to a coupling between (Y_F, Y_G) and $(\tilde{Y}_F, \tilde{Y}_G)$.

For each $x \in \mathbb{R}^q$, let f_x be the density of Y_F conditional on $Y_E = x$ and on $\mathcal{G}$, and let $\tilde{f}_x$ be the density of $\tilde{Y}_F$ conditional on $\tilde{Y}_E = x$ and on $\mathcal{G}$. We note that the conditional distribution of $\tilde{Y}_F$, given $\tilde{Y}_E = x$ and $\mathcal{G}$, is $N(Jx, H)$ where $J = H_F H_E^{-1}$ and $H = \frac{1}{2} H_F H_E^{-1} H_G$. So $\tilde{f}_x$ is the density function of $N(Jx, H)$.

We need to find a coupling between Y_F and $\tilde{Y}_F$, conditional on $Y_E = x$ and $\tilde{Y}_E = \tilde{x}$. To do this we need a coupling between the distributions with densities f_x and $\tilde{f}_{\tilde{x}}$. We shall in fact construct a coupling between f_x and $\tilde{f}_x$, then use the fact that $\tilde{f}_{\tilde{x}}$ is just $\tilde{f}_x$ translated by $J(\tilde{x} - x)$.

First we note that $f_x(y) = \frac{2^{1/2} f_F(y) f_G(2^{1/2}x - y)}{f_E(x)}$. Then the provided the condition

$$\|G_r\| < 2^{\eta n/6} \quad \text{and} \quad \|(G_r G_r^t)^{-1}\| < 2^{\eta n/3} \quad \text{for each } r \in E \tag{27}$$

holds, applying Lemma 4 to each of E, F, G gives

$$\left| \frac{f_x(y)}{\tilde{f}_x(y)} - 1 - p_x(y) \right| < C 2^{(16\eta - 2)n} \tag{28}$$

for $|x|, |y| \leq 2^{n\eta/3}$, where $p_x(y) = p_F(y) + p_G(2^{1/2}x - y) - p_E(x)$.

Let $\Omega = \{x \in \mathbb{R}^q : \mathbb{E}(|Y_F|^3 \chi_{|Y_F| \geq 2^{\eta n/3}} | Y_E = x \,\&\, \mathcal{G}) > 2^{-2n}\}$, and let $p = 60/\eta$. Then we see that, provided (27) holds,

$$\begin{aligned} \mathbb{P}(Y_E \in \Omega | \mathcal{G}) &\leq 2^{2n} \mathbb{E}(|Y_F|^3 \chi_{|Y_F| \geq 2^{\eta n/3}} | \mathcal{G}) \leq 2^{-18n} \mathbb{E}(|Y_F|^{p+3} | \mathcal{G}) \\ &\leq C 2^{-18n} \|H_E\|^{(p+3)/2} \leq C 2^{-2n} \end{aligned} \tag{29}$$

Let $\mathcal{E}$ denote the event that (27) holds, $|Y_E| \leq 2^{n\eta/6 - 1}$ and $Y_E \notin \Omega$. Write $x = Y_E$. In order to apply Corollary 3 to the conditional distribution of Y_F, we make the change of variable $y = Jx + H^{1/2}u$, noting that then $\tilde{f}_x(y) = (\det H)^{-1/2} \phi(u)$. We define $g_x(u) = (\det H)^{1/2} f_x(Jx + H^{1/2}u)$. Then, provided $\mathcal{E}$ holds, (28) gives

$$\int_{\mathcal{A}} (1 + |u|)^3 \left| \left\{ 1 + p_x(Jx + H^{1/2}u) \right\} \phi(u) - g_x(u) \right| du < C 2^{(16\eta - 2)n} \tag{30}$$

where $\mathcal{A} = \{u \in \mathbb{R}^q : |Jx + H^{1/2}u| < 2^{n\eta/3}\}$. Also, writing $y = Jx + H^{1/2}u$, if $|y| \geq 2^{n\eta/3}$ we have $|H^{1/2}u| \leq 2|y|$ so $|u| \leq 2^{1 + n\eta/6}|y|$ and then, using $x \notin \Omega$, we find $\int_{\mathcal{A}^c} (1 + |u|)^3 g_x(u) \leq 2^{(\eta - 2)n}$. We also easily get $\int_{\mathcal{A}} (1 + |u|)^3 |1 + p_x(Jx + H^{1/2}u) \phi(u)| du < C 2^{-2n}$. Putting these bounds together, we obtain

$$\int_{\mathbb{R}^q} (1 + |u|)^3 \left| \left\{ 1 + p_x(Jx + H^{1/2}u) \right\} \phi(u) - g_x(u) \right| du < C 2^{(16\eta - 2)n} \tag{31}$$

The polynomial $p_x(Jx + H^{1/2}u)$ has coefficients bounded by $C 2^{(5\eta - 1)n}$ and then applying Corollary 3 we deduce that $\mathbb{W}_3(g_x, \phi) \leq C 2^{(16\eta - 2)n/3}$. Then $\mathbb{W}_3(f_x, \tilde{f}_x) \leq$

$\|H^{1/2}\|\mathbb{W}_3(g_x, \phi) \leq C2^{(17\eta-2)n/3}$. In other words, conditional on $Y_E = x$ and assuming $\mathcal{E}$, we can find a random variable Y_F^* with density $\tilde{f}_x$ such that $\mathbb{E}|Y_F^* - Y_F|^3 \leq C2^{(17\eta-2)n}$. If $\mathcal{E}$ fails then we find an arbitrary variable Y^* with density $\tilde{f}_x$. One easily finds that $\mathbb{P}(\mathcal{E}) \leq C2^{-6n}$ and then taking expectation over $\mathcal{G}$ and Y_E we find that unconditionally

$$\mathbb{E}|Y_F^* - Y_F|^3 \leq C2^{(17\eta-2)n} \tag{32}$$

We can now complete the recursive step by defining

$$\tilde{Y}_F = Y_F^* + H_F H_E^{-1}(\tilde{Y}_E - Y_E) \tag{33}$$

which has the correct conditional density $\tilde{f}_{\tilde{x}}$ with $\tilde{x} = \tilde{Y}_E$. Then we must have $\tilde{Y}_G = 2^{1/2}\tilde{Y}_E - \tilde{Y}_F$. It is natural to define $Y_G^* = 2^{1/2}Y_E - Y_F^*$; then one sees that (32) and (33) both hold with F replaced by G.

Conclusion of proof. We apply the recursive procedure as described above, starting with E_0 (initial step), then using the recursive step to proceed from dyadic sets of size 2^n to dyadic sets of size 2^{n-1}, to generate a coupling for every dyadic set. The result can be summarised as follows: conditional on $\mathcal{G}$ we have constructed a coupling between the sets of random variables (Y_E) and $(\tilde{Y}_E)$, each ranging over the dyadic sets E, such that (32) and (33) hold whenever F is a dyadic set of size 2^{n-1} contained in a dyadic set E of size 2^n.

Now consider a given dyadic set E of size 2^n. We can write in a unique way $E = E_k \subseteq E_{k-1} \subseteq \cdots \subseteq E_0$ where $k = \kappa - n$ and, for each j, E_j is a dyadic set of size $2^{\kappa-j}$. Then from (33) we obtain $\tilde{Y}_E - Y_E = \sum_{j=1}^k H_{E_k} H_{E_j}^{-1}(Y_{E_j}^* - Y_{E_j}) + H_{E_k} H_{E_0}^{-1}(\tilde{Y}_{E_0} - Y_{E_0})$. From this, using (26) and (32) along with the L^p bounds for $H_{E_j}^{-1}$ and H_{E_k}, and using Hölder's inequality, we obtain $\|\tilde{Y}_E - Y_E\|_{5/2} \leq C2^{3(17\eta-2)n/10}$. Thus

$$\|\tilde{\gamma}_E - \gamma_E\|_{5/2} \leq C2^{(51\eta-1)n/10}h \tag{34}$$

holds whenever E is a dyadic set of size 2^n. We now apply this to (19). If $1 \leq j \leq N$ we can write $\{0, 1, \ldots, j-1\}$ as a union of dyadic sets $E_1 \cup \cdots \cup E_k$ where $E_1, \ldots, E_k$ have different sizes. Then (19) gives $\alpha^{(j)} - \tilde{\alpha}^{(j)} = (B^{(j)})^{-1} \sum_{i=1}^k (\tilde{\gamma}_{E_i} - \gamma_{E_i})$. Provided $\eta < \frac{1}{51}$, (34) then gives (18) using Hölder's inequality and an L^{10} bound for $(B^{(j)})^{-1}$. This completes the proof of the theorem.

Remark A natural question is whether the theorem is true without the nondegeneracy condition. Without this condition the KMT-type argument faces considerable technical difficulties, but I would conjecture that the theorem is still true.

The splitting $V^{(j)} = Q^{(j)} + R^{(j)}$ is introduced in order to allow the vectors b^k as well as the Lie brackets be included in the nondegeneracy condition. If we have the nondegeneracy condition with the brackets only (i.e. the term in r_k is omitted from the definition of $L_{t,x}$) then we can prove the theorem without this splitting - but this condition is considerable stronger (e.g. it cannot hold if $q = d = 2$).

Note that our result is slightly weaker than than the bound (5) for Milstein which has a max over j. Equation (5) is deduced from the bound for individual j using Doob's martingale inequality; however we cannot apply this to our scheme because our coupling does not preserve the filtrations, so the error $\tilde{x}^{(j)} - x(jh)$ is not a martingale. The following example shows that the analogue of (5) fails for scheme (6), whatever coupling is used.

Example We consider the SDE with $q = 3$ and $d = 2$ given by

$$dx_1 = dW_1, \quad dx_2 = dW_2, \quad dx_3 = x_1 dW_2 - x_2 dW_1$$

on the time interval [0, 1], with initial condition $x_i(0) = 0$.

This SDE has solution $x_1 = W_1$, $x_2 = W_2$, $x_3(t) = \int_0^t (W_1(s)dW_2(s) - W_2(s)dW_1(s))$. We find that $\rho_{312}(x) = 1$, $\rho_{321}(x) = -1$ and all other ρ_{ikl} are zero. It is then easy to check that the hypotheses of Theorem 1 are satisfied. We also note that the Milstein approximation is exact, in that $x^{(j)} = x(jh)$ for each j.

We claim that there is a constant $c > 0$ such that, for any $N \in \mathbb{N}$ the approximation using scheme (6) with $h = \frac{1}{N}$ and any coupling between the random variables $\tilde{V}_k^{(j)}, z_k^{(j)}, \lambda_{12}^{(j)}$ used by (6) and the Brownian path W, we have

$$\mathbb{P}(\max_{0 \le j < N} |\tilde{x}^{(j)} - x(jh)| \ge cN^{-1} \log N) > \frac{1}{2} \tag{35}$$

To prove this claim we first note that

$$x_3^{(j+1)} - x_3^{(j)} = x_1^{(j)} V_2^{(j)} - x_2^{(j)} V_1^{(j)} + I_{12}^{(j)} - I_{21}^{(j)} \tag{36}$$

and

$$\tilde{x}_3^{(j+1)} - \tilde{x}_3^{(j)} = \tilde{x}_1^{(j)} \tilde{V}_2^{(j)} - \tilde{x}_2^{(j)} \tilde{V}_1^{(j)} + 2(z_1^{(j)} \tilde{V}_2^{(j)} - z_2^{(j)} \tilde{V}_1^{(j)} + \lambda_{12}^{(j)}) \tag{37}$$

We also define random variables $M = \max_{0 \le j < N} |\tilde{x}^{(j)} - x^{(j)}|$, $K = \max_{1 \le j \le N} |W(jh)|$ and $\tilde{K} = \max_{1 \le j \le N} |(\tilde{x}_1^{(j)}, \tilde{x}_2^{(j)})|$. And we set $X^j = h^{-1}(I_{12}^{(j)} - I_{21}^{(j)})$, $Y^{(j)} = \frac{2}{h}(z_1^{(j)} \tilde{V}_2^{(j)} - z_2^{(j)} \tilde{V}_1^{(j)})$ and $Z^{(j)} = \frac{2}{h}\lambda_{12}^{(j)}$. Then subtracting (36) from (37) and using the above definitions we find that

$$h|X^{(j)} - Y^{(j)} - Z^{(j)}| \le 2M(1 + 2K + 2\tilde{K}) \tag{38}$$

The idea is to use (38) to get a lower bound for M. For this we need the distributions of the random variables on the LHS of (38). First note that, from the known distribution of the Lévy area, $X^{(j)}$ has density $\frac{1}{2}\text{sech}(\pi x/2)$ so $\mathbb{P}(|X^{(j)}| \ge \lambda) \ge C_1 e^{-\pi\lambda/2}$ for $\lambda > 0$. And $Y^{(j)}$ can be expressed as $\frac{1}{2\sqrt{3}}(P^2 - Q^2 + R^2 - S^2)$ where P, Q, R, S are independent $N(0, 1)$, so that $P^2 + R^2$ and $Q^2 + S^2$ have exponential distributions, and then a simple calculation shows that $Y^{(j)}$ has a symmetric exponential distribution with $\mathbb{P}(|Y^{(j)}| > \lambda) = e^{-\sqrt{3}\lambda}$. Moreover $Z^{(j)}$ has $N(0, \frac{1}{3})$ distribution, from which one finds easily that $\mathbb{P}(|Y^{(j)} + Z^{(j)}| > \lambda) \leq C_2 e^{-5\lambda/3}$ (using $\frac{5}{3} < \sqrt{3}$).

Now fix α and β with $\frac{3}{5} < \beta < \alpha < \frac{2}{\pi}$. Then we have

$$\mathbb{P}(\max_{0\leq j<N} |X^{(j)}| \leq \alpha \log N) \leq (1 - C_1 e^{-\pi\alpha \log N/2})^N \leq \exp(C_1 N^{1-\frac{\pi\alpha}{2}})$$

and

$$\mathbb{P}(\max_{0\leq j<N} |Y^{(j)} + Z^{(j)}| \geq \beta \log N) \leq C_2 N^{1-\frac{3\beta}{5}}$$

So if N is large enough we have $\mathbb{P}(\max_{0\leq j<N} |X^{(j)}| \leq \alpha \log N) \leq \frac{1}{8}$ and $\mathbb{P}(\max_{0\leq j<N} |Y^{(j)} + Z^{(j)}| \geq \beta \log N) \leq \frac{1}{8}$. Moreover we can find a constant C_3 so that $\mathbb{P}(K \geq C_3) \leq \frac{1}{8}$ and $\mathbb{P}(\tilde{K} \geq C_3) \leq \frac{1}{8}$. Then, with probability at least $\frac{1}{2}$, we have

$$\begin{aligned} &\max_{0\leq j<N} |X^{(j)}| \geq \alpha \log N, \quad \max_{0\leq j<N} |Y^{(j)} + Z^{(j)}| \leq \beta \log N, \\ &K \leq C_3, \text{ and } \tilde{K} \leq C_3 \end{aligned} \tag{39}$$

Finally, using (38), (39) implies $2M(1+4C_3) \geq (\alpha-\beta)h \log N$, giving the required result.

References

1. A. Alfonsi, B. Jourdain, A. Kohatsu-Higa, Pathwise optimal transport bounds between a one-dimensional diffusion and its Euler scheme. Ann. Appl. Probab. **24**, 1049–1080 (2014)
2. A.M. Davie, Pathwise approximation of stochastic differential equations using coupling, preprint. www.maths.ed.ac.uk/~adavie/coum.pdf
3. U. Einmahl, Extensions of results of Komlós, Major and Tusnády to the multivariate case. J. Multivar. Anal. **28**, 20–68 (1989)
4. P.E. Kloeden, E. Platen, *Numerical Solution of Stochastic Differential Equations* (Springer, Heidelberg, 1995)
5. J. Komlós, P. Major, G. Tusnády, An approximation of partial sums of independent RV's and the sample DF. I. Z. Wahr. und Wer. Gebiete **32**, 111–131 (1975)
6. S.T. Rachev, L. Ruschendorff, Mass Transportation Problems, Volume 1, Theory; Volume 2, Applications. Springer, Heidelberg (1998)
7. L.N. Vaserstein, Markov processes over denumerable products of spaces describing large system of automata, Problemy Peredaci Informacii **5**, 64–72 (Russian). (Translation in Problems of Information Transmission) **5**(1969), 47–52 (1969)
8. C. Villani, *Topics in Optimal Transportation* (AMS, New York, 2003)

9. M. Wiktorsson, Joint characteristic function and simultaneous simulation of iterated Itô integrals for multiple independent Brownian motions. Ann. Appl. Probab. **11**, 470–487 (2001)
10. A. Zaitsev, Multidimensional version of a result of Sakhanenko on the invariance principle for vectors with finite exponential moments. I, II, III Teor. Veroyatnost. i Primenen, 45 (2000), 718–738; 46 (2001), 535–561 and 744–769. (Translations in Theory Probab. Appl.) 45 (2002), 624–641; 46 (2003), 490–514 and 676–698

Regularity Theory for Rough Partial Differential Equations and Parabolic Comparison Revisited

Joscha Diehl, Peter K. Friz and Harald Oberhauser

Abstract Partial differential equations driven by rough paths are studied. We return to the investigations of [Caruana, Friz and Oberhauser: A (rough) pathwise approach to a class of non- linear SPDEs, Annales de l'Institut Henri Poincaré/Analyse Non Linéaire 2011, 28, pp. 27–46], motivated by the Lions–Souganidis theory of viscosity solutions for SPDEs. We continue and complement the previous (uniqueness) results with general existence and regularity statements. Much of this is transformed to questions for deterministic parabolic partial differential equations in viscosity sense. On a technical level, we establish a refined parabolic theorem of sums which may be useful in its own right.

Keywords Rough partial differential equations · Regularity · Existence · Parabolic comparison and theorem of sums

1991 Mathematics Subject Classification 35R99 · 60H15

1 Introduction

In Ref. [5], and then [6], inspired by earlier works of Lions–Souganidis on stochastic viscosity theory [24, 25, 27, 28], rough path stability (with respect to the multi-dimensional driving differential $dz = \dot{z}dt$) was established for the parabolic Cauchy

J. Diehl
Technische Universität Berlin, Berlin, Germany

P.K. Friz (✉)
Technische Universität Berlin and WIAS Berlin, Berlin, Germany
e-mail: friz@math.tu-berlin.de

H. Oberhauser
Oxford-MAN, University of Oxford, Oxford, UK

D. Crisan et al. (eds.), *Stochastic Analysis and Applications 2014*,
Springer Proceedings in Mathematics & Statistics 100,
DOI 10.1007/978-3-319-11292-3_8

problem with spatial domain $\mathbb{R}^n$,

$$du = F\left(t, x, Du, D^2u\right) dt - \sum_{i=1}^{d} Du\,(t, x) \cdot V_i\,(x)\, dz^i, \quad u\,(0, \cdot) = u_0.$$

Loosely speaking, if a family (z^ϵ) is Cauchy in rough path metric, with rough path limit $\mathbf{z}$ (example: typical mollifications of Brownian motion will satisfy this condition with probability one; the rough path limit is identified as Brownian motion and Lévy's stochastic area; see e.g. [15]), then the resulting PDE solutions of the above problem, (u^ϵ) will converge locally uniformly to a limit which is seen to depend only on $\mathbf{z}$, and not on the particular sequence (z^ϵ). In particular, this allows a (rough)pathwise and robust view on stochastic partial differential equations (SPDEs). Immediate (probabilistic) benefits of this approach include support theorems, large deviations and a variety of limit theorems for SPDEs; see [6] for a discussion. Another nice application of robustness in the driving signal is that it quickly leads to *splitting* results for such (rough, and then stochastic) PDEs; see [17]. There are various extensions to noise other than $H(x, Du) = \sum_{i=1}^{d} Du \cdot V_i\,(x)$. The "fully non-linear" case $H = H(Du)$, with non-linear dependence on Du, is quite intricate and discussed in Ref. [25], and then [26] with regard to applications; adding x-dependence i.e. $H = H(x, Du)$ is a difficult problem, cf. the forthcoming book by Lions–Souganidis and also forthcoming joint work with P. Gassiat. The case $H = H(x, u)$ is considered in Refs. [12, 27]; the later reference makes a link to backward stochastic differential equations *with rough drivers*. Even the "fully linear" case, with both F and $H = H(x, u, Du)$ linear (in u and its derivatives), is interesting as it covers the Zakai equation from filtering theory (e.g. [1]): robust dependence on z (the "observation" path) is a classical problem, which goes back to the engineering literature of the late seventies; the rough path point of view has recently led to resolution of this problem; [11, 16].

Having commented on the importance of such classes of (rough) partial differential equations, let us describe the contribution of this paper. We complement the stability result of Ref. [6], which settled uniqueness, with general existence and regularity results, giving conditions for a space-time modulus of continuity (en passant, this justifies regarding solutions as elements in BUC-spaces, as is common in viscosity theory) and also spatial Lipschitz regularity. Our conditions are satisfied for large classes of F; for instance, infima and/or suprema over linear weakly elliptic operators, as is typical in stochastic control resp. differential games. As for the noise term, we have focused on $H(x, Du) = Du \cdot V\,(x)$, but adaptions to (linear) $H = H(x, u, Du)$ or the setting of Ref. [12] are not difficult.

As a matter of fact, after reduction of the RPDE problem to a (classical) viscosity problem, we are in the need of fairly general (parabolic) comparison results on $\mathbb{R}^n$. Unfortunately, we failed to find this in the literature which forced us to revisit and adapt some results from viscosity theory. This is a common situation, of course, but since the necessary work seems to go well beyond a routine exercise, and may be of independent interest, we opted to include a reasonably self-contained discussion (Sect. 2) which may be read independently from the rough path/RPDE considerations of Sect. 3.

2 Parabolic Comparison Revisited

2.1 Structural Conditions on F

Let $F = F(t, x, u, p, X) : [0, T] \times \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}^n \times S^n \to \mathbb{R}$ be continuous and degenerate elliptic i.e. non-decreasing in X. Here S^n denotes the space of symmetric $n \times n$ matrices. Assume also that there exists γ such that, uniformly in t, x, p, X,

$$\gamma(u - v) \leq F(t, x, v, p, X) - F(t, x, u, p, X) \text{ whenever } v \leq u. \tag{1}$$

When $\gamma \geq 0$ such Fs are called *proper*. Since we will be interested in parabolic problems of the form $\partial_t - F$ a suitable change of variable ($u \leftrightarrow e^{\gamma t}u$) shows that $\gamma < 0$ does not cause trouble. Assume furthermore that there exists, for all $R > 0$, a function $\theta_R : [0, \infty] \to [0, \infty]$ with $\theta_R(0+) = 0$, such that

$$\begin{aligned} &F(t, x, r, \alpha(x - \tilde{x}), X) - F(t, \tilde{x}, r, \alpha(x - \tilde{x}), Y) \\ &\quad \leq \theta_R\left(\alpha|x - \tilde{x}|^2 + |x - \tilde{x}|\right) \end{aligned} \tag{2}$$

for all $t \in [0, T],\ x, \tilde{x} \in \mathbb{R}^n,\ r \in [-R, R], \alpha > 0$ and $X, Y \in S^n$ (the space of $n \times n$ symmetric matrices) which satisfy

$$-3\alpha\begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix} \leq \begin{pmatrix} X & 0 \\ 0 & -Y \end{pmatrix} \leq 3\alpha\begin{pmatrix} I & -I \\ -I & I \end{pmatrix}. \tag{3}$$

Under these conditions, comparison for the Cauchy–Dirichlet problem $\partial_t - F = 0$ on $(0, T) \times \Omega$, with Ω bounded, holds (User's Guide, Chap. 8). We shall be interested in comparison for bounded (semi-continuous, sub- and super-) solutions on $(0, T] \times \mathbb{R}^n$. In particular, the unboundedness of $\mathbb{R}^n$ leads us to the following additional assumption: assume $F = F(t, x, u, p, X)$ is *uniformly continuous* (UC) whenever u, p, X remain bounded; i.e.

$$\forall R > 0 : F|_{[0,T]\times\mathbb{R}^n\times[-R,R]\times B_R\times M_R} \text{ is uniformly continuous} \tag{4}$$

where B_R, M_R denote (open) balls of radius R in $\mathbb{R}^n,\ S^n$ respectively.[1] Remark that these structural conditions are satisfied when $F = F(u, p, X)$ is proper (no t, x dependence).

2.2 Statement of Theorems

We write BC, BUC, BUSC, BLSC for bounded continuous, bounded uniformly continuous and bounded upper- resp. lower semi-continuous functions.

[1] Using any of the equivalent norms on S^n.

Theorem 1 (Comparison and spatial regularity) *Assume F satisfies the assumptions of Sect. 2.1. Consider $u \in \mathrm{BUSC}\left([0,T) \times \mathbb{R}^n\right)$, $v \in \mathrm{BLSC}\left([0,T) \times \mathbb{R}^n\right)$, extended to $[0,T] \times \mathbb{R}^n$ via their semi-continuous envelopes; i.e.*[2]

$$u(T,x) = \limsup_{\substack{(t,y)\in[0,T)\times\mathbb{R}^n: \\ t\to T, y\to x}} u(t,y), \quad v(T,x) = \liminf_{\substack{(t,y)\in[0,T)\times\mathbb{R}^n: \\ t\to T, y\to x}} v(t,y). \tag{5}$$

Assume that, in the sense of parabolic viscosity sub- and super-solutions [3]

$$\partial_t u - F\left(t,x,u,Du,D^2u\right) \le 0 \le \partial_t v - F\left(t,x,v,Dv,D^2v\right) \quad \text{on } (0,T)\times\mathbb{R}^n. \tag{6}$$

Then the following statements hold true.
(i) The validity of (6) extends to $Q := (0,T] \times \mathbb{R}^n$.
(ii) If $u_0 := u(0,)\ \ v_0 := v(0,) \in \mathrm{BUC}(\mathbb{R}^n)$ and $u_0 \le v_0$ on $\mathbb{R}^n$ one has the "key" estimate

$$u(t,x) - v(t,y) \le \inf_{\alpha}\left[\frac{\alpha}{2}|x-y|^2 + l(\alpha)\right],$$

valid for all $(t,x,y) \in [0,T] \times \mathbb{R}^n \times \mathbb{R}^n$, where $l(\alpha)$ tends to 0 as $\alpha \uparrow \infty$, uniformly in $t \in [0,T]$.

Remark 1 Since $\tilde{u}(t,x) = e^{-\gamma t}u(t,x)$ [resp. $\tilde{v}(t,x) = e^{-\gamma t}v(t,x)$] is a sub- [resp. super-]solution to $\left(\partial_t - \tilde{F}\right)\tilde{u} + \gamma\tilde{u} = 0$ with

$$\tilde{F}(t,x,p,X) = e^{-\gamma t}F\left(t,x,e^{\gamma t}\tilde{u},e^{\gamma t}D\tilde{u},e^{\gamma t}D^2\tilde{u}\right)$$

we can always reduce to the case that $\gamma > 0$. In particular, we shall give the proof under this assumption.

Remark 2 The key estimate immediately implies comparison (take $x = y$)

$$u \le v \text{ on } [0,T] \times \mathbb{R}^n.$$

By a 2ϵ argument, it also yields a spatial modulus for any solution u; uniform in $t \in [0,T]$. Indeed, for fixed $t \le T$ pick α large enough so that $l(\alpha) < \epsilon/2$; for any x,y : $|x-y|$ small enough (only depending on α and hence ϵ) we have $u(t,x) - u(t,y) < \epsilon$. By switching the roles of x and y, if necessary, we see $|u(t,x) - u(t,y)| < \epsilon$.

[2] If one assumes that given $u \in \mathrm{BUSC}\left([0,T] \times \mathbb{R}^n\right)$, $v \in \mathrm{BLSC}\left([0,T] \times \mathbb{R}^n\right)$ satisfy (6) on $(0,T] \times \mathbb{R}^n$ then it already follows that (5) holds true. This follows from the so-called Accessibility Theorem [7].

[3] As is well-known, the precise meaning of (6) is expressed (equivalently) in terms of "touching" test-functions or in term of sub- and super-jets. We shall switch between these points of view without further comments.

Theorem 2 (Spatial Lipschitz regularity) *Assume F satisfies the assumptions of Sect.* 2.1 *with the strengthening that the modulus θ_R is linear, i.e. $\theta_R(x) = \theta_R x$ for a constant $\theta_R > 0$. Let $u \in BC([0,T] \times \mathbb{R}^n)$ be a solution to*

$$\partial_t u - F\left(t, x, u, Du, D^2 u\right) = 0 \text{ on } (0,T] \times \mathbb{R}^n. \tag{7}$$

and assume $u(0,\cdot)$ to be Lipschitz with Lipschitz constant L_{u_0}. Then, for all $t \in (0,T]$, $u(t,\cdot)$ is Lipschitz uniformly in $t \in [0,T]$, with Lipschitz constant $e^{\bar{\gamma}t}/2 + 2L_{u_0}^2 + \theta_R^2/\gamma$ where $\bar{\gamma} := 2(\theta_R + 1)$.

Theorem 3 (Time-space regularity) *Assume F satisfies the assumptions of Sect.* 2.1*, with the strengthening*

$$\forall R > 0 : F|_{[0,T]\times\mathbb{R}^n\times[-R,R]\times B_R\times M_R} \text{ is bounded, uniformly continuous.} \tag{8}$$

(i) Let $u \in \mathrm{BC}([0,T] \times \mathbb{R}^n)$ be a viscosity solution to $\partial_t - F = 0$ on $(0,T] \times \mathbb{R}^n$ with intial data $u_0 = u(0,\cdot) \in \mathrm{BUC}(\mathbb{R}^n)$. Then

$$u = u(t,x) \in \mathrm{BUC}\left([0,T] \times \mathbb{R}^n\right).$$

(ii) If, in addition, θ_R is linear and if F has also linear growth in the Hessian, i.e. there exists an $M > 0$ such that

$$|F(t,x,r,p,X)| \le M(1 + |X|),$$

for all $t \in [0,T], x \in \mathbb{R}^n$, $r \in \mathbb{R}, p \in \mathbb{R}^n$ and $X, Y \in S^n$, and $u_0 = u(0,\cdot) \in \mathrm{BC} \cap \mathrm{Lip}(\mathbb{R}^n)$ then u is $1/2$-Hö lder in time (uniformly in space) and Lipschitz in space (uniformly in $t \in [0,T]$).

Theorem 4 (Existence) *Assume F satisfies the assumptions of Sect.* 2.1*, with the strengthening* (8)*, as above. Let $u_0 \in \mathrm{BUC}(\mathbb{R}^n)$. Then there exists a unique bounded viscosity solution to the initial value problem*

$$\partial_t u - F\left(t, x, u, Du, D^2 u\right) = 0 \text{ on } (0,T] \times \mathbb{R}^n,$$
$$u(0,\cdot) = u_0$$

and $u = u(t,x) \in \mathrm{BUC}([0,T] \times \mathbb{R})$.

2.3 Comments on the Existing Literature

The above ensemble of results gives, under natural conditions, a fairly complete picture of the (model) case of bounded solutions to parabolic problems on $\mathbb{R}^n$. While it is not harder to think of further generalizations (e.g. unbounded solutions, discontinuous F, Dirichlet problem on unbounded domains …), the setting here is appropriate for the study of stochastic viscosity[4] and rough partial differential equations. Moreover, and this is the *raison d′être* of the present section, one is hard-pressed to find these results in a similar form in the literature.

Let us be specific and point to the closest we are aware of: results in the spirit of Theorem 1, part (ii) and 3, part (i), are found in Ref. [19]. As for Theorem 1, part (ii), related results are found in Ref. [7, 18, 22] (and the references therein). Regularity results relating to Theorem 3, part (ii), are found in Ref. [20]. (The works [20, 21] are based on a useful continuous dependence result whose proof is a variant of the comparison principle proof[5] If one restricts to the first order case, so that $F = F(t, x, u, Du)$, the theory is less involved and there is a fairly complete text book literature [2, 3]. For instance, the natural analogue of condition 4 in a first oder setting appear in Ref. [2], p. 49, as condition (H11) and p. 136, as condition (H23).

On a technical level, the deepest result of this section is an extension of the *Theorem of Sums* (TOS), cf. Sect. 2.9, and it does not appear that one can get uniform continuity in space-time easily without it. To elaborate on this point, recall that almost every modern treatise of second order comparison relies in one way or another on the TOS, also known as *Crandall-Ishii Lemma* [9]. A parabolic version of the TOS on $(0, T) \times \mathbb{R}^n$ then underlies most second order (parabolic) comparison results; such as those in [10, Chap. 8] or [14, Chap. 5]. As is well-known, its application requires a barrier at time T; e.g. replace a subsolution u by $u^\gamma := u - \gamma/(T - t)$ or so, followed by $\gamma \downarrow 0$ in the end. The downside is that an initially bounded function u is turn into an unbounded function u^γ; consequently various localizations of the non-linearity are necessary to deal with the resulting unboundedness. (An example of the resulting complication was also seen in Ref. [12].) But then, establishing a spatial modulus of solutions with the (standard) form of the parabolic Theorem of Sums leads to an (apriori) dependence of the spatial modulus in time; establishing the (desired) uniformity in $t \in [0, T]$, cf. (iii) above, would then entail, if possible at all, a painstaking checking of uniformity in γ for all double limits in the technical Lemma 6 below. All these difficulties, as will be seen, are avoided by our extension of the (parabolic) TOS on $Q := (0, T] \times \mathbb{R}^n$.[6] Since, in fact, the TOS is often used as a "black-box" theorem, in particular in the stochastic control literature such as [14, Chap. 5]), there seems to be every reason to use it in its strongest possible form; in this sense our extension seems to be of independent interest. Concluding this short subsection, it appears that Theorems 1–4, which are otherwise proved by more or less well-known techniques, have not been obtained previously in this form for lack of a suitable TOS, valid at terminal time.

---

[4] All results of Lions–Souganidis on stochastic viscosity theory are stated in BUC-spaces with spatial domain $\mathbb{R}^n$.

[5] These ideas may prove useful in establishing rates of convergence for equations driven by (z^ϵ), convergent in rough path sense.

[6] The point here, of course, is to handle appropriately terminal time T which is a well-documented subtlety in (parabolic) viscosity theory; some mistakes in the early literature were corrected in Ref. [7].

2.4 Proof of Theorem 1: Parabolic Comparison, Part (i)

Proof Assume $u \in \text{BUSC}([0,T) \times \mathbb{R}^n)$ solves $\partial_t u - F(t,x,u,Du,Du) \leq 0$ with "properness" $\gamma \geq 0$; with initial data $u(0,\cdot)$ on $(0,T) \times \mathbb{R}^n$. By assumption, u is extended to $\text{BUSC}([0,T] \times \mathbb{R}^n)$ by setting

$$u(T,x) = \lim \sup_{t \uparrow T, y \to x} u(t,y).$$

Assume $u - \phi$ has a (strict) max at $(T,\bar{x})$, relative to $[0,T] \times \mathbb{R}^n$. (The test function ϕ is defined in an open neighbourhood of $(T,\bar{x}) \in [0,T] \times \mathbb{R}^n$.) In the remainder of this section we establish

$$\partial_t \phi(T,\bar{x}) - F\left(T,\bar{x},u(T,\bar{x}),D\phi(T,\bar{x}),D^2\phi(T,\bar{x})\right) \leq 0.$$

To this end, start by taking $(t^n,x^n) \in (0,T) \times \mathbb{R}^n$ s.t. $(t^n,x^n) \to (T,\bar{x})$ and $u(t^n,x^n) \to u(T,\bar{x})$. Set $\alpha_n := T - t^n \downarrow 0$. Then take

$$(t_n,x_n) \in \arg\max\left(u - \phi - \frac{\alpha_n^2}{T-t}\right) \equiv \arg\max \psi_n.$$

over $[0,T] \times \mathbb{R}^n$. In order to guarantee that the sequence $(t_n,x_n) \in [0,T) \times \mathbb{R}^n$ remains in a compact, say $[T/2,T] \times \bar{B}_1(\bar{x})$, we make the assumption (without loss of generality) that $\phi(T,\bar{x}) = 0$ and $\phi(t,x) > 3|u|_\infty$ for $(t,x) \notin [T/2,T] \times \bar{B}_1(\bar{x})$; this implies $(t_n,x_n) \in [T/2,T] \times \bar{B}_1(\bar{x})$ for n large enough, as desired. By compactness, $(t_n,x_n) \to \left(\tilde{t},\tilde{x}\right)$ at least along a subsequence $n(k)$. We shall run through the other sequence (t^n,x^n) along the same subsequence and relabel both to keep the same notation. Note $\psi_n(t_n,x_n)$ is non-decreasing and bounded, hence $\psi_n(t_n,x_n) \to l$. Since $\psi_n(t_n,x_n) \leq (u-\phi)(t_n,x_n)$ it follows (using USC of $u-\phi$) that $l \leq (u-\phi)\left(\tilde{t},\tilde{x}\right)$. On the other hand,

$$\psi_n(t_n,x_n) \geq \psi_n\left(t^n,x^n\right) = (u-\phi)\left(t^n,x^n\right) - \underbrace{\frac{\alpha_n^2}{T-t^n}}_{=\alpha_n}$$

and hence $l \geq (u-\phi)(T,\bar{x})$. Since $(T,\bar{x})$ was a strict maximum point for $u-\phi$ conclude that $\left(\tilde{t},\tilde{x}\right) = (T,\bar{x})$ is the common limit of the sequences $(t^n,x^n), (t_n,x_n)$. Now we note that

$$(u-\phi)(t_n,x_n) \geq \psi_n(t_n,x_n) \geq (u-\phi)\left(t^n,x^n\right) - \alpha_n$$

which implies that $u(t_n,x_n) \geq u(t^n,x^n) + o(1)$ where $o(1) \to 0$ as $n \to \infty$. By definition of a subsolution,

$$\partial_t \phi(t_n,x_n) - F\left(t_n,x_n,u(t_n,x_n),D\phi(t_n,x_n),D^2\phi(t_n,x_n)\right) \leq 0$$

and hence, using properness of F (more specifically, (1) applied with $u = u(t_n, x_n)$ and $v = u(t^n, x^n) + o(1)$),

$$\begin{aligned}-F(u(t_n, x_n)) &\geq -F\left(u\left(t^n, x^n\right) + o(1)\right) + \gamma\left(u(t_n, x_n) - \left(u\left(t^n, x^n\right) + o(1)\right)\right)\\ &\geq -F\left(u\left(t^n, x^n\right)\right) + o(1),\end{aligned}$$

also using uniform continuity of F as function of u over compacts, we obtain

$$\partial_t \phi(t_n, x_n) - F\left(t_n, x_n, u\left(t^n, x^n\right), D\phi(t_n, x_n), D^2\phi(t_n, x_n)\right) \leq o(1).$$

Sending $n \to \infty$ yields (use continuity of ϕ and F)

$$\partial_t \phi(T, \bar{x}) - F\left(T, \bar{x}, u(T, \bar{x}), D\phi(T, \bar{x}), D^2\phi(T, \bar{x})\right) \leq 0,$$

as desired. □

2.5 Proof of Theorem 1: Parabolic Comparison, Part (ii)

Proof By assumption, $u(t, x) - v(t, y)$ is bounded on $[0, T] \times \mathbb{R}^n \times \mathbb{R}^n$. Let $\left(\hat{t}, \hat{x}, \hat{y}\right)$ be a maximum point of

$$\phi(t, x, y) := u(t, x) - v(t, y) - \frac{\alpha}{2}|x - y|^2 - \varepsilon\left(|x|^2 + |y|^2\right) \tag{9}$$

over $[0, T] \times \mathbb{R}^n \times \mathbb{R}^n$ where $\alpha > 0$ and $\varepsilon > 0$; such a maximum exists since $\phi \in \mathrm{USC}([0, T] \times \mathbb{R}^n \times \mathbb{R}^n)$ and $\phi \to -\infty$ as $|x|, |y| \to \infty$. (The presence $\varepsilon > 0$ amounts to a barrier at ∞ in space). The plan is to show a "key estimate" of the form

$$u(t, x) - v(t, y) \leq \inf_\alpha \left[\frac{\alpha}{2}|x - y|^2 + l(\alpha)\right], \tag{10}$$

valid on $[0, T] \times \mathbb{R}^n \times \mathbb{R}^n$, where $l(\alpha)$ tends to 0 as $\alpha \uparrow \infty$. Thanks to the very definition of $\left(\hat{t}, \hat{x}, \hat{y}\right)$ as arg max of $\phi = \phi(t, x, y)$, we obtain the estimate

$$u(t, x) - v(t, y) \leq \frac{\alpha}{2}|x - y|^2 + \varepsilon\left(|x|^2 + |y|^2\right) + \phi\left(\hat{t}, \hat{x}, \hat{y}\right).$$

Note that $\left(\hat{t}, \hat{x}, \hat{y}\right)$ depends on α, ε. We shall consider the cases $\hat{t} = 0$ and $\hat{t} \in (0, T]$ separately. In the first case $\hat{t} = 0$ we have

$$\phi\left(0, \hat{x}, \hat{y}\right) = \sup_{x,y}\left[u_0(x) - v_0(y) - \frac{\alpha}{2}|x - y|^2 - \varepsilon\left(|x|^2 + |y|^2\right)\right] =: A_{\alpha,\varepsilon}$$

and Lemma 5 below asserts that $A_{\alpha,\varepsilon} \to \sup_x [u_0(x) - v_0(x)] \le 0$ as $(\varepsilon, \alpha) \to (0, \infty)$. The second case is $\hat{t} \in (0, T]$ and we will show

$$\phi\left(\hat{t}, \hat{x}, \hat{y}\right) \le B_{\alpha,\varepsilon} \text{ where } \left(\limsup_{\varepsilon\to 0} B_{\alpha,\varepsilon}\right) \to 0 \text{ as } \alpha \to \infty; \tag{11}$$

it is here that we will use Theorem of Sums and viscosity properties. (Since

$$\phi\left(\hat{t}, \hat{x}, \hat{y}\right) \le u\left(\hat{t}, \hat{x}\right) - v\left(\hat{t}, \hat{y}\right)$$

we can and will use the fact that it is enough to consider the case $u\left(\hat{t}, \hat{x}\right) - v\left(\hat{t}, \hat{y}\right) \ge 0$.) Leaving the details of this to below, let us quickly complete the argument: our discussion of the two cases above gives $\phi\left(\hat{t}, \hat{x}, \hat{y}\right) \le A_{\alpha,\varepsilon} \vee B_{\alpha,\varepsilon}$ and hence

$$u(t, x) - v(t, y) \le \frac{\alpha}{2}|x - y|^2 + \varepsilon\left(|x|^2 + |y|^2\right) + A_{\alpha,\varepsilon} \vee B_{\alpha,\varepsilon};$$

we emphasize that this estimate is valid for all $t, x, y \in [0, T] \times \mathbb{R}^n \times \mathbb{R}^n$ and $\alpha, \varepsilon > 0$. Take now $\limsup_{\varepsilon\to 0}$ on the right hand side, then optimize over $\alpha > 0$, to obtain the key estimate

$$u(t, x) - v(t, y) \le \inf_\alpha \left\{\frac{\alpha}{2}|x - y|^2 + l(\alpha)\right\}$$

where we may take $l(\alpha) := \limsup_{\varepsilon\to 0} A_{\alpha,\varepsilon} \vee \limsup_{\varepsilon\to 0} B_{\alpha,\varepsilon}$, noting that $l(\alpha)$ indeed tends to 0 as $\alpha \to \infty$. It remains to prove the estimate (11). To this end, rewrite ϕ as

$$\phi(t, x, y) = u^\varepsilon(t, x) - v^\varepsilon(t, y) - \frac{\alpha}{2}|x - y|^2$$

where $u^\varepsilon(t, x) = u(t, x) - \varepsilon|x|^2$ and $v^\varepsilon(t, y) = v(t, y) + \varepsilon|y|^2$. Since u^ε (resp. v^ε) are upper (resp. lower) semi-continuous we can apply the (parabolic) Theorem of Sums as given in the appendix at $\left(\hat{t}, \hat{x}, \hat{y}\right) \in (0, T] \times \mathbb{R}^n \times \mathbb{R}^n$ to learn that there are numbers a, b and $X, Y \in S^n$ such that

$$\left(a, \alpha\left(\hat{x} - \hat{y}\right), X\right) \in \bar{\mathcal{P}}_Q^{2,+} u^\varepsilon\left(\hat{t}, \hat{x}\right), \quad \left(b, \alpha\left(\hat{x} - \hat{y}\right), Y\right) \in \bar{\mathcal{P}}_Q^{2,-} v^\varepsilon\left(\hat{t}, \hat{y}\right) \tag{12}$$

such that $a - b \ge 0$ (equality if $\hat{t} \in (0, T)$, although this does not matter), and such that one has the two-sided matrix estimate (3). It is easy to see (cf. [10, Remark 2.7]) that (12) is equivalent to

$$\left(a, \alpha\left(\hat{x} - \hat{y}\right) + 2\varepsilon\hat{x}, X + 2\varepsilon I\right) \in \bar{\mathcal{P}}_Q^{2,+} u\left(\hat{t}, \hat{x}\right),$$
$$\left(b, \alpha\left(\hat{x} - \hat{y}\right) - 2\varepsilon\hat{y}, Y - 2\varepsilon I\right) \in \bar{\mathcal{P}}_Q^{2,-} v\left(\hat{t}, \hat{y}\right).$$

Using the viscosity sub- and super-solution properties (and part (i) in the case that $\hat{t} = T$) we then see that

$$
\begin{aligned}
a - F\left(\hat{t}, \hat{x}, u\left(\hat{t}, \hat{x}\right), \alpha\left(\hat{x} - \hat{y}\right) + 2\varepsilon\hat{x}, X + 2\varepsilon I\right) &\leq 0, \\
b - F\left(\hat{t}, \hat{y}, v\left(\hat{t}, \hat{y}\right), \alpha\left(\hat{x} - \hat{y}\right) - 2\varepsilon\hat{y}, Y - 2\varepsilon I\right) &\geq 0.
\end{aligned}
$$

Note that (using $a - b \geq 0$)

$$
\begin{aligned}
&F\left(\hat{t}, \hat{y}, v\left(\hat{t}, \hat{y}\right), \alpha\left(\hat{x} - \hat{y}\right) - 2\varepsilon\hat{y}, Y - 2\varepsilon I\right) \\
&\quad - F\left(\hat{t}, \hat{x}, u\left(\hat{t}, \hat{x}\right), \alpha\left(\hat{x} - \hat{y}\right) + 2\varepsilon\hat{x}, X + 2\varepsilon I\right) \leq 0
\end{aligned} \tag{13}
$$

Trivially, (recall it is enough to consider the case $u\left(\hat{t}, \hat{x}\right) \geq v\left(\hat{t}, \hat{y}\right)$)

$$
\begin{aligned}
\gamma\phi\left(\hat{t}, \hat{x}, \hat{y}\right) &\leq \gamma\left(u\left(\hat{t}, \hat{x}\right) - v\left(\hat{t}, \hat{y}\right)\right) \\
&\leq F\left(\hat{t}, \hat{y}, v\left(\hat{t}, \hat{y}\right), \alpha\left(\hat{x} - \hat{y}\right) - 2\varepsilon\hat{y}, Y - 2\varepsilon I\right) \\
&\quad - F\left(\hat{t}, \hat{y}, u\left(\hat{t}, \hat{x}\right), \alpha\left(\hat{x} - \hat{y}\right) - 2\varepsilon\hat{y}, Y - 2\varepsilon I\right) \\
&\leq F\left(\hat{t}, \hat{x}, u\left(\hat{t}, \hat{x}\right), \alpha\left(\hat{x} - \hat{y}\right) + 2\varepsilon\hat{x}, X + 2\varepsilon I\right) \\
&\quad - F\left(\hat{t}, \hat{y}, u\left(\hat{t}, \hat{x}\right), \alpha\left(\hat{x} - \hat{y}\right) - 2\varepsilon\hat{y}, Y - 2\varepsilon I\right)
\end{aligned}
$$

where we used (13) in the last estimate. If ε were absent (e.g. set $\varepsilon = 0$ throughout) we would estimate, with $R := |u|_\infty \vee |v|_\infty$,

$$
\begin{aligned}
&F\left(\hat{t}, \hat{x}, u\left(\hat{t}, \hat{x}\right), \alpha\left(\hat{x} - \hat{y}\right), X\right) - F\left(\hat{t}, \hat{y}, u\left(\hat{t}, \hat{x}\right), \alpha\left(\hat{x} - \hat{y}\right), Y\right) \\
&\quad \leq \theta_R\left(\alpha\left|\hat{x} - \hat{y}\right|^2 + \left|\hat{x} - \hat{y}\right|\right) =: B_\alpha
\end{aligned}
$$

and since $\alpha\left|\hat{x} - \hat{y}\right|^2 + \left|\hat{x} - \hat{y}\right| \leq 2\alpha\left|\hat{x} - \hat{y}\right|^2 + 1/\alpha \to 0$ as $\alpha \to \infty$, thanks to [10, Lemma 3.1], we see that $B_\alpha \to 0$ with $\alpha \to \infty$, which is enough to conclude. The present case where $\varepsilon > 0$ is essentially reduced to the case $\varepsilon = 0$ by adding/subtracting

$$
F\left(\hat{t}, \hat{x}, u\left(\hat{t}, \hat{x}\right), \alpha\left(\hat{x} - \hat{y}\right), X\right) - F\left(\hat{t}, \hat{y}, u\left(\hat{t}, \hat{x}\right), \alpha\left(\hat{x} - \hat{y}\right), Y\right),
$$

but we need some refined properties of $\left(\hat{t}, \hat{x}, \hat{y}\right)$ as collected in Lemma 6: (a) $p = \alpha\left(\hat{x} - \hat{y}\right)$ remains, for fixed α, bounded as $\varepsilon \to 0$, (b) $2\varepsilon\left|\hat{x}\right|$ and $2\varepsilon\left|\hat{y}\right|$ tend to zero as as $\varepsilon \to 0$ for fixed (large enough) α; this follows from the fact, that for α large enough we must have $\limsup_{\varepsilon\to 0} \varepsilon|\hat{x}|^2 = c_\alpha < \infty$ (after all, c_α tends to zero with $\alpha \to \infty$) and by rewriting $\limsup_{\varepsilon\to 0} \varepsilon|\hat{x}| \leq \sqrt{c_\alpha} \limsup_{\varepsilon\to 0} \sqrt{\varepsilon} = 0$, (c) that $\limsup_{\varepsilon\to 0}\left(\frac{\alpha}{2}\left|\hat{x} - \hat{y}\right|^2 + \left|\hat{x} - \hat{y}\right|\right) \to 0$ as $\alpha \to \infty$. We also note that (3) implies (d): any matrix norm of X, Y is bounded by a constant times α, independent of ε.

We can now return to the estimate of ϕ and clearly have

$$\phi\left(\hat{t},\hat{x},\hat{y}\right) \leq \frac{1}{\gamma}\left[(i)+(ii)+(ii)\right] =: B_{\alpha,\varepsilon}$$

where

$$\begin{aligned}
(i) &= \left|F\left(\hat{t},\hat{x},u\left(\hat{t},\hat{x}\right),\alpha\left(\hat{x}-\hat{y}\right)+2\varepsilon\hat{x},X+2\varepsilon I\right)-F\left(\hat{t},\hat{x},u\left(\hat{t},\hat{x}\right),\alpha\left(\hat{x}-\hat{y}\right),X\right)\right| \\
(ii) &= \left|F\left(\hat{t},\hat{y},u\left(\hat{t},\hat{x}\right),\alpha\left(\hat{x}-\hat{y}\right)-2\varepsilon\hat{y},Y-2\varepsilon I\right)-F\left(\hat{t},\hat{y},u\left(\hat{t},\hat{x}\right),\alpha\left(\hat{x}-\hat{y}\right),Y\right)\right| \\
(iii) &= \theta_R\left(\alpha\left|\hat{x}-\hat{y}\right|^2+\left|\hat{x}-\hat{y}\right|\right).
\end{aligned}$$

From (a), (d) above the gradient and Hessian argument in F as seen in (i) , (ii), i.e.

$$\alpha\left(\hat{x}-\hat{y}\right) \pm 2\varepsilon\hat{x} \text{ and } X+2\varepsilon I, Y-2\varepsilon I,$$

remain in a bounded set, for fixed α, uniformly as $\varepsilon \to 0$. From (b) above and the assumed uniform continuity properties of F (i.e. (4)), it then follows that for fixed (large enough) α

$$(i)\,,(ii) \to 0 \text{ as } \varepsilon \to 0.$$

On the other hand, continuity of θ_R at $0+$ together with (c) above shows that also $(iii) \to 0$ as $\varepsilon << \frac{1}{\alpha} \to 0$. We conclude that

$$B_{\alpha,\varepsilon} \to 0 \text{ as } \varepsilon << \frac{1}{\alpha} \to 0,$$

which implies (11), as desired. The proof is now finished. □

Lemma 5 *Assume* $u_0, v_0 \in \mathrm{BUC}\left(\mathbb{R}^n\right)$. *Then*

$$\sup_{x,y}\left[u_0(x)-v_0(y)-\frac{\alpha}{2}|x-y|^2-\varepsilon\left(|x|^2+|y|^2\right)\right]$$
$$\to \sup_x\left[u_0(x)-v_0(x)\right] \; as \; (\varepsilon,\frac{1}{\alpha}) \to (0,0)\,.$$

Proof (Similar to [2, Lemme 2.9] but we include full details for the reader's convenience.)

Without loss of generality $M := \sup_x\left[u_0(x)-v_0(x)\right] > 0$; for otherwise replace u_0 by $u_0+2|M|$. Write $M_{\alpha,\varepsilon}$ for the achieved maximum (at $\hat{x},\hat{y}$, say) of the left-hand-side. Obviously, $u_0(x)-v_0(x)-2\varepsilon|x|^2 \leq M_{\alpha,\varepsilon}$ for any x and so

$$M \leq \lim\inf_{\substack{\varepsilon\to 0\\ \alpha\to\infty}} M_{\alpha,\varepsilon}.$$

It follows that we can and will consider ε (α) small (large) enough so that $M_{\alpha,\varepsilon} > 0$. On the other hand, $|u_0|, |v_0| \leq R < \infty$ and so

$$0 \leq M_{\alpha,\varepsilon} \leq 2R - \frac{\alpha}{2}\left|\hat{x} - \hat{y}\right|^2 - \varepsilon\left(\left|\hat{x}\right|^2 + \left|\hat{y}\right|^2\right)$$

from which we deduce $\frac{\alpha}{2}\left|\hat{x} - \hat{y}\right|^2 \leq 2R$, or $\left|\hat{x} - \hat{y}\right| \leq \sqrt{4R/\alpha}$. By omitting the (positive) penalty terms, we can also estimate

$$\begin{aligned} M_{\alpha,\varepsilon} &\leq u_0\left(\hat{x}\right) - v_0\left(\hat{y}\right) \\ &\leq u_0\left(\hat{x}\right) - v_0\left(\hat{x}\right) + \sigma_{v_0}\left(\sqrt{4R/\alpha}\right) \\ &\leq M + \sigma_{v_0}\left(\sqrt{4R/\alpha}\right) \end{aligned}$$

where σ_{v_0} denotes the modulus of continuity of v_0. It follows that

$$\lim \sup_{\substack{\varepsilon\to 0 \\ \alpha\to\infty}} M_{\alpha,\varepsilon} \leq M$$

which shows that the $\lim M_{\alpha,\varepsilon}$ (as $\varepsilon \to 0$, $\alpha \to \infty$) exists and is equal to M. □

Lemma 6 *Let* $u \in \mathrm{BUSC}\left([0,T] \times \mathbb{R}^n\right)$ *and* $v \in \mathrm{BLSC}\left([0,T] \times \mathbb{R}^n\right)$. *Consider a maximum point* $\left(\hat{t}, \hat{x}, \hat{y}\right) \in (0,T] \times \mathbb{R}^n \times \mathbb{R}^n$ *of*

$$\phi\left(t,x,y\right) = u\left(t,x\right) - v\left(t,y\right) - \frac{\alpha}{2}\left|x-y\right|^2 - \varepsilon\left(\left|x\right|^2 + \left|y\right|^2\right).$$

where $\alpha, \varepsilon > 0$. *Then*

$$\limsup_{\varepsilon\to 0} \alpha\left(\hat{x} - \hat{y}\right) = C\left(\alpha\right) < \infty, \tag{14}$$

$$\limsup_{\alpha\to\infty} \limsup_{\varepsilon\to 0} \varepsilon\left(\left|\hat{x}\right|^2 + \left|\hat{y}\right|^2\right) = 0, \tag{15}$$

$$\limsup_{\alpha\to\infty} \limsup_{\varepsilon\to 0} \left(\frac{\alpha}{2}\left|\hat{x} - \hat{y}\right|^2 + \left|\hat{x} - \hat{y}\right|\right) = 0. \tag{16}$$

The order in which limits are taken is important and suggests the notation

$$\limsup_{\varepsilon << \frac{1}{\alpha} \to 0} (\ldots) := \limsup_{\alpha\to\infty} \limsup_{\varepsilon\to 0} (\ldots), \quad \liminf_{\varepsilon << \frac{1}{\alpha} \to 0} (\ldots) := \liminf_{\alpha\to\infty} \liminf_{\varepsilon\to 0} (\ldots).$$

Proof A similar Lemma (without t dependence) is found in Barles' book [2, Lemme 4.3]. Again, we include details for the reader's convenience.

We start with some notation, where unless otherwise stated $t \in [0, T]$ and $x, y \in \mathbb{R}^n$,

$$
\begin{aligned}
M_{\alpha,\varepsilon} &:= \sup_{t,x,y} \phi(t, x, y) = u\left(\hat{t}, \hat{x}\right) - v\left(\hat{t}, \hat{y}\right) - \frac{\alpha}{2}\left|\hat{x} - \hat{y}\right|^2 - \varepsilon\left(\left|\hat{x}\right|^2 + \left|\hat{y}\right|^2\right); \\
M(h) &:= \sup_{t,x,y:|x-y|\le h} \left[u(t, x) - v(t, y)\right] \ge \sup_{t,x}\left[u(t, x) - v(t, x)\right] \\
M' &:= \downarrow \lim_{h\to 0} M(h),
\end{aligned}
$$

where M' exists as limit of $M(h)$, non-increasing in h and bounded from below.

Step 1: Take $t = x = y = 0$ as argument of $\phi(t, x, y)$. Since $M_{\alpha,\varepsilon} = \sup\phi$ we have

$$
c = u(0, 0) - v(0, 0) \le M_{\alpha,\varepsilon} = u\left(\hat{t}, \hat{x}\right) - v\left(\hat{t}, \hat{y}\right) - \frac{\alpha}{2}\left|\hat{x} - \hat{y}\right|^2 - \varepsilon\left(\left|\hat{x}\right|^2 + \left|\hat{y}\right|^2\right)
$$

and hence, for a suitable constant C (e.g. $C^2 := \sup u + \sup(-v) + c$)

$$
\frac{\alpha}{2}\left|\hat{x} - \hat{y}\right|^2 + \varepsilon\left(\left|\hat{x}\right|^2 + \left|\hat{y}\right|^2\right) \le C^2
$$

which implies

$$
\left|\hat{x} - \hat{y}\right| \le C\sqrt{2/\alpha} \tag{17}
$$

and hence $\alpha\left|\hat{x} - \hat{y}\right| \le \sqrt{2\alpha}C$ which is the first claimed estimate (14).

Step 2: We first argue that it is enough to show the two estimates

$$
\limsup_{\varepsilon << \frac{1}{\alpha} \to 0} \left[u\left(\hat{t}, \hat{x}\right) - v\left(\hat{t}, \hat{y}\right)\right] \le M' \le \liminf_{\varepsilon << \frac{1}{\alpha} \to 0} M_{\alpha,\varepsilon}. \tag{18}
$$

Indeed, from $\frac{\alpha}{2}\left|\hat{x} - \hat{y}\right|^2 + \varepsilon\left(\left|\hat{x}\right|^2 + \left|\hat{y}\right|^2\right) = u\left(\hat{t}, \hat{x}\right) - v\left(\hat{t}, \hat{y}\right) - M_{\alpha,\varepsilon}$ it readily follows that

$$
\begin{aligned}
&\limsup_{\varepsilon << \frac{1}{\alpha} \to 0} \frac{\alpha}{2}\left|\hat{x} - \hat{y}\right|^2 + \varepsilon\left(\left|\hat{x}\right|^2 + \left|\hat{y}\right|^2\right) \\
&\quad \le \limsup_{\varepsilon << \frac{1}{\alpha} \to 0} \left[u\left(\hat{t}, \hat{x}\right) - v\left(\hat{t}, \hat{y}\right) - M_{\alpha,\varepsilon}\right] \\
&\quad = \limsup_{\varepsilon << \frac{1}{\alpha} \to 0} \left[u\left(\hat{t}, \hat{x}\right) - v\left(\hat{t}, \hat{y}\right)\right] - \liminf_{\varepsilon << \frac{1}{\alpha} \to 0} M_{\alpha,\varepsilon} \\
&\quad \le 0 \text{ (and hence } = 0).
\end{aligned}
$$

This, together with (17), already gives (15) and also (16).

We are left to show (18). For the first estimate, it suffices to note that, from (17) and the definition of $M(h)$ applied with $h = C\sqrt{2/\alpha}$,

$$\limsup_{\varepsilon << \frac{1}{\alpha} \to 0} \left[u\left(\hat{t}, \hat{x}\right) - v\left(\hat{t}, \hat{y}\right)\right] \leq \limsup_{\varepsilon << \frac{1}{\alpha} \to 0} M\left(\sqrt{\frac{2}{\alpha}} C\right)$$
$$= \lim_{\alpha \to \infty} M\left(\sqrt{\frac{2}{\alpha}} C\right) = M'.$$

We now turn to the second estimate in (18). From the very definition of M' as $\lim_{h \to 0} M(h)$, there exists a family (t_h, x_h, y_h) so that

$$|x_h - y_h| \leq h \text{ and } u(t_h, x_h) - v(t_h, x_h) \to M' \text{ as } h \to 0 \tag{19}$$

For every α, ε we may take (t_h, x_h, y_h) as argument of ϕ; since $M_{\alpha,\varepsilon} = \sup \phi$ we have

$$u(t_h, x_h) - v(t_h, y_h) - \frac{\alpha}{2} h^2 - \varepsilon(|x_h|^2 + |y_h|^2) \leq M_{\alpha,\varepsilon} \tag{20}$$

Take now $\varepsilon = \varepsilon(h) \to 0$ with $h \to 0$; fast enough so that $\varepsilon(|x_h|^2 + |y_h|^2) \to 0$; for instance $\varepsilon(h) := h / \left(1 + (|x_h|^2 + |y_h|^2)\right)$ would do. It follows that

$$\begin{aligned} M' &= \lim_{h \to 0} u(t_h, x_h) - v(t_h, y_h) \\ &= \liminf_{h \to 0} u(t_h, x_h) - v(t_h, y_h) - \frac{\alpha}{2} h^2 - \varepsilon(|x_h|^2 + |y_h|^2) \\ &\leq \liminf_{h \to 0} M_{\alpha,\varepsilon_h} = \liminf_{\varepsilon \to 0} M_{\alpha,\varepsilon} \text{ by monotonicity of } M_{\alpha,\varepsilon} \text{ in } \varepsilon. \end{aligned}$$

Since this is valid for every α, we also have

$$M' \leq \liminf_{\alpha \to \infty} \liminf_{\varepsilon \to 0} M_{\alpha,\varepsilon}.$$

This is precisely the second estimate in (18) and so the proof is finished. □

2.6 Proof of Theorem 2: Lipschitz Regularity in Space

Remark 3 The argument will be similar to the regularity proof in [20]. Like there, it can be adapted to show Hölder continuity in x for Hölder initial data and a corresponding condition on the modulus θ_R.

Proof We modify the proof of Theorem 1 (ii) by adding a specific t-dependence to the test function. This idea can for example also be found in Ref. [20]. Since the proof is similar to the proof of Theorem 1 (ii) we will omit redundant arguments.

Let $\left(\hat{t},\hat{x},\hat{y}\right)$ be a maximum point of

$$\phi\left(t,x,y\right):=u\left(t,x\right)-u\left(t,y\right)-\frac{\alpha}{2}e^{\bar{\gamma}t}\left|x-y\right|^{2}-\varepsilon\left(\left|x\right|^{2}+\left|y\right|^{2}\right) \tag{21}$$

over $[0,T]\times\mathbb{R}^{n}\times\mathbb{R}^{n}$ where $\alpha>0$ and $\varepsilon>0$. The plan is to show the estimate

$$u\left(t,x\right)-u\left(t,y\right)\leq\inf_{\alpha}\left[e^{\bar{\gamma}t}\frac{\alpha}{2}\left|x-y\right|^{2}+(2L_{u0}^{2}+\frac{\theta_{R}^{2}}{\gamma})\frac{1}{\alpha}.\right], \tag{22}$$

valid on $[0,T]\times\mathbb{R}^{n}\times\mathbb{R}^{n}$. It then immediately follows, that

$$u\left(t,x\right)-u\left(t,y\right)\leq(e^{\bar{\gamma}t}\frac{1}{2}+2L_{u0}^{2}+\frac{\theta_{R}^{2}}{\gamma})|x-y|.$$

Thanks to the very definition of $\left(\hat{t},\hat{x},\hat{y}\right)$ as arg max of $\phi\left(t,x,y\right)=u\left(t,x\right)-u\left(t,y\right)-\frac{\alpha}{2}\left|x-y\right|^{2}-\varepsilon\left(\left|x\right|^{2}+\left|y\right|^{2}\right)$, we obtain the estimate

$$u\left(t,x\right)-u\left(t,y\right)\leq\frac{\alpha}{2}e^{\bar{\gamma}t}\left|x-y\right|^{2}+\varepsilon\left(\left|x\right|^{2}+\left|y\right|^{2}\right)+\phi\left(\hat{t},\hat{x},\hat{y}\right).$$

Note that $\left(\hat{t},\hat{x},\hat{y}\right)$ depends on α,ε. We shall consider the cases $\hat{t}=0$ and $\hat{t}\in(0,T]$ separately. In the first case $\hat{t}=0$ we have

$$\phi\left(0,\hat{x},\hat{y}\right)=\sup_{x,y}\left[u_{0}\left(x\right)-u_{0}\left(y\right)-\frac{\alpha}{2}\left|x-y\right|^{2}-\varepsilon\left(\left|x\right|^{2}+\left|y\right|^{2}\right)\right]=:A_{\alpha,\varepsilon}$$

and Lemma 7 below asserts that

$$\limsup_{\varepsilon\to0}A_{\alpha,\varepsilon}\leq2L_{u0}^{2}\frac{1}{\alpha}. \tag{23}$$

The second case is $\hat{t}\in(0,T]$ and we will show

$$\phi\left(\hat{t},\hat{x},\hat{y}\right)\leq B_{\alpha,\varepsilon}\text{ where }\left(\limsup_{\varepsilon\to0}B_{\alpha,\varepsilon}\right)\leq\frac{\theta_{R}^{2}}{\gamma}\frac{1}{\alpha}; \tag{24}$$

it is here that we will use Theorem of Sums and viscosity properties. From (23) and (24) we can then deduce (22) just as in the proof of Theorem (1) (ii).

It remains to prove the estimate (24). To this end, as before, we can deduce the existence of

$$\left(a,\alpha e^{\bar{\gamma}\hat{t}}\left(\hat{x}-\hat{y}\right)+2\varepsilon\hat{x},X+2\varepsilon I\right)\in\bar{\mathcal{P}}_{Q}^{2,+}u\left(\hat{t},\hat{x}\right),$$
$$\left(b,\alpha e^{\bar{\gamma}\hat{t}}\left(\hat{x}-\hat{y}\right)-2\varepsilon\hat{y},Y-2\varepsilon I\right)\in\bar{\mathcal{P}}_{Q}^{2,-}u\left(\hat{t},\hat{y}\right),$$

with the difference, that in this case $a-b \geq \frac{\alpha}{2}\bar{\gamma}e^{\bar{\gamma}\hat{t}}\left|\hat{x}-\hat{y}\right|^2$ (since the time derivative of the test function does not vanish) and such that one has the two-sided matrix estimate

$$-3\alpha e^{\bar{\gamma}\hat{t}}\begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix} \leq \begin{pmatrix} X & 0 \\ 0 & -Y \end{pmatrix} \leq 3\alpha e^{\bar{\gamma}\hat{t}}\begin{pmatrix} I & -I \\ -I & I \end{pmatrix}. \tag{25}$$

Using the viscosity sub- and super-solution properties we then see that

$$\begin{aligned} &F\left(\hat{t},\hat{y},u\left(\hat{t},\hat{y}\right),\alpha e^{\bar{\gamma}\hat{t}}\left(\hat{x}-\hat{y}\right)-2\varepsilon\hat{y},Y-2\varepsilon I\right) \\ &\quad -F\left(\hat{t},\hat{x},u\left(\hat{t},\hat{x}\right),\alpha e^{\bar{\gamma}\hat{t}}\left(\hat{x}-\hat{y}\right)+2\varepsilon\hat{x},X+2\varepsilon I\right) \leq -\frac{\alpha}{2}\bar{\gamma}e^{\bar{\gamma}\hat{t}}\left|x-y\right|^2. \end{aligned} \tag{26}$$

Trivially, (recall it is enough to consider the case $u\left(\hat{t},\hat{x}\right) \geq u\left(\hat{t},\hat{y}\right)$)

$$\begin{aligned} \gamma\phi\left(\hat{t},\hat{x},\hat{y}\right) &\leq \gamma\left(u\left(\hat{t},\hat{x}\right)-u\left(\hat{t},\hat{y}\right)\right) \\ &\leq F\left(\hat{t},\hat{y},u\left(\hat{t},\hat{y}\right),\alpha e^{\bar{\gamma}\hat{t}}\left(\hat{x}-\hat{y}\right)-2\varepsilon\hat{y},Y-2\varepsilon I\right) \\ &\quad -F\left(\hat{t},\hat{y},u\left(\hat{t},\hat{x}\right),\alpha e^{\bar{\gamma}\hat{t}}\left(\hat{x}-\hat{y}\right)-2\varepsilon\hat{y},Y-2\varepsilon I\right) \\ &\leq F\left(\hat{t},\hat{x},u\left(\hat{t},\hat{x}\right),\alpha e^{\bar{\gamma}\hat{t}}\left(\hat{x}-\hat{y}\right)+2\varepsilon\hat{x},X+2\varepsilon I\right) \\ &\quad -F\left(\hat{t},\hat{y},u\left(\hat{t},\hat{x}\right),\alpha e^{\bar{\gamma}\hat{t}}\left(\hat{x}-\hat{y}\right)-2\varepsilon\hat{y},Y-2\varepsilon I\right)-\frac{\alpha}{2}\bar{\gamma}e^{\bar{\gamma}\hat{t}}\left|\hat{x}-\hat{y}\right|^2 \end{aligned}$$

where we used (26) in the last estimate. So

$$\phi\left(\hat{t},\hat{x},\hat{y}\right) \leq \frac{1}{\gamma}\left[(i)+(ii)+(iii)\right] =: B_{\alpha,\varepsilon}$$

where (remember the choice of $\bar{\gamma} = 2(\theta_R+1)$)

$$\begin{aligned} (i) &= \left|F\left(\hat{t},\hat{x},u\left(\hat{t},\hat{x}\right),\alpha\left(\hat{x}-\hat{y}\right)+2\varepsilon\hat{x},X+2\varepsilon I\right)-F\left(\hat{t},\hat{x},u\left(\hat{t},\hat{x}\right),\alpha\left(\hat{x}-\hat{y}\right),X\right)\right| \\ (ii) &= \left|F\left(\hat{t},\hat{y},u\left(\hat{t},\hat{x}\right),\alpha\left(\hat{x}-\hat{y}\right)-2\varepsilon\hat{y},Y-2\varepsilon I\right)-F\left(\hat{t},\hat{y},u\left(\hat{t},\hat{x}\right),\alpha\left(\hat{x}-\hat{y}\right),Y\right)\right| \\ (iii) &= \theta_R(\alpha\left|\hat{x}-\hat{y}\right|^2+\left|\hat{x}-\hat{y}\right|)-\frac{\alpha}{2}\bar{\gamma}e^{\bar{\gamma}\hat{t}}\left|\hat{x}-\hat{y}\right|^2 \\ &= \theta_R\left|\hat{x}-\hat{y}\right|+\alpha\left(\theta_R-\frac{1}{2}\bar{\gamma}e^{\bar{\gamma}\hat{t}}\right)\left|\hat{x}-\hat{y}\right|^2 \\ &\leq \theta_R\left|\hat{x}-\hat{y}\right|-\alpha\left|\hat{x}-\hat{y}\right|^2 \\ &\leq \theta_R^2\frac{1}{\alpha}. \end{aligned}$$

The last inequality follows from $\sup_r\left[c_1r-c_2r^2\right] \leq c_1^2c_2^{-1}$, for $c_1, c_2 > 0$. As before we deduce for fixed (large enough) α that (i), $(ii) \to 0$ as $\varepsilon \to 0$. Hence

$$\limsup_{\varepsilon\to 0} B_{\alpha,\varepsilon} \leq \frac{\theta_R^2}{\gamma}\frac{1}{\alpha}.$$

□

Lemma 7 *Assume u_0 is bounded and Lipschitz with Lipschitz constant L_{u_0}. Then*

$$\limsup_{\varepsilon\to 0}\sup_{x,y}\left[u_0(x)-u_0(y)-\frac{\alpha}{2}|x-y|^2-\varepsilon\left(|x|^2+|y|^2\right)\right]\le 2L_{u_0}^2\frac{1}{\alpha}.$$

Proof Write $M_{\alpha,\varepsilon}$ for the achieved maximum (at $\hat{x}$, $\hat{y}$, say) of the left-hand-side (for fixed α and ε). By comparing to evaluation at $x=y=0$ we have

$$0\le u_0(\hat{x})-u_0(\hat{y})-\frac{\alpha}{2}\left|\hat{x}-\hat{y}\right|^2-\varepsilon\left(\left|\hat{x}\right|^2+\left|\hat{y}\right|^2\right)$$

from which we deduce $\frac{\alpha}{2}\left|\hat{x}-\hat{y}\right|^2-L_{u_0}\left|\hat{x}-\hat{y}\right|\le 0$. Hence $\left|\hat{x}-\hat{y}\right|\le 2\frac{L_{u_0}}{\alpha}$. By omitting the (positive) penalty terms, we can also estimate

$$M_{\alpha,\varepsilon}\le u_0\left(\hat{x}\right)-u_0\left(\hat{y}\right)\le L_{u_0}\left|\hat{x}-\hat{y}\right|\le 2\frac{L_{u_0}^2}{\alpha}.$$

It follows that $\limsup_{\varepsilon\to 0}M_{\alpha,\varepsilon}\le 2\frac{L_{u_0}^2}{\alpha}$. □

2.7 Proof of Theorem 3: Regularity in Time-Space

Remark 4 The linear growth condition in (ii) is especially satisfied for linear problems. Even in this case the Hölder regularity cannot be improved, as maybe seen in the standard heat equation started with initial data $u_0(x)=\max\{|x|,1\}$.

Proof (i): We adapt the argument from [4, Lemma 9.1]; see also [3]. From Theorem 1, there exists a spatial modulus m for $u(t,\cdot)$, uniform over $t\in[0,T]$. Given $0\le t_0<t\le T$ and $x_0,x\in\mathbb{R}^n$ we now estimate, using the triangle inequality,

$$|u(t,x)-u(t_0,x_0)|\le m(|x_0-x|)+|u(t,x_0)-u(t_0,x_0)|.$$

We shall show that $|u(t,x_0)-u(t_0,x_0)|$ goes to zero as $t\downarrow t_0$, *uniformly in $x_0\in\mathbb{R}^n$ and $t_0\in[0,T)$*. We will show a little more. Fix $x_0\in\mathbb{R}^n$ and $R\in(0,\infty)$; for instance $R=1$ would do (and there is no need to track dependence in R). We claim that for every $\eta>0$ one can find constants $C=C(\eta)$, $K=K(\eta)$, *not dependent on x_0 and* t_0, such that, for all $x\in B_{R/2}(x_0)$ and $y\in B_R(x_0)$ and all $t\in[t_0,T]$

$$u(t,y)-u(t_0,x)\le\eta+C|y-x|^2+K(t-t_0)\tag{27}$$

and

$$u(t,y)-u(t_0,x)\ge-\eta-C|y-x|^2-K(t-t_0).\tag{28}$$

(Choosing $x = y = x_0$ in these estimates shows that $|u(t, x_0) - u(t_0, x_0)| \leq \inf\{\eta + K(\eta)(t - t_0) : \eta > 0\}$ which immediately gives the desired uniform continuity in time, uniformly in x_0.) We only prove (27), (28) being proved in an analogous way. In the sequel, x is fixed in $B_{R/2}(x_0)$. Rewrite (27) as

$$u - \chi \leq 0 \text{ on } [t_0, T] \times B_R(x_0)$$

where $\chi(t, y) := u(t_0, x) + \eta + C|y - x|^2 + K(t - t_0)$. We shall see below we can find C, the choice of which only depends on η (and in a harmless way on $|u|_{\infty;[0,T]\times\mathbb{R}^n}$, R and $m(\cdot)$ *but not on K and not on* x_0, t_0), such that $u - \chi \leq 0$ on the parabolic boundary of $[t_0, T] \times B_R(x_0)$. The extension to the interior is then based on the maximum principle. More precisely, we can chose K depending on η (and again in a harmless way $|u|_{\infty;[0,T]\times\mathbb{R}^n}$, R and $m(\cdot)$) such that χ is a (smooth) strict supersolution of $\partial_t - F$ on $(t_0, T) \times B_R(x_0)$;

$$K - F(t, y, \chi(t, y), 2C(y - x), 2CI) > 0 \text{ on } (t_0, T) \times B_R(x_0).$$

Indeed, by properness we have

$$K - F(t, y, \chi(t, y), 2C(y - x), 2CI) > K - F(t, y, -|u|_\infty, 2C(y - x), 2CI);$$

noting $|y - x| \leq 2R$ so that $p := 2C(y - x), X := 2CI$ remain in a bounded set whose size may depend on η through C, it then follows by our structural assumption on the non-linearity [7] that we can pick $K = K(\eta)$ large enough such as to achieve the claimed strict inequality. (Note that this choice of K is uniformly in t_0 provided we can find C with the correct dependences.) Since, on the other hand, u is a viscosity solution (hence subsolution), it follows from the very definition of a subsolution that

$$K - F\left(\hat{t}, \hat{y}, \chi(t, y), 2C\left(\hat{y} - x\right), 2CI\right) \leq 0$$

whenever $\left(\hat{t}, \hat{y}\right) \in (t_0, T] \times B_R(x_0)$ is a maximum point of $u - \chi$. (Note that $\hat{t} = T$ is possible here, we then rely on part (i) of Theorem 1.) This contradiction shows that maximum points of $u - \chi$ over $[t_0, T] \times \bar{B}_R(x_0)$ are necessarily achieved on the parabolic boundary

$$(t, y) \in [t_0, T] \times \partial B_R(x_0) \cup \{t_0\} \times \bar{B}_R(x_0).$$

The remainder of the proof is thus concerned with showing that $u - \chi \leq 0$ on this parabolic boundary. Consider first the case that $t \in [t_0, T]$ and $|y - x_0| = R$. Since $x \in B_{R/2}(x_0)$ we must have $|y - x| \geq R/2$ and it thus suffices to take $C \geq 8|u|_{\infty;[0,T]\times\mathbb{R}^n}/R^2$ to ensure that

$$u(t_0, y) \leq u(t_0, x) + \eta + C|y - x|^2 + K(t - t_0) \tag{29}$$

[7] ... notably boundedness of $F(\cdot, \cdot, y, p, X)$ when y, p, X remain in a bounded set ...

for all $t \in [t_0, T]$ and $y \in B_R(x_0)$, and any $\eta, K \geq 0$. The second case to be considered is $t = t_0$ and $y \in \bar{B}_R(x_0)$. We want to see that for every η there exists $C = C(\eta)$ such that

$$u(t_0, y) \leq u(t_0, x) + \eta + C|y-x|^2 \text{ for all } y \in \bar{B}_R(x_0); \tag{30}$$

but this follows immediately from the fact (cf. Theorem 1) that $u(t_0, \cdot)$ has a spatial modulus m. Indeed: If there were $\eta > 0$ such that for all C there are points y_C so that $u(t_0, y) > u(t_0, x) + \eta + C|y-x|^2$, then $|y_C - x|^2 \leq 2|u|_{\infty;[0,T]\times\mathbb{R}^n}/C \to 0$ with $C \to \infty$ and a contradiction to

$$m(|y_C - x|) \geq u(t_0, y) - u(t_0, x) \geq \eta > 0.$$

is obtained as soon as C is chosen large enough and this choice depends only on η, $|u|_{\infty;[0,T]\times\mathbb{R}^n}$ and m. Since all these quantities are independent of t_0, so is our choice of C.

(ii): By Theorem 2 we get for every $t \in [0, T]$ that $u(t, \cdot)$ is Lipschitz continuous with the same Lipschitz constant, say L.

For the regularity in time, let η be given. Choose $C = \frac{L^2}{4\eta}$, then (30) holds. If we then fix R large enough, (29) is also fulfilled.

We can now choose K explicitly. Indeed, with $K \geq M(1+2C)$, χ is a supersolution. Using this in (27), we get

$$u(t,x) - u(t_0,x) \leq \eta + M\left(1 + \frac{L^2}{4\eta}\right)t,$$

for all $t \in [t_0, T]$. Optimization with respect to η leads to

$$u(t,x) - u(t_0,x) \leq \tilde{L}t^{1/2},$$

where $\tilde{L}$ depends only on M and L. □

2.8 *Proof of Theorem 4: Existence*

At last, we discuss existence via Perron's Method; the only difficulty in the proof is to produce subsolutions and supersolutions.

Proof Step 1: Assume u_0 is Lipschitz continuous with Lipschitz constant L. Define for $z \in \mathbb{R}^n$, $\varepsilon > 0$

$$\psi_{\varepsilon,z}(x) := u_0(z) - L\left(|x-z|^2 + \varepsilon\right)^{1/2}.$$

We will show that there exists $A_\varepsilon \leq 0$ (non-positive, yet to be chosen) such that

$$u_{\varepsilon,z}(t,x) := A_\varepsilon t + \psi_{\varepsilon,z}(x)$$

is a (classical) subsolution of $\partial_t - F = 0$. To this end we first note that $Du_{\varepsilon,z} = D\psi_{\varepsilon,z}$ and $D^2 u_{\varepsilon,z} = D^2\psi_{\varepsilon,z}$ are bounded by LC_ε where C is a constant dependent on ε. We also note that (for any non-positive choice of A_ε)

$$u_{\varepsilon,z}(t,x) \le u_{\varepsilon,z}(0,x) = \psi_{\varepsilon,z}(x) \le u_0(z) - L|x-z| \le u_0(x),$$

thanks to L-Lipschitzness of u_0. Since $F = F(t,x,u,p,X)$ is assumed to be proper, and thus in particular anti-monotone in u, we have

$$\begin{aligned} &\partial_t u_{\varepsilon,z} - F\left(t,x,u_{\varepsilon,z},Du_{\varepsilon,z},D^2u_{\varepsilon,z}\right) \\ &= A_\varepsilon - F\left(t,x,u_{\varepsilon,z},D\psi_{\varepsilon,z},D^2\psi_{\varepsilon,z}\right) \\ &\le A_\varepsilon - F\left(t,x,|u_0|_\infty,D\psi_{\varepsilon,z},D^2\psi_{\varepsilon,z}\right). \end{aligned}$$

Since $|u_0|_\infty < \infty$ and $\left|D\psi_{\varepsilon,z}\right|, \left|D^2\psi_{\varepsilon,z}\right| \le LC_\varepsilon$ we can use the assumed boundedness of F over sets where u, p, X remain bounded. In particular, we can pick A_ε negative, large enough, such that

$$\partial_t u_{\varepsilon,z} - F\left(t,x,u_{\varepsilon,z},Du_{\varepsilon,z},D^2u_{\varepsilon,z}\right) \le \cdots \le 0.$$

We now define the sup of all these subsolutions,

$$\hat{u}(t,x) := \sup_{\varepsilon\in(0,1],z\in\mathbb{R}^n} u_{\varepsilon,z}(t,x) \le u_0(x) \le |u_0|_\infty < \infty.$$

Then

$$-\infty < A_1T - L \le A_1t - L = u_{1,x}(t,x) \le \hat{u}(t,x) \le u_0(x) \le |u_0|_\infty < \infty,$$

so that $\hat{u}$ is bounded. We moreover have

$$\hat{u}(0,x) = \sup_{\varepsilon\in(0,1],z\in\mathbb{R}^n} \psi_{\varepsilon,z}(x) = \sup_{\varepsilon\in(0,1]} u_0(x) - L\varepsilon^{1/2} = u_0(x).$$

The upper semicontinuous envelope $\underline{u}(t,x) := \hat{u}^*$ is then (cf. Proposition 8.2 in [8] for instance) also a bounded subsolution to $\partial_t - F = 0$.

Step 2: We show that $\hat{u}(t,x)$ is continous at $t=0$; this implies that

$$\underline{u}(0,x) := \hat{u}(0,x) = u_0(x)$$

and thus yields a sub-solution with the correct initial data. Let $(t^n,x^n) \to (0,x)$. First we show lower semicontinuity, i.e.

$$\liminf_{n\to\infty} \hat{u}(t^n,x^n) \ge \hat{u}(0,x).$$

Let $\delta > 0$. Choose $\tilde{\varepsilon}, \tilde{z}$ such that

$$u_{\tilde{\varepsilon},\tilde{z}}(0, x) \geq \hat{u}(0, x) - \delta.$$

Let M be a bound for $|Du_{\tilde{\varepsilon},\tilde{z}}|$ (and hence for $|D\psi_{\tilde{\varepsilon},\tilde{z}}|$). Choose N such that for $n \geq N$

$$|t^n|, |x^n - x| \leq \min\left\{\frac{\delta}{A_{\tilde{\varepsilon}}}, \frac{\delta}{M}\right\}.$$

Then

$$\begin{aligned}
\hat{u}(t^n, x^n) &\geq u_{\tilde{\varepsilon},\tilde{z}}(t^n, x^n) \\
&= u_{\tilde{\varepsilon},\tilde{z}}(t^n, x^n) - u_{\tilde{\varepsilon},\tilde{z}}(0, x) + u_{\tilde{\varepsilon},\tilde{z}}(0, x) \\
&= A_{\tilde{\varepsilon}} t^n + \psi_{\tilde{\varepsilon},\tilde{z}}(x^n) - \psi_{\tilde{\varepsilon},\tilde{z}}(x) + u_{\tilde{\varepsilon},\tilde{z}}(0, x) \\
&\geq \hat{u}(0, x) - 3\delta,
\end{aligned}$$

which proves the lower semicontinuity.

For upper semicontinuity, notice that

$$\begin{aligned}
u_{\varepsilon,z}(s, y) &= A_\varepsilon s + \psi_{\varepsilon,z}(y) \\
&\leq A_\varepsilon s + u_0(y) \\
&\leq u_0(y),
\end{aligned}$$

where we have used that $A_\varepsilon \leq 0$ and that $\psi_{\varepsilon,z}(y) \leq u_0(y)$, as shown above. Hence, $\hat{u}(s, y) \leq u_0(y)$, and then for $(t^n, x^n) \to (0, x)$, we have

$$\limsup_n \hat{u}(t^n, x^n) \leq \limsup_n u_0(x^n) = u_0(x) = \hat{u}(0, x).$$

Hence $\hat{u}$ is also upper semicontinuous at $(0, x)$ and hence continuous at $(0, x)$.

Step 3: Similarly, one constructs a bounded super-solution with correct (bounded, Lipschitz) initial data u_0. Perron's method then applies and yields a bounded viscosity solution to $\partial_t - F = 0$ with bounded, Lipschitz initial data.

Step 4: Let now $u_0 \in \mathrm{BUC}(\mathbb{R}^n)$ and u_0^n be a sequence of bounded Lipschitz functions such that $|u_0^n - u_0|_\infty \to 0$. By the previous step there exists a bounded solution u^n to $\partial_t - F = 0$ with initial data $u^n(0, \cdot) = u_0^n$. (It is also unique by comparison.) Since F is proper ($\gamma \geq 0$), the solutions form a contraction in the sense

$$|u^n - u^m|_{\infty;[0,T]\times\mathbb{R}^n} \leq |u_0^n - u_0^m|_{\infty;\mathbb{R}^n}$$

(This follows immediately from comparison and properness.). Hence u^n is Cauchy in supremum norm and converges to a continuous bounded function $u : [0, T] \times \mathbb{R}^n \to \mathbb{R}$. By Lemma 6.1 in the User's Guide we then have that u is a bounded solution to $\partial_t - F = 0$ with BUC($\mathbb{R}^n$) initial data. By comparison, it is the unique (bounded) solution with this initial data. At last, Corollary 3 shows that the solution is BUC in time space. $\square$

2.9 Parabolic Theorem of Sums Revisited

We start some recalls on parabolic jets. If $u : (0, T) \times \mathbb{R}^n \to \mathbb{R}$ its parabolic semijet $\mathcal{P}^{2,+}u$ is defined by $(b, p, X) \in \mathbb{R} \times \mathbb{R}^n \times \mathcal{S}^n$ lies in $\mathcal{P}^{2,+}u(s, z)$ if $(s, z) \in (0, T) \times \mathbb{R}^n$ and

$$u(t, x) \leq u(s, z) + b(t - s) + \langle p, x - z\rangle + \frac{1}{2}\langle X(x - z), x - z\rangle + o\left(|t - s| + |x - z|^2\right)$$

as $(0, T) \times \mathbb{R}^n \ni (t, x) \to (s, z)$. Consider now $u : Q \to \mathbb{R}$ where $Q = (0, T] \times \mathbb{R}^n$. The parabolic semijet relative to Q, write $\mathcal{P}_Q^{2,+}u$, as used in Ref. [19] for instance, is defined by $(b, p, X) \in \mathbb{R} \times \mathbb{R}^n \times \mathcal{S}^n$ lies in $\mathcal{P}_Q^{2,+}u(s, z)$ if $(s, z) \in (0, T) \times \mathbb{R}^n$ and

$$u(t, x) \leq u(s, z) + b(t - s) + \langle p, x - z\rangle + \frac{1}{2}\langle X(x - z), x - z\rangle + o\left(|t - s| + |x - z|^2\right)$$

as $Q \ni (t, x) \to (s, z)$. Note that $\mathcal{P}_Q^{2,+}u(s, z) = \mathcal{P}^{2,+}u(s, z)$ for $(s, z) \in (0, T) \times \mathbb{R}^n$. Note also the special behaviour of the semijet at time T in the sense that

$$(b, p, X) \in \mathcal{P}_Q^{2,+}u(T, z) \implies \forall b' \leq b : \left(b', p, X\right) \in \mathcal{P}_Q^{2,+}u(T, z). \tag{31}$$

Closures of these jets are defined in the usual way; e.g.

$$(b, p, X) \in \bar{\mathcal{P}}_Q^{2,+}u(T, z)$$

iff $\exists\, (t_n, z_n; b_n, p_n, X_n) \in Q \times \mathbb{R} \times \mathbb{R}^n \times \mathcal{S}^n : (b_n, p_n, X_n) \in \bar{\mathcal{P}}_Q^{2,+}u(t_n, z_n)$ and

$$(t_n, z_n; u(t_n, z_n); b_n, p_n, X_n) \to (T, z; u(T, z); b, p, X).$$

Recall the classical parabolic Theorem of Sums.

Theorem 8 [9, Theorem 7] *Let* $u_1, u_2 \in \mathrm{USC}((0, T) \times \mathbb{R}^n)$ *and* $w \in \mathrm{USC}\left((0, T) \times \mathbb{R}^{2n}\right)$ *be given by*

$$w(t, x) = u_1(t, x_1) + u_2(t, x_2)$$

Suppose that $s \in (0,T)$, $z = (z_1, z_2) \in \mathbb{R}^{2n}$, $b \in \mathbb{R}$, $p = (p_1, p_2) \in \mathbb{R}^{2n}$, $A \in \mathcal{S}^{2n}$ *with*

$$(b, p, A) \in \mathcal{P}^{2,+} w\,(s,z)\,. \tag{32}$$

Assume moreover that there is an $r > 0$ *such that for every* $M > 0$ *there is a* C *such that for* $i = 1, 2$

$$b_i \le C \text{ whenever } (b_i, q_i, X_i) \in \mathcal{P}^{2,+} w\,(t, x_i)\,, \tag{33}$$

$$|x_i - z_i| + |s-t| < r \text{ and } |u_i\,(t,x_i)| + |q_i| + \|X_i\| \le M.$$

Then for each $\varepsilon > 0$ *there exists* $(b_i,\ X_i) \in \mathbb{R} \times \mathcal{S}^n$ *such that*

$$(b_i, p_i, X_i) \in \bar{\mathcal{P}}^{2,+} u_i\,(s, z_i)$$

and

$$-\left(\frac{1}{\varepsilon} + \|A\|\right) I \le \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix} \le A + \varepsilon A^2 \text{ and } b_1 + b_2 = b. \tag{34}$$

The proof of the above Theorem is usually reduced (cf. Lemma 8 in [9]) to the case $b = 0, z = 0, p = 0$ and $v_1\,(s,0) = v_2\,(s,0) = 0$, where (in order to avoid confusion) we write v_i instead of u_i. Condition (32) translates than to

$$v_1\,(t,x_1) + v_2\,(t,x_2) - \frac{1}{2}\langle Ax, x\rangle \le 0 \text{ for all } (t,x) \in (0,T) \times \mathbb{R}^{2n}; \tag{35}$$

this also means that the left-hand-side as a function of (t, x_1, x_2) has a global maximum at $(s, 0, 0)$. The assertion of the (reduced) Theorem is then the existence of $(b_i,\ X_i) \in \mathbb{R} \times \mathcal{S}^n$ such that $(b_i, 0, X_i) \in \bar{\mathcal{P}}^{2,+} v_i\,(s,0)$ for $i = 1, 2$ and (34) holds with $b = 0$.

We now give the main result of this subsection.

Theorem 9 *Assume that* u_i *has a finite extension to* $(0,T] \times \mathbb{R}^n$, $i = 1, 2$, *via its semi-continuous envelopes, that is,*

$$u_i\,(T,x) = \limsup_{\substack{(t,y) \in (0,T) \times \mathbb{R}^n: \\ t \uparrow T,\, y \to x}} u_i\,(t,y) < \infty.$$

Then the above Theorem remains valid at $s = T$ *if*

$$\mathcal{P}^{2,+} w\,(s,z) \text{ and } \bar{\mathcal{P}}^{2,+} u_i\,(s, z_i)$$

is replaced by

$$\mathcal{P}_Q^{2,+} w\,(T,z) \text{ and } \bar{\mathcal{P}}_Q^{2,+} u_i\,(T, z_i)$$

and the final equality in (34) is replaced by

$$b_1 + b_2 \geq b. \tag{36}$$

Remark 5 If we knew (but we don't!) that the final conclusion is $(b_i, p_i, X_i) \in \mathcal{P}^{2,+}u\,(T, z_i)$, rather than just being an element in the closure $\bar{\mathcal{P}}_Q^{2,+}u\,(T, z_i)$, then we could trivially diminish the b_i's such as to have $b_1 + b_2 = b$; cf. (31).

Remark 6 If one knows that w is left-accessible in the sense of [7] it should be possible to simplify the proof by avoiding doubling in t (and thus the penalty parameter m) and by using the classical parabolic (instead of elliptic) Theorem of Sums on $(0, T)$ in step 1 of the proof below. Left-accessibility is actually given in our application since then $u_1, -u_2$ are viscosity sub-/supersolutions. In that case condition (33) is also immediately satisfied, see e.g. Lemma V.6.1 in [14].

Proof Step 1: We focus on the reduced setting (and thus write v_i instead of u_i) and (following the proof of Lemma 8 in [9]) redefine $v_i\,(t_i, x_i)$ as $-\infty$ when $|x_i| > 1$ or $t_i \notin [T/2, T]$. We can also assume that (35) is strict if $t < s = T$ or $x \neq 0$. For the rest of the proof, we shall abbreviate (t_1, t_2) , (x_1, x_2) etc. by (t, x). With this notation in mind we set

$$w\,(t, x) = v_1\,(t_1, x_1) + v_2\,(t_2, x_2) - \frac{1}{2}\,\langle Ax, x\rangle\,.$$

By the extension via semi-continuous envelopes, there exist a sequence $(t^n, x^n) \in (0, T)^2 \times (\mathbb{R}^n)^2$, such that

$$\left(t^n, x^n\right) \equiv \left(t^{1,n}, t^{2,n}, x^{1,n}, x^{2,n}\right) \to (T, T, 0, 0)$$

and

$$w(t^{1,n}, t^{2,n}, x^{1,n}, x^{2,n}) \to w(T, T, 0, 0).$$

We now consider w with a penalty term for $t_1 \neq t_2$ and a barrier at time T for both t_1 and t_2.

$$\psi_{m,n}\,(t, x) = w\,(t, x) - \left\{\frac{m}{2}\,|t_1 - t_2|^2 + \sum_{i=1}^{2}\left(T - t^{i,n}\right)^2 / \,(T - t_i)\right\},$$

indexed by $(m, n) \in \mathbb{N}^2$, say. By assumption w has a maximum at $(T, T, 0, 0)$ which we may assume to be strict (otherwise subtract suitable forth order terms ...). Define now

$$\left(\hat{t}, \hat{x}\right) \in \arg\max \psi_{m,n} \text{ over } [T - r, T]^2 \times \bar{B}_r\,(0)^2$$

where $r = T/2$ (for instance). When we want to emphasize dependence on m, n we write $\left(\hat{t}_{m,n}, \hat{x}_{m,n}\right)$. We shall see below (Step 2) that there exists increasing sequences $m = m(k), n = n(k)$ so that

$$\left(\hat{t}, \hat{x}\right)|_{m=m(k),n=n(k)} \to (T, T, 0, 0). \tag{37}$$

Using the (elliptic) Theorem of Sums in the form of [9, Theorem 1] we find that there are

$$(b_i, p_i, X_i) \in \bar{\mathcal{P}}^{2,+} v_i \left(\hat{t}_i, \hat{x}_i\right)$$

(where $\hat{t}_i \to T, \hat{x}_i \to 0$ as $k \to \infty$) such that the first part of (34) holds and

$$A\begin{pmatrix}\hat{x}_1\\ \hat{x}_2\end{pmatrix} = \begin{pmatrix}p_1\\ p_2\end{pmatrix}, \quad b_i = m\left(t_i - t_{3-i}\right) + \left(T - t^{i,\varepsilon}\right)^2 / \left(T - t_i\right)^2.$$

for $i = 1, 2$. Note that

$$b_1 + b_2 = m\left(t_1 - t_2\right) + m\left(t_2 - t_1\right) + \text{(positive terms)} \geq 0;$$

since each b_i is bounded above by the assumptions and the estimates on the X_i it follows that the b_i lie in precompact sets. Upon passing to the limit $k \to \infty$ we obtain points

$$(b_i, p_i, X_i) \in \bar{\mathcal{P}}^{2,+} v_i (T, 0), \quad i = 1, 2;$$

with $b_1 + b_2 \geq 0$.

Step 2: We still have to establish (37). We first remark that for arbitrary (strictly) increasing sequences $m(k), n(k)$, compactness implies that

$$\left\{\left(\hat{t}_{m(k),n(k)}, \hat{x}_{m(k),n(k)}\right) : k \geq 1\right\} \in [T - r, T]^2 \times \bar{B}_r(0)^2$$

has limit points. Note also $\hat{t}_1, \hat{t}_2 \in [T - r, T)$ thanks to the barrier at time T. The key technical ingredient for the remained of the argument is and we postpone details of these to Step 3 below:

$$w\left(\hat{t}, \hat{x}\right) - \psi_{m,n}\left(\hat{t}, \hat{x}\right) = \left\{\frac{m}{2}\left|\hat{t}_1 - \hat{t}_2\right|^2 + \sum_{i=1}^{2}\left(T - t^{i,n}\right)^2 / \left(T - \hat{t}_i\right)\right\}$$
$$\to 0 \text{ as } \frac{1}{n} << \frac{1}{m} \to 0. \tag{38}$$

In particular, for every $k > 0$ there exists $m(k)$ such that for all $m \geq m(k)$

$$\lim\sup_{n\to\infty} \{...\} < \frac{1}{k}.$$

By making $m(k)$ larger if necessary we may assume that $m(k)$ is (strictly) increasing in k. Furthermore there exists $n(m(k), k) = n(k)$ such that for all $n \geq n(k) : \{...\} < 2/k$. Again, we may make $n(k)$ larger if necessary so that $n(k)$ is strictly increasing. Recall $t^{1,n(k)} - t^{2,n(k)} \to T - T = 0$ as $k \to \infty$. For reasons that will become apparent further below, we actually want the stronger statement that

$$\frac{m(k)}{2} \left| t^{1,n(k)} - t^{2,n(k)} \right|^2 \to 0 \text{ as } k \to \infty \tag{39}$$

which we can achieve by modifying $n(k)$ such as to run to ∞ even faster. Note that the so-constructed $m = m(k)$, $n = n(k)$ has the property

$$\left[w\left(\hat{t}, \hat{x}\right) - \psi_{m,n}\left(\hat{t}, \hat{x}\right) \right] |_{m=m(k),n=(k)} = \{...\} |_{m=m(k),n=(k)} \to 0 \text{ as } k \to \infty. \tag{40}$$

By switching to a subsequence (k_l) if necessary we may also assume (after relabeling) that

$$\left(\hat{t}_{m(k),n(k)}, \hat{x}_{m(k),n(k)}\right) \to \left(\tilde{t}, \tilde{x}\right) \in [T - r, T]^2 \times \bar{B}_r(0)^2 \text{ as } k \to \infty.$$

In the sequel we think of $\left(\hat{t}, \hat{x}\right)$ as this sequence indexed by k. We have

$$\begin{aligned} w\left(\tilde{t}, \tilde{x}\right) &\geq \limsup_{k \to \infty} w\left(\hat{t}, \hat{x}\right) |_{m=m(k),n=(k)} \text{ by upper-semi-continuity} \\ &= \limsup_{k \to \infty} \psi_{m,n}\left(\hat{t}, \hat{x}\right) |_{m=m(k),n=(k)} \text{ thanks to (40).} \end{aligned} \tag{41}$$

On the other hand, thanks to the particular form of our time-T barrier,

$$\begin{aligned} \psi_{m,n}\left(\hat{t}, \hat{x}\right) &\geq \psi_{m,n}\left(t^n, x^n\right) \\ &= w\left(t^n, x^n\right) - \left\{ \frac{m}{2} \left| t^{1,n} - t^{2,n} \right|^2 + \sum_{i=1}^{2} \left(T - t^{i,n} \right) \right\}. \end{aligned}$$

Take now $m = m(k)$, $n = n(k)$ as constructed above. Then

$$\begin{aligned} \psi_{m,n}\left(\hat{t}, \hat{x}\right) |_{m=m(k),n=(k)} \geq\ & w\left(t^{n(k)}, x^{n(k)}\right) \\ & - \left\{ \frac{m(k)}{2} \left| t^{1,n(k)} - t^{2,n(k)} \right|^2 + \sum_{i=1}^{2} \left(T - t^{i,n(k)} \right) \right\} \end{aligned}$$

The first term in the curly bracket goes to zero (with $k \to \infty$) thanks to (39), the other term goes to zero since $t^{i,n} \to T$ with $n \to \infty$, and hence also along $n(k)$. On the other hand (recall $x^{i,n} \to 0$)

$$w\left(t^{n(k)}, x^{n(k)}\right) \to v_1(T, 0) + v_2(T, 0) - \frac{1}{2} \langle A0, 0 \rangle = 0 \text{ as } k \to \infty.$$

(In the reduced setting $v_1(T, 0) = v_2(T, 0) = 0$.) It follows that

$$\liminf_{k\to\infty} \psi_{m,n}\left(\hat{t}, \hat{x}\right)|_{m=m(k),n=(k)} = 0.$$

Together with (41) we see that $w\left(\tilde{t}, \tilde{x}\right) \geq 0$. But $w(T, T, 0, 0) = 0$ was a strict maximum in $[T - r, T]^2 \times \bar{B}_r(0)^2$ and so we must have $\left(\tilde{t}, \tilde{x}\right) = (T, T, 0, 0)$.

Step 3: Set

$$M(h) = \sup_{\substack{(t,x)\in[T-r,T)^2\times\bar{B}_r(0)^2 \\ |t_1-t_2|<h}} w(t_1, t_2, x_1, x_2) \text{ and } M' = \lim_{h\to 0} M(h)$$

It is enough to show

$$\limsup_{\frac{1}{n}<<\frac{1}{m}\to 0} w\left(\hat{t}, \hat{x}\right) \leq M' \leq \liminf_{\frac{1}{n}<<\frac{1}{m}\to 0} \psi_{m,n}\left(\hat{t}, \hat{x}\right). \tag{42}$$

since the claimed

$$w\left(\hat{t}, \hat{x}\right) - \psi_{m,n}\left(\hat{t}, \hat{x}\right) = \left\{\frac{m}{2}\left|\hat{t}_1 - \hat{t}_2\right|^2 + \sum_{i=1}^{2}\left(T - t^{i,n}\right)^2 / \left(T - \hat{t}_i\right)\right\}$$
$$\to 0 \text{ as } \frac{1}{n} << \frac{1}{m} \to 0.$$

follows from

$$\limsup_{\frac{1}{n}<<\frac{1}{m}\to 0} \{...\} \leq \limsup_{\frac{1}{n}<<\frac{1}{m}\to 0} w\left(\hat{t}, \hat{x}\right) - \liminf_{\frac{1}{n}<<\frac{1}{m}\to 0} \psi_{m,n}\left(\hat{t}, \hat{x}\right)$$
$$\leq 0 \text{ (and hence } = 0).$$

Note that $w\left(\hat{t}, \hat{x}\right)$ is bounded on $[T - r, T]^2 \times \bar{B}_r(0)^2$ so that

$$\left|\hat{t}_1 - \hat{t}_2\right|^2 = O(1/m) \implies w\left(\hat{t}, \hat{x}\right) \leq M\left(\text{const}/\sqrt{m}\right).$$

On the other hand, from the very definition of M' as $\lim_{h\to 0} M(h)$, there exists a family (t_h, x_h) so that

$$|t_{1,h} - t_{2,h}| \leq h \text{ and } w(t_h, x_h) \to M' \text{ as } h \to 0 \tag{43}$$

For every m, n we may take (t_h, x_h) as argument of $\psi_{m,n}$ (which itself has a maximum at $\hat{t}, \hat{x}$); hence

$$w(t_h, x_h) - \frac{m}{2}h^2 - \sum_{i=1}^{2}\left(T - t^{i,n}\right)^2 / \left(T - t_{i,h}\right) \leq \psi_{m,n}\left(\hat{t}, \hat{x}\right). \tag{44}$$

Take now a sequence $n = n(h)$, fast enough increasing as $h \searrow$ such that $\left(T - t^{i,n}\right)^2 / \left(T - t_{i,h}\right) \to 0$ with $h \to 0$. It follows that

$$
\begin{aligned}
M' &= \lim_{h\to 0} w(t_h, x_h) \\
&= \liminf_{h\to 0} \left(w(t_h, x_h) - \frac{m}{2} h^2 - \sum_{i=1}^{2} \left(T - t^{i,n(h)}\right)^2 / \left(T - t_{i,h}\right) \right) \\
&\le \liminf_{h\to 0} \psi_{m,n(h)}\left(\hat{t}, \hat{x}\right) = \liminf_{n\to\infty} \psi_{m,n}\left(\hat{t}, \hat{x}\right) \text{ by monotonicity of } \sup \psi_{m,n} \text{ in } n.
\end{aligned}
$$

(In the last equality we used that $t^{i,n} \uparrow T$; this shows that $\sup \psi_{m,n}$ is indeed monoton in n.) The proof is now finished. □

3 RPDEs: Existence, Uniqueness and Regularity

We start with a brief recall on *rough path theory*; [15, 29–31]. Given a collection $(V_1, \ldots, V_d)$ of (sufficiently nice) vector fields on $\mathbb{R}^n$ and $z \in C^1\left([0, T], \mathbb{R}^d\right)$ one considers the (unique) solution y to the ordinary differential equation

$$
\dot{y}(t) = \sum_{i=1}^{d} V_i(y)\, \dot{z}^i(t), \quad y(0) = y_0 \in \mathbb{R}^n. \tag{45}
$$

The question is, if the output signal y depends in a stable way on the driving signal z. The answer, of course, depends strongly on how to measure distance between input signals. If one uses the supremum norm, so that the distance between driving signals $z, \tilde{z}$ is given by $|z - \tilde{z}|_{\infty;[0,T]}$, then the solution will in general *not* depend continuously on the input. If $|z - \tilde{z}|_{\infty;[0,T]}$ is replaced by the (much) stronger distance ($z_{s,t} := z_t - z_s$)

$$
|z - \tilde{z}|_{1-\text{Höl};[0,T]} = \sup_{0\le s<t\le T} \frac{\left|z_{s,t} - \tilde{z}_{s,t}\right|}{|t-s|},
$$

it is elementary to see that now the solution map is continuous (in fact, locally Lipschitz); however, this continuity does not lend itself to push the meaning of (45): the closure of smooth paths in this metric yields precisely C^1. Lyons' theory of rough paths exhibits an entire cascade of α-Hölder type rough path (or, as a variation on the scheme: $(1/\alpha)$-variation) metrics, for each $\alpha \in (0, 1]$ on path-space under which such ODE solutions are continuous (and even locally Lipschitz) functions of their driving signal. For instance, the "rough path" α-Hölder distance between two smooth $\mathbb{R}^d$-valued paths $z, \tilde{z}$ is given by

$$
\max_{j=1,\ldots,[1/\alpha]} \sup_{0\le s<t\le T} \sum \frac{\left|z^{(j)}_{s,t} - \tilde{z}^{(j)}_{s,t}\right|}{|t-s|^{j\alpha}}
$$

where $z_{s,t}^{(j)} = \int dz_{r_1} \otimes \cdots \otimes dz_{r_j}$ with integration over the j-dimensional simplex $\{s < r_1 < \cdots < r_j < t\}$. This allows to extend the very meaning of (45), in a unique and continuous fashion, to driving signals which live in the abstract completion of smooth $\mathbb{R}^d$-valued paths (with respect to rough path α-Hölder metric). The space of so-called α-Hölder rough paths[8] is precisely this abstract completion. In fact, this space can be realized as genuine path space,

$$C^{0,\alpha-\text{Höl}}\left([0,T], G^{[1/\alpha]}\left(\mathbb{R}^d\right)\right)$$

where $G^{[1/\alpha]}\left(\mathbb{R}^d\right)$ is the free step-$[1/\alpha]$ nilpotent group over $\mathbb{R}^d$, equipped with Carnot–Caratheodory metric; realized as a subset of $1 + \mathfrak{t}^{[1/\alpha]}\left(\mathbb{R}^d\right)$ where

$$\mathfrak{t}^{[1/\alpha]}\left(\mathbb{R}^d\right) = \mathbb{R}^d \oplus \left(\mathbb{R}^d\right)^{\otimes 2} \oplus \cdots \oplus \left(\mathbb{R}^d\right)^{\otimes [1/\alpha]}$$

is the natural state space for (up to $[1/\alpha]$) iterated integrals of a smooth $\mathbb{R}^d$-valued path. For instance, almost every realization of d-dimensional Brownian motion B *enhanced with its iterated stochastic integrals in the sense of Stratonovich,* i.e. the matrix-valued process given by

$$B^{(2)} := \left(\int_0^{\cdot} B^i \circ dB^j\right)_{i,j\in\{1,\dots,d\}} \tag{46}$$

yields a path $\mathbf{B}(\omega)$ in $G^2\left(\mathbb{R}^d\right)$ with finite α-Hölder, for any $\alpha < 1/2$. ($\mathbf{B}$ is known as *Brownian rough path.*) We remark that $B^{(2)} = \frac{1}{2} B \otimes B + A$ where $A := \text{Anti}\left(B^{(2)}\right)$ is known as *Lévy's stochastic area;* in other words $\mathbf{B}(\omega)$ is determined by (B, A), i.e. Brownian motion *enhanced with Lévy's area.* Turning to the main topic of this section, we follow [25, 26, 28] in considering a real-valued function of time and space $u = u(t,x) \in \text{BUC}([0,T] \times \mathbb{R}^n)$ which solves the nonlinear partial differential equation

$$\begin{aligned} du &= F\left(t,x,Du,D^2u\right)dt + \sum_{i=1}^d H_i(x,Du)\,dz^i \\ &\equiv F\left(t,x,Du,D^2u\right)dt + H(x,Du)\,dz \end{aligned} \tag{47}$$

in viscosity sense. When $z : [0,T] \to \mathbb{R}^d$ is C^1 then, subject to suitable conditions on F and H, this falls in the standard setting of viscosity theory as discussed above. This can be pushed further to $z \in W^{1,1}$ (see e.g. [25, Remark 4] and the references given there) but, as was pointed out by various authors, the case when $z = z(t)$ has

[8] In the strict terminology of rough path theory: geometric α-Hölder rough paths.

only "Brownian" regularity (just below 1/2-Hölder, say) falls dramatically outside the scope of the standard theory. The reader can find a variety of examples (drawing from fields as diverse as stochastic control theory, pathwise stochastic control, interest rate theory, front propagation and phase transition in random media, ...) in the articles [24, 26] justifying the need of a theory of (non-linear) *stochastic partial differential equations* (SPDEs) in which z in (47) is taken as a Brownian motion.[9] In the same series of articles a satisfactory theory is established for the case of non-linear Hamiltonian with no spatial dependence, i.e. $H = H(Du)$. Our present discussion follows [6] in that we consider non-linear F and $H = H(x, Du)$, linear in Du. The following condition may be considered as a global version of the corresponding definition put forward by Lions–Souganidis.

Definition 1 Let ϕ denote the solution flow to the RDE $dy = V(y)\, d\mathbf{z}(t)$. (As is well known, this yields is a C^3-flow of diffeomorphisms provided $V = (V_1, \ldots, V_d)$ is a collection of $\mathrm{Lip}^{\gamma+2}(\mathbb{R}^n; \mathbb{R}^n)$ vector fields with $\gamma > 1/\alpha$ and if

$$\mathbf{z} \in C^{0,\alpha-H\ddot{o}l}\left([0, T], G^{[1/\alpha]}\left(\mathbb{R}^d\right)\right).$$

A continuous function u is called a *rough viscosity solution* to

$$du = F\left(t, x, Du, D^2u\right) dt - Du(t, x) \cdot V(x)\, d\mathbf{z}(t)$$

if $v(t, \cdot) := u(t, \phi_t(\cdot))$ is a viscosity solution to

$$\partial_t v - \tilde{F}\left(t, x, Dv, D^2v\right) = 0$$

where $\tilde{F}(t, x, p, X)$ is given by

$$F\left(t, \phi_t(x), \left\langle p, D\phi_t^{-1}|_{\phi_t(x)}\right\rangle, \left\langle X, D\phi_t^{-1}|_{\phi_t(x)} \otimes D\phi_t^{-1}|_{\phi_t(x)}\right\rangle \right.$$
$$\left. + \left\langle p, D^2\phi_t^{-1}|_{\phi_t(x)}\right\rangle\right). \tag{48}$$

It should be noted that in the case when $\mathbf{z}$ arises from a smooth $\mathbb{R}^d$-valued path z, the definition is consistent with the interpretation

$$\partial_t u = F\left(t, x, Du, D^2u\right) - Du(t, x) \cdot V(x)\, \dot{z}(t);$$

this is verified in Ref. [6], for instance. The following result adds existence and regularity to the main result of Ref. [6].

[9] ... in which case (47) is understood in Stratonovich form.

Theorem 10 *Let $\alpha \in (0,1]$ and $(z^\varepsilon) \subset C^\infty\left([0,T],\mathbb{R}^d\right)$ be Cauchy in (α-Hölder) rough path topology with rough path limit $\mathbf{z} \in C^{0,\alpha-\text{Höl}}\left([0,T],G^{[1/\alpha]}\left(\mathbb{R}^d\right)\right)$. Assume*

$$u_0^\varepsilon \in \mathrm{BUC}\left(\mathbb{R}^n\right) \to u_0 \in \mathrm{BUC}\left(\mathbb{R}^n\right),$$

uniformly as $\varepsilon \to 0$. Let $F = F(t,x,p,X)$ be continuous, degenerate elliptic, and assume that $\partial_t - \tilde{F}$ where $\tilde{F}$ is given in (48) satisfies the assumptions of Sect. 2.1 with assumption (2.1) strengthened to [10]

$$\forall R > 0 : \tilde{F}|_{[0,T]\times\mathbb{R}^n\times[-R,R]\times B_R\times M_R} \text{ is bounded, uniformly continuous.}$$

Assume that $V = (V_1,\ldots,V_d)$ is a collection of $\mathrm{Lip}^{\gamma+2}(\mathbb{R}^n;\mathbb{R}^n)$ vector fields with $\gamma > 1/\alpha$. Then
(i) ***Existence, uniqueness****: there exists unique* BUC *-solutions to the approximate problems*

$$du^\varepsilon = F\left(t,x,Du^\varepsilon,D^2u^\varepsilon\right)dt - Du^\varepsilon(t,x)\cdot V(x)\,dz^\varepsilon(t), \tag{49}$$

$$u^\varepsilon(0,\cdot) = u_0^\varepsilon \tag{50}$$

and the locally uniform limit $u = \lim_{\varepsilon\to 0} u^\varepsilon$ exists, and is the unique BUC $([0,T]\times\mathbb{R}^n)$ *rough viscosity solution (Definition 1) to*

$$du = F\left(t,x,Du,D^2u\right)dt - Du(t,x)\cdot V(x)\,d\mathbf{z}(t),$$
$$u(0,\cdot) = u_0 \in \mathrm{BUC}\left(\mathbb{R}^n\right).$$

In particular, $u = u^{\mathbf{z}}$ only depends on $\mathbf{z}$ and u_0 but not on the particular approximating sequence $\{z^\varepsilon\}$.
(ii) ***Robustness****: The solution map $(\mathbf{z},u_0) \mapsto u^{\mathbf{z}}$ from*

$$C^{0,\alpha-\text{Höl}}\left([0,T],G^{[p]}\left(\mathbb{R}^d\right)\right) \times \mathrm{BUC}\left(\mathbb{R}^n\right) \to \mathrm{BUC}\left([0,T]\times\mathbb{R}^n\right)$$

is continuous.
(iii) ***Regularity****: Make the additional assumption that $\tilde{F}$ has modulus θ_R which is actually linear, and that $\tilde{F}$ has linear growth in the Hessian. Then, if $u_0 \in \mathrm{Lip}^1(\mathbb{R}^n)$, i.e. bounded and Lipschitz continuous, then u is Lipschitz in space and* $\min(\alpha,1/2)$-*Hölder in time.*

Proof By assumption, $F(\cdot,\phi.(*)\cdot,0,0) = \tilde{F}(\cdot,*,0,0)$ is bounded on $[0,T]\times\mathbb{R}^n$, and the assumption that $u_0^\varepsilon \to u_0$ uniformly, as can be seen by comparison with function of the type $(t,x) \mapsto \pm C(t+1)$, with sufficiently large C, independent of

[10] This may be expressed in terms of F; in particular F then satisfies $\Phi^{(3)}$-invariant comparison as introduced in Ref. [6]; there it was also checked that these structural assumptions are satisfied by many examples ; notably those arising from pathwise stochastic control theory.

ε, that the family

$$\{u^{\varepsilon}\}$$

is locally uniformly bounded in ε. We also note that our structural assumptions imply both existence and comparison (hence uniqueness) of the approximate problems. Since RDE solution flows (as C^3-flows of diffeomorphisms) depend continuously on the rough driving signals [15, Theorem 11.12,11.13], the corresponding $\tilde{F}(\varepsilon)$ given by (48), with ϕ give as the solution flow to $dy = V(y)\,dz^{\varepsilon}$, converge locally uniformly to $\tilde{F}$ based on the RDE solution flow to $dy = V(y)\,d\mathbf{z}$. Thanks to the Barles-Perthame procedure (which we apply to the transformed equation, cf (iii) below) we see that

$$u = \lim_{\varepsilon \to 0} u^{\varepsilon}$$

exists locally uniformly and is a rough viscosity solution in the sense of Definition 1. Since solutions to $\partial_t - \tilde{F}$ with BUC initial data, are also BUC in space time, we see that $u \in \mathrm{BUC}([0,T] \times \mathbb{R}^n)$. This settles (i).
The argument for (ii) is identical, one just considers approximate RDE problems $dy = V(y)\,d\mathbf{z}^{\varepsilon}$ where each $\mathbf{z}^{\varepsilon}$ may be a genuine rough paths (rather than a smooth, approximating path).
(iii) Since $v(t,\cdot) := u(t, \phi_t(\cdot))$ is a viscosity solution to

$$\partial_t v - \tilde{F}\left(t, x, Dv, D^2 v\right) = 0$$

it follows from the results in Sect. 2 that v is Lipschitz in space and (1/2)-Hölder in time. Since the time-space regularity

$$(x,t) \mapsto \phi_t(\cdot) \text{ resp. } \phi_t^{-1}(\cdot)$$

is well-understood [15], namely α-Hölder in time and Lipschitz (and more) in space, the claimed regularity of the rough viscosity solution u follows. □

4 Examples

Let us verify in some detail that Theorem 10 applies to two concrete rough resp. stochastic partial differential equations of interest.

4.1 Stochastic HJB Equation

Following [25] consider the following stochastic Hamilton–Jacobi–Bellman equation

$$du = \inf_{\alpha \in A} L_{\alpha}\left(t, x, Du, D^2 u\right) dt + Du(t,x) \cdot V_i(x) \circ dB.$$

The case of constant vector fields $V = (V_1, \ldots, V_d)$ was treated in the aforementioned reference; one then has a truely pathwise theory, i.e. for every continuous path $B : [0, T] \to \mathbb{R}^d$. Moreover, the solution is robust in the sense that it is the "Wong–Zakai" limit of approximate problems where B is replaced by (piecewise) smooth B^ε, uniformly convergent to B as $\varepsilon \to 0$. This was extended with the aid of rough path theory to non-constant vector fields in Ref. [6]. It was assumed that all approximate problems have a unique (viscosity) solution, convergence then takes places if (B^ε) is Cauchy in rough path metric. Although existence is not an issue here (one could use stochastic control theory and then the stability result of [6]), Theorem 10 applies on purely analytical grounds and also gives provides information about the regularity of the solution. To be a little more specific, one assumes

$$L_\alpha(t, x, p, X) = \operatorname{Tr}\left[\sigma_\alpha(t, x)\, \sigma_\alpha(t, x)^T X\right] + b_\alpha(t, x) \cdot p,$$

where $\sigma_\alpha(t, x) : [0, T] \times \mathbb{R}^n \to \mathbb{R}^{n \times n'}$ and $b_\alpha(t, x) : [0, T] \times \mathbb{R}^n \to \mathbb{R}^n$ are bounded, continuous, and Lipschitz continuous in x, uniformly in $t \in [0, T]$; with all these properties valid uniformly in $\alpha \in A$. Assume also that $V_1, \ldots, V_d$ are in $\mathrm{Lip}^5(\mathbb{R}^n)$, say, reflecting the regularity need for a α-Hölder driving rough path, any $\alpha < 1/2$, such as Brownian motion enhanced to a rough path. A rigorous application of Theorem 10 is easily justified, along the lines of [6] example 3 and 5 on pages 33, 35 respectively. (The only structural aspect of $\tilde{F}$ which was not verified explicitly in that paper was boundedness of $\tilde{F}$, but this an easy consequence of the fact that the auxilary (rough) flows have bounded derivatives.)

4.2 Zakai Equation

As is well-known, the filtering problem for non-linear diffusion leads to the following stochastic partial differential equation. To be precise, consider

$$\begin{aligned} dX_t &= b(X_t)\, dt + \sigma(X_t)\, dB_t + V(X_t)\, d\tilde{B}_t \\ dY_t &= h(X_t)\, dt + d\tilde{B}_t \end{aligned} \tag{51}$$

where B and $\tilde{B}$ are independent, multidimensional Brownian motions. Note that the generator of X is of the form

$$\mathcal{L}u = \frac{1}{2}\operatorname{Tr}\left[\sigma(x)\, \sigma(x)^T D^2 u\right] + \frac{1}{2}\operatorname{Tr}\left[V(x)\, V(x)^T D^2 u\right] + b(x) \cdot Du$$

Define also the first order operators, $V = (V_1, \ldots, V_d)$

$$\mathcal{N}_k u = V_k(x) \cdot Du + h(x)\, u, \quad k = 1, \ldots, d$$

and note that the (formal) adjoint $\mathcal{N}_k^* u$ is of the form $\mathcal{N}_k^* u = H(x, u, Du)$ with H linear in u and Du. The goal is to compute for a given real-valued function φ

$$\pi_t(\varphi) = \mathbb{E}[\varphi(X_t) | \sigma(Y_r, r \leq t)]$$

and from basic principles it follows that there exists a map $\phi_t^\varphi : C([0, T], \mathbb{R}^{d_Y}) \to \mathbb{R}$ such that

$$\phi_t^\varphi \left(Y|_{[0,t]}\right) = \pi_t(\varphi) \quad \mathbb{P} - a.s. \tag{52}$$

As pointed out by M. Clark in the late seventies, there is a problem here since a basic (practical!) requirement of a filter is robustness, say in the form of continuous dependence of the observation path. He then showed that in the case of uncorrelated noise ($\sigma \equiv 0$ in (51)) (which corresponds to the case $V = 0$ above) there exists a unique $\overline{\phi}_t^\varphi : C([0, T], \mathbb{R}^n) \to \mathbb{R}$ which is continuous in uniform norm and fulfills (52), thus providing a version of the conditional expectation $\pi_t(\varphi)$ which is robust under approximations in uniform norm of the observation Y. Unfortunately in the correlated noise case this is no longer true! In Ref. [11] it was recently shown that in this case robustness still holds in a rough path sense. Now recall that under well-known conditions (e.g. [1]), π_t can be written in the form

$$\pi_t(\varphi) = \int_{\mathbb{R}^{d_X}} \varphi(x) \frac{u_t(x)}{\int u_t(\tilde{x})\, d\tilde{x}} dx \tag{53}$$

where $u_t \in L^1(\mathbb{R}^n)$ a.s. and (u_t) is the L^2-solution of the (dual) Zakai SPDE[11]

$$\begin{aligned} du_t &= \mathcal{L}^* u_t dt + \sum_k \mathcal{N}_k^* u_t dY_t^k \\ &= \left(\mathcal{L}^* - \frac{1}{2} \sum_k \mathcal{N}_k^* \mathcal{N}_k^*\right) u_t dt + \sum_k \mathcal{N}_k^* u_t \circ dY_t^k \\ &\equiv F\left(x, u, Du, D^2 u\right) dt + H(x, u, Du) \circ dY_t^i. \end{aligned}$$

Note that $\mathcal{L}^*$ has the "stochastic" parabolicity property which here means that F is degenerate elliptic. The resulting stochastic PDE does not quite fall in the class considered in Theorem 10, for H has additional dependence on u, but the method still works. A more complicated transformation, detailed in Ref. [16], allows to given (rough)pathwise meaning to the above equation. Indeed, one can show that $v(t, x) = \psi^{-1}(t, u(t, \phi(t, x)), x)$ satisfies a parabolic equation of the form

$$\partial_t v - \tilde{F}\left(x, v, Dv, D^2 v\right) = 0$$

[11] Consistency of L^2-theory with RPDE theory has been established in Ref. [16].

with initial data $v(0, \cdot) = u(0, \cdot)$. Moreover, with this definition u will depend continuously on the driving signal (i.e. the observation path Y) in rough path topology; thereby solving the robustness problem for the (un-normalized) conditional density.

In fact, $\tilde{F}$ turns out to be linear in $v, Dv, D^2 v$ and there is an explicit, if complicated, expression. In particular then, if b, σ are bounded and Lipschitz, $V \in \mathrm{Lip}^{\gamma+2}$ and $h \in \mathrm{Lip}^{\gamma+1}$ for some $\gamma > 2$, $\tilde{F}$ is seen to meet the assumption of Sect. 2, necessary to arrive at the same conclusions as those stated in Theorem 10. An appealing feature here is that one can immediately handle degenerate situations (including the case when F degenerates to a purely first order operator) and also that one gets automatically continuous versions of solutions to the Zakai equation, without requiring dimension-dependent regularity assumptions on the coefficients (as pointed out by Krylov [23], a disadvantage of the L^2 theory of SPDEs). On the other hand, our regularity assumption (in particular in the noise terms) are more stringent[12] than what is needed to ensure existence and uniqueness in the L^2 theory of SPDEs.

Acknowledgments J. Diehl is supported by DFG Priority Program 1324; P. Friz and H. Oberhauser received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013)/ERC grant agreement nr. 258237. P. Friz would like to thank G. Barles for a helpful email exchange, E. Jakobsen for pointing us to [20, 21], M. Soner and T. Souganidis for discussions during the Oberwolfach meetings on *Stochastic Analysis in Finance and Insurance* and *Rough Paths and PDEs*.

References

1. A. Bain, D. Crisan, *Fundamentals of Stochastic Filtering* (Springer, New York, 2009)
2. G. Barles, *Solutions de viscosité des équations de Hamilton-Jacobi* (Springer, Berlin, 2004)
3. M. Bardi, I. Capuzzo-Dolcetta, *Optimal Control and Viscosity Solutions of Hamilton-Jacobi-Bellman Equations* (Birkhaeuser, Boston, 1997)
4. G. Barles, S. Biton, O. Ley, A geometrical approach to the study of unbounded solutions of quasilinear parabolic equations. Arch. Ration. Mech. Anal. **162**(4), 287–325 (2002)
5. M. Caruana, P. Friz, Partial differential equations driven by rough paths. J. Differ. Equ. **247**(1), 140–173 (2009)
6. M. Caruana, P. Friz, H. Oberhauser, A (rough) pathwise approach to a class of nonlinear SPDEs, Annales de l'Institut Henri Poincaré/Analyse non linéaire vol. 28, pp. 27–46 (ISSN: 0294–1449) (2011)
7. Y.-G. Chen, Y. Giga, S. Goto, Remarks on viscosity solutions for evolution equations. Proc. Japan Acad. Ser. A Math. Sci. **67**(10), 323–328 (1991)
8. M.G. Crandall, *Viscosity Solutinos: A Primer*, Lecture Notes in Mathematics (Springer, New York, 1996)
9. M.G. Crandall, H. Ishii, The maximum principle for semicontinuous functions. Differ. Integral Equ. **3**(6), 1001–1014 (1990)
10. M.G. Crandall, H. Ishii, P.-L. Lions, User's guide to viscosity solutions of second order partial differential equations. Bull. Amer. Math. Soc. (N.S.) **27**(1), 1–67 (1992)
11. D. Crisan, J. Diehl, P. Friz, Harald Oberhauser Robust filtering: correlated noise and multidimensional observation, Ann Appl Probab. **23**(5), 2139–2160 (2013). doi:10.1214/12-AAP896

[12] Above regularity assumptions can be weakened with the "joint lift" approach carried out in Ref. [13].

12. J. Diehl, F. Peter, Backward stochastic differential equations with rough drivers. The Annals of Probability 2012, Vol. 40, (4) 1715–1758, Publisher: Institute of Mathematical Statistics
13. J. Diehl, H. Oberhauser, S. Riedel, A Levy-area between Brownian motion and rough paths with applications to robust non-linear filtering and RPDEs. Stoch Process Appl. **125**(1), 161–181 January 2015. http://dx.doi.org/10.1016/j.spa.2014.08.005
14. H.W. Fleming, H.M. Soner, *Controlled Markov Processes and Viscosity Solutions, Volume 25 of Stochastic Modelling and Applied Probability*, 2nd edn. (Springer, New York, 2006)
15. P.K. Friz, N.B. Victoir, *Multidimensional Stochastic Processes as Rough Paths: Theory and Applications, Cambridge Studies in Advanced Mathematics* (Cambridge University Press, Cambridge, 2010)
16. P. Friz, H. Oberhauser, Rough path stability of (semi-)linear SPDEs. Probab Theory Relat Fields. **158**(1-2), 401–434 February 2014
17. P. Friz, H. Oberhauser, On the splitting-up method for rough (partial) differential equations, J. Differ. Equ., Vol. 251 (2) pp. 316–338 ISSN: 0022–0396 (2011)
18. Y. Giga, *Surface Evolution Equations: A Level Set Approach* (Birkhauser, Basel, 2006)
19. Y. Giga, S. Goto, H. Ishii, M.-H. Sato, Comparison principle and convexity preserving properties for singular degenerate parabolic equations on unbounded domains. Indiana Univ. Math. J. **40**(2), 443–470 (1991)
20. E.R. Jakobsen, K.H. Karlsen, Continuous dependence estimates for viscosity solutions of fully nonlinear degenerate parabolic equations. J. Differ. Equ. **183**(2), 497–525 (2002). (ISSN: 0022-0396)
21. E.R. Jakobsen, K.H. Karlsen, Continuous dependence estimates for viscosity solutions of integro-PDEs. J. Differ. Equ. **212**(2), 278–318 (2002)
22. Petri Juutinen, On the definition of viscosity solutions for parabolic equations. Proc. Amer. Math. Soc. **129**, 2907–2911 (2001)
23. N.V. Krylov, An analytic approach to SPDEs, in *Stochastic partial differential equations: six perspectives of Math. Surveys Monogr.,*, Amer. Math. Soc., Providence, vol .64 pp. 185–242 (1999)
24. P.-L. Lions, P.E. Souganidis, Viscosity solutions of fully nonlinear stochastic partial differential equations. Sūrikaisekikenkyūsho Kōkyūroku, (1287):58–65. Viscosity solutions of differential equations and related topics (Japanese) (Kyoto, 2001) (2002)
25. P.-L. Lions, P.E. Souganidis, Fully nonlinear stochastic partial differential equations. C. R. Acad. Sci. Paris Sér. I Math. **326**(9), 1085–1092 (1998)
26. P.-L. Lions, P.E. Souganidis, Fully nonlinear stochastic partial differential equations: non-smooth equations and applications. C. R. Acad. Sci. Paris Sér. I Math. **327**(8), 735–741 (1998)
27. P.-L. Lions, P.E. Souganidis, Fully nonlinear stochastic pde with semilinear stochastic dependence. C. R. Acad. Sci. Paris Sér. I Math. **331**(8), 617–624 (2000)
28. P.-L. Lions, P.E. Souganidis, Uniqueness of weak solutions of fully nonlinear stochastic partial differential equations. C. R. Acad. Sci. Paris Sér. I Math. **331**(10), 783–790 (2000)
29. T. Lyons, Differential equations driven by rough signals. Rev. Mat. Iberoam. **14**(2), 215–310 (1998)
30. T. Lyons, Z. Qian, *System Control and Rough Paths, Oxford Mathematical Monographs* (Oxford University Press, Oxford, 2002)
31. T.J. Lyons, M. Caruana, T. Lévy, Differential equations driven by rough paths, volume 1908 of Lecture Notes in Mathematics. Springer, Berlin, 2007. Lectures from the 34th Summer School on Probability Theory held in Saint-Flour, July 6–24, 2004, With an introduction concerning the Summer School by Jean Picard

Time-Inconsistent Portfolio Investment Problems

Yidong Dong and Ronnie Sircar

Abstract The explicit results for the classical Merton optimal investment/consumption problem rely on the use of constant risk aversion parameters and exponential discounting. However, many studies have suggested that individual investors can have different risk aversions over time, and they discount future rewards less rapidly than exponentially. While state-dependent risk aversions and non-exponential type (e.g. hyperbolic) discountings align more with the real life behavior and household consumption data, they have tractability issues and make the problem time-inconsistent. We analyze the cases where these problems can be closely approximated by time-consistent ones. By asymptotic approximations, we are able to characterize the equilibrium strategies explicitly in terms of the corrections to solutions for the base problems with constant risk aversion and exponential discounting. We also explore the effects of hyperbolic discounting under proportional transaction costs.

Keywords Time-inconsistency · Portfolio optimization · Asymptotic methods · Stochastic control · Stochastic risk aversion

1 Introduction and Background

1.1 The Merton Problem of Portfolio Optimization

The portfolio optimization problem in a continuous-time diffusion model was first introduced by Merton in the 1960s, with the original papers reprinted later in [24], where he was able to derive explicit solutions for the value functions and optimal

---

Partially supported by NSF grant DMS-1211906.

Y. Dong (✉) · R. Sircar
Department of Operations Research and Financial Engineering, Princeton University,
Princeton 08544, NJ, USA
e-mail: yidongd@princeton.edu

R. Sircar
e-mail: sircar@princeton.edu

D. Crisan et al. (eds.), *Stochastic Analysis and Applications 2014*,
Springer Proceedings in Mathematics & Statistics 100,
DOI 10.1007/978-3-319-11292-3_9

strategies in cases with geometric Brownian motions and special types of utility functions. Ever since then, there has been plenty of development aimed at generalizing Merton's results in different ways. To deal with market incompleteness is a direction that a large proportion of the works have been dedicated to. For example, Campbell and Viceira [4] and Wachter [31] studied the problem with stochastic drift returns. For problems of partial hedging with a non-traded asset, as well as utility indifference pricing, one could refer to the collection [5]. Meanwhile, stochastic volatility and transaction cost are two topics that have received much attention and popularity. We refer the reader to Chacko and Viceira [6] and Kraft [22] for some explicit results in cases with stochastic volatility. For transaction costs, things are much more subtle as the problem becomes less tractable. Davis and Norman [10] were able to solve the problem numerically as a free-boundary ODE system, and Shreve and Soner [28] treated it using the viscosity solution approach.

More recently, asymptotic methods have been used widely to solve the extensions of Merton's problem around their classical and well-established counterpart problems. For example, Fouque et al. [15] have used multiscale expansions to approximate the case with stochastic volatility around the constant volatility case. Bouchard et al. [3] used asymptotics for small transaction costs to derive tractable models for a partial hedging problem under expected loss constraints.

1.2 Time Inconsistency

The key to solving Merton's problem is the use of the Dynamic Programming Principle (hereinafter DPP) in order to formulate the Hamilton-Jacobi-Bellman (hereinafter HJB) equation. In a typical dynamic programming problem setup, when an agent wants to optimize an objective function by choosing the best plan, he is only required to decide his current action. This is because DPP assumes that one's future selves are going to solve the remaining part of today's problem and act optimally when future comes. However, in many problems, the DPP does not hold, and an agent does not have such "commitment power" on their future selves, which is the ability to enforce a course of plans obtained by repeatedly optimizing the same objective function over time. In such problems, the future selves may have changed preferences or tastes, or would want to make decisions based on different objective functions, effectively acting as opponents of the current self.

The dilemma described above is called dynamic inconsistency, which has been noted and studied by economists for many years, mainly in the context of non-exponential type discount functions. In [29], Strotz demonstrated that when a discount function was applied to consumption plans, one could favor a certain plan at the beginning, but later switch preference to another plan. This would hold true for most types of discount functions, the only exception being the exponential. Nevertheless, exponential discounting is "by default" in most literatures, as none of the other types could produce explicit solutions. Results from experimental studies contradict this assumption (see, for example, Loewenstein and Prelec [23]), indicating that the discount rates for the near future are much lower than the discount rates for the time further away in future, and therefore a hyperbolic type discount function would be more realistic.

Other types of time-inconsistency do exist as well. Bjork and Murgoci [1] listed out three possible scenarios where time inconsistency would occur in typical Markovian stochastic control problems. More specifically, given an objective function of the following form:

$$J(t, x, \pi) = \mathbb{E}\left[\int_t^T \varphi(s-t)F(X_s^\pi, x)ds + G(X_T^\pi, x)\right] + H(x, \mathbb{E}\left[X_T^\pi\right]),$$

where X^π is some controlled diffusion process with $X_t^\pi = x$ and π being our control, the optimization for $J(t, x, \pi)$ is a time-inconsistent problem if:

1. the discount function $\varphi(s-t)$ is not of exponential type, e.g. a hyperbolic discount function;
2. x appears in the objective function, e.g. a utility function that depends on the initial wealth x;
3. $H()$ is a nonlinear function of $\mathbb{E}\left[X_T^\pi\right]$, e.g. continuous-time mean-variance optimization on X_T^π.

In all the three cases, the standard HJB equations cannot be derived since the usual formulation requires an argument about the value function (process) being a super-martingale for arbitrary controls and being a martingale at optimum, which does not hold here. In light of the non-applicability of DPP on these problems, some have turned in a game-theoretic direction. By treating the problem as a game played with one's future selves, it is possible to derive a sub-game Nash equilibrium. In the next section, we will discuss recent works on deriving equilibrium strategies in some of the time-inconsistent problems.

1.3 Recent Literature on Time-Inconsistent Portfolio Optimization

One of the earlier advances was made by Harris and Laibson who discussed the existence and uniqueness of an equilibrium consumption path in the case of hyperbolic [16] and quasi-hyperbolic [17] discount functions in a discrete-time setup. They also derived the Euler relation for the equilibrium path using the recursive property of an equilibrium consumption plan. Ekeland and Lazrak [11] studied the problem in continuous time with a more general non-exponential type discount function and derived an equation for the equilibrium value function process, which was comparable to Harris and Laibson's results and resembles the classical HJB equation plus a non-local term. Later, Ekeland et al. [13] looked at an investment/consumption problem from life insurance with time-inconsistent discount functions. They solved the non-local HJB-type equation numerically and were able to obtain a hump-shaped consumption path that agreed with the household consumption data, as opposed to the monotone shape path produced by exponential discount functions.

Progresses have been made on other types of time inconsistent problems as well. Bjork et al. [2] used the same technique to study the continuous time mean-variance optimization problem with a state-dependent risk aversion parameter. They obtained a system of HJB-type equations which they were able to reduce to an ODE system and solve numerically if risk aversion had a special form. Their equilibrium strategy was comparable to the utility-maximizing strategy in a Merton model statically, but was able to capture some horizon effect as opposed to the Merton optimal strategy which was constant in time. On the other hand, Hu et al. [18] derived an open-loop equilibrium strategy, characterized by a system of forward-backward stochastic differential equations, to solve a time-inconsistent stochastic linear quadratic control problem, which is the generalized version of the mean-variance problem. Pirvu and Zhang [26] have studied the problem of utility indifference pricing under a discrete time model with a state-dependent risk aversion modelled by a two-state regime-switching Markov chain.

The remaining part of this article is organized as follows. In Sect. 2, we study the portfolio optimization problem with time-varying risk aversions that depend on the wealth or volatility factor. A discrete-time example will be given to illustrate the time inconsistency, followed by the derivation of the HJB-type equation in continuous time. We will use asymptotic methods to derive the equilibrium strategies up to first order. In Sect. 3, we look at hyperbolic discounting problems and use similar methodologies to obtain tractable solutions in this case. An extension with proportional transaction costs is also studied, and we provide some numerical results for this problem. Section 4 concludes.

2 Utility Maximization with Time-Varying Risk Aversion

In this section, we look at the classical Merton problem of portfolio optimization, but with the risk aversion parameter being state-dependent.[1] Our motivation is that, in the classical case, we need to, at time 0, fix a (constant) risk aversion parameter for expected utility at terminal time T. This value reflects our present conjecture about our future attitude towards risk, and thus it would be unnatural for this conjecture to be independent of the current state of the world, for example the wealth level and economic conditions. There are many indicators in the market that can, at least partially, measure investors' risk aversion. As mentioned in Coudert and Gex [9], the movement of risk aversion is often correlated with market indices, for example the gold price and VIX. There are also aggregate indicators of risk aversion created by financial institutions such as JP Morgan's Liquidity, Credit and Volatility Index. The consequence of incorporating such dependence is that the problem now becomes

[1] These models can be seen as a particular example of the studies on state-dependent utility/preference by Karni [20]. In this case the dependency has an explicit functional form as $\gamma(\cdot)$.

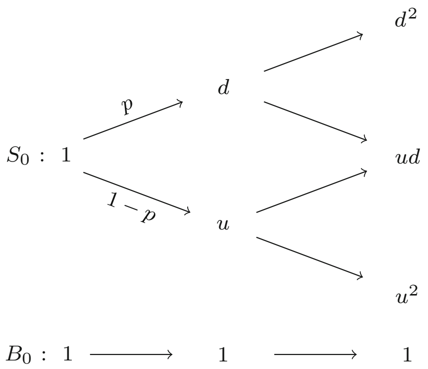
time-inconsistent, as the risk aversion will likely to be different at a later time leading to a different objective function to optimize. An example is provided in the next section as an illustration. We will follow closely the methodology described in [1]. As we will see later, a system of equations of the HJB type can be obtained in this manner, which admits the equilibrium solution via first order conditions. And when the risk aversion is constant, this system will degenerate to the classical HJB equation.

2.1 Time-Inconsistency and Wealth-Dependent Risk Aversion

To keep the dimension small, we start by describing the time-inconsistency problem with the risk aversion being dependent on the current wealth level, X_t. Since the current wealth level is an indicator on how much loss (downside risk) one is able to bear, we believe this dependence is natural. We will illustrate the derivation of the HJB-type system of equations in this case, which can be easily extended to cases where risk aversion depends on other state variables.

2.1.1 An Illustration

We can use a simple two-period binomial tree to illustrate the time inconsistency that results from the wealth-dependent risk aversion. Let $k \in \{0, 1, 2\}$ denote the time periods. Suppose there are two assets S_k and B_k, B_k being the risk free asset and S_k being the risky one with $u > 1$, $d < 1$ and $p \in [0,\ 1]$ as the usual parameters in a binomial tree model. We also assume both assets have value equal to 1 at time 0 and we have zero interest rate so $B_k = 1\ \forall\, k$.



Let X_k denote our wealth at time k and suppose $X_0 = 1$ for simplicity. We use an exponential utility function $U(x) = -e^{-\gamma x}$ here, and we let the risk aversion γ be a function of the current wealth level (denoted as x here):

$$\gamma(x) = \begin{cases} a & \text{if } \ x > 1 \\ 1 & \text{if } \ x = 1 \\ b & \text{if } \ x < 1, \end{cases}$$

for $a, b > 0$.

Let $0 < \pi < 1$ denote the amount of wealth invested in S_k at time 0. At time $k = 0$, with wealth $X_0 = 1$, the expected utility of terminal wealth $X_2^\pi := \pi S_2 + (1-\pi)B_2$ can be written as:

$$\begin{aligned} \mathbb{E}[U(X_2^\pi)] &= -e^{-1}\left\{p^2 e^{\pi(1-u^2)} + 2p(1-p)e^{\pi(1-ud)} + (1-p)^2 e^{\pi(1-d^2)}\right\} \\ &=: -e^{-1} f_1(\pi), \end{aligned}$$

where the risk aversion $\gamma = 1$. At time $k = 1$, depending on whether the stock price goes up or down, the risk aversion will becomes $\gamma = a$ or b because we have either $X_1 > 1$ or $X_1 < 1$. The expected utility of X_2^π at time 1 is either

$$\mathbb{E}[U(X_2^\pi) \mid S_1 = u] = -e^{-1}\left\{pe^{\pi(1-u^2)a} + (1-p)e^{\pi(1-ud)a}\right\} =: -e^{-1} f_2(\pi),$$

or

$$\mathbb{E}[U(X_2^\pi) \mid S_1 = d] = -e^{-1}\left\{pe^{\pi(1-ud)b} + (1-p)e^{\pi(1-d^2)b}\right\} =: -e^{-1} f_3(\pi).$$

Remark 2.1 It is possible to choose $p, \ u, \ d, \ \pi$ such that

$$\frac{\partial}{\partial\pi} f_1(\pi) > 0, \quad \frac{\partial}{\partial\pi} f_2(\pi) < 0 \quad \text{and} \quad \frac{\partial}{\partial\pi} f_3(\pi) < 0.$$

For instance, if $a = 0.5$ and $b = 2$, then by choosing $u = 2$, $d = 0.5$, $p = 0.5$ and $\pi = 0.5$ we can obtain the desired inequalities.

Suppose we have Portfolio #1 that has π in the stock and $1-\pi$ in the bank, and Portfolio #2 that has $\pi - \epsilon$ in stock and $1 - \pi + \epsilon$ in the bank for an infinitesimal positive amount ϵ. The signs of the first derivatives in Remark 2.1 tell us that, at the second period, Portfolio #1 is always favored over Portfolio #2. However, at time 0, Portfolio #2 is the better one. We recall the definition of time consistent utility function, such as in [7, 21]:

Definition 2.2 A dynamic utility function $(U_t)_{t=0}^T$ is time-consistent if for all $X, \ Y \in L(\mathcal{F}_T)$ and $t \in 0, \ldots, T-1$,

$$U_{t+1}(X) \geq U_{t+1}(Y) \qquad \text{implies} \ \ U_t(X) \geq U_t(Y).$$

We can see that in our case the preference between Portfolio #1 and #2 is "flipped" in the two periods, which clearly violates the definition of time consistency.

2.1.2 Formal Problem Setup in Continuous Time

We use the standard two-asset framework where we have a risky stock S_t and a risk-free bond that we can invest our wealth in. By assuming zero interest rate or working under the discounted unit, we only need to define the stock dynamics:

$$dS_t = \mu S_t \, dt + \sigma S_t \, dW_t, \tag{1}$$

where W_t denotes a standard Brownian motion, so S_t is a geometric Brownian motion (henceforth GBM). For the time being, we assume constant drift μ and volatility σ. The case where the volatility is stochastic will be discussed in Sect. 2.3 when we consider the volatility-dependent risk aversion as an extension of this problem.

Let X_t denote our wealth at time t, which consists of the cash amount π_t invested in the risky stock as well as the remaining part invested in the riskless bond. The wealth process X_t follows the controlled diffusion:

$$dX_t = \pi_t \frac{dS_t}{S_t} = \pi_t [\mu \, dt + \sigma \, dW_t]. \tag{2}$$

The optimization problem is to maximize the expected utility of terminal wealth at time T among all admissible strategies $\boldsymbol{\pi}$, given the wealth level being $X_t = x$ at time t. This problem can be represented using the value function $V(t, x)$

$$\begin{aligned} V(t, x) &:= \sup_{\boldsymbol{\pi} \in \boldsymbol{\Pi}} \mathbb{E}_{t,x} \left[U(X_T^{\boldsymbol{\pi}}, \ \gamma(X_t)) \right] \\ &= \sup_{\boldsymbol{\pi} \in \boldsymbol{\Pi}} \mathrm{J}(t, x, \boldsymbol{\pi}), \end{aligned} \tag{3}$$

where $\boldsymbol{\Pi}$ is the set of all admissible strategies that are adapted to the filtration $(\mathcal{F}_s)$ generated by the stock price process and which satisfy $\mathbb{E}\left[\int_t^T \pi_s^2 ds < \infty\right]$, and $\gamma(X_t)$ is the risk aversion that we fix for our future self at time T based on the current wealth level X_t. Here $U(x_1, \gamma(x_2))$ denotes the von Neumann-Morgenstern utility function which is a twice differentiable, concave function in $x_1 \in \mathbb{R}^+$.

Remark 2.3 Note that in $U(x_1, \gamma(x_2))$, x_1 is the wealth at future for which we want to compute the utility, with risk aversion computed using current wealth level x_2. In order to retain the differentiability and concavity at terminal time T when x_1 and x_2 coincide, we also require $\gamma(x_2)$ to be chosen such that

- $U(x_1, \gamma(x_2))$ is twice differentiable in x_2;
- $\left(U_{x_1x_1} + U_{x_2x_2} + 2U_{x_1x_2}\right) |_{x_2=x_1} < 0$ for all $x_1 > 0$.

In the classical case, the risk aversion γ is constant so we can suppress its argument by denoting the utility function as $U = U^{\gamma}(x)$. Then the optimal strategy can be computed from the HJB equation associated with the value function, and the DPP, from which the HJB equation is derived, guarantees that the optimal strategy $\boldsymbol{\pi}^*$ computed at the initial time will remain optimal at a later time. A rigorous proof of the derivation of HJB equation from DPP can be found in, for example, Pham [25]. In this case, the optimal strategy takes the form:

$$\pi_t^* = \begin{cases} \dfrac{\lambda}{\sigma}\dfrac{1}{\gamma} & \text{for exponential utility functions} \\ \dfrac{\lambda}{\sigma}\dfrac{x}{\gamma} & \text{for log and power utility functions,} \end{cases}$$

where $X_t = x$ and λ is the Sharpe ratio defined by $\lambda := \frac{\mu}{\sigma}$.

Now, as we have made the risk aversion wealth dependent, intuitively the optimal strategy might be obtained by replacing the constant γ with $\gamma(x)$ in the above expressions. Is it really the case? It turns out that this is not so trivial, since we cannot even formulate the HJB equation (in the classical DPP sense) once we allow the risk aversion to change with the current wealth level. As our objective function changes constantly, our future selves will not solve the "remaining" part of the optimization problem that our current self is facing now.

Using a game-theoretic approach, we can think of it as a game played by a number of ordered players (our selves at different times), each of whom has his own utility function and has temporary control over the resource (wealth). For a particular player, when the resource is in his possession (obtained from the previous player), he can choose the strategy to be applied to the resource at this particular moment. After that, the player has to pass on the resource to the next player and he will no longer be able to apply strategies to it or control what other players' strategies will be. As the game is played by a continuum of players in the continuous time setting, each player would have to play against all his future selves.

To define the equilibrium in this game, we follow the explanation given in [11], in which it was assumed that the current self has the ability to commit all future selves to his decision up to a small period $\epsilon > 0$. Thus the player can form a small coalition with players in the near future. Now suppose $\boldsymbol{\pi} \equiv (\pi_s)_{s \in [t,T]}$ is an admissible policy (all strategies over time). Define another policy $\boldsymbol{\pi}_\epsilon$ as:

$$\boldsymbol{\pi}_\epsilon = \begin{cases} \pi, & s \in [t,\ t+\epsilon] \\ \pi_s, & s \in (t+\epsilon,\ T], \end{cases} \tag{4}$$

where π can be any strategy that makes $\boldsymbol{\pi}_\epsilon$ admissible. Then the following from [11] gives the definition of the equilibrium policy.

Definition 2.4 A policy $\bar{\boldsymbol{\pi}} : (t, x) \to \mathbb{R}$ is an equilibrium one if for any $t,\ x > 0$ and any arbitrary π,

$$\lim_{\epsilon \downarrow 0} \frac{\mathrm{J}(t, x, \bar{\boldsymbol{\pi}}) - \mathrm{J}(t, x, \bar{\boldsymbol{\pi}}_\epsilon)}{\epsilon} \geq 0$$

where J is our objective function.

This definition means that, if we are using the equilibrium policy $\bar{\boldsymbol{\pi}}$, then we will not be better off by committing the immediate future selves to our action instead of letting them choose the best strategy in their views. This also means that the equilibrium policy computed at one time should coincide, from the next period onward, with the equilibrium policy computed at the next period. The equilibrium policy is therefore time-consistent as the future selves have no incentives to deviate from this path. We refer the readers to the paragraphs following Definition 1 in Ekeland and Lazrak [12] for a detailed explanation about the equilibrium strategy in discrete time setting. The definition leads to the following result as appeared in, e.g. [1]:

Proposition 2.5 *Assuming sufficient regularity, the equilibrium value function and Markovian policy satisfy the following extended HJB-type system:*

$$\begin{cases} \sup_{\pi_t \in \mathbb{R}} \left(\mathcal{A}^{\pi_t} V(t,x) - \mathcal{A}^{\pi_t} f(t,x,x) + \mathcal{A}^{\pi_t} f^w(t,x) \right) = 0 \\ \mathcal{A}^{\pi_t^*} f^w(t,x) = 0 \\ V(T,x) = U(x, \gamma(x)) \\ f(T,x,w) = U(x,\gamma(w)), \end{cases} \tag{5}$$

where $\mathcal{A}^{\pi_t}$ contains the infinitesimal generator of the wealth process taking the form:

$$\mathcal{A}^{\pi_t} g(t,x) = g_t + \mu \pi_t g_x + \frac{1}{2}\sigma^2 \pi_t^2 g_{xx}$$

$$\begin{aligned} \mathcal{A}^{\pi_t} h(t,x,x) = h_t + \mu\pi_t h_x + \mu\pi_t h_{w|w=x} + \frac{1}{2}\sigma^2\pi_t^2 h_{xx} \\ + \frac{1}{2}\sigma^2\pi_t^2 h_{ww|w=x} + \sigma^2\pi_t^2 h_{xw|w=x}, \end{aligned}$$

and $f^w(t,x)$ means fixing the w variable of $f(t,x,w)$ as constant.

Proof We need to define the following "auxiliary value function":

$$f(t,x,w) = \mathbb{E}_{t,x}\left[U(X_T^{\boldsymbol{\pi}^*},\ \gamma(w)) \right],$$

which is made from $V(t,x)$ by making the initial wealth value in $\gamma(\cdot)$ vary independently from X_t, and where $\boldsymbol{\pi}^*$ denotes the equilibrium strategy. For every fixed w, $\gamma(w)$ can be treated as a constant, as w and x are independent. Thus $f(t,x,w)$ is the value function for a Merton problem with constant risk aversion parameter $\gamma(w)$.

By construction of $\boldsymbol{\pi}_\epsilon$ from the definition, we have the following equality:

$$\begin{aligned}
\mathbb{E}_{t,x}\left[\mathrm{J}(t+\epsilon, X_{t+\epsilon}, \boldsymbol{\pi}_\epsilon)\right] &= \mathbb{E}_{t,x}\left[f(t+\epsilon, X_{t+\epsilon}, X_{t+\epsilon})\right] \\
&= \mathbb{E}_{t,x}\left[f(t+\epsilon, X_{t+\epsilon}, X_{t+\epsilon})\right] + \mathrm{J}(t, x, \boldsymbol{\pi}) \\
&\quad - \mathbb{E}_{t,x}\left[U(X_T^{\pi}, \gamma(x))\right] \\
&= \mathbb{E}_{t,x}\left[f(t+\epsilon, X_{t+\epsilon}, X_{t+\epsilon})\right] + \mathrm{J}(t, x, \boldsymbol{\pi}) \\
&\quad - \mathbb{E}_{t,x}\left[\mathbb{E}\left[U(X_T^{\pi}, \gamma(x)) \mid X_{t+\epsilon}, t\right]\right] \\
&= \mathbb{E}_{t,x}\left[f(t+\epsilon, X_{t+\epsilon}, X_{t+\epsilon})\right] + \mathrm{J}(t, x, \boldsymbol{\pi}) \\
&\quad - \mathbb{E}_{t,x}\left[f(t+\epsilon, X_{t+\epsilon}, x)\right].
\end{aligned}$$

Since $\mathrm{J}(t+\epsilon, X_{t+\epsilon}, \boldsymbol{\pi}_\epsilon) = V(t+\epsilon, X_{t+\epsilon})$, we can write the above equation as:

$$\begin{aligned}
\mathbb{E}_{t,x}\left[V(t+\epsilon, X_{t+\epsilon})\right] &= \mathbb{E}_{t,x}\left[f(t+\epsilon, X_{t+\epsilon}, X_{t+\epsilon})\right] \\
&\quad + \mathrm{J}(t, x, \boldsymbol{\pi}) - \mathbb{E}_{t,x}\left[f(t+\epsilon, X_{t+\epsilon}, x)\right].
\end{aligned}$$

Taking the supremum and rearranging the equation, we get

$$\begin{aligned}
&\sup_{\boldsymbol{\pi}\in\boldsymbol{\Pi}} \Big(\mathbb{E}_{t,x}\left[V(t+\epsilon, X_{t+\epsilon})\right] - V(t,x) + \mathbb{E}_{t,x}\left[f(t+\epsilon, X_{t+\epsilon}, x)\right] \\
&\qquad - \mathbb{E}_{t,x}\left[f(t+\epsilon, X_{t+\epsilon}, X_{t+\epsilon})\right]\Big) = 0.
\end{aligned}$$

Now we take the limit $\epsilon \to 0$,

$$\sup_{\pi_t}\left(\mathcal{A}^{\pi_t} V(t,x) - \mathcal{A}^{\pi_t} f(t,x,x) + \mathcal{A}^{\pi_t} f^{w|w=x}(t,x)\right) = 0. \tag{6}$$

Meanwhile, for every fixed w, $f(t, X_t, w)$ corresponds to a martingale process and thus it must satisfy the PDE

$$A^{\pi_t^*} f^w(t,x) = 0, \tag{7}$$

where π^* is the equilibrium policy appeared in the definition of $f(t,x,w)$. In addition, there are two terminal conditions for V and f:

$$\begin{cases} V(T,x) = U(x, \gamma(x)) \\ f(T,x,w) = U(x, \gamma(w)). \end{cases} \tag{8}$$

We get the extended HJB-type system by combining Eqs. (6), (7) and (8).

The verification theorem provided by [1] holds here, we shall quote:

Theorem 2.6 (Bjork and Murgoci) *Assume that* $(V(t,x), f(t,x,w))$ *is a solution of the system defined in* (5)*, and that the strategy path* $\boldsymbol{\pi}^*$ *realizes the supremum* *in the equation. Then* π^* *is an equilibrium policy, and* $V(t, x)$ *is the corresponding value function.*

Proof See [1].

When writing out the first two equations in (5) explicitly, we get the following two PDEs:

$$\begin{cases} V_t + \sup_{\pi} \left\{ \mu\pi(V_x - f_w) + \frac{1}{2}\sigma^2\pi^2(V_{xx} - f_{ww} - 2f_{xw}) \right\} = 0 \\ f_t + \mu\pi^* f_x + \frac{1}{2}\sigma^2\pi^{*2} f_{xx} = 0. \end{cases} \tag{9}$$

Note that all w partial derivatives are evaluated at the point $w = x$.

We can find the equilibrium strategy by the first order condition:

$$\pi^* = -\frac{\lambda}{\sigma} \frac{V_x - f_w}{V_{xx} - f_{ww} - 2f_{xw}} \mid_{w=x}, \tag{10}$$

where λ denotes the constant Sharpe ratio. Inserting (10) back into (9) and we obtain the following two PDEs to solve for $V(t, x)$ and $f(t, x, w)$

$$\begin{cases} V_t - \frac{1}{2}\lambda^2 \frac{V_x^2 - 2V_x f_y + f_w^2}{V_{xx} - f_{ww} - 2f_{xw}} = 0 \\ f_t + \lambda^2 \left[\frac{f_x(f_w - V_x)}{V_{xx} - f_{ww} - 2f_{xw}} + \frac{1}{2} \frac{(V_x^2 - 2V_x f_w + f_w^2) f_{xx}}{(V_{xx} - f_{ww} - 2f_{xw})^2} \right] = 0. \end{cases} \tag{11}$$

2.1.3 A Remark: Why Two Equations Instead of One?

As we can see from the above, we now face an HJB-type system of two equations instead of solving one single HJB equation as in the time-consistent case. It turns out this is essential for characterizing the equilibrium strategy and value function. In the definition of the equilibrium strategy, coalition is allowed for an infinitesimal period, during which we are actually solving a Merton problem with constant risk aversion. That is what the function $f(t, x, w)$ represents when setting $w = x$ and it is the time-consistent part of the problem. After this infinitesimal period, however, the evolution of the value function cannot be characterized by this function $f(t, x, w)$ any more, as the problem now is time-inconsistent. This is the reason we need $V(t, x)$ as our value function.

In Harris and Laibson [17], the dynamic consumption choice problem with quasi-hyperbolic discounting was also solved by the equilibrium strategy and value function which were defined similarly using two functions. There is a continuation-value function characterizing the dynamics of the true time-inconsistent value function and there is another current-value function. The current-value function is used locally at the current point (t, x) to derive the equilibrium strategy. Our HJB-type system has a strong analogy to the two functions in their work.

Another way of describing this is that the true value function $V(t, x)$ cannot be solved alone. For each point (t, x), the value of $V(t, x)$ is determined by another non-local function $f(t, x, w)$ by setting $w = x$ in its third argument, i.e. $V(t, x) = f(t, x, x)$. The first equation in (5) can be considered as a PDE of the non-local type. For the non-exponential discounting problem in [11], a non-local integro-PDE was obtained, where the dynamics of the value function depends on an integral of the value function at all future time. In general, non-local PDEs are very difficult to solve.

2.1.4 Asymptotic Expansions

If the time-inconsistent problem is close to a time-consistent one, we can approximate the first problem very effectively using the latter by asymptotic methods. Here we look at the special case where the risk aversion only varies slowly with the wealth level, i.e. it is close to the case of constant risk aversion. Mathematically, this corresponds to

$$\gamma(x) = \gamma_0 + \epsilon\gamma_1(x) + \cdots$$

for positive $\epsilon \ll 1$. We look for an expansion of the form

$$V(t, x) = V_0(t, x) + \epsilon V_1(t, x) + \cdots ,$$

and

$$f(t, x, w) = f_0(t, x) + \epsilon f_1(t, x, w) + \cdots ,$$

for the equations in (11).

We first introduce a few notations.

Definition 2.7 We define the risk tolerance to be

$$R := -\frac{V_{0,x}}{V_{0,xx}};$$

and use $\mathcal{D}_k$ to denote

$$\mathcal{D}_k := R^k \frac{\partial^k}{\partial x^k};$$

finally define the linear operator $\mathcal{L}_{t,x}$ as

$$\mathcal{L}_{t,x} := \partial_t + \lambda^2\mathcal{D}_1 + \frac{1}{2}\lambda^2\mathcal{D}_2.$$

Collecting zeroth order terms in (11) we get

$$\begin{cases} \mathcal{L}_{t,x} V_0 = 0 \\ \mathcal{L}_{t,x} f_0 = 0, \end{cases} \tag{12}$$

with terminal conditions $V_0(T, x) = U(x, \gamma_0)$ and $f_0(T, x) = U(x, \gamma_0)$. Since f_0 and V_0 have the same terminal condition, we find $V_0(t, x) = f_0(t, x)$ which is the classical Merton value function with constant risk aversion parameter γ_0.

At the first order, we have:

$$\begin{cases} \mathcal{L}_{t,x} V_1 = \lambda^2 R\, f_{1,w} + \dfrac{\lambda^2}{2} R^2 \left(f_{1,ww} + 2 f_{1,xw}\right) \\ \\ \mathcal{L}_{t,x} f_1 = 0, \end{cases} \tag{13}$$

with terminal conditions $V_1(T, x) = \frac{\partial U}{\partial \gamma}(x, \gamma_0)\gamma_1(x)$ and $f_1(T, x, w) = \frac{\partial U}{\partial \gamma}(x, \gamma_0)$ $\gamma_1(w)$. The following proposition will be useful for solving the order ϵ PDEs.

Lemma 2.8 *We have*

$$\frac{\partial}{\partial \gamma} \mathcal{L}_{t,x} V_0 = \mathcal{L}_{t,x} \left(\frac{\partial V_0}{\partial \gamma} \right). \tag{14}$$

Proof For any function v,

$$\begin{aligned} \frac{\partial}{\partial \gamma} \mathcal{L}_{t,x} v &= \frac{\partial}{\partial \gamma} \left(v_t + \lambda^2 R v_x + \frac{1}{2} \lambda^2 R^2 v_{xx} \right) \\ &= \mathcal{L}_{t,x} \left(\frac{\partial v}{\partial \gamma} \right) + \lambda^2 \frac{\partial R}{\partial \gamma} v_x + \lambda^2 \left(R \frac{\partial R}{\partial \gamma} \right) v_{xx}. \end{aligned}$$

The last two terms cancel out when $v = V_0$ since $R = -\frac{V_{0,x}}{V_{0,xx}}$.

Lemma 2.8 will lead us to the solutions of f_1 and V_1.

Proposition 2.9 *The solution to the second equation in* (13) *is*

$$f_1(t, x, w) = \gamma_1(w) \frac{\partial f_0}{\partial \gamma}.$$

Therefore the order ϵ value function is

$$V_1(t, x) = \gamma_1(x) \frac{\partial V_0}{\partial \gamma}. \tag{15}$$

Proof By direct substitution and verification.

2.1.5 Effect on the Trading Strategy

Recall the equilibrium strategy from (10)

$$\pi^* = -\frac{\lambda}{\sigma}\frac{V_x - f_w}{V_{xx} - f_{ww} - 2f_{xw}}.$$

We plug in $V = V_0 + \epsilon V_1$ and $f = f_0 + \epsilon f_1$,

$$\begin{aligned}\pi^* &= -\frac{\lambda}{\sigma}\frac{V_{0,x} + \epsilon h_x \gamma_1(x)}{V_{0,xx} + \epsilon h_{xx}\gamma_1(x)} + O(\epsilon^2)\\ &= \frac{\lambda}{\sigma}R(1+\epsilon\frac{h_x\gamma_1(x)}{V_{0,x}})(1-\epsilon\frac{h_{xx}\gamma_1(x)}{V_{0,xx}}) + O(\epsilon^2)\\ &= \frac{\lambda}{\sigma}R\left[1+\epsilon\gamma_1(x)\left(\frac{h_x}{V_{0,x}} - \frac{h_{xx}}{V_{0,xx}}\right)\right] + O(\epsilon^2), \end{aligned} \tag{16}$$

where we denote $h := \frac{\partial V_0}{\partial \gamma}$. Thus the equilibrium strategy will deviate from the optimal strategy in the case of constant risk aversion γ_0 by a fraction given by $\epsilon\gamma_1(x)\left(\frac{h_x}{V_{0,x}} - \frac{h_{xx}}{V_{0,xx}}\right)$.

2.1.6 Power Utility Case

Recall that the power utility function with constant risk aversion parameter γ is:

$$U(x) = \frac{x^{1-\gamma}}{1-\gamma}.$$

For the Merton problem with power utility and constant risk aversion, the value function $V(t,x)$ satisfies

$$V_t - \frac{1}{2}\lambda^2\frac{V_x^2}{V_{xx}} = 0,$$

with terminal condition $V(T,x) = \frac{x^{1-\gamma}}{1-\gamma}$. The solution for the PDE above is given by

$$V(t,x) = \frac{x^{1-\gamma}}{1-\gamma}e^{\frac{\lambda^2}{2}\left(\frac{1-\gamma}{\gamma}\right)(T-t)}.$$

This is our zeroth order value function $V_0(t,x)$ once we replace γ with γ_0. We can find the first order correction by taking the partial derivative w.r.t. γ and multiplying by $\gamma_1(x)$:

$$V_1(t,x) = \gamma_1(x)\left[\frac{1}{1-\gamma_0} - \log(x) - \frac{\lambda^2(T-t)}{2\gamma_0^2}\right]V_0(t,x). \tag{17}$$

The equilibrium trading strategy up to the first order is thus given by:

$$\begin{aligned}\pi^* &= \frac{\lambda}{\sigma\gamma_0}\left[1+\epsilon\gamma_1(x)\left(\frac{h_x}{V_{0,x}} - \frac{h_{xx}}{V_{0,xx}}\right)\right]\\ &= \frac{\lambda}{\sigma\gamma_0}\left[1+\epsilon\frac{\gamma_1(x)}{\gamma_0}\right].\end{aligned}$$

We have provided some plots for the power utility case. Figure 1a compares a power utility function with constant risk aversion $\gamma = 2$ to the one with risk aversion slowly decreasing in wealth. For this illustration, we have chosen $\gamma_1(x) = -\tan^{-1}(x-10)$ and $\epsilon = 0.01$ which retain the twice differentiability and concavity of the utility function. Figure 1b compares the Merton optimal strategy with the equilibrium strategy up to the first order correction.

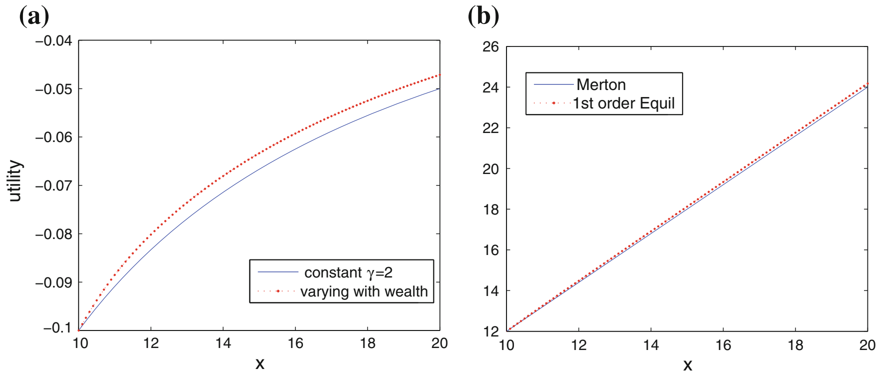


Fig. 1 Plots of utility functions and equilibrium trading strategies (up to 1st order correction) against wealth level in the case of power utility function. We chose $\mu = 0.15$, $r = 0$, $\sigma = 0.25$, $\gamma_0 = 2$, $\gamma_1(x) = -\tan^{-1}(x-10)$ and $\epsilon = 0.01$. Risk aversion is modeled as slowly decreasing with wealth level here and the corresponding utility function is still concave and slightly above the utility function with constant risk aversion. Moreover, we see that the equilibrium strategy is slightly above the Merton strategy due to a lower risk aversion. **a** Utility functions. **b** Equilibrium Investment Strategies

2.2 Utility-Indifference Pricing with Wealth-Dependent Risk Aversion

One of the immediate applications of wealth-dependent risk aversion is indifference pricing, where the (buyer's) price of the option is set such that the buyer has the same expected utility no matter he chooses to invest in a portfolio without the option or to invest in another portfolio with the option but paying a price at the beginning. In scenarios where the option is likely to cost a significant portion of the investor's wealth, for example constructing a power plant or start an R&D project as often considered in real option valuation problems it is possible for the investor to become more risk averse after he purchases the option. Wealth-dependent risk aversion can be used to capture this change. Here we look at the indifference pricing of an option written on a non-traded asset. The controlled wealth process follows

$$dX_t^{\pi,x} = \pi_t \, dS_t^{(1)} + r(X_t^{\pi,x} - \pi_t S_t^{(1)})dt,$$

where the price $S_t^{(1)}$ of the traded asset follows the geometric Brownian motion with drift μ

$$dS_t^1 = \mu S_t^{(1)} \, dt + \sigma S_t^{(1)} \, dW_t^{(1)}.$$

The option written on the non-traded asset $S_t^{(2)}$ has payoff $C(S_T^{(2)})$ at terminal time T. And $S_t^{(2)}$ follows the SDE

$$dS_t^{(2)} = p \, dt + q \, dW_t^{(2)},$$

where $W_t^{(1)}$ and $W_t^{(2)}$ have correlation ρ. Now assuming $r = 0$, the value function for the Merton problem without the option is

$$\begin{aligned} V(x, 0) &= - e^{-\gamma_0 x - \lambda^2 T/2} + \epsilon \gamma_1(x) x e^{-\gamma_0 x - \lambda^2 T/2} + o(\epsilon^2) \\ &= - e^{-\gamma_0 x - \lambda^2 T/2} (1 - \epsilon \gamma_1(x) x) + o(\epsilon^2). \end{aligned} \tag{18}$$

Note that we are using exponential utility here to simplify the calculations. Now the value function with a long position in k units of the option is given by

$$\begin{aligned} V(x - p_k, k) = &- e^{-\gamma_0 (x - p_k) - \lambda^2 T/2} (1 - \epsilon \gamma_1(x) x) \\ &\left(E^{Q_0} [e^{-k\gamma_0 (1-\rho^2) C(S_T^{(2)})} (1 - k \epsilon \gamma_1(x) (1 - \rho^2) C(S_T^{(2)}))] \right)^{1/(1-\rho^2)} \\ = &- e^{-\gamma_0 (x - p_k) - \lambda^2 T/2} (1 - \epsilon \gamma_1(x) x) \left(E^{Q_0} [e^{-k\gamma_0 (1-\rho^2) C(S_T^{(2)})}] \right)^{1/(1-\rho^2)} \\ &+ \epsilon k \gamma_1(x) e^{-\gamma_0 (x - p_k) - \lambda^2 T/2} \left(E^{Q_0} [e^{-k\gamma_0 (1-\rho^2) C(S_T^{(2)})}] \right)^{\rho^2/(1-\rho^2)} \end{aligned}$$

$$E^{Q_0}[C(S_T^{(2)})e^{-k\gamma_0(1-\rho^2)C(S_T^{(2)})}], \tag{19}$$

where Q_0 is the probability measure under which $S_t^{(2)}$ has a new drift $p - \rho\lambda q$ but the same diffusion q.

For the solution of p_k we seek the following expansion:

$$p_k = p_k^{(0)} + \epsilon p_k^{(\epsilon)} + O(\epsilon^2)(\epsilon^2).$$

Consequently,

$$V(x - p_k^{(0)} - \epsilon p_k^{(\epsilon)}, k) = V^0(x,0)e^{\gamma_0 p_k^{(0)}}\Theta_T^{\frac{1}{1-\rho^2}}\Bigg[1 + \epsilon\Bigg(1 - \gamma_1(x - p_k^{(0)})x \\ - \frac{k\gamma_1(x - p_k^{(0)})E^{Q_0}\left[C(S_T^{(2)})\Lambda_T\right]}{\Theta_T}\Bigg)\Bigg] + O(\epsilon^2) \tag{20}$$

where we have used the following notation

$$\begin{aligned} V^0(x,0) &:= -e^{-\gamma_0 x - \lambda^2 T/2}, \\ \Lambda_T &:= e^{-k\gamma_0(1-\rho^2)C(S_T^{(2)})}, \\ \Theta_T &:= E^{Q_0}[\Lambda_T], \end{aligned}$$

and that $\gamma_1(x - p_k^{(0)} - \epsilon p_k^{(1)}) \approx \gamma_1(x - p_k^{(0)})$. Now we just need to equate (18) and (20). At the zeroth order,

$$-e^{-\gamma_0 x - \lambda^2 T/2} = -e^{-\gamma_0(x - p_k^{(0)}) - \lambda^2 T/2}\Theta_T^{1/(1-\rho^2)},$$

from which we can find the zeroth order indifference price:

$$p_k^{(0)} = -\frac{1}{(1-\rho^2)\gamma_0}\log\Theta_T.$$

At order ϵ, after substituting in $p_k^{(0)}$, we have

$$\begin{aligned} -\gamma_1^+ x V^0(x,0) = p_k^{(\epsilon)} V^0(x,0) - \gamma_1^- x V^0(x,0) \\ - k\gamma_1^- \Theta_T^{-1} E^{Q_0}\left[C(S_T^{(2)})\Lambda_T\right] V^0(x,0), \end{aligned}$$

from which we can get the order ϵ correction to the indifference price

$$p_k^{(\epsilon)} = (\gamma_1^- - \gamma_1^+)x + \frac{k\gamma_1^- E^{Q_0}[C(S_T^{(2)})\Lambda_T]}{\Theta_T}.$$

As we have assumed the investor will become more risk averse when holding the option, i.e. $\gamma_1^- \geq \gamma_1^+$, the correction $p_k^{(\epsilon)}$ above tells us that the true indifference price will be higher than the constant risk aversion case for $k > 0$ and $C(\cdot) \geq 0$.

2.3 Stochastic Volatility Models with Volatility-Dependent Risk Aversion

Since first introduced by Hull and White [19] for pricing options, stochastic volatility models have gained wide popularity as they could reproduce features about the implied volatility surface which are missing in the standard Black-Scholes models. Incorporating the stochastic volatility framework is also one of the many extensions being studied recently for the classical Merton portfolio optimization problems.

One way to look at the problem is to use the timescale stochastic volatility asymptotics, which has been applied to many option pricing problems, see Fouque et al. [14] and the references therein. In this framework, the volatility is assumed to have a fast mean-reverting factor following a speeded-up diffusion process and/or a slow factor following a slowed-down diffusion. The empirical evidence to support the multiscale stochastic volatility model can be found in Chernov [8]. When Fouque et al. [15] treated the Merton problem in this way, they obtained in explicit forms both the fast and slow scale corrections to the value function, which resemble the stochastic volatility corrections for pricing European-style options.

A natural question to ask is whether volatility, an indicator of investment risk, would affect the risk aversion parameter, a measure of investor's attitude towards risk. This question is nontrivial when the assumption of constant volatility has been dropped. Intuitively the answer should be yes, as investors would usually focus on preserving capitals when the risky assets have high volatility, thus becoming more risk averse. Empirical results also support this argument, which can be found in Scheicher [27] and Tarashev et al. [30]. Scheicher [27] discovered a positive relationship between the implied risk aversion in German equity market and the implied volatility of the US market, and Tarashev et al. [30] concluded from results obtained in different equity and fixed income markets that higher risk aversion is linked to higher volatilities and this is more noticeable in the equity markets.

In the next section, we will look at the case where risk aversion is a function of the slow scale volatility factor so it slowly varies. Our rationale is that the effect from the fast scale factor would likely be averaged out over the investment horizon if it is long enough; and it is the general trend of the volatility that would reflect the change in investors' risk aversion. We will derive the extended HJB system and carry out the slow scale asymptotic expansion in this case to approximate the value function and equilibrium strategy. Using power utility function, we will show that it is possible to obtain results similar to Fouque et al. [15] but with an additional correction term to account for change in risk aversion.

2.3.1 Slow Scale Stochastic Volatility Model

Suppose that the volatility slowly fluctuates as a general diffusion process and the stock price follows a geometric Brownian Motion:

$$\begin{cases} dS_t = \mu S_t\, dt + \sigma(Z_t) S_t\, dW_t \\ dZ_t = \delta c(Z_t)\, dt + \sqrt{\delta} g(Z_t)\, dW_t', \end{cases} \tag{21}$$

where W_t and W_t' are Brownian motions with correlation $\rho' \in (-1, 1)$. We have the wealth process:

$$dX_t = \pi_t \mu dt + \pi_t \sigma(Z_t) dW_t, \tag{22}$$

and the associated infinitesimal generator:

$$\begin{aligned} \mathcal{A}^{\pi_t}(t, x, z) = \partial_t + \pi_t \mu(z) \partial_x + \frac{1}{2}\pi_t^2 \sigma(z)^2 \partial_x^2 + \delta c(z) \partial_z \\ + \frac{1}{2}\delta g(z)^2 \partial_z^2 + \sqrt{\delta} \pi_t \rho g(z) \sigma(z) \partial_{xz}^2. \end{aligned}$$

The portfolio optimization problem we consider here is:

$$V(t, x, z) = \sup_{\pi} \mathbb{E}_{t,x,z}\left[U(X_T^{\pi}, \gamma(Z_t))\right], \tag{23}$$

where we make the risk aversion dependent on current level of the slow factor Z_t. The extended HJB system for the value function can be derived in the same way as the wealth-dependent risk aversion case (with a two dimensional state process now), which is given by

$$\begin{cases} \sup_{\pi_t}\{V_t + \pi_t \mu(z) V_x + \frac{1}{2}\pi_t^2 \sigma(z)^2 V_{xx} + \sqrt{\delta}\rho \pi_t g(z) \sigma(z)(V_{xz} - f_{xw}) \\ +\delta c(z)(V_z - f_w) + \frac{1}{2}\delta g(z)^2 (V_{zz} - f_{ww} - 2 f_{wz})\} = 0 \\ f_t + \pi_t^* \mu(z) f_x + \frac{1}{2}\pi_t^{*2} \sigma(z)^2 f_{xx} + \delta c(z) f_z + \frac{1}{2}\delta g(z)^2 f_{zz} \\ \quad +\sqrt{\delta}\rho \pi_t^* g(z) \sigma(z) f_{xz} = 0, \end{cases} \tag{24}$$

with terminal conditions:

$$\begin{cases} V(T,x,z) = U(x,\gamma(z)) \\ f(T,x,z,w) = U(x,\gamma(w)). \end{cases} \tag{25}$$

By the first order condition, the equilibrium strategy takes the form:

$$\pi_t^* = -\frac{\mu(z)V_x + \sqrt{\delta}\rho g(z)\sigma(z)[V_{xz} - f_{xw}]}{\sigma(z)^2 V_{xx}} \mid_{w=z}. \tag{26}$$

Plugging this equilibrium strategy back to the extended HJB system, we get:

$$\begin{cases} V_t - \dfrac{\{\mu(z)V_x + \sqrt{\delta}\rho g(z)\sigma(z)[V_{xz} - f_{xw}]\}^2}{2\sigma(z)^2 V_{xx}} \\ \quad + \delta\{c(z)(V_z - f_w) + \dfrac{g(z)^2}{2}[V_{zz} - f_{ww} - 2f_{zw}]\} = 0 \\ f_t - \dfrac{\mu(z)V_x + \sqrt{\delta}\rho g(z)\sigma(z)[V_{xz} - f_{xw}]}{\sigma(z)^2 V_{xx}}[\mu(z)f_x + \sqrt{\delta}\rho g(z)\sigma(z)f_{xz}] \\ \quad + \dfrac{\{\mu(z)V_x + \sqrt{\delta}\rho g(z)\sigma(z)[V_{xz} - f_{xw}]\}^2}{2\sigma(z)^2 V_{xx}^2} f_{xx} + \delta[c(z)f_z + \dfrac{g(z)^2}{2} f_{zz}] = 0. \end{cases} \tag{27}$$

Now we assume the risk aversion $\gamma(Z_t)$ takes the form:

$$\gamma(Z_t) = \gamma_0 + \sqrt{\delta}\gamma_1(Z_t) + o(\delta), \tag{28}$$

thus slowly varies with the slow scale volatility factor. And we expand V and f as:

$$\begin{aligned} V(t,x,z) &= V_0(t,x,z) + \sqrt{\delta}V_1(t,x,z) + \delta V_2(t,x,z) + o(\delta^{\frac{3}{2}}) \\ f(t,x,z,w) &= f_0(t,x,z,w) + \sqrt{\delta}f_1(t,x,z,w) \\ &\quad + \delta f_2(t,x,z,w) + o(\delta^{\frac{3}{2}}). \end{aligned} \tag{29}$$

Now introduce the risk tolerance function:

$$R := R(t,x,z) = -\frac{V_{0,x}(t,x,z)}{V_{0,xx}(t,x,z)},$$

and the differential operator:

$$\mathcal{D}_k := R^k \frac{\partial^k}{\partial x^k},$$

as well as the linear operator:

$$\mathcal{L}_{t,x,z} = \partial_t + \lambda(z)^2 \mathcal{D}_1 + \frac{1}{2}\lambda(z)^2 \mathcal{D}_2,$$

where $\lambda(z) := \frac{\mu}{\sigma(z)}$ denotes the Sharpe ratio.

We can find that $V_0(t,x,z) = f_0(t,x,z,z)$ is the Merton value function with Sharpe ratio fixed at $\lambda(z)$ and risk aversion parameter fixed at γ_0. As a result, order $\sqrt{\delta}$ equations become:

$$\begin{cases} \mathcal{L}_{t,x,z} V_1 + \rho g(z)\lambda(z)\mathcal{D}_1(V_{0,z} - \cancel{f_{0,w}}) = 0 \\ \mathcal{L}_{t,x,z} f_1 + \rho g(z)\lambda(z)\mathcal{D}_1 f_{0,z} = 0, \end{cases} \tag{30}$$

with terminal condition given by:

$$\begin{cases} V_1(T,x,z) = \gamma_1(z)\dfrac{\partial U}{\partial \gamma}(x,\gamma_0) \\ f_1(T,x,z,w) = \gamma_1(w)\dfrac{\partial U}{\partial \gamma}(x,\gamma_0). \end{cases}$$

Proposition 2.10 *The solution to* (30) *is given by:*

$$V_1(t,x,z) = \gamma_1(z)V_{0,\gamma} + \frac{1}{2}(T-t)\rho\lambda(z)g(z)\mathcal{D}_1 V_{0,z}. \tag{31}$$

Proof Using the result from Lemma 2.8, we can see that $\gamma_1(w) f_{0,\gamma}$ is a solution to the PDE problem below:

$$\begin{cases} \mathcal{L}_{t,x,z} f_1 = 0, \\ f_1(T,x,z,w) = \gamma_1(w)\dfrac{\partial U}{\partial \gamma}(x,\gamma_0), \end{cases}$$

which is the original order $\sqrt{\delta}$ PDE problem without the source term. Now if we can find the solution to the full PDE problem with zero terminal condition, we can find the full solution satisfying the original terminal condition by combining the two partial solutions. This can be done by making use of the "Vega-Gamma" relation in Lemma 3.1 of Fouque et al. [15], which states

$$f_{0,z} = -(T-t)\lambda(z)\lambda'(z)\mathcal{D}_2 f_0.$$

The second problem can be rewritten as follows

$$\begin{cases} \mathcal{L}_{t,x,z} f_1 = (T-t)\rho g(z)\lambda(z)^2\lambda'(z)\mathcal{D}_1\mathcal{D}_2 f_0 \\ f_1(T,x,z,w) = 0. \end{cases}$$

Using the commutativity property of $\mathcal{L}_{t,x,z}$ with $\mathcal{D}_1$ and the equality $\mathcal{D}_1 f_0 = -\mathcal{D}_2 f_0$ from [15], we find that

$$\begin{aligned} f_1(t,x,z,w) &= -\frac{1}{2}(T-t)^2 \rho g(z)\lambda(z)^2\lambda'(z)\mathcal{D}_1\mathcal{D}_2 f_0 \\ &= \frac{1}{2}(T-t)\rho\lambda(z)g(z)\mathcal{D}_1 f_{0,z} \end{aligned}$$

is the solution. We get the solution for $V(t,x,z)$ by combining the two partial solutions and replacing the w variable with z.

Once we get the solution of V_1, the equilibrium trading strategy up to order $o(\sqrt{\delta})$ can be computed:

$$\pi^* = \frac{\lambda(z)}{\sigma(z)}R - \sqrt{\delta}\left\{\frac{\rho g(z)}{\sigma(z)}\frac{V_{0,xz}}{V_{0,xx}} + \frac{\lambda(z)}{\sigma(z)}\left[\frac{V_{1,x}}{V_{0,xx}} + R\frac{V_{1,xx}}{V_{0,xx}}\right]\right\}. \tag{32}$$

Power Utility Case

For a power utility function:

$$U(x,\gamma(z)) = \frac{1}{1-\gamma(z)}x^{1-\gamma(z)}, \tag{33}$$

we have the zeroth order value function given by:

$$V_0(t,x,z) = \frac{x^{1-\gamma_0}}{1-\gamma_0}e^{\frac{\lambda(z)^2}{2}\frac{1-\gamma_0}{\gamma_0}(T-t)}. \tag{34}$$

Thus the explicit form of the first order value function is:

$$\begin{aligned} V_1(t,x,z) = \Bigg\{ & \frac{(T-t)^2\rho g(z)\lambda(z)^2\lambda'(z)(1-\gamma_0)^2}{2\gamma_0^2} \\ & + \gamma_1(z)\left[\frac{1}{1-\gamma_0} - \log(x) - \frac{\lambda(z)^2(T-t)}{2\gamma_0^2}\right]\Bigg\} V_0(t,x,z). \end{aligned} \tag{35}$$

The strategy is given by:

$$\pi^* = \frac{\lambda(z)x}{\sigma(z)\gamma_0} + \sqrt{\delta}\left[\underbrace{\frac{\rho g(z)\lambda'(z)\lambda(z)(1-\gamma_0)(T-t)}{\sigma(z)\gamma_0^2}}_{\text{slow factor adjustment in [15]}} - \underbrace{\frac{\lambda(z)}{\sigma(z)}\frac{\gamma_1(z)}{\gamma_0}(1+\frac{1}{\gamma_0})}_{\text{risk aversion adjustment}}\right]x. \tag{36}$$

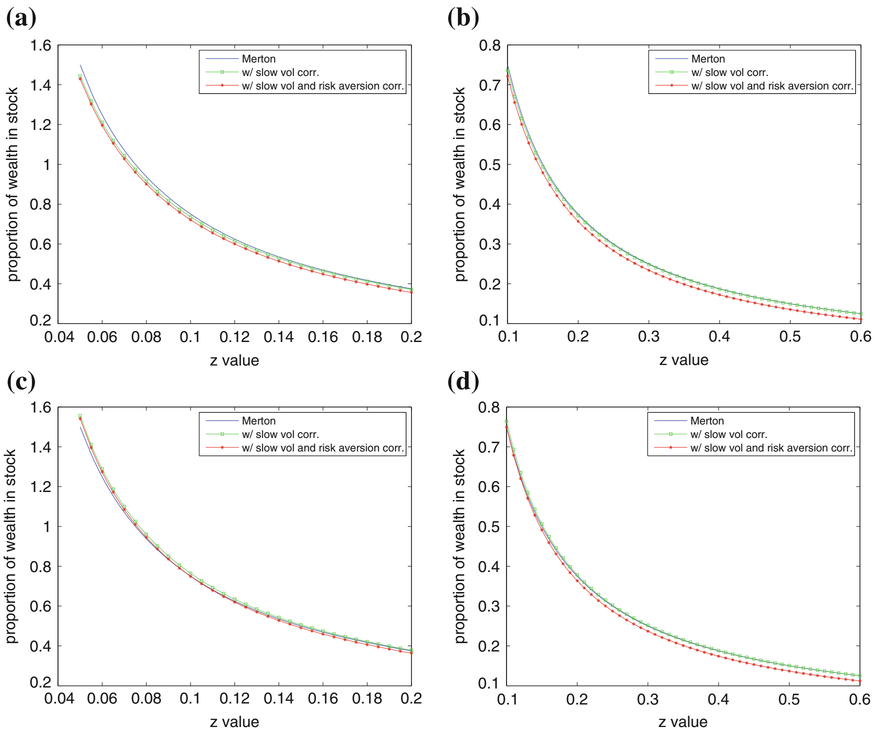


Fig. 2 Plots of the equilibrium strategies, in terms of the proportion of total wealth, against the slow stochastic volatility parameter z, in the case of power utility function. We chose the stochastic volatility model to be slow scale CIR(Heston), with $\mu = 0.15$, $r = 0$, $\sigma(z) = \sqrt{z}$, $\gamma(z) = \frac{\sqrt{z}}{2}$, $\rho = 0.2$, $T = 5$, $\gamma_0 = 2$, $\gamma_1 = \tan^{-1}(z)$ and the time scale is $\delta = 0.1$

We now compare the Merton optimal strategy, the optimal strategy with first order correction for the slow volatility factor appeared in [15] and our equilibrium strategy with first order correction. Note that the second strategy is equivalent to (36) with only the first fractional term inside the square bracket. We notice that for different levels of the slow factor, the proportions that the two adjustment factors would contribute to the first order correction are different. Figure 2 contains the plots of the three strategies for different ranges of the slow factor. Figure 2a, c show that for small z, the main contributor of the first order correction is the first fractional term inside the square bracket of (71), whereas for larger values of z, as Fig. 2b, d suggest, an increasing risk aversion plays the major role instead. The direction to which the first adjustment factor affects the strategy depends on the sign of the correlation factor ρ.

2.4 Comparison with Mixture of Power Utility Functions

A mixture of power utility functions takes the following form:

$$U^{mix}(x) = c_1 \frac{x^{1-\gamma_1}}{1-\gamma_1} + c_2 \frac{x^{1-\gamma_2}}{1-\gamma_2},$$

as introduced in Fouque et al. [15], where $\gamma_1 \neq \gamma_2$ and $c_1,\ c_2$ are positive constants. Under this utility function, the relative risk aversion is not constant any more but decreases in x. Now let us look at a power utility function with wealth-dependent risk aversion:

$$U(x) \ = \frac{x^{1-\gamma(x)}}{1-\gamma(x)}. \tag{37}$$

We can choose $\gamma(x)$ to make $U^{mix}(x) \ = \ U(x)$, but in the case of power utility the solution will be a complex-valued function due to the presence of $\gamma(x)$ in the exponent of x. (In contrast, for a mixture of exponential utility functions, $\gamma(x)$ will be real-valued).

For the case of small wealth-dependence, we have the following expansion:

$$\begin{aligned}
U(x) &= \frac{x^{1-(\gamma_0+\epsilon\gamma_1+O(\epsilon^2))}}{1-(\gamma_0+\epsilon\gamma_1+O(\epsilon^2))} \\
&= \frac{x^{1-\gamma_0} - \epsilon\gamma_1 \log(x) x^{1-\gamma_0} + O(\epsilon^2)}{1-\gamma_0}(1+\epsilon\frac{\gamma_1}{1-\gamma_0} + O(\epsilon^2)) \\
&= \frac{x^{1-\gamma_0}}{1-\gamma_0} + \epsilon\left\{-\frac{x^{1-\gamma_0}}{1-\gamma_0}\gamma_1\log(x) + \frac{x^{1-\gamma_0}}{1-\gamma_0}\frac{\gamma_1}{1-\gamma_0}\right\} + O(\epsilon^2)
\end{aligned}$$

where $\gamma_1 \equiv \gamma_1(x)$ can be chosen in such a way that the expansion is also a mixture of power utility functions. For example, we can set $\gamma_1(x)$ to be:

$$\gamma_1(x) = \frac{c_1 x^{k_1} + c_2 x^{k_2}}{-\log(x) + \frac{1}{1-\gamma_0}}$$

then for x belonging to the region where $\gamma_1(x)$ is bounded, the expansion above becomes:

$$\begin{aligned}
&\frac{x^{1-\gamma_0}}{1-\gamma_0} + \epsilon\frac{x^{1-\gamma_0}}{1-\gamma_0}\left(c_1 x^{k_1} + c_2 x^{k_2}\right) + O(\epsilon^2) \\
&= \frac{x^{1-\gamma_0}}{1-\gamma_0} + \epsilon c_1 \frac{x^{1-\gamma_0+k_1}}{1-\gamma_0} + \epsilon c_2 \frac{x^{1-\gamma_0+k_2}}{1-\gamma_0} + O(\epsilon^2)
\end{aligned}$$

$$= \frac{x^{1-\gamma_0}}{1-\gamma_0} + \epsilon c_1 \frac{1-\gamma_0+k_1}{1-\gamma_0} \frac{x^{1-\gamma_0+k_1}}{1-\gamma_0+k_1} + \epsilon c_2 \frac{1-\gamma_0+k_2}{1-\gamma_0} \frac{x^{1-\gamma_0+k_2}}{1-\gamma_0+k_2} + O(\epsilon^2)$$

i.e. a mixture of three power utility functions up to order ϵ.

Despite the similarity between the two types of utility functions, i.e. the terminal conditions for the two problems, the portfolio optimization under a mixture of power utility functions remains time-consistent as the risk aversion always depends on the terminal wealth, which is a random variable revealed at time T. In our problem here, we have made $\gamma(\cdot)$ dependent on the instantaneous level of wealth which becomes the source of time inconsistency.

3 Investment/Consumption Problems with Non-exponential Discounting

In the previous section we have looked at the utility maximization for terminal wealth with time-varying risk aversions by using the method of asymptotic expansions. Here we want to study the investment/consumption problem under non-exponential discounting. We adopt the same two-asset diffusion model (1) for this problem thus we have our wealth process being

$$dX_t = [\pi_t(\mu - r)X_t + (rX_t - c_t)]\, dt + \pi_t \sigma X_t\, dW_t, \tag{38}$$

where the additional term c_t denotes our instantaneous consumption rate and π_t is the proportion of wealth invested in the risky asset. We define the objective function as:

$$J(t, x, \pi, \mathbf{c}) = \mathbb{E}_{t,x}\left[\int_t^T \varphi(s-t)U(c_s)ds + \varphi(T-t)U(X_T^{\pi,\mathbf{c}})\right], \tag{39}$$

where $U(\cdot)$ is some appropriate utility function to be chosen and $\varphi(\cdot)$ is the discount function for the utility derived from consumption. We do not require $\varphi(\cdot)$ to be exponential which is the source of time inconsistency for this problem. As usual, the value function is defined as:

$$V(t, x) = \sup_{\pi,\, \mathbf{c}} J(t, x, \pi, \mathbf{c}). \tag{40}$$

Similar to the utility maximization for terminal wealth case, we have the following result as a consequence of Definition 2.4:

Proposition 3.1 *The value function $V(t, x)$ satisfies the following HJB-type equation:*

$$
\begin{cases}
\sup\limits_{\pi,c\in\mathbb{R}\times\mathbb{R}_+}\{\frac{\partial V}{\partial t}+[\pi x(\mu-r)+(rx-c)]\frac{\partial V}{\partial x}+\frac{\pi^2}{2}\sigma^2x^2\frac{\partial^2 V}{\partial x^2}+U(c)\} \\
\quad =-\mathbb{E}_{t,x}[\int\limits_t^T\varphi'(s-t)U(c_s^*)ds+\varphi'(T-t)U(X_T^{\pi,\mathbf{c}^*})] \\
\quad V(T,x)=0 \\
\quad V(t,0)=0,
\end{cases}
\tag{41}
$$

where c_s^* denotes the equilibrium consumption in the future time $s\geq t$.

Proof For this proof we ignore the $\varphi(T-t)U(X_T^{\pi,\mathbf{c}^*})$ term for simplicity. Using the definition of equilibrium strategies in (4), let us define:

$$
\pi_s^\epsilon=\begin{cases}\pi & \text{for } s\in[t,t+\epsilon]\\ \pi^* & \text{for } s\in(t+\epsilon,T]\end{cases}\quad\text{and}\quad \mathbf{c}_s^\epsilon=\begin{cases}c & \text{for } s\in[t,t+\epsilon]\\ \mathbf{c}^* & \text{for } s\in(t+\epsilon,T].\end{cases}
$$

i.e. our policy $\mathbf{u}:=\left(\pi_s^\epsilon,\ c_s^\epsilon\right)_{s\in[t,T]}$ is defined such that it is a uniform and arbitrary perturbation from $\mathbf{u}^*$ for the period $[t,t+\epsilon]$ and the two strategy will coincide after $t+\epsilon$. Therefore we have

$$
J(t+\epsilon,X_{t+\epsilon},\mathbf{u})=V(t+\epsilon,X_{t+\epsilon}),
$$

which we take the expectation conditional on (t,x) and plug into the following inequality:

$$
\begin{aligned}
V(t,x)&\geq J(t,x,\mathbf{u})\\
&=J(t,x,\mathbf{u})-\mathbb{E}_{t,x}[J(t+\epsilon,X_{t+\epsilon},\mathbf{u})]+\mathbb{E}_{t,x}\left[V(t+\epsilon,X_{t+\epsilon})\right]\\
&=\mathbb{E}_{t,x}\left[\int\limits_t^T\varphi(s-t)U(c_s^\epsilon)ds-\int\limits_{t+\epsilon}^T\varphi(s-t-\epsilon)U(c_s^*)ds\right]\\
&\quad+\mathbb{E}_{t,x}\left[V(t+\epsilon,X_{t+\epsilon})\right]\\
&\approx\epsilon\mathbb{E}_{t,x}\left[U(c_{t+\epsilon}^\epsilon)-\int\limits_{t+\epsilon}^T\varphi'(s-t-\epsilon)U(c_s^\epsilon)ds\right]+\mathbb{E}_{t,x}[V(t+\epsilon,X_{t+\epsilon})],
\end{aligned}
$$

which in turn is a result of the following simple Taylor expansion for point t around $(t+\epsilon)$:

$$
\int\limits_t^T\varphi(s-t)U(c_s^\epsilon)ds\approx\int\limits_{t+\epsilon}^T\varphi(s-t-\epsilon)U(c_s^\epsilon)ds+(-\epsilon)
$$

$$\left(-U(c^{\epsilon}_{t+\epsilon}) - \int\limits_{t+\epsilon}^{T} \varphi'(s-t-\epsilon)U(c^{\epsilon}_s)ds)\right) + o(\epsilon^2).$$

Dividing the inequality by ϵ and taking the limit $\epsilon \to 0$, we obtain:

$$\mathcal{G}^{\pi,c}V(t,x) + U(c_t) + \mathbb{E}_{t,x}\left[\int\limits_{t}^{T} \varphi'(s-t)U(c^*_s)ds\right] \leq 0,$$

where $\mathcal{G}^{\pi,c}$ denotes the infinitesimal generator for $V(t,x)$. If we take the supremum over π and c, the inequality above becomes equality and we recover the HJB-type equation for $V(t,x)$ less the $\mathbb{E}[\varphi'(T-t)U(X_T^{\pi,\mathbf{c}^*})]$ term, which can be obtained using the same argument as above. The boundary conditions are straightforward.

Remark 3.2 A first look may suggest that the result (41) above contradicts the remark made in Sect. 2.1.3 regarding the two-equation characteristics for time inconsistency, since this time we only have one HJB-type equation. In fact, the two-equation feature is masked in the term $\mathbb{E}_{t,x}[\int_t^T \varphi'(s-t)U(c^*_s)ds]$, which characterizes the difference between how one's current self and his immediate future self would value future consumption. This is equivalent to saying the derivative characterizes the difference on the current value function and the continuation value function. If we take the discounting function to be of exponential type, then the term $\mathbb{E}_{t,x}[\int_t^T \varphi'(s-t)U(c^*_s)ds]$ will simply reduce to $-rV(t,x)$ where r is the exponential discount rate, and the HJB-type equation will reduce to the classical HJB equation for an investment/consumption problem. However, for all non-exponential-type discounting functions, $\mathbb{E}_{t,x}[\int_t^T \varphi'(s-t)U(c^*_s)ds]$ makes the equation non-local and thus hard to solve. See Ekeland et al. [13] for a numerical treatment of a similar problem using backward integration.

3.1 Approximating a Hyperbolic Discount Function

On one hand, the exponential discounting produces explicit solutions but is less realistic. On the other hand, a hyperbolic discount function becomes less tractable but will be more in accordance with how people behave. There is a clear trade-off between tractability and realisticity. Consider the following discount function:

$$\varphi^{\alpha}(\tau) = e^{(\alpha-1)\delta_0\tau - \alpha\log(1+\delta_1\tau)} \tag{42}$$

for $\alpha \in [0, 1]$. When $\alpha = 0$, this is an exponential discount function with discount rate δ_0. When $\alpha = 1$, this is a hyperbolic discount function with rate δ_1. For $\alpha \in (0, 1)$, the discount function will have partial amount of the features that a hyperbolic discount function has.

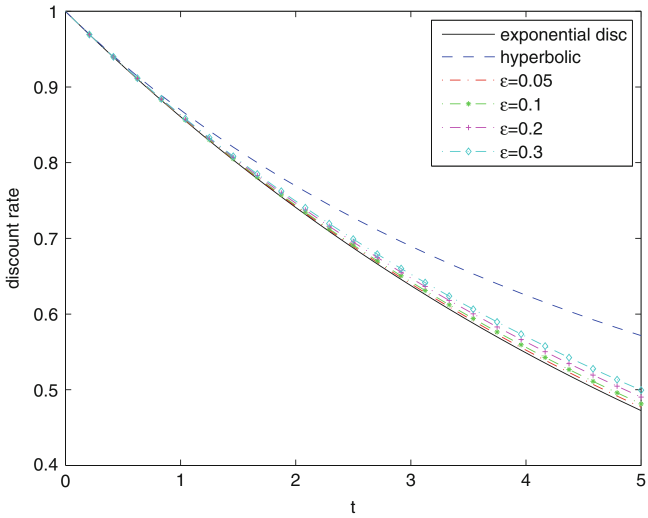


Fig. 3 A comparison of discount functions $\exp(-\delta_0\tau)$, $1/(1+\delta_1\tau)$ and the one defined by (43) with $\delta_0 = \delta_1 = 0.15$ for various values of ϵ

Now we consider the case where $\alpha = \epsilon > 0$ is very small, then

$$\varphi^{\epsilon}(\tau) \approx e^{-\delta_0\tau}\left(1 + \epsilon\Delta(\tau) + o(\epsilon^2)\right), \tag{43}$$

where $\Delta(\tau) = \delta_0\tau - \log(1+\delta_1\tau)$ (we can choose other forms of $\Delta(\tau)$ as well). This discount function will allow us to solve the HJB-type equation (41) using asymptotic expansions in the following subsection. Figure 3 illustrates that this discount function is close to the exponential discounting case for small ϵ while it bends towards the hyperbolic discount function. Thus it mimics the hyperbolic discounting feature by a small amount controlled by ϵ.

3.2 *Solving the HJB-type Equation Using Asymptotic Expansions*

Let us go back to the HJB type equation (41). Using the first order conditions, the maximizations over π and c can be done separately:

$$\pi^* = -\frac{\mu - r}{x\sigma^2}\frac{V_x}{V_{xx}} \quad \text{and} \quad c^* = (U')^{-1}(V_x), \tag{44}$$

where V_x denotes the first derivative w.r.t x and so on. We can see that c^* is the Legendre transform of the utility function at V_x. From now on we will adopt a power utility function with risk aversion γ:

$$U(c) = \frac{c^{1-\gamma}}{1-\gamma},$$

thus we have $c^* = (V_x)^{-1/\gamma}$. We plug π^* and c^* into (41) to obtain the following nonlinear non-local PDE:

$$\begin{aligned} V_t - \frac{\lambda^2}{2}\ \frac{V_x^2}{V_{xx}} + \frac{\gamma}{1-\gamma}(V_x)^{\frac{\gamma-1}{\gamma}} + rxV_x \\ = -\mathbb{E}_{t,x}\left[\int_t^T \varphi'(s-t)U(c_s^*)ds + \varphi'(T-t)U(X_T^{\pi,\mathbf{c}^*})\right], \end{aligned} \tag{45}$$

with boundary conditions $V(T,x) = 0$ and $V(t,0) = 0$.

As a consequence of the expansion (43) for the discount function, we seek a similar expansion for the value function:

$$V(t,x) = V_0(t,x) + \epsilon V_1(t,x) + o(\epsilon^2), \tag{46}$$

which we plug into (45). After grouping terms of different orders, we have the following PDEs for the first two orders:

$$\begin{aligned} &V_{0,t} - \frac{\lambda^2}{2}\frac{V_{0,x}^2}{V_{0,xx}} + \frac{\gamma}{1-\gamma}(V_{0,x})^{\frac{\gamma-1}{\gamma}} + rxV_{0,x} - \delta_0 V_0 = 0, \\ &V_{1,t} - \left(\lambda^2\frac{V_{0,x}}{V_{0,xx}} + (V_{0,x})^{-\frac{1}{\gamma}} - rx\right)V_{1,x} + \frac{\lambda^2}{2}\frac{V_{0,x}^2}{V_{0,xx}^2}V_{1,xx} - \delta_0 V_1 \\ &= -\mathbb{E}_{t,x}\left[\int_t^T \Delta'(s-t)\ e^{-\delta_0(s-t)}\frac{[c_{0,s}^*(X_s^{(0)})]^{1-\gamma}}{1-\gamma}ds + \Delta'(T-t)\frac{(X_T^{(0)})^{1-\gamma}}{1-\gamma}\right], \end{aligned} \tag{47}$$

where $X_s^{(0)}$ denotes the wealth process under the zeroth order equilibrium investment and consumption strategies π_0^* and c_0^*. The detail of the decomposition of (45) into (47) can be found in the Appendix.

Note: The first equation in (47) can be solved in a fairly standard way with the appropriate boundary conditions. Once this is solved, we obtain the zeroth order value function as well as the zeroth order strategies that will give explicit forms for the parameters of the second equation. As we will see later, the solution to the second PDE can be found explicitly. We have therefore managed to bypass the "nonlocal" issue in the HJB-type PDE by using asymptotic expansions. This allows us to avoid the usual numerical procedures as seen for example, in [13].

3.2.1 Zeroth Order Solution

The solution to the zeroth order equation with zero terminal/boundary conditions is very well-known. Using separation of variables method, we seek solution $V_0(t,x)$ of the following form:

$$V_0(t,x) = \frac{x^{1-\gamma}}{1-\gamma}\,[f(t)]^{\gamma}\,. \tag{48}$$

The original PDE problem reduces to the following ODE problem

$$f'(t) + \frac{1-\gamma}{\gamma}\left(\frac{\lambda^2}{2\gamma}+r\right)f(t) + e^{\frac{\delta_0}{\gamma}t} = 0, \tag{49}$$

with $f(T)=1$. Thus we have

$$f(t) = \frac{-e^{A_2 t} + A_3 e^{A_2 T + A_1(T-t)}}{A_1 + A_2} \tag{50}$$

where $A_1 = \frac{1-\gamma}{\gamma}\left(\frac{\lambda^2}{2\gamma}+r\right)$, $A_2 = \frac{\delta_0}{\gamma}$ and $A_3 = \frac{A_1+A_2+e^{A_2T}}{e^{A_2T}}$. Therefore, We can also compute the zeroth order equilibrium strategies:

$$\pi_0^* = \frac{\lambda}{\sigma\gamma} \quad \text{and} \quad c_0^* = f(t)^{-1}x. \tag{51}$$

3.2.2 First Order Solution

Using the preceding result, we can simplify the first order PDE from (47) into:

$$\begin{aligned} &V_{1,t} + \left(\frac{\lambda^2}{\gamma}+r-1\right)xV_{1,x} + \frac{\lambda^2}{2\gamma^2}x^2 V_{1,xx} - \delta_0 V_1 \qquad (52)\\ &= \mathbb{E}_{t,x}\left[\int_t^T \Delta'(s-t)e^{-\delta_0(s-t)}\frac{[c_{0,s}^*(X_s^{(0)})]^{1-\gamma}}{1-\gamma}ds + \Delta'(T-t)\frac{(X_T^{(0)})^{1-\gamma}}{1-\gamma}\right]. \end{aligned}$$

In order to deal with the expectation term on the right side, we need the dynamics of the zeroth order wealth process $X_t^{(0)}$ under zeroth order equilibrium strategies:

$$dX_t^{(0)} = \left(\pi_0^*(\mu-r)+r-f(t)^{-1}\right)X_t^{(0)}\,dt + \pi_0^*\sigma X_t^{(0)}\,dW_t, \tag{53}$$

which we notice is a lognormal process and we can write out the expectation term explicitly.

It follows that

$$\mathbb{E}_{0,x}\left[\frac{(X_t^{(0)})^{1-\gamma}}{1-\gamma}\right] = \frac{x^{1-\gamma}}{1-\gamma} e^{(1-\gamma)\left[\pi_0^*(\mu-r)+r-f(t)^{-1}-\frac{\gamma}{2}\pi_0^{*2}\sigma^2\right]t}. \tag{54}$$

Therefore, (52) becomes

$$V_{1,t} + \left(\frac{\lambda^2}{\gamma} + r - 1\right) x V_{1,x} + \frac{\lambda^2}{2\gamma^2} x^2 V_{1,xx} = \delta_0 V_1 + \frac{x^{1-\gamma}}{1-\gamma} F(t), \tag{55}$$

where $F(t)$ denotes the integral:

$$F(t) := \int_t^T \Delta'(s-t) e^{-\delta_0(s-t)} e^{(1-\gamma)\left[\pi_0^*(\mu-r)+r-f(s)^{-1}-\frac{\gamma}{2}\pi_0^{*2}\sigma^2\right]s} ds.$$

The ansatz $V_1(t,x) = \frac{x^{1-\gamma}}{1-\gamma} g(t)$ reduces (55) to a first order ODE problem:

$$g'(t) + \left[\left(\frac{\lambda^2}{2\gamma} + r - 1\right)(1-\gamma) - \delta_0\right] g(t) = F(t), \tag{56}$$

with terminal condition $g(T) = 0$, which has a solution given by:

$$g(t) = \int_t^T F(s) e^{B_1(s-t)} ds, \tag{57}$$

where $B_1 := \left(\frac{\lambda^2}{2\gamma} + r - 1\right)(1-\gamma) - \delta_0$.

3.2.3 First Order Corrections for Equilibrium Strategies

Proposition 3.3 *We have the following respective first order corrections (to multiply by ϵ) to the equilibrium strategies:*

$$\pi_1^* = 0 \quad \text{and } c_1^* = -\frac{1}{\gamma}\frac{g(t)}{f(t)} c_0^*. \tag{58}$$

Proof We have

$$V_0(t,x) = U(x) f(t) \quad \text{and } V_1(t,x) = U(x) g(t)$$

where $U(x)$ is the power utility function with risk aversion γ. For the equilibrium proportion of wealth invested in the risky asset, we have

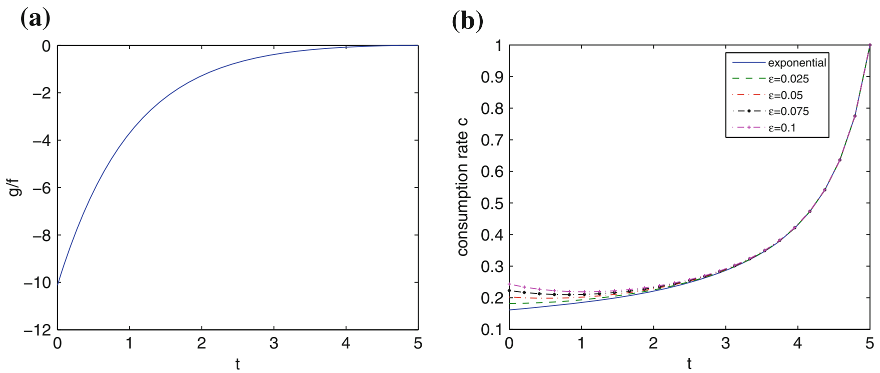


Fig. 4 $\mu = 0.15$, $r = 0.05$, $\sigma = 0.25$, $\gamma = 2$, $\delta_0 = \delta_1 = 0.15$ and $T = 5$. We have included the utility of wealth at time T here to fix the unbounded consumption rate near T

$$\pi^* \approx -\frac{\mu - r}{x\sigma^2}\frac{V_{0,x} + \epsilon V_{1,x}}{V_{0,xx} + \epsilon V_{1,xx}} = -\frac{\mu - r}{x\sigma^2}\frac{U'(x)}{U''(x)}\frac{(f(t) + \epsilon g(t))}{(f(t) + \epsilon g(t))} = \frac{\mu - r}{\gamma\sigma^2} \equiv \pi_0^*,$$

whereas for the equilibrium consumption rate, we have

$$c^* \approx \left(V_{0,x} + \epsilon V_{1,x}\right)^{-\frac{1}{\gamma}} = \left(V_{0,x}\right)^{-\frac{1}{\gamma}}\left[1 - \frac{\epsilon}{\gamma}\frac{V_{1,x}}{V_{0,x}} + o(\epsilon^2)\right] = c_0^*\left(1 - \frac{\epsilon}{\gamma}\frac{g(t)}{f(t)}\right).$$

We have found that adding a small amount of hyperbolic-discounting feature to the discount function does not change the proportion of wealth invested in the risky asset, while it will affect the consumption rate by a fraction depending on the ratio $\frac{g(t)}{f(t)}$. Figure 4 illustrates how the approximated equilibrium strategies change over time compared to the optimal one in the exponential discounting case. In general, we find that hyperbolic discounting would encourage one to consume at a faster rate. The fact that $g(t)$ is negative also means that the value function is more negative compared to the exponential discounting case, indicating a loss of welfare. For relatively larger values of ϵ, the equilibrium strategy is clearly non-monotonic. More precisely, the ideal consumption speed starts at some higher level compared to the exponential discounting case and it has a decreasing trend at the beginning. But eventually the consumption speed will start to increase monotonically once we are sufficiently far away from the commencing point $t = 0$. In fact, this non-monotonicity feature agrees with the consumption pattern observed in real-life household data, which is one of the main reasons economists support the use of hyperbolic discounting. We also note that similar results were obtained in [13] in which the authors made use of backward numerical integration techniques to solve the full extended HJB equation analogous to (41).

3.3 A Bound for the Value Function: Infinite Horizon Case

In this section we want to illustrate some characteristics of the hyperbolic discounting problem using Laplace transform. Suppose we have an infinite horizon investment/consumption problem instead:

$$V(x) = \sup_{\mathbf{c}} \mathbb{E}\left[\int_0^\infty \frac{1}{1+\delta t}\frac{c_t^{1-\gamma}}{1-\gamma}dt\right]. \tag{59}$$

The following equation holds for the hyperbolic discount function by Laplace transform:

$$\frac{1}{1+\delta t} = \int_0^\infty \frac{e^{-\tau(t+\frac{1}{\delta})}}{\delta}d\tau.$$

Therefore we have

$$\begin{aligned} V(x) &= \sup_{\mathbf{c}} \int_0^\infty \mathbb{E}\left[\int_0^\infty e^{-\tau t}\frac{c_t^{1-\gamma}}{1-\gamma}dt\right]\frac{e^{-\frac{\tau}{\delta}}}{\delta}d\tau \\ &= \sup_{\mathbf{c}} \int_0^\infty \bar{J}(x,c,\tau)\frac{e^{-\frac{\tau}{\delta}}}{\delta}d\tau \\ &\le \int_0^\infty \sup_{\mathbf{c}} \bar{J}(x,\mathbf{c},\tau)\frac{e^{-\frac{\tau}{\delta}}}{\delta}d\tau \\ &= \beta C(x)\int_0^\infty \frac{e^{-\beta\tau}}{(\tau+\alpha)^\gamma}d\tau, \end{aligned} \tag{60}$$

where $C(x) := \frac{\gamma^\gamma x^{1-\gamma}}{1-\gamma}$, $\alpha := -\delta(1-\gamma) - \frac{\lambda^2(1-\gamma)}{2\gamma}$, $\beta = \frac{1}{\delta}$ and $\bar{J}(x,c,\tau)$ denotes the objective function for the infinite horizon investment problem under consumption c and exponential discount rate τ, in which case the value function has an explicit solution.

The second line of (60) best illustrates how time inconsistency arises from hyperbolic discounting. Loosely speaking, the integral can be seen as the weighted average of a continuum of optimization problems parameterized by the (exponential-type) discount rate τ. If there is a policy $\mathbf{c}^*$ that can maximize all the objective functions, then the inequality becomes an equality and we can say $\mathbf{c}^*$ is the optimal-for-all policy. Unfortunately, the optimal-for-all policy does not exist most of the time. Nevertheless we can still find a policy $\mathbf{c}^{**}$ that maximizes the integral, i.e. a linearization of the objectives. And it turns out this particular policy $\mathbf{c}^{**}$ is a Pareto optimum that corresponds to a point on the Pareto front of the multi-objective optimization problem. Consequently, the difference between two sides of the inequality corresponds to the distance between a strict optimal value and the Pareto-optimal value under the particular linearization given.

The integral in the last line of (60) can be solved for positive-integer-valued $\gamma = n \in \mathbb{Z}_+$

$$\int_0^\infty \frac{e^{-\beta\tau}}{(\tau+\alpha)^\gamma} d\tau = -n! \sum_{j=0}^{n} \frac{\alpha^j}{j!}.$$

Thus we have produced a bound for the value function in case γ is a positive integer

$$V(x) \leq -\beta n! \sum_{j=0}^{n} \frac{\alpha^j}{j!} C(x). \tag{61}$$

3.4 Extension with Proportional Transaction Costs

We extend our study to the situation where proportional transaction cost exists. The dynamics of the portfolio can be represented as below:

$$\begin{aligned} dX_t^{(b)} &= (rX_t^{(b)} - \hat{c}_t)dt - (1+\kappa)d\hat{L}_t + (1-\lambda)d\hat{M}_t \\ dX_t^{(s)} &= \mu X_t^{(s)} dt + \sigma X_t^{(s)} dW_t + d\hat{L}_t - d\hat{M}_t, \end{aligned} \tag{62}$$

where $X_t^{(b)}$ and $X_t^{(s)}$ represent the wealth in the risk-free bank account and in the risky asset (stock) respectively. Again $\hat{c}_t$ is the rate of consumption and $d\hat{L}_t := \hat{l}_t dt$ and $d\hat{M}_t := \hat{m}_t dt$ denote the purchase and sell of the risky asset which will incur proportional transaction costs κ and λ respectively.

Our objective function has now been modified into maximizing consumption utility over an infinite horizon because we want to make the analysis simpler. The objective function is given by

$$J(x^{(b)}, x^{(s)}, \hat{c}, \hat{l}, \hat{m}) = \mathbb{E}\left[\int_0^\infty \varphi(s) U(c_s) ds \mid X_0^{(b)} = x^{(b)}, \quad X_0^{(s)} = x^{(s)}\right], \tag{63}$$

given the current level of wealth $x^{(b)}$ in the bank account and $x^{(s)}$ in the stock as well as the admissible controls $\hat{c}$, $\hat{m}$, $\hat{l}$, where the utility function $U(.)$ is still chosen to be the power type. Now define the value function:

$$V(x^{(b)}, x^{(s)}) = \sup_{\hat{c},\hat{m},\hat{l}} J(x^{(b)}, x^{(s)}, \hat{c}, \hat{l}, \hat{m}). \tag{64}$$

Almost identical to the result from Proposition 3.1, the value function satisfies the HJB-type equation:

$$\sup_{\hat{c},\hat{m},\hat{l}} (rx^{(b)} - \hat{c})V_{x^{(b)}} + \mu x^{(s)} V_{x^{(1)}} + \frac{1}{2}\sigma^2 (x^{(s)})^2 V_{x^{(s)}x^{(s)}} + [(1-\lambda)V_{x^{(b)}} - V_{x^{(s)}})]\hat{m}$$

$$+ [V_{x^{(1)}} - (1+\kappa)V_{x^{(b)}}]\hat{l} + U(\hat{c}) = -\mathbb{E}_{x^{(b)}x^{(s)}} \left[\int_0^\infty \varphi'(s) U(\hat{c}_s^*) ds \right], \tag{65}$$

only this time there is no time derivative. When $\varphi(\cdot)$ is exponential type, this becomes the HJB equation that was probably first derived by Davis and Norman [10], who noticed that the desirable strategies for purchase and sell were "bang-bang" type which only took place on the boundaries of the no-transaction region at maximum possible rates.

The homothetic property holds for the value function since we have chosen to use a power utility function, meaning that

$$V(\rho x^{(b)}, \rho x^{(s)}) = \rho^{1-\gamma} V(x^{(b)}, x^{(s)}), \tag{66}$$

for any positive constant ρ. Thus we can write the value function $V(x^{(b)}, x^{(s)})$ into

$$V(x^{(b)}, x^{(s)}) = (x^{(s)})^{1-\gamma} V(x^{(b)}/x^{(s)}, 1) := (x^{(s)})^{1-\gamma} \Phi(x^{(b)}/x^{(s)}). \tag{67}$$

As a consequence, it is sufficient to study the transformed value function $\Phi(z)$ where we use z to denote the ratio $x^{(b)}/x^{(s)}$.

The problem reduces to a free boundary ODE problem:

$$(\mu - \frac{1}{2}\sigma^2\gamma)(1-\gamma)\Phi(z) + (r - \mu + \sigma^2\gamma)z\Phi'(z) + \frac{1}{2}\sigma^2 z^2 \Phi''(z)$$

$$+ \frac{\gamma}{1-\gamma}\left[\Phi'(z)\right]^{-(1-\gamma)/\gamma} + \mathbb{E}_z \left[\int_0^\infty \varphi'(s) \frac{\left[\Phi'(Z_s)\right]^{-(1-\gamma)/\gamma}}{1-\gamma} ds \right] = 0, \tag{68}$$

with free boundary conditions:

$$\begin{aligned} \Phi'(l)(1-\lambda+l) - (1-\gamma)\Phi(l) &= 0 \\ \Phi'(u)(1+\kappa+u) - (1-\gamma)\Phi(u) &= 0, \end{aligned} \tag{69}$$

where the upper and lower boundaries u and l are to be determined. The ODE (68) is difficult to solve because it involves a free boundary as well as a non-local term $\mathbb{E}_z \left[\int_0^\infty \varphi'(s) \frac{[\Phi'(Z_s)]^{-(1-\gamma)/\gamma}}{1-\gamma} ds \right]$ that is the source of time inconsistency. Again let us deal with it using the asymptotic approximation method. We assume the same expansion for the discount function $\varphi(\cdot)$ as in (43). And we seek an expansion for the solution $\Phi(z)$ of the following form:

$$\Phi(z) = \Phi_0(z) + \epsilon\,\Phi_1(z) + o(\epsilon^2). \tag{70}$$

At zeroth order, we need to solve the free boundary ODE:

$$\begin{cases} \beta_0\Phi_0 + \beta_1 z\Phi_0' + \beta_2 z^2\Phi_0'' + \dfrac{\gamma}{1-\gamma}\left[\Phi_0'\right]^{\frac{\gamma-1}{\gamma}} = 0 \\ \Phi_0'(l_0)(1-\lambda+l_0) - (1-\gamma)\Phi_0(l_0) = 0 \\ \Phi_0'(u_0)(1+\kappa+u_0) - (1-\gamma)\Phi_0(u_0) = 0, \end{cases} \tag{71}$$

with $l_0,\ u_0$ to be determined, where β_0, β_1 and β_2 are constant parameters defined as

$$\beta_0 := (\mu - \frac{1}{2}\sigma^2\gamma)(1-\gamma) - \delta_0, \quad \beta_1 := r - \mu + \sigma^2\gamma, \quad \beta_2 := \frac{1}{2}\sigma^2.$$

At first order, we need to solve a fixed boundary ODE problem, but with a nonlocal term:

$$\begin{cases} \beta_0\Phi_1 + \left[\beta_1 z + \dfrac{\gamma}{1-\gamma}(\Phi_0')^{-\frac{1}{\gamma}}\right]\Phi_1' + \beta_2 z^2\Phi_1'' \\ \quad + \mathbb{E}_z\left[\displaystyle\int_0^\infty e^{-\delta_0 s}\Delta'(s)\frac{(\Phi_0')^{\frac{\gamma-1}{\gamma}}}{1-\gamma}ds\right] = 0 \\ \Phi_1'(l_0)(1-\lambda+l_0) - (1-\gamma)\Phi_1(l_0) = 0 \\ \Phi_1'(u_0)(1+\kappa+u_0) - (1-\gamma)\Phi_1(u_0) = 0, \end{cases} \tag{72}$$

from which we can compute the first order corrections to the NT boundary as

$$\begin{aligned} l_1 &= -\frac{(1-\lambda+l_0)\Phi_1''(l_0) + \gamma\Phi_1'(l_0)}{(1-\lambda+l_0)\Phi_0'''(l_0) + (1+\gamma)\Phi_0''(l_0)} \\ u_1 &= -\frac{(1+\kappa+u_0)\Phi_1''(u_0) + \gamma\Phi_1'(u_0)}{(1+\kappa+u_0)\Phi_0'''(u_0) + (1+\gamma)\Phi_0''(u_0)}, \end{aligned} \tag{73}$$

which are derived from the original boundary equations.

3.4.1 Zeroth Order Solution

The zeroth order problem (71) is exactly the original problem in [10], which has been shown to have a solution that can be written as

$$\Phi_0(z) = \frac{1}{1-\gamma}\left[\frac{1-\gamma}{\gamma}h_1(z)\right]^{-\gamma}\left(\frac{z}{h_2(z)}\right)^{1-\gamma}, \tag{74}$$

where $h_2(z)$ and $h_1(z)$ solve the system below

$$\begin{aligned} h_2'(z) &= \frac{1}{\beta_2 z}\left[R(h_2(z)) - h_1(z)\right] \\ h_1'(z) &= \frac{1-\gamma}{\gamma}\frac{h_1(z)}{\beta_2 z h_2(z)}\left[h_1(z) - Q(h_2(z))\right], \end{aligned} \tag{75}$$

with boundary conditions

$$h_2(l_0) = \frac{l_0}{l_0+1-\lambda}, \quad h_1(l_0) = Q\left(\frac{l_0}{l_0+1-\lambda}\right), \quad h_2(u_0) = \frac{u_0}{u_0+1+\kappa},$$
$$h_1(u_0) = Q\left(\frac{u_0}{u_0+1+\kappa}\right),$$

where we define $Q(x) := -\frac{\beta_0}{1-\gamma} - \beta_1 x + \beta_2\gamma x^2$ and $R(x) := Q(x) + \beta_2(1-x)x$. This ODE system (75) can be solved numerically using a shooting method as suggested by Davis and Norman [10].

3.4.2 First Order Solution

Recall (72), in order to obtain $\Phi_1(z)$, we need to solve a fixed boundary ODE, which is numerically straightforward except for the source term

$$\mathbb{E}_z\left[\int_0^\infty e^{-\delta_0 s}\Delta'(s)\frac{(\Phi_0'(Z_s))^{\frac{\gamma-1}{\gamma}}}{1-\gamma}ds\right],$$

which involves a path integral depending on the process $Z_t \equiv \frac{X_t}{Y_t}$. Note that the major issue here is that we do not have an explicit form for Φ_0 as it is computed numerically, whereas the nonlocality issue has disappeared similar to the case without transaction cost because of the expansion we have used. To approximate the source term we reply on Monte Carlo method to generate a large number of sample paths for Z_t up to some time T and evaluate the truncated integral for each of these paths using Riemann-sum approximation, after which the estimated expectation can be obtained by taking the average. We first use Ito's Lemma to get the dynamics for the process Z_s under the zeroth order equilibrium strategies c_0^*, dL_0^* and dM_0^*:

$$\begin{aligned} dZ_t = {} & [(r-\mu+\frac{\sigma^2}{2})Z_t - c_{0,t}^*]dt - \sigma Z_t dW_t \\ & - (1+\kappa+Z_t)dL_{0,t}^* + (1-\lambda+Z_t)dM_{0,t}^*. \end{aligned} \tag{76}$$

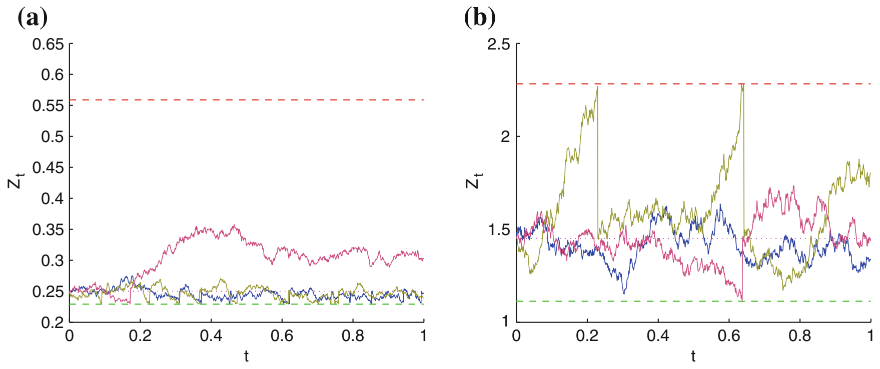


Fig. 5 Some realizations of Z_s with zeroth order optimal consumption rate c_0^* and boundaries l_0 and u_0. Note that each time the process hits the boundaries, it will be pushed back to the Merton line inside the NT region. **a**. Z_s, $\sigma = 0.25$. **b** Z_s, $\sigma = 0.35$

To further simplify the problem we put restrictions on the "Bang-Bang" type strategies dL_0^* and dM_0^* so that the process Z_t diffuses within the zeroth order NT region but whenever it hits the boundary l_0 or u_0, it will be pushed back to the Merton ratio line in the NT region. Figure 5 gives a few sample path of the controlled process X_s. We repeat the approximations for a grid of initial values z and we can smooth out the results using Fourier-type curve fitting method.

We are left with a second-order ODE with a mixed-type boundary condition to solve. Numerical discretization makes it a linear system of equations $Ax = b$ with A being a tridiagonal matrix. Once we solve this, we can compute the first-order corrections for the NT boundaries as well as for the equilibrium strategies.

3.4.3 Numerical Results

We have numerically solved the zeroth and first order ODE problems using the following set of parameter values: $r = 0.05$, $\mu = 0.15$, $\gamma = 2$, $\kappa = \lambda = 0.01$, $\delta_0 = \delta_1 = 0.15$ or 0.3 and $\sigma = 0.25 : 0.02 : 0.35$. Figure 6 gives illustrations for the zeroth order value function $\Phi_0(z)$ and the zeroth order equilibrium consumption rate $c_0^*(z)$. For different volatility σ, the NT boundaries are different. Figure 7 illustrates the NT region with/without first order corrections. We can see that hyperbolic discounting has the effect of shrinking the NT region, which leads to more frequent trading and rebalancing. This result matches the behavior of typical individual investors who tend to be myopic and impatient and are therefore prone to excessive rebalancing of their investment portfolios. However, whether this is a good or bad thing requires further investigation on this problem.

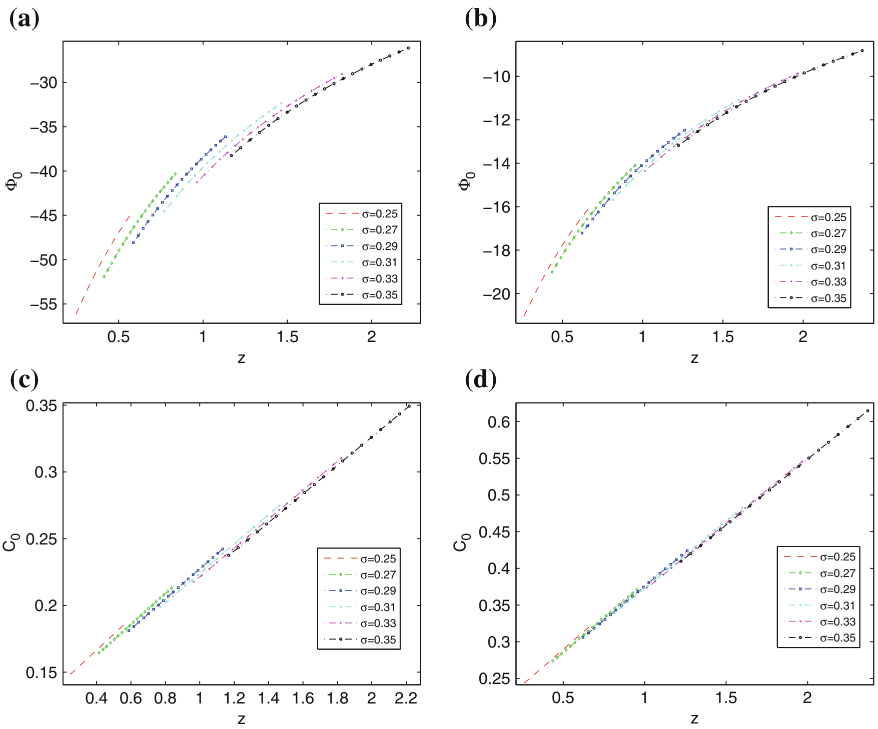


Fig. 6 Plots of the zeroth value function $\Phi_0(z)$ and optimal consumption rate c_0^* for parameter values $r = 0.05$, $\mu = 0.15$, $\gamma = 2$ and $k = \lambda = 0.01$. **a** $\Phi_0(z)$, $\delta_0 = \delta_1 = 0.15$. **b** $\Phi_0(z)$, $\delta_0 = \delta_1 = 0.3$. **c** c_0^*, $\delta_0 = \delta_1 = 0.15$. **d** c_0^*, $\delta_0 = \delta_1 = 0.3$

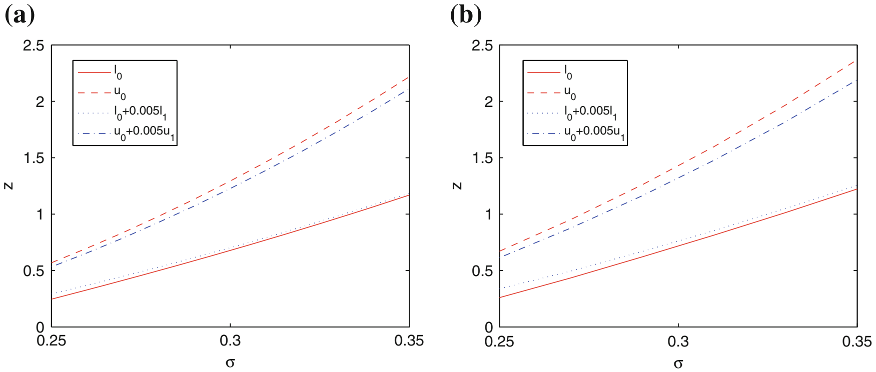


Fig. 7 NT regions as function of volatility σ. **a** $\delta_0 = \delta_1 = 0.15$. **b** $\delta_0 = \delta_1 = 0.3$

4 Conclusion

In this article, we have studied several time-inconsistent problems related to portfolio optimization. By using asymptotic methods, we can handle the nonlocality issue that arises from the game-theoretic methodology framework introduced to tackle time-inconsistency. Tractable solutions have been obtained in situations where the time-inconsistent problems can be closely approximated by time-consistent ones, which can also provide a qualitative/directional characterization of the equilibrium investment strategies in more general cases. Our results are intuitive and can describe how differently investors behave in reality and in time-consistent settings.

5 Derivation of (47)

This part is to demonstrate that the expansion for (45) will lead to (47). The expansion for the left hand of the equation is straightforward and therefore omitted. The main challenge of the expansion is the term:

$$\mathbb{E}_{t,x}\left[\int_t^T \varphi'(s-t)U(c_s^*)ds\right].$$

We start by introducing the following expansions:

$$\begin{aligned}
&c^*(\cdot) = c_0^*(\cdot) + \epsilon c_1^*(\cdot) + o(\epsilon^2), \quad X_t = X_t^{(0)} + \epsilon X_t^{(1)} + o(\epsilon^2),\\
&\varphi(\tau) = e^{-\delta_0\tau} + \epsilon\Delta(\tau)e^{-\delta_0\tau} + o(\epsilon^2),
\end{aligned}$$

which will be plugged into the equilibrium value function for power utility function:

$$\begin{aligned}
V(t,x) &= \mathbb{E}_{t,x}\left[\int_t^T \varphi(s-t)U(c^*(X_s))ds\right]\\
&\approx \mathbb{E}_{t,x}\left[\int_t^T \varphi(s-t)\left(\frac{c_0^*(X_s^{(0)})^{1-\gamma}}{1-\gamma} + \epsilon c_0^*(X_s^{(0)})^{-\gamma}c_{0,x}^*(X_s^{(0)})\,X_s^{(1)}\right.\right.\\
&\qquad\qquad\left.\left. + \epsilon c_0^*(X_s^{(0)})^{-\gamma}c_1^*(X_s^{(0)})\right)ds\right]\\
&= \mathbb{E}_{t,x}\left[\int_t^T e^{-\delta_0(s-t)}\frac{c_0^*(X_s^{(0)})^{1-\gamma}}{1-\gamma}\,ds\right]
\end{aligned}$$

$$
\begin{aligned}
&+\epsilon\mathbb{E}_{t,x}\left[\int_t^T \Delta(s-t)e^{-\delta_0(s-t)}\frac{c_0^*(X_s^{(0)})^{1-\gamma}}{1-\gamma}\right.\\
&\qquad + e^{-\delta_0(s-t)}c_0^*(X_s^{(0)})^{-\gamma}c_{0,x}^*(X_s^{(0)})\,X_s^{(1)}\\
&\qquad \left. + e^{-\delta_0(s-t)}c_0^*(X_s^{(0)})^{-\gamma}c_1^*(X_s^{(0)})ds\right]\\
&=: V_0(t,x)+\epsilon V_1(t,x).
\end{aligned}
$$

This leads to the expansion:

$$
\begin{aligned}
&\mathbb{E}_{t,x}\left[\int_t^T \varphi'(s-t)U(c_s^*)ds\right]\\
&\approx -\delta_0\mathbb{E}_{t,x}\left[\int_t^T e^{-\delta_0(s-t)}\frac{c_0^*(X_s^{(0)})^{1-\gamma}}{1-\gamma}\,ds\right]\\
&\quad -\epsilon\delta_0\mathbb{E}_{t,x}\left[\int_t^T \Delta(s-t)e^{-\delta_0(s-t)}\frac{c_0^*(X_s^{(0)})^{1-\gamma}}{1-\gamma}\right.\\
&\qquad + e^{-\delta_0(s-t)}c_0^*(X_s^{(0)})^{-\gamma}c_{0,x}^*(X_s^{(0)})\,X_s^{(1)}\\
&\qquad \left. + \; e^{-\delta_0(s-t)}c_0^*(X_s^{(0)})^{-\gamma}c_1^*(X_s^{(0)})ds\right]\\
&\quad +\epsilon\mathbb{E}_{t,x}\left[\int_t^T \Delta'(s-t)e^{-\delta_0(s-t)}\frac{[c_{0,s}^*(X_s^{(0)})]^{1-\gamma}}{1-\gamma}ds\right]\\
&= -\delta_0 V_0-\epsilon\delta_0 V_1+\epsilon\mathbb{E}_{t,x}\left[\int_t^T \Delta'(s-t)e^{-\delta_0(s-t)}\frac{[c_{0,s}^*(X_s^{(0)})]^{1-\gamma}}{1-\gamma}ds\right],
\end{aligned}
$$

of which the zeroth order term will go into the RHS of the first equation in (47) and the remaining two terms will go into the V_1 equation.

References

1. T. Bjork, A. Murgoci, A general theory of markovian time inconsistent stochastic control problems. Preprint (2010)
2. T. Bjork, A. Murgoci, X. Zhou, Mean-variance portfolio optimization with state-dependent risk aversion. Math. Financ. (2012)

3. B. Bouchard, L. Moreau, M.H. Soner, Hedging under an expected loss constraint with small transaction costs. arXiv preprint arXiv:1309.4916 (2013)
4. J.Y. Campbell, L.M. Viceira, Consumption and portfolio decisions when expected returns are time varying. Q. J. Econ. **114**(2), 433–495 (1999)
5. R. Carmona, *Indifference Pricing: Theory and Applications*, Series in Financial Engineering (Princeton University Press, Princeton, 2008)
6. G. Chacko, L.M. Viceira, Dynamic consumption and portfolio choice with stochastic volatility in incomplete markets. Rev. Financ. Stud. **18**(4), 1369–1402 (2005)
7. P. Cheridito, M. Kupper, Composition of time-consistent dynamic monetary risk measures in discrete time. Int. J. Theor. Appl. Financ. **14**(01), 137–162 (2011)
8. M. Chernov, A.R. Gallant, E. Ghysels, G. Tauchen, Alternative models for stock price dynamics. J. Econom. **116**(1), 225–257 (2003)
9. V. Coudert, M. Gex, Does risk aversion drive financial crises? testing the predictive power of empirical indicators. J. Empir. Financ. **15**(2), 167–184 (2008)
10. M.H. Davis, A.R. Norman, Portfolio selection with transaction costs. Math. Oper. Res. **15**(4), 676–713 (1990)
11. I. Ekeland, A. Lazrak, Being serious about non-commitment: subgame perfect equilibrium in continuous time. arXiv preprint arXiv:math/0604264 (2006)
12. I. Ekeland, A. Lazrak. Equilibrium policies when preferences are time inconsistent. arXiv preprint arXiv:0808.3790 (2008)
13. I. Ekeland, O. Mbodji, T.A. Pirvu, Time-consistent portfolio management. SIAM J. Financ. Math. **3**(1), 1–32 (2012)
14. J.-P. Fouque, G. Papanicolaou, R. Sircar, K. Solna, *Multiscale Stochastic Volatility for Equity, Interest Rate, and Credit Derivatives* (Cambridge University Press, Cambridge, 2011)
15. J.-P. Fouque, R. Sircar, T. Zariphopoulou, Portfolio optimization and stochastic volatility asymptotics. Preprint (2013)
16. C. Harris, D. Laibson, Dynamic choices of hyperbolic consumers. Econometrica **69**(4), 935–957 (2001)
17. C.J. Harris, D. Laibson. Hyperbolic discounting and consumption, in *Advances in Economics and Econometrics: Theory and Applications, vol. 1*, in ed. by M. Dewatripont, L.P. Hansen, S. Turnovsky (Eighth World Congress, 2002), pp. 258–298
18. Y. Hu, H. Jin, X.Y. Zhou, Time-inconsistent stochastic linear-quadratic control. SIAM J. Control Optim. **50**(3), 1548–1572 (2012)
19. J. Hull, A. White, The pricing of options on assets with stochastic volatilities. J. Financ. **42**(2), 281–300 (1987)
20. E. Karni, Risk aversion for state-dependent utility functions: measurement and applications. Int. Econ. Rev. **24**(3), 637–647 (1983)
21. T.C. Koopmans, Stationary ordinal utility and impatience. Econom. J. Econom. Soc. **28**, 287–309 (1960)
22. H. Kraft, Optimal portfolios and Heston's stochastic volatility model: an explicit solution for power utility. Quant. Financ. **5**(3), 303–313 (2005)
23. G. Loewenstein, D. Prelec, Anomalies in intertemporal choice: evidence and an interpretation. Q. J. Econ. **107**(2), 573–597 (1992)
24. R. Merton, *Continuous-Time Finance* (Wiley, New York, 1992)
25. H. Pham, *Continuous-time Stochastic Control and Optimization with Financial Applications* (Springer, Berlin, 2009)
26. T.A. Pirvu, H. Zhang, Utility indifference pricing: a time consistent approach. Appl. Math. Financ. **20**(4), 304–326 (2013)
27. M. Scheicher, What drives investor risk aversion? daily evidence from the german equity market. BIS Q. Rev. 67–74 (2003)
28. S. Shreve, H. Soner, Optimal investment and consumption with transaction costs. Ann. Appl. Prob. **4**, 609–692 (1994)
29. R.H. Strotz, Myopia and inconsistency in dynamic utility maximization. Rev. Econ. Stud. **23**, 165–180 (1955)

30. N. Tarashev, K. Tsatsaronis, D. Karampatos, Investors' attitude towards risk: what can we learn from options? BIS Q. Rev. **6**, 57–65 (2003)
31. J.A. Wachter, Portfolio and consumption decisions under mean-reverting returns: an exact solution for complete markets. J. Financ. Quant. Anal. **37**(01), 63–91 (2002)

Decompositions of Diffusion Operators and Related Couplings

David Elworthy

Abstract Results by Cranston, Greven, and Feng-Yu Wang, on relationships between coupling and shift coupling, and harmonic functions and space time harmonic functions are reviewed. These lead to extensions of a result by Freire on the separate harmonicity of bounded harmonic functions on certain product manifolds. The extensions are to situations where a diffusion operator is decomposed into the sum of two other commuting diffusion operators. This is shown to arise for a class of foliated Riemannian manifolds with totally geodesic leaves. A form of skew product decomposition of Brownian motions on these foliated manifolds is obtained, as are gradient estimates in leaf directions. Relationships between stochastic completeness of the manifold itself and stochastic completeness of its leaves are established. Baudoin and Garafola's "sub-Riemannian manifolds with transverse symmetries" are shown to be examples.

Keywords Foliations · Stochastic analysis · Coupling · Bounded harmonic functions · Commuting diffusion operators · Non-explosion · Hypo-elliptic diffusions

1 Introduction

Let M be a C^∞, connected, n-dimensional manifold without boundary. By a *diffusion operator* on M we will mean a smooth semi-elliptic operator acting on functions on M with no zero order term. The standard example, and the one of most interest here, is the Laplace-Beltrami operator Δ^M when M is a Riemannian manifold. We take $\Delta^M f := \operatorname{div} \operatorname{grad} f$, so Brownian motion on the Riemannian manifold M has generator $\frac{1}{2}\Delta^M$. The smoothness of the coefficients ensure that such an operator has associated to it a unique diffusion process, which is a solution of the corresponding martingale problem, see [20].

Given a diffusion operator $\mathcal{L}$ on M we will consider "attainability" subsets, "constancy" subsets, and the related notion of anisotropic gradient estimates.

D. Elworthy (✉)
University of Warwick, Coventry, UK
e-mail: kde@maths.warwick.ac.uk

D. Crisan et al. (eds.), *Stochastic Analysis and Applications 2014*,
Springer Proceedings in Mathematics & Statistics 100,
DOI 10.1007/978-3-319-11292-3_10

Definition 1.1 For a fixed diffusion operator $\mathcal{L}$ on M, when $x, y \in M$ write $x \sim_{sc} y$ if there is a successful shift coupling (defined below) between the $\mathcal{L}$-diffusions starting from x and y, and write $x \sim_{b\mathcal{H}} y$ if every bounded $\mathcal{L}$-harmonic function h has $h(x) = h(y)$. A subset A of M will be called an *attainability set* if all its elements are equivalent under $\sim_{sc}$ and a *constancy set* if they are for $\sim_{b\mathcal{H}}$.

An extreme example of anisotropic gradient estimates arises when M has a foliation by immersed manifolds, "leaves", each one of which is a constancy set, so giving trivial gradients in the directions of the leaves.

We shall discuss these subsets in the context of decompositions $\mathcal{L} = \mathcal{A} + \mathcal{B}$ of $\mathcal{L}$ into the sum of two diffusions, for example as in skew product decompositions. In the first sections we review known results about coupling, harmonic functions and such decompositions, bringing out their interpretation in terms of the notions above. Then we consider a class of foliations of Riemannian manifolds, Riemannian foliations with totally geodesic leaves, for which the Laplace-Beltrami operator decomposes into commuting diffusion operators, the Laplacian along the leaves of the foliation and a transverse diffusion operator. For such foliations we show how a coupling for the leaf Laplacian diffusions starting from points on the same leaf gives a coupling for the Brownian motions of M starting at these points with the same coupling time.

On the way, in Theorem 3.10, we give results on relationships between stochastic completeness of the manifold and stochastic completeness of its leaves.

In two appendices we relate our situation to that considered by Baudoin and collaborators from the viewpoint of Bakry-Emery theory; see [4–6, 10, 31].

Most of the observations about invariant diffusions on foliated manifolds come from joint work with XueMei Li and Yves LeJan.

2 Shift Coupling, Coupling, and Harmonic Maps

Let $\mathcal{C}(M^+)$ be the usual space of continuous paths into the one point compactification M^+ of M which once at infinity stay there. As usual $\{\theta_t : t \geq 0\}$ denotes the shift flow on $\mathcal{C}(M^+)$, given by $\theta_t(\sigma)(s) = \sigma(t+s)$. The Borel sigma-algebra of $\mathcal{C}(M^+)$ has the two subalgebras: the invariant sigma-algebra $\mathcal{I}$ and the tail sigma-algebra $\mathcal{T}$ defined by

$$\mathcal{I} = \{B \in \mathrm{Borel}(\mathcal{C}(M^+)) : \theta_t^{-1} B = B, \quad \text{for all } t \geq 0\}$$

and

$$\mathcal{T} = \bigcap_{t \geq 0} \theta_t^{-1} \mathrm{Borel}(\mathcal{C}(M^+)).$$

We recall the definition of shift couplings for general continuous time processes from Thorisson [28].

Definition 2.1 Let Z^1 and Z^2 be continuous processes with values in M. By a *coupling* of Z^1 and Z^2 we mean a pair of continuous processes $\hat{Z}^1$, $\hat{Z}^2$ defined on the same probability space $\{\Omega, \mathcal{F}, \mathbf{P}\}$, with the same laws as Z^1 and Z^2 respectively. Treating $\hat{Z}^1$, $\hat{Z}^2$ as $\mathcal{C}(M^+)$-valued random variables we then say a pair (T_1, T_2) of random times $T_j : \Omega \to [0, \infty]$ is a *shift-coupling* if

1. $\{T_1 < \infty\} = \{T_2 < \infty\}$
2. $\theta_{T_1}\hat{Z}^1 = \theta_{T_2}\hat{Z}^2$ on $\{T_1 < \infty\}$.

The shift coupling is *successful* if $\mathbf{P}\{T_1 < \infty\} = 1$. When $T_1 = T_2$ the common value T, say, is called a *coupling time* and if this is almost surely finite we have a *successful coupling*.

A basic result from [28], which only requires M to be a Polish space, is

Theorem 2.2 (Thorisson 1994) *The following statements are equivalent:*

1. *There is a successful shift coupling of Z^1 and Z^2*
2. *For $A \in \mathcal{I}$ we have $\mathbf{P}\{Z^1 \in A\} = \mathbf{P}\{Z^2 \in A\}$.*

For couplings the corresponding result holds with the tail sigma-algebra replacing the invariant sigma-algebra.

Now fix a diffusion operator $\mathcal{L}$ on M. Let $\{P_t : t \geq 0\}$ be its associated semi-group, and $\mathbf{P}_x$ the law of the $\mathcal{L}$-diffusion starting at a point $x \in M$. For simplicity assume that $\mathcal{L}$ is conservative, i.e $P_t 1 = 1$, or equivalently the $\mathcal{L}$-diffusions do not explode. A universally measurable function $h : M \to \mathbf{R}$ is $\mathcal{L}$-harmonic if $P_t h = h$ for all $t \geq 0$. When h is bounded this holds if and only if $\{h(\sigma_t) : t \geq 0\}$ is a martingale. From Dynkin [14], such functions satisfy $\mathcal{L}h = 0$. Let $b\mathcal{H}$, or $b\mathcal{H}^{\mathcal{L}}$, be the vector space of bounded $\mathcal{L}$-harmonic functions.

The fundamental result relating bounded harmonic functions to the invariant sigma-algebra as stated and proved in [27, p. 423], is

Theorem 2.3 *The formula $h(x) = \mathbf{E}_x(F)$ gives a bijection between $b\mathcal{H}$ and $\mathbf{P}_x$-almost sure equivalence classes of bounded $\mathcal{I}$-measurable $F : \mathcal{C}(M) \to \mathbf{R}$. Moreover $F(\sigma) = \lim_{t\to\infty} h(\sigma_t)$ almost surely.*

Recall that the corresponding result holds for bounded *space-time harmonic* functions, i.e. solutions of $\frac{\partial}{\partial t}h + \mathcal{L}h = 0$, using the tail sigma-algebra.

From these theorems we immediately see that $x \sim_{sc} y$ if and only if $\sim_{b\mathcal{H}}$ and *constancy and attainability are equivalent conditions*.

Although we shall not use it we note a quantitative sharpening of Thorisson's theorem for the case of transient diffusions by Cranston and Greven [12]:
For $x, y \in M$ there exists a shift coupling (T_1, T_2) between $\mathcal{L}$-diffusions from x and y such that

$$\text{Prob}\{T_1 = \infty\} + \text{Prob}\{T_2 = \infty\} = \sup_{h \in b\mathcal{H}, |h| \leq 1} |h(x) - h(y)|.$$

2.1 Examples

1. For an elliptic $\mathcal{L}$, if M has a non-empty open attainability set then M is attainable, since a harmonic function constant on a non-trivial open set is constant.
2. Let M be $\mathbf{R}^3$ with Heisenberg group structure and $\mathcal{L}$ the corresponding hypo-elliptic diffusion operator. Then $b\mathcal{H}$ is trivial, [12].
3. Suppose M has the structure of a product of complete Riemannian manifolds, $M = F \times Q$. Take $\mathcal{L} = \frac{1}{2}\triangle^M$ and observe

$$\triangle^M f(x, y) = \triangle^F(f(-, y))(x) + \triangle^Q(f(x, -))(y).$$

Assuming that F admits only constant harmonic functions, or equivalently is attainable for $\triangle^F$, we easily see that $F \times \{y\}$ is an attainable set for each $y \in Q$ and consequently every bounded harmonic function on M is constant on each $F \times \{y\}$. See [16]. This example will be generalised below.

2.2 Regular Diffusions

Under mild conditions there is the pleasant situation that one does not need to be concerned about the difference between shift-coupling and coupling nor between tail and invariant sigma-algebras. We will say that a diffusion operator $\mathcal{L}$ is *regular* if all bounded space-time $\mathcal{L}$-harmonic functions are constant in time and so $\mathcal{L}$-harmonic. In other words if every bounded solution $f : [0, \infty) \times M \to \mathbf{R}$ to $\frac{\partial}{\partial t}h + \mathcal{L}h = 0$, or more precisely which satisfies $P_s f(t + s, -) = f(t, -)$ for $s, t > 0$, is independent of time, and so $\mathcal{L}$-harmonic. From above this is equivalent to the tail and invariant sigma-algebras agreeing up to sets of measure zero. From [13] we have:

Theorem 2.4 (M. Cranston and F-Yu Wang) *Suppose the operator $\mathcal{L}$ satisfies the parabolic Harnack inequality*

$$P_t f \le \Phi(P_{t+h} f) \quad 0 \le f \le 1, \tag{1}$$

for some $t, h > 0$ and some continuous increasing $\Phi : [0, 1] \to \mathbf{R}$ with $\Phi(0) < 1$. Then a successful shift coupling exists for any pair of initial distributions if and only if so does a successful coupling. Equivalently $\mathcal{L}$ is regular.

Cranston and Wang give a detailed discussion of the inequality (1) in [13]. In particular they show that it holds for a class of sub-elliptic operators, for $\triangle^M$ for M complete Riemannian with Ricci curvature bounded below, and in the following cases:

1. $\mathcal{L} = \triangle^M + Z$ for a complete Riemannian M, where Z is a smooth vector field on M and we have the Bakry-Emery curvature dimension inequality $CD(\rho, k)$:

$$\Gamma_2(f,f) \geq \rho\Gamma(f,f) + \frac{1}{k}(\mathcal{L}f)^2 \tag{2}$$

for all smooth compactly supported $f : M \to \mathbf{R}$ and some $\rho \in \mathbf{R}$, some $0 < k < \infty$. (See [3] or [2] for Γ_2.)

2. $\mathcal{L}$ as in 1 above on a complete Riemannian manifold such that $CD(\rho, \infty)$ holds for some $\rho \in \mathbf{R}$ together with

$$\inf_{x \in M} P_t \mathbf{1}_{B(x,r)}(x) > 0 \quad \text{for some } t > 0 \text{ and some } r > 0.$$

Here $B(x, r)$ is the closed ball about x of radius r.

3. For M a unimodular Lie group and $\mathcal{L} = \frac{1}{2}\sum_{k=1}^{k=m}(X^k)^2$ with $X^1, \ldots, X^k$ left invariant vector fields.

Also they show that if complete Riemannian manifolds M and N are roughly isometric, e.g. see [21], and $\triangle^N$ is regular, then so is $\triangle^M$.

2.3 Decompositions of Diffusion Operators

The following, simple but seemingly not well known result, was suggested by a result of Freire [18]. Freire considered positive harmonic functions as well as bounded ones, on products of Riemannian manifolds.

Theorem 2.5 *Suppose $\mathcal{L}$ is the sum of two commuting smooth diffusion operators $\mathcal{L} = \mathcal{A}^1 + \mathcal{A}^2$ with $P_t = P_t^1 P_t^2$ and $P_t^2 P_s^1 = P_s^1 P_t^2$, all $s, t \geq 0$, for the associated diffusion semi-groups. Assume $\mathcal{A}^1$ is regular. Then every bounded $\mathcal{L}$-harmonic h is both $\mathcal{A}^1$-harmonic and $\mathcal{A}^2$-harmonic.*

Proof Suppose h is bounded and $\mathcal{L}$-harmonic. Then for $s, t > 0$

$$P_t^1 P_{t+s}^2 h = P_s^2 P_t^1 P_t^2 h = P_s^2 P_t h = P_s^2 h.$$

Thus $(s, x) \mapsto P_s^2 h$ is space-time $\mathcal{A}^1$-harmonic. Since it is bounded and $\mathcal{A}^1$ is regular it is independent of time and $\mathcal{A}^1$-harmonic. Thus $P_s^2 h = h$ for all $s \geq 0$, i.e. h is $\mathcal{A}^2$-harmonic, and also h is $\mathcal{A}^1$-harmonic, as required. □

We go on to give examples of such decompositions.

3 Submersions and Foliations

We will describe a class of decomposable diffusion operators in increasing generality, starting with now "classical" results.

For any C^2 map $p : M \to N$ we say a diffusion operator $\mathcal{L}$ on M *lies over* a diffusion operator $\mathcal{A}$ on N when $\mathcal{L}(f \circ p) = \mathcal{A}(f) \circ p$ for smooth $f : N \to \mathbf{R}$.

If so we say $\mathcal{L}$ and $\mathcal{A}$ are *intertwined* by p. Following [17], a diffusion operator $\mathcal{B}$ on M is *along* a sub-bundle S of TM if $\delta^{\mathcal{B}}(\phi) = 0$ for any C^1 section ϕ of the annihilator sub-bundle S^0 of S in T^*M. Here $\delta^{\mathcal{B}}$ is the first order operator from one forms to functions canonically associated to $\mathcal{B}$, so $\mathcal{B}f = \delta^{\mathcal{B}} df$, see [17]. If $\mathcal{B}$ has a Hörmander form then $\mathcal{B}$ is along S iff all the vector fields involved are sections of S. Also $\mathcal{B}$ is *cohesive* if its principle symbol $\sigma^{\mathcal{B}} : T^*M \to TM$ has constant rank, and so has image a sub-bundle E of TM, and moreover $\mathcal{B}$ is along E. Thus elliptic diffusion operators are cohesive, with $E = TM$. Cohesive diffusion operators have smooth Hörmander form representations, $\mathcal{B} = \frac{1}{2}\sum_{j=1}^{k} X^j X^j + A$ say, and then for each $x \in M$,

$$\operatorname{span}\{X^1(x), \dots, X^k(x)\} = E_x \quad \text{and } A(x) \in E_x.$$

3.1 Riemannian Submersions

A smooth surjective map $p : M \to N$ of Riemannian manifolds is a *Riemannian submersion* if its derivative maps $T_u p : T_u M \to T_{p(u)} N$ are orthogonal projections, i.e. for each $u \in M$ the map $T_u p (T_u p)^* : T_{p(u)} N \to T_{p(u)} N$ is the identity.

Recall that a submanifold of M, such as a fibre $p^{-1}(y)$, is *totally geodesic* if any geodesic of the submanifold with its induced Riemannian structure is also a geodesic of M. This holds if and only if the second fundamental form of the submanifold vanishes identically. The submanifold is *minimal* if the trace of the second fundamental form, the 'mean curvature', vanishes identically.

We have the following, see [17] for a detailed discussion,

1. Δ^M lies over Δ^N iff p has minimal fibres.
   Equivalently p maps Brownian motion on M to Brownian motion on N iff p has minimal fibres.
2. If the fibres are minimal there is a decomposition $\Delta^M = \Delta^V + (\Delta^N)^H$ where $(\Delta^N)^H$ is horizontal, i.e along the horizontal sub-bundle of TM and over Δ^N, and Δ^V is vertical, i.e. along the kernel of Tp. These commute iff Δ^V commutes with Lie differentiation by all horizontal lifts of vector fields on N, which holds iff the fibres are totally geodesic. Herman [19], Berard-Bergery and Bourguignon [8].
3. For complete manifolds there is a skew-product decomposition of BM on M. This is used to see the fibres are constancy sets for Δ^M if they are availability sets for Δ^V and totally geodesic. Elworthy-Kendall [16]. See also [23].

3.2 Intertwined Diffusions

Suppose $p : M \to N$ is smooth and surjective with $\mathcal{L}$ on M over $\mathcal{A}$ on N. Assume $\mathcal{A}$ is cohesive. From [17] we know that the principal symbols of $\mathcal{L}$ and $\mathcal{A}$ determine *horizontal lift* maps (a non-linear semi-connection) $\mathfrak{h}_u : E_{p(u)} \to T_u M$, smooth in $u \in M$, linear, and with $T_u p \circ \mathfrak{h}_u : E_{p(u)} \to E_{p(u)}$ the identity. The image determines the horizontal sub-bundle H of TM. Sections of H are called *horizontal vector fields*. A diffusion operator is *horizontal* if it is along H and is a *lift of* $\mathcal{A}$ if it lies over $\mathcal{A}$. We have the following extension by Elworthy et al. [17], of the results of Hermann and Berard-Bergery and Bourguignon.

Theorem 3.1 *For smooth $p : M \to N$ intertwining $\mathcal{L}$ with a cohesive $\mathcal{A}$ on N:*

[i] *There is a unique decomposition $\mathcal{L} = \mathcal{A}^H + \mathcal{L}^V$ into the sum of two smooth diffusion operators $\mathcal{A}^H$, the horizontal lift of A, and $\mathcal{L}^V$, a vertical diffusion operator.*

[ii] *Assuming completeness and strong stochastic completeness of the semi-connection determined by $\mathfrak{h}$, in this decomposition the operators commute and so do their semigroups if and only if $\mathcal{L}^V$ commutes with Lie differentiation along horizontal lifts of all vector fields in any Hörmander form representation of $\mathcal{A}$.*

By *completeness* of the connection we mean that the horizontal lifts of C^1 curves in N along E can be defined for all time, while the *strong stochastic completeness* means that horizontal lifts of the $\mathcal{A}$-diffusion paths $\{x_t\}_t$ from $x_0 \in N$ to M starting at arbitrary points on $p^{-1}(x_0)$ can not only be defined for all time but can be chosen to give continuous maps of $p^{-1}(x_0)$ into M which are diffeomorphisms of $/\!/_t : p^{-1}(x_0) \to p^{-1}(x_t)$ almost surely. We consider this as a stochastic parallel translation.

In the commuting case of part [ii] we have $P_t = P_t^V P_t^H$ and $P_s^V P_t^H = P_t^H P_s^V$ all $s, t \geq 0$, in an obvious notation. This comes [17], from the *stochastic holonomy invariance* of $\mathcal{L}^V$:

$$/\!/_s^*(\mathcal{L}^V) = \mathcal{L}^V \qquad s \geq 0.$$

If we have this we also obtain a skew-product decomposition of $\mathcal{L}$-diffusions: $u_t := /\!/_t(y_t)$ is an $\mathcal{L}$-diffusion if $\{y_t\}_t$ is a $\mathcal{L}^V$-diffusion on the fibre $p^{-1}(x_0)$.

When $\mathcal{L}^V$ is regular and we have commutation we can apply Theorem 2.5 to conclude that each fibre of p is a constancy set for $\mathcal{L}$ if all bounded $\mathcal{L}^V$-harmonic functions are constant on such fibres. Alternatively one can use the skew-product decomposition as in [16] and use a coupling argument, as we will do below for foliations.

3.3 Invariant Diffusions on Foliated Manifolds

Definition 3.2 A C^r codimension s foliation $\mathcal{F}$ of M is defined by a maximal collection of pairs $\{(U_\alpha, p_\alpha), \alpha \in I\}$ of open subsets U_α of M and submersions $p_\alpha : U_\alpha \to U_\alpha^0$ onto open subsets of $\mathbf{R}^s$ satisfying

- $\cup_{\alpha \in I} U_\alpha = M$

- if $U_\alpha \cap U_\beta \neq \emptyset$, there exists a local C^r diffeomorphism $g_{\alpha\beta}$ of R^s such that $p_\alpha = g_{\alpha\beta} \circ p_\beta$ on $U_\alpha \cap U_\beta$.

The map (p_α) are called *disintegrating maps* of $\mathcal{F}$. The connected components of the sets $p_\alpha^{-1}(c)$, $c \in R^s$ are called the *plaques* of the foliation.

A foliation arises from an integrable sub-bundle of TM, to be denoted by TL. These are the tangent vectors to the *leaves*, the maximal connected integral sub-manifolds of the sub-bundle. The leaves are unions of plaques, and are immersed manifolds of dimension $n - s$. We shall restrict ourselves to smooth foliations.

Submersions $p : M \to N$ are foliations on M with disintegrating maps the compositions of charts of N with restrictions of p to the inverse images under p of the domains of those charts. The leaves are the connected components of the fibres of p. Other examples to bear in mind include the foliations with one dimensional leaves arising from irrational flows on a torus, with dense leaves, and the foliation of a Möbius band by longitudinal circles. For the latter, note for future reference, that one of these circles has half the circumference of the others.

Definition 3.3 Let M be a foliated manifold given by disintegrating maps $p_\alpha : U_\alpha \to U_\alpha^0 \subset \mathbf{R}^s$. A diffusion operator $\mathcal{L}$ on M is said to be *invariant* for the foliation $\mathcal{F}$ if there exists $\mathcal{A}_\alpha$ on U_α^0 such that the restriction $\mathcal{L}_\alpha$ of $\mathcal{L}$ over U_α lies over $\mathcal{A}_\alpha$, that is

$$\mathcal{L}_\alpha(f \circ p_\alpha) = \mathcal{A}_\alpha(f) \circ p_\alpha.$$

Suppose $\mathcal{L}$ is invariant for $\mathcal{F}$ and for any α the rank of the principle symbol $\sigma^{\mathcal{A}_\alpha}(x)$ is constant, giving vector sub-bundles $E^\alpha \to U_\alpha^0$ of the trivial bundles $U_\alpha^0 \times R^s$ over each U_α^0. As described in Sect. 3.2 we have horizontal lift maps $\mathfrak{h}_y^\alpha : E^\alpha_{p_\alpha(y)} \to T_yM$, with $T_yp_\alpha \circ \mathfrak{h}_y^\alpha = Id$. Set $H_y = \text{Image}(\mathfrak{h}_y^\alpha)$. It is independent of α, see Proposition 2.1 of [17], and we obtain a sub-bundle H of TN. This will be transverse to TL in TM if the $\mathcal{A}_\alpha$ are elliptic. There is the following extension of Theorem 3.1[i] to this situation:

Proposition 3.4 *Suppose the $\mathcal{A}_\alpha$ are cohesive. Then there is a unique diffusion operator $\mathcal{A}$ along H whose restrictions to each U_α lie over $\mathcal{A}_\alpha$, and a decomposition $\mathcal{L} = \mathcal{A} + \mathcal{L}^{\mathcal{F}}$ where $\mathcal{L}^{\mathcal{F}}$ is along the leaves of $\mathcal{F}$, i.e. along TL. Moreover $\mathcal{A}$ is cohesive.*

Proof This comes from the uniqueness of the horizontal lifts of the $\mathcal{A}_\alpha$ over p_α given by Theorem 3.1[i]. □

Remark 3.5 The leaf component $\mathcal{L}^{\mathcal{F}}$ restricts to give a smooth diffusion operator on any leaf L. The corresponding diffusion process starting at a point x_0 of L is seen, via uniqueness of the martingale problems, to be an $\mathcal{L}^{\mathcal{F}}$-diffusion when mapped by the inclusion into M. The explosion time on the leaf is the same as that for the $\mathcal{L}^{\mathcal{F}}$-diffusion as can be seen by taking disintegrating charts.

When each $\mathcal{A}_\alpha$ is cohesive we will call $\mathcal{L}$ *transversely cohesive* and the operators $\mathcal{A}$ and $\mathcal{L}^{\mathcal{F}}$ of Sect. 3.4 its *transversal and leaf wise components*. We would like conditions under which they and their semi-groups commute. For the operators themselves to commute when acting on C^2 functions is a local property, and so may be shown by working in disintegrating charts. Commutation of their semigroups does not follow and is not local, as can be seen from Nelson's example, see [17]. It does follow under certain global regularity conditions on the operators and their semigroups, [17, Proposition 6.1], [7]. Rather than proceeding that way we will seek an analogue of the stochastic holonomy invariance used for submersions. We will not pursue this in generality but will restrict ourselves to the Riemannian situation.

3.4 Riemannian Foliations

A foliation on M is said to be *Riemannian* when there is given a Riemannian metric $\{\langle,\rangle_y, y \in M\}$ on the quotient bundle TM/TL such that there is an open cover $\{U_\alpha\}_{\alpha=1}^\infty$ of M with disintegrating maps $p_\alpha : U_\alpha \to U_\alpha^0 \subset R^s$ and a Riemannian metric $\{\langle,\rangle_z^\alpha, z \in U_\alpha^0\}$ on U_α^0, for which $T_y p_\alpha$ induces an isometry of $TM_y/TL_y, \langle,\rangle_y$ with $T_{p_\alpha(y)}U_\alpha^0, \langle,\rangle_{p_\alpha(y)}^\alpha$, each $y \in U_\alpha$.

We follow the books by Molino [25], and by Tondeur [29, 30], to describe the basic properties of Riemannian foliations.

We will be mainly concerned with foliations of Riemannian manifolds whose metrics induce a Riemannian metric on the bundle E orthogonal to TL which when identified with TM/TL gives a Riemannian foliation. These will be called *bundle-like metrics*, following Bruce Reinhart. For these the disintegrating maps p_α will be Riemannian submersions onto U_α^0 with its given Riemannian structure. In particular if the leaves of the foliation are minimal submanifolds of M the Laplacian, $\triangle^M$, will be invariant and locally lies over the Laplacians of the U_α^0. If the leaves are totally geodesic we can apply the results of Herman and Berard-Bergery and Bourguignon described in Sect. 3.1 to see that the leaf and transverse components of $\triangle^M$ commute on C^2 functions.

From the local description it is easy to see that $\mathcal{L}^{\mathcal{F}}$ restricts to the Laplacian on each leaf. Write it as $\triangle^{\mathcal{F}}$.

To see when the semigroups of the two components commute we would like to obtain an analogue of the parallel translation $/\!/_t$ mapping leaf to leaf as in the submersion case of Sect. 3.2. It would need to give isometries between the leaves in order to leave the leaf-wise component $\triangle^{\mathcal{F}}$ invariant. However this is not in general possible, as we see from the observation about the Möbius band above. This problem will not arise on lifting to the transverse frame bundle, as we will now do.

3.4.1 The Transverse Frame Bundle and Its Connection

For a Riemannian foliation there is a natural connection on the orthonormal frame bundle of the quotient bundle TM/TL which extends the so called Bott connection. This can be defined over a disintegrating chart. For such a chart $p_\alpha : U_\alpha \to U_\alpha^0 \subset \mathbf{R}^s$ observe that the derivative $Tp_\alpha : TU_\alpha \to TU_\alpha^0$ determines a mapping of orthonormal frame bundles

$$O(p_\alpha) : O(TM/TL)|_{U_\alpha} \to O(TU_\alpha^0).$$

When M is Riemannian with a bundle-like metric we can replace TM/TL by the transverse bundle E, (which corresponds to the H of Proposition 3.4), and then $O(p_\alpha)(u) = Tp_\alpha \circ u$ for $u : \mathbf{R}^s \to E_x$ a frame at $x \in U_\alpha$.

If we pull back $O(TU_\alpha^0)$ by p_α to get $p_\alpha^*(O(TU_\alpha^0))$ we get from $O(p_\alpha)$ an isomorphism of principal bundles $O(TM/TL) \to p_\alpha^*(O(TU_\alpha^0))$. This can be used to transfer the pulled back Levi-Civita connections of TU_α^0 for each α, to obtain a metric connection on $O(TM/TL)$ or equivalently on TM/TL, or on E in the bundle-like metric case. We will call this the *transverse Levi-Civita connection*, or just the *transverse connection*.

The map $O(p_\alpha)$ is a submersion and gives a disintegrating chart for a foliation $\mathcal{F}^O$ of $O(TM/TL)$ or $O(E)$. The leaves of this foliation are horizontal for the transverse Levi-Civita connection and form coverings of the leaves of $\mathcal{F}$.

The transverse connection determines the bundle mapping

$$X : O(TM/TL) \times \mathbf{R}^s \to T\left(O(TM/TL)\right) \tag{3}$$

defined by $X(u)(e) = h_u(u(e))$ where $h_u : T_xM/T_xL \to T_u\left(O(TM/TL)\right)$ is the horizontal lift for the transverse connection, for u a frame at $x \in M$. We will use this later, in Eq. (6), for the canonical SDE on $O(E)$. As usual, by the equivariance of h_- under the right action of the orthogonal group $O(s)$ on $O(TM/TL)$, or on E, it satisfies

$$X(u.g)(e) = TR_g X(u)(ge) \quad \text{for } e \in \mathbf{R}^s, \quad g \in O(s), \tag{4}$$

where $TR_g : T_u\left(O(TM/TL)\right) \to T_{ug}\left(O(TM/TL)\right)$ is the derivative of right multiplication by $g \in O(s)$.

Note that, by construction of the transverse connection, over any disintegrating chart we have,

$$T_u\left(O(p_\alpha)\right)\left(X(u)(e)\right) = X_\alpha^0(O(p_\alpha)(u))(e) \quad u \in O(TM/TL)|_{U_\alpha}, \quad e \in \mathbf{R}^s \tag{5}$$

where $X_\alpha^0 : O(TU_\alpha^0) \times \mathbf{R}^s \to TO(TU_\alpha^0)$ gives the canonical SDE for the Levi-Civita connection on $O(TU_\alpha^0)$, so

$$X_\alpha^0(u)(e) = h_{\alpha,u}^0(u(e))$$

for a frame u over $y \in U^0_\alpha$, where $h^0_{\alpha,u} : T_y U^0_\alpha \to T_u \left(O(TU^0_\alpha)\right)$ is the horizontal lift given by the Levi-Civita connection of the Riemannian structure on U^0_α.

Definition 3.6 The transverse connection is said to be *complete* if the vector fields $X(-)(e)$ on $O(TM/TL)$ are complete (i.e. have trajectories for all time) for each $e \in \mathbf{R}^s$. It is said to be *stochastically complete* if the canonical SDE

$$du_t = X(u_t) \circ dB_t \tag{6}$$

for $\{B_t\}_{t \geq 0}$ a Brownian motion on $\mathbf{R}^s$, has solutions for all time.

It is a standard result, [30], which follows from the equality (5), that the flows of each of the vector fields $X(-)(e), e \in \mathbf{R}^s$, map leaves of the foliation $\mathcal{F}^O$ to leaves, and so if the connection is complete the leaves are diffeomorphic to each other. Consequently, in this situation, any two leaves of $\mathcal{F}$ have covers which are diffeomorphic; a result of Reinhart [26].

3.4.2 The Transverse Diffusion and "Parallel Translation"

Suppose now that M is Riemannian and complete and $\triangle^M$ is invariant by the foliation. In order for its leaf component $\triangle^{\mathcal{F}}$ to commute on C^2 functions with its transverse component $\mathcal{A}$ we have seen that the leaves of $\mathcal{F}$ must be totally geodesic. In this situation, by the corresponding result for Riemannian submersions we know that when $f : M \to \mathbf{R}$ is C^2

$$\mathcal{A}f = \triangle^E f := \operatorname{trace} \nabla^E df|_E. \tag{7}$$

Consequently we shall call a $\frac{1}{2}\triangle^E$-diffusion process on M a *transversal Brownian motion*.

Since the leaves of $\mathcal{F}^{O(E)}$ cover the leaves of $\mathcal{F}$ there is a uniquely defined lift of $\triangle^{\mathcal{F}}$ to $O(E)$ and it restricts to the Laplacian on each leaf of $\mathcal{F}^O$ given their induced covering Riemannian structure. Write it as $\triangle^{\mathcal{F}^O}$. Write $\frac{1}{2}\triangle^H$ for the generator $\frac{1}{2}\sum_{k=1}^s X^j X^j$ where X^j is the vector field $X(-)(e_j)$ on $O(E)$ for a basis $\{e_j\}_{j=1}^s$ of $\mathbf{R}^s$. Then $\mathcal{L} := \frac{1}{2}\triangle^H + \frac{1}{2}\triangle^{\mathcal{F}^O}$ is invariant by $\mathcal{F}^O$ and $\frac{1}{2}\triangle^H$ is its transverse component. In a disintegrating chart $O(p_\alpha)$ for $\mathcal{F}^{O(E)}$ we are in the situation discussed in [17] with the operator $\mathcal{L}$ lying over the horizontal Laplacian of $O(U^0_\alpha)$.

Our key observations come next. For them we first need to observe that each leaf of $\mathcal{F}^O$ inherits a Riemannian metric from the leaf of $\mathcal{F}$ which it covers and there is the following standard Lemma:

Lemma 3.7 *Suppose M is a Riemannian manifold with a bundle-like metric for a foliation $\mathcal{F}$ with totally geodesic leaves. Then the, possibly only local, flows of the vector fields $X(-)(e)$ on $O(E)$ defined above give local isometries between the leaves of $\mathcal{F}^O$.*

Proof We have already observed that the flows map leaf to leaf. The isometry property is local so we can work over a disintegrating chart $p_\alpha : U_\alpha \to U^0_\alpha$, reducing the situation to that of a Riemannian submersion with totally geodesic fibres.

For a detailed proof take a Hörmander form representation $\sum Y^{0,j}Y^{0,j}$ for $\Delta^{U^0_\alpha}$ and let Y^j be the horizontal lift of $Y^{0,j}$ over p_α. Then $\sum Y^jY^j$ is a Hörmander form of Δ^E. From the Riemannian submersion theory we know that the local flows of the Y^j map portions of leaves isometrically into leaves. It follows that their horizontal lifts to $O(E)|_{U_\alpha}$ map portions of leaves of $\mathcal{F}^O$ isometrically to leaves of $\mathcal{F}^O$, since these leaves isometrically cover those of $\mathcal{F}$.

Thus if $\tilde{Y}^j$ denotes the horizontal lift of Y^j to $O(E)|_{U_\alpha}$ using the transverse connection, the $\tilde{Y}^j$ commute with $\Delta^{\mathcal{F}^O}$. These are horizontal lifts, over $O(p_\alpha)$ of the corresponding lifts of the $Y^{0,j}$ to $O(U^0_\alpha)$ and they give another Hörmander form representation of $\Delta^H := \sum X^jX^j$, for $X^j = X(-)(e_j)$ as usual. Now Δ^H is the horizontal lift over $O(p_\alpha)$ of the usual horizontal Laplacian on $O(U^0_\alpha)$ and we can apply Theorem 6.0.2 of [17] to see that Δ^H commutes with $\Delta^{\mathcal{F}^O}$ and so each of the basic, for $O(p_\alpha)$, vector fields $X(-)(e)$ commutes with $\Delta^{\mathcal{F}^O}$, giving the required result. □

Parts [i] and [iii] of the following theorem came out of discussions with XueMei Li and Yves LeJan. Part [iii] gives the appropriate analogue, for foliations, of the skew product decomposition for submersions in [16].

Theorem 3.8 *Suppose M is a complete Riemannian manifold with a bundle-like metric for a foliation $\mathcal{F}$ with totally geodesic leaves.*

[i] *The solutions to the canonical SDE (6), project down to form transversal Brownian motions on M.*

[ii] *If $\frac{1}{2}\Delta^E$ is conservative, i.e the transverse Brownian motions exist for all time, then the connection is strongly stochastically complete in the sense that for each leaf L of $\mathcal{F}^O$ there is a version of the solution flow of the canonical SDE which gives a continuous family of smooth maps*

$$//_t : L \to O(E)$$

and maps L isometrically onto leaves of $\mathcal{F}^O$ at each time $t \geq 0$.

[iii] *If $\{\tilde{x}_t : 0 \leq t < \zeta\}$ is the lift to a leaf of $\mathcal{F}^O$ of a leaf Brownian motion $\{x_t : 0 \leq t < \zeta\}$ on some leaf L_{x_0} of $\mathcal{F}$ which is independent of $\{//_t\}_{t\geq 0}$, and if $\frac{1}{2}\Delta^E$ is conservative, then the $O(E)$-valued process $\{//_t(\tilde{x}_t) : 0 \leq t < \zeta\}$ projects to a Brownian motion on M.*

Proof It is enough to check [i] over a disintegrating chart $p_\alpha : U_\alpha \to U^0_\alpha$, which is a Riemannian submersion. Since the canonical SDE on $O(E)|_{U_\alpha}$ lies over that of $O(TU^0_\alpha)$ via $O(p_\alpha)$, the solutions to that SDE get mapped to the solutions of the one for $O(TU^0_\alpha)$. However it is standard that the latter projects down to Brownian motion on U^0_α. See [1, p. 144]. The projection of the solutions on $O(E)|_{U_\alpha}$ down to U_α are horizontal with respect to the map p_α, in the sense of Riemannian submersions, see Sect. 3.2. We now know they are lifts of Brownian motions on U^0_α. Therefore their generator is the horizontal lift via p_α of $\frac{1}{2}\triangle^{U_\alpha^0}$, which is the transverse component $\frac{1}{2}\triangle^E$ of $\frac{1}{2}\triangle^M$ as required.

For part [ii] we first fix a point $u_0 \in O(E)$ above $x_0 \in M$. By part [i] the solution $\{u_t\}_t$ of the canonical SDE, (6), starting at u_0 is the horizontal lift of a transverse Brownian motion which we are to assume exists for all time. It follows that the canonical SDE (6) is complete. We will choose a local flow for it following [15], as [1, 22]. For this there is a subset Ω_0 of full measure in the probability space Ω for the driving Brownian motion $B_.$ of our SDE (6), and stopping times η^u : $\Omega \to [0, \infty]$, $u \in O(E)$, together with a version of the (measurable) solution flow $/\!\!/_t : t \geq 0$ such that if we set $O(E)(t, \omega) = \{u \in O(E)\ s.t.\ \eta^u > t\}$ then for each $(t, \omega) \in [0, \infty) \times \Omega_0$ we have

(a) $O(E)(t, \omega)$ is open in $O(E)$
(b) $/\!\!/_t(\omega) : O(E)(t, \omega) \to O(E)$ is a smooth diffeomorphism onto an open subset of $O(E)$.

Moreover if for any $K \subset O(E)$ we set $\eta^K(\omega) = \inf\{\eta^u(\omega)\ s.t.\ u \in K\}$, then
(c) when K is compact, $u_0 \in O(E)$ and $d(-, -)$ is a complete metric on $O(E)$,

$$\sup_{u \in K} d(u_0, /\!\!/_t(u, \omega)) \to \infty \quad \text{as} \quad t \uparrow \eta^K(\omega) \quad \text{almost surely on} \quad \eta^K < \infty. \tag{8}$$

Since the Riemannian metric on M is assumed complete we can take $d(-, -)$ to come from the natural equivariant Riemannian metric of $O(E)$ which lies over that of M, using the splitting $TO(E) = HTO(E) \bigoplus VTO(E)$ of the tangent space of the frame bundle into horizontal and vertical parts given by the transverse connection. Then if u, v lie in a leaf L of $\mathcal{F}^O$ we have $d(u, v) \leq d^L(u, v)$ for d^L the leaf metric.

By Wong-Zakai approximation [15], Lemma 3.7 implies that $/\!\!/_t^*(\triangle^{O(\mathcal{F})}) = \triangle^{O(\mathcal{F})}$ for $t \geq 0$, almost surely, where defined, c.f. [17, p. 109]. Also each $/\!\!/_t(\omega)$ maps the intersections of leaves of $\mathcal{F}^O$ with $O(E)(t, \omega)$ to leaves because the X^j are basic, as described after Definition 3.6. It follows that it does this isometrically.

Now fix $u_0 \in O(E)$ and let K be a compact subset of the leaf L_{u_0} of $\mathcal{F}^O$ through u_0, with $u_0 \in K$. Let $\{u_t\}_{0 \leq t < \eta^{u_0}}$ be the solution to our SDE from u_0. Then if $0 \leq t < \eta^K(\omega)$ we have, almost surely,

$$\begin{aligned} \sup_{u \in K} d(u_0, /\!\!/_t(u, \omega)) &\leq d(u_0, u_t(\omega)) + \sup_{u \in K} d(u_t(\omega), /\!\!/_t(u, \omega)) \\ &\leq d(u_0, u_t(\omega)) + \sup_{u \in K} d^{L_{u_t}}(u_t(\omega), /\!\!/_t(u, \omega)) \\ &\leq d(u_0, u_t(\omega)) + R \end{aligned}$$

where R is the diameter of K in the leaf metric.
From [c] above we see that $\eta^K \geq \eta^{u_0}$ almost surely and so η^K is almost surely infinite if Δ^E is conservative. Since L_{u_0} is a countable union of such compact sets K we see [ii] holds.

To prove [iii] it suffices to show that $\{/\!/_t(\tilde{x}_t)\}_t$ solves the martingale problem for $\frac{1}{2}(\triangle^H + \triangle^{\mathcal{F}^O})$ since that operator lies over $\frac{1}{2}\triangle^M$. For this take $f : O(E) \to \mathbf{R}$ which is smooth and of compact support. Then, for $t < \zeta$, writing $u_t := /\!/_t(\tilde{x}_t)$

$$
\begin{aligned}
f(u_t) &= f(u_0) + \int_0^t (df)_{u_s} T /\!/_s \circ \, d\tilde{x}_s + \int_0^t (df)_{u_s} X(u_s) \circ \, dB_s \\
&= \text{ martingale} + \int_0^t \frac{1}{2} \triangle^{\mathcal{F}^O} (f \circ /\!/_s)(\tilde{x}_s)\, ds + \int_0^t \frac{1}{2} \triangle^H (f)(u_s)\, ds \\
&= \int_0^t \frac{1}{2} (\triangle^{\mathcal{F}^O} + \triangle^H)(f)(u_s)\, ds,
\end{aligned}
$$

because $\triangle^{\mathcal{F}^O}(f \circ /\!/_s) = \triangle^{\mathcal{F}^O}(f) \circ /\!/_s$ by part [ii].

Remark 3.9 If we only assume the transversal BM from x_0 exists up to an explosion time ξ^{x_0} then the decomposition of Brownian motion on M given by part [iii] holds up until $\xi^{x_0} \wedge \zeta$. Note also that, from the proof of part [ii] of the theorem, ξ^{x_0} can be chosen to depend only on the leaf of $\mathcal{F}$ containing x_0, not x_0 itself.

The following result is useful because the transversal Laplacian may be hypoelliptic, as in Examples C below, or more degenerate, and it is generally easier to check non-explosion for elliptic diffusions. The first part seems to be a new result even for the special case of Riemannian submersions with totally geodesic fibres: for them the transversal process is just the horizontal lift of Brownian motion on the base space which is itself the projection of Brownian motion on M. Metric completeness of M is needed as can be seen from considering a standard projection from $\mathbf{R}^3$ to $\mathbf{R}$ and restricting it to obtain $p : \mathbf{R}^3 - \{0\} \to \mathbf{R}$.

Theorem 3.10 *Suppose M is a complete Riemannian manifold with a bundle-like metric for a foliation $\mathcal{F}$ with totally geodesic leaves. Then the stochastic completeness of M implies that of the leaves of $\mathcal{F}$ and also that the transversal process exists for all time.*

As a partial converse: if there is a dense leaf L of $\mathcal{F}$ which is stochastically complete then so is M. Moreover on any stochastically complete leaf L of $\mathcal{F}$ the heat semigroup $\{P_t^M\}_{t\geq 0}$ of M has $P_t^M(1)$ constant on L for each t.

Proof We will use the notation of the proof of Theorem 3.8. Set $z_t = /\!/_t \tilde{x}_t$ and $\eta := \eta^{u_0} = \eta^{L_{u_0}}$, the explosion time from any point on the leaf of $\mathcal{F}^O$ through $u_0 = \tilde{x}_0$. Assume M is stochastically complete. We will show that η is almost surely equal to the explosion time ξ^{u_0} of $\tilde{x}_{\cdot}$ and so equal to the explosion time of the leaf Brownian motion on L_{u_0} starting at any point on that leaf. Since η is independent of these explosion times it follows that they are all equal to some non-random constant

c say. By Remark 3.5 we know this implies that Brownian motion on the leaf has explosion time c for each starting point. By considering the heat semigroup for L_{u_0} acting on the identically one function we see that we must have $c = \infty$, giving the main result. To show equality of η and ξ^{u_0} first suppose $0 \leq t < \eta < \xi^{u_o} \leq \infty$. By the previous theorem, using the leaf metric d^L:

$$d^L(u_t, z_t) = d^L(u_t, /\!/_t \tilde{x}_t) = d^L(u_0, \tilde{x}_t). \tag{9}$$

But as $t \nearrow \eta$ we have $d^L(u_t, z_t) \to \infty$ since $z_t \to z_\eta$, while

$$d^L(u_0, \tilde{x}_t) \to d^L(u_0, \tilde{x}_\eta) < \infty,$$

so $\xi^{u_o} \leq \eta$ almost surely. Here we have used the fact that if M is stochastically complete then any horizontal lift of its Brownian motion to $O(E)$ exists for all time; for example see [15, Theorem 13C, p. 175].

On the other hand suppose $0 \leq t < \xi^{u_0} < \eta \leq \infty$. Again the equalities (9) hold. As $t \nearrow \xi^{u_0}$ this time

$$d^L(u_t, z_t) \to d^L(u_{\xi^{u_0}}, z_{\xi^{u_0}}) < \infty$$

while $d^L(u_0, \tilde{x}_t) \to \infty$. Therefore $\eta \leq \xi^{u_0}$ almost surely, as required.

For the partial converse suppose that the leaf L^{u_0} of $\mathcal{F}^O$ covers a leaf L of $\mathcal{F}$ which is dense in M and is stochastically complete with its induced Riemannian structure. From part [iii] of the previous theorem we know that Brownian motion on M starting from x_0 has lifetime ζ^{x_0}, say, with $\zeta \geq \eta$. However using the Eq. (9) consider $t \nearrow \eta$ on the event $\eta < \zeta^{x_0}$:

$$d^L(u_t, z_t) \to \infty \;\text{ while }\; d^L(u_0, \tilde{x}_t) \to d^L(u_0, \tilde{x}_\eta).$$

By Eq. (9) this shows the event has measure zero so $\eta = \zeta^{x_0}$ almost surely. Since we know $\eta = \eta^{u_0}$ is independent of the point u_0 of L_{u_o} this shows that ζ^{x_0} is independent of $x_0 \in L$. Since $P_t^M(1)(x_0) = \mathbf{P}\{t < \zeta^{x_0}\}$ we see $P_t^M(1)$ is constant on L. If L is dense then $P_t^M(1)$ will be constant on M and so identically one, giving stochastic completeness of M.

Theorem 3.11 *Suppose M is a complete Riemannian manifold with a bundle-like metric for a foliation $\mathcal{F}$ with totally geodesic leaves, and for which the transverse Brownian motion exists for all time. Then the leaf and transverse Laplacians, $\frac{1}{2}\triangle^{\mathcal{F}}$ and $\frac{1}{2}\triangle^E$ commute together with their semigroups $P^{\mathcal{F}}_\cdot$ and $P^E_\cdot$: for $s, t \geq 0$*

$$P_s^E P_t^{\mathcal{F}} = P_t^{\mathcal{F}} P_s^E \tag{10}$$

$$P_t^E P_t^{\mathcal{F}} = P_t^M. \tag{11}$$

Indeed the corresponding results hold for the semigroups $P^{\mathcal{F}^O}_{\cdot}$ *and* $P^H_{\cdot}$ *generated by* $\frac{1}{2}\triangle^{\mathcal{F}^O}$ *and* $\frac{1}{2}\triangle^H$.

Proof By part [ii] of the previous theorem we see $P_t^{\mathcal{F}^O}(f \circ //_s) = P_t^{\mathcal{F}^O}(f) \circ //_s$ so taking expectations gives the result for the semigroups acting on functions $f : O(E) \to \mathbf{R}$. Applying this to functions which factorize through the projection $O(E) \to M$ yields the commutation (10). □

From Theorem 2.5 we immediately have:

Corollary 3.12 *Under the conditions of the Theorem, if also one of* $\triangle^{\mathcal{F}}$ *or* $\triangle^E$ *is regular in the sense of Sect. 2.2, then any bounded harmonic function on* M *is both leaf harmonic and transversely harmonic. In particular if further every bounded leaf harmonic function is constant on the leaves of* $\mathcal{F}$ *then the leaves are constancy sets. Also every compact leaf is a constancy set.*

3.4.3 Examples

(A) One of the simplest non-trivial examples of a Riemannian manifold with bundle-like metric for a foliation with totally geodesic fibres is that of the open Möbius band foliated by "horizontal circles". Thus

$$M = \mathbf{R} \times [0,1]/\sim \quad \text{where} \;\; (x,0) \sim (\tau(x),1) \quad \text{for} \;\; \tau(x) = -x,$$

and the leaves are given by $\{a\} \times [0,1]$.
There is then a transverse foliation consisting of the "vertical lines". Looked at the other way round, as is more common, we have a Riemannian submersion, with totally geodesic fibres, the vertical lines, and integrable horizontal subspaces, the circles.

(B) A more interesting example potential theoretically is

$$M = \text{ Hyperbolic 3-space} \times [0,1]/\sim$$

where $(x,0) \sim (\tau(x),1)$ for $\tau(x) = -x$ using the disc model of hyperbolic space. Again we take can take $\mathcal{F}$ to consist of the "horizontal" circles. The bounded harmonic functions on M are $h : M \to \mathbf{R}$ of the form $h(x,\theta) = \tilde{h}(x)$ for $\tilde{h}$ bounded harmonic on hyperbolic 3-space with $\tilde{h}(\tau(x)) = \tilde{h}(x)$. Such harmonic functions correspond to bounded measurable functions on the sphere at infinity of hyperbolic space which are invariant under the antipodal map.

(C) There are wide classes of examples when the leafs are one dimensional, so we have a flow. An important but familiar one is when $M = R^3$ with its Heisenberg group structure. It can be given a left invariant Riemannian metric. Take the foliation by vertical lines $\{(a,b,z)\}_{z\in\mathbf{R}}$. The transverse operator is then the usual sub-Riemannian operator associated to the Heisenberg group, e.g. see

[17, Example 2.2.11]. Generalisations of this are the manifolds with K-contact structures, including the Sasakian manifolds [9]. The transverse Laplacians of these have been analysed in [6], with Jing Wang discussing the case of the CR hyperbolic space, a circle bundle over complex hyperbolic space in [31], and Bonnefont that of $SL(2, \mathbf{R})$ considered as a circle bundle over the hyperbolic plane, and its universal cover [10].

In general the trajectories in a complete Riemannian manifold of a Killing vector field of unit length form the leaves of a Riemannian foliation with totally geodesic leaves, see [29, p. 137].

(D) Baudoin and Garofalo [5] give a generalisation of the examples in (C) above. These they call *sub-Riemannian manifolds with transverse symmetries*. In Appendix B below we show that they also give rise to Riemannian foliations with totally geodesic fibres.

(E) A general discussion of foliations with totally geodesic leaves can be found in Tondeur's book [29].

Remark 3.13 It would be interesting to find a useful definition of a non-trivial decomposition of a diffusion operator into commuting diffusion operators, and examine the resulting geometry to see whether such decompositions necessarily come from, possibly non-integrable, versions of Riemannian foliations with totally geodesic fibres.

4 Coupling and Gradient Estimates for Riemannian Foliations

We now can see how couplings of Brownian motions on a leaf of $\mathcal{F}$ determine couplings of Brownian motions on M, extending the construction in [16].

Proposition 4.1 *Assume that M is a complete Riemannian manifold with a bundle-like metric for a foliation $\mathcal{F}$ with totally geodesic leaves, and for which the transverse Brownian motion exists for all time.*

Let $x^1_\cdot, x^2_\cdot$ be coupled leaf Brownian motions on a leaf L of $\mathcal{F}$ with a coupling time $T \leq \infty$ so that on $T < \infty$ we have $x^1_T = x^2_T$. Take lifts $\tilde{x}^1_\cdot, \tilde{x}^2_\cdot$ of them to a leaf $\tilde{L}$ of $\mathcal{F}^O$ above L. Choose transverse flows $/\!/_t : \tilde{L} \to O(E)$ as in Theorem 3.8, and independent of $x^1_\cdot, x^2_\cdot$. Then, using the right action of $O(s)$ on $O(E)$, there is a (random) $\gamma \in O(s)$ with the equality in law

$$/\!/_T(\tilde{x}^1_T) =_{\mathcal{L}aw} /\!/_T(\tilde{x}^2_T) \cdot \gamma \qquad T < \infty \tag{12}$$

Consequently the projections of $/\!/_\cdot(\tilde{x}^1_\cdot)$ and $/\!/_\cdot(\tilde{x}^1_\cdot)$ on M are coupled Brownian motions on M having the same law at time T, on $T < \infty$.

In fact the map γ takes values in the holonomy group, the group of covering transformations of $\tilde{L} \to L$.

Proof We can suppose that our canonical SDE on $O(E)$, Eq. (6), is driven by the canonical Brownian motion $B_t(\omega) = \omega_t$ of $\mathbf{R}^s$. Then the transverse flow $/\!/_t : \tilde{L} \to O(E)$, $t \geq 0$, can be chosen to satisfy

$$/\!/_t(u.g, \omega) = /\!/_t(u, g\omega) \cdot g \text{ for all } s \geq 0 \quad \text{and all covering transformations } g \in O(s). \tag{13}$$

This can be seen, using Eq. (4), from its construction in the proof of Theorem 3.8. It is a standard result for such canonical SDE on frame bundles, see for example the proof of Theorem 4C, Chap. 2 of [1].

Since $\tilde{x}^1_\cdot, \tilde{x}^2_\cdot$ cover $x^1_\cdot, x^2_\cdot$ we know there exists a random $\gamma \in O(E)$, measurable with respect to $\sigma\{x^1_s, x^2_s, 0 \leq s \leq T\}$ with values in the covering transformations of $\tilde{L} \to L$ such that

$$\tilde{x}^1_T = \tilde{x}^2_T \cdot \gamma \qquad T < \infty.$$

Then by (13)

$$/\!/_T(\tilde{x}^1_T, \omega) = /\!/_T(\tilde{x}^2_T \cdot \gamma, \omega) = /\!/_T(\tilde{x}^2_T, \gamma\omega) \cdot \gamma \qquad T < \infty \tag{14}$$

giving (12).

Also from Theorem 3.8 we know that, if $\pi : O(E) \to M$ is the projection, the processes $\pi/\!/_\cdot(\tilde{x}^1_\cdot)$ and $\pi/\!/_\cdot(\tilde{x}^2_\cdot)$ are Brownian motions on M. Now observe that on $T < \infty$

$$\pi/\!/_T(\tilde{x}^1_T, \omega) = \pi/\!/_T(\tilde{x}^2_T, \gamma\omega) \cdot \gamma = \pi/\!/_T(\tilde{x}^2_T, \gamma\omega) =_{\mathcal{L}aw} \pi/\!/_T(\tilde{x}^2_T, \omega),$$

proving our claim. □

A key result concerning couplings of Brownian motions concerns the "reflection coupling" described by Kendall:

Theorem 4.2 (Kendall-Cranston [11]) *For a complete Riemannian manifold N with Ricci curvature Ric^N bounded below by $-K$ some $K \geq 0$ there is a coupling of Brownian motions from given points x^1_0, x^2_0 with coupling time T satisfying*

$$\mathbf{P}\{T = \infty\} \leq 2\sqrt{K(\dim N - 1)}\, d^N(x^1_0, x^2_0). \tag{15}$$

From this we are able to apply Proposition 4.1 and follow Cranston's argument in [11] to get

Theorem 4.3 *Suppose M is a complete and stochastically complete n-dimensional Riemannian manifold with a bundle-like metric for a codimension s foliation $\mathcal{F}$ with totally geodesic leaves and whose leaves have Ricci curvatures bounded below by $-K$ some $K \geq 0$, using the leaf metric. Let $h : M \to \mathbf{R}$ be a non-negative bounded harmonic function. Then at each $x \in M$ its gradient in leaf directions $\nabla^{\mathcal{F}} h$ satisfies*

$$|\nabla^{\mathcal{F}} h(x)| \leq 2\sqrt{K(n - s - 1)}\,|h|_\infty. \tag{16}$$

Proof Since the leaves have Ricci curvatures bounded below the leaf Brownian motions exist for all time. By Remark 3.9 the same holds for the transverse Brownian motion, since M is assumed stochastically complete. We can therefore apply Proposition 4.1 to a Cranston-Kendall coupling for a given leaf L of $\mathcal{F}$, as in Theorem 4.2, starting from points x_0^1, x_0^2 in L.

In the notation of Proposition 4.1 we can assume that $\pi /\!/_{\cdot}(\tilde{x}_{\cdot}^1)$ and $\pi /\!/_{\cdot}(\tilde{x}_{\cdot}^2)$ have the same laws after $t \geq T$. Therefore following the standard argument:

$$\begin{aligned} |h(x_0^1) - h(x_0^2)| &= |\mathbf{E}\{h(\pi /\!/_t(\tilde{x}_t^1)) - h(\pi /\!/_t(\tilde{x}_t^2))\}| \\ &= |\mathbf{E}\{h(\pi /\!/_t(\tilde{x}_t^1)) - h(\pi /\!/_t(\tilde{x}_t^2)),\ t < T\}| \\ &\leq |h|_\infty \mathbf{P}\{t < T\} \\ &\to |h|_\infty \mathbf{P}\{T = \infty\} \ \text{ as } t \to \infty. \end{aligned}$$

The result follows by Eq. (15).

Alternatively the theorem is seen to hold by applying Cranston's result [11] directly to each leaf of $\mathcal{F}$ using Corollary 3.10. □

Remark 4.4 For a Riemannian foliation whose leaves are not necessarily totally geodesic there is still a skew-product decomposition of its Brownian motion, as in the foliation of $\mathbf{R}^2 - \{0\}$ by spheres about the origin. To obtain a coupling of Brownian motions on the manifolds using couplings on the leaves would then require considering couplings on manifolds with varying metrics. A special case of this is discussed by Lindvall and Rogers in the last section of [24].

5 Appendix A: $\Gamma^{\mathcal{A}+\mathcal{B}}$

Recall that for a diffusion operator $\mathcal{L}$ on M in Bakry-Emery theory there is the square field operator $\Gamma^{\mathcal{L}}$ given by

$$2\Gamma^{\mathcal{L}}(f, g) := \mathcal{L}(fg) - f\mathcal{L}g - g\mathcal{L}f \ \text{ for } C^2 \text{ functions } f, g : M \to \mathbf{R}.$$

Thus if $\mathcal{L}$ has a decomposition $\mathcal{L} = \mathcal{A} + \mathcal{B}$ we have

$$\Gamma^{\mathcal{L}} = \Gamma^{\mathcal{A}} + \Gamma^{\mathcal{B}}. \tag{17}$$

Now suppose $\mathcal{L}$ is invariant by a foliation $\mathcal{F}$ for which it is transversely cohesive as in Proposition 3.4 and that the corresponding decomposition $\mathcal{L} = \mathcal{A} + \mathcal{L}^{\mathcal{F}}$ is such that the transversal and leaf-wise operators $\mathcal{A}$ and $\mathcal{L}^{\mathcal{F}}$ commute on C^2 functions. For example we could be in the situation of a Riemannian foliation with totally geodesic fibres considered above. For completeness we show that the basic assumption, e.g. Hypothesis 1.2 of Baudoin and Garofalo [5], used by Baudoin and collaborators to augment the square field operator of certain sub-Riemannian operators, holds in our more general situation. See also [4, 6]. Write $\mathcal{B}$ for $\mathcal{L}^{\mathcal{F}}$.

Proposition 5.1 *For any C^2 function $f : M \to \mathbf{R}$ we have*

$$\Gamma^{A}(f, \Gamma^{\mathcal{B}}(f,f)) = \Gamma^{\mathcal{B}}(f, \Gamma^{\mathcal{A}}(f,f))$$

and

$$\Gamma^{A}(f, \Gamma^{\mathcal{L}}(f,f)) = \Gamma^{\mathcal{L}}(f, \Gamma^{A}(f,f)). \tag{18}$$

Proof From Eq. (17)

$$\Gamma^{\mathcal{A}}(f, \Gamma^{\mathcal{L}}(f,f)) - \Gamma^{\mathcal{L}}(f, \Gamma^{\mathcal{A}}(f,f)) = \Gamma^{A}(f, \Gamma^{\mathcal{B}}(f,f)) - \Gamma^{\mathcal{B}}(f, \Gamma^{\mathcal{A}}(f,f)) \tag{19}$$

and so it suffices to prove the first equation of the proposition.

We can take Hörmander form representations

$$\mathcal{A} = \sum_{j=1}^{k} X^j X^j + A$$

$$\mathcal{B} = \sum_{b=1}^{K} Y^b Y^b + B$$

though the Y^b may only be locally Lipschitz, e.g. see [17, Sect. 9.2].

Then $\Gamma^{\mathcal{A}}(f,f) = \sum_j (X^j f)^2$ and similarly for $\mathcal{B}$. Thus

$$\Gamma^{A}(f, \Gamma^{\mathcal{B}}(f,f)) = \sum_{j,b} (X^j f) X^j (Y^b f)^2 \tag{20}$$

$$= 2 \sum_{j,b} (X^j f)(Y^b f)(X^j Y^b f), \tag{21}$$

giving

$$\Gamma^{A}(f, \Gamma^{\mathcal{B}}(f,f)) - \Gamma^{\mathcal{B}}(f, \Gamma^{\mathcal{A}}(f,f)) = 2 \sum_{j,b} (X^j f)(Y^b f)[X^j, Y^b](f). \tag{22}$$

The claim is local so we can work in a disintegrating chart for $\mathcal{F}$ and so assume our foliation comes from a submersion. Then we can take the vector fields X^j to be basic, i.e. horizontal lifts of vector fields on the base space, in the sense of Theorem 6.0.2 of [17]. By that theorem X^j then commutes with $\mathcal{B}$ for each j. Therefore

$$2X^j \Gamma^{\mathcal{B}}(f,f) = \mathcal{B}(X^j f^2) - 2(X^j f)\mathcal{B}f - 2f(\mathcal{B}X^j f) \tag{23}$$

$$= 4\Gamma^{\mathcal{B}}(f, X^j f) \tag{24}$$

as can also be seen from the fact that $\mathcal{B}$ is invariant under the flow of X^j. Equivalently for each j

$$X^j \sum_b (Y^b f)^2 = 2 \sum_b (Y^b f)(Y^b X^j f) \tag{25}$$

giving

$$\sum_b (Y^b f)[X^j, Y^b]f = 0 \quad \text{for each } j. \tag{26}$$

proving our result via Eq. (22). □

Remark 5.2 Commutativity was not used to prove Eq. (22).

6 Appendix B: Diffusion Operators with Transverse Symmetries

Following Baudoin and Garofalo [5] consider a manifold M with a sub-bundle E of TM, and a smooth cohesive diffusion operator A along E whose principal symbol determines a Riemannian metric on E. Then A will be said to have *transverse symmetries* if there is a Lie algebra $\mathcal{V}$ of vector fields which are infinitesimal isometries of E with its given metric i.e. if $Z \in \mathcal{V}$ then

1. Lie differentiation by Z maps C^1 sections of E to sections of E and
2. Lie differentiation by Z annihilates the metric on E

and which has the property that there is the direct sum decomposition for each $x \in M$

$$E_x \bigoplus \mathcal{V}(x) = T_x M,$$

where $\mathcal{V}(x)$ denotes the vector space obtained by evaluating the elements of $\mathcal{V}$ at $x \in M$.

It follows that if we let $\mathcal{F}$ be the sub-bundle of TM with fibres $\mathcal{V}(x)$ then $\mathcal{F}$ is integrable and so determines a foliation transverse to E. Here we do not make Baudoin and Garofalo's assumption that E is bracket generating, so $\mathcal{A}$ may not be hypo-elliptic.

Let $\mathbf{Z} : M \times \mathcal{V} \to TM$ be the evaluation map $\mathbf{Z}(x)V = V(x)$. It is a surjection onto $\mathcal{F}$. (In [4] it is assumed to be injective also.) Take any inner product on $\mathcal{V}$ and use $\mathbf{Z}$ to project it to a Riemannian metric on $\mathcal{F}$. Give TM the metric which is the direct sum of this and the one given on E. Let $e^1, \dots, e^k$ be an orthonormal base for $\mathcal{V}$ and set $Z^a(x) = \mathbf{Z}(x)(e^a)$ for $x \in M$ and $a = 1, \dots, k$. Take a local orthonormal framing by smooth local sections $X^1, \dots, X^p$ of E. Continuing to follow [5] we have commutation relations:

$$[X^i, X^j] = \sum_\ell \omega_\ell^{ij} X^\ell + \sum_a \gamma_a^{ij} Z^a \tag{27}$$

$$[X^j, Z^a] = \sum_\ell \delta_\ell^{ja} X^\ell \tag{28}$$

for smooth functions $\omega_\ell^{ij}, \gamma_a^{ij}, \delta_\ell^{ja}$ with $\omega_\ell^{ij} = -\omega_\ell^{ji}$ and $\gamma_a^{ij} = -\gamma_a^{ji}$. The infinitesimal isometry property of the Z^a gives:

$$\delta_\ell^{ja} = -\delta_j^{\ell a}. \tag{29}$$

We can now quickly deduce the following proposition.

Proposition 6.1 *The vertical foliation $\mathcal{F}$ of a manifold with a cohesive $\mathcal{A}$ with transverse symmetries, furnished with a Riemannian metric as above, is a Riemannian foliation with bundle-like metric and totally geodesic leaves.*

Proof According to [29, Theorem 5.19, p. 56], $\mathcal{F}$ is Riemannian and the metric is bundle-like if and only if

$$\langle [Z, X], X\rangle = 0 \quad \text{for all sections } Z \in C^1\Gamma\mathcal{F}, X \in C^1\Gamma E, ||X|| = 1. \tag{30}$$

Now from formulae (28) and (29) above

$$\langle [Z^a, X^j], X^\ell\rangle = -\delta_\ell^{ja}\langle X^j, X^\ell\rangle = 0.$$

Equation (30) follows when $Z = Z^a$ for any $a = 1, \dots, k$, since we can take $X^1 = X$ if $||X|| = 1$. It therefore holds for any $Z \in C^1\Gamma\mathcal{F}$.

On the other hand a condition given in [29, Theorem 5.23, p. 58], for $\mathcal{F}$ to have totally geodesic fibres is that

$$\langle [Z, X], Z\rangle = 0 \quad \text{for all } Z \in C^1\Gamma\mathcal{F} \quad \text{with } ||Z|| = 1, X \in C^1\Gamma E. \tag{31}$$

We can assume $X = X^j$ above. Then from formula (28) above

$$\langle [Z^a, X], Z^b\rangle = 0$$

For a general $Z \in C^1\Gamma\mathcal{F}$ with $||Z|| = 1$ write $Z = \sum_a f_a Z^a$ for C^1 functions $f_a : M \to \mathbf{R}$. Then

$$\begin{aligned}\langle [Z, X], Z\rangle &= \sum_{a,b} f_a f_b \langle [Z^a, X], Z^b\rangle - \sum_{a,b} df_a(X) f_b \langle Z^a, Z^b\rangle && (32)\\ &= 0 - \sum_a df_a(X) f_a && (33)\\ &= 0 && (34)\end{aligned}$$

since $1 = ||Z|| = \sum_a (f_a)^2$. □

References

1. A. Ancona, K.D. Elworthy, M. Émery, H. Kunita, Stochastic differential geometry at Saint-Flour, *Probability at Saint-Flour* (Springer, Heidelberg, 2013)
2. D. Bakry and Michel Émery. Propaganda for Γ_2. In From local times to global geometry, control and physics (Coventry, 1984/85), volume 150 of Pitman Res. Notes Math. Ser., pp. 39–46. Longman Sci. Tech., Harlow, 1986.
3. D. Bakry, I. Gentil, M. Ledoux, *Analysis and Geometry of Markov diffusion Operators, volume 348 of Die Grundlehren der Mathematischen Wissenschaften* (Springer, Berlin, 2014)
4. F. Baudoin, Stochastic analysis on sub-riemannian manifolds with transverse symmetries. arXiv:1402.4490v1 [math.PR], 2014
5. F. Baudoin, N. Garafola, Curvature-dimension inequalities and ricci lower bounds for sub-riemannian manifolds with transverse symmetries. arXiv:1101.3590v4 [math.DG], 2012
6. Fabrice Baudoin, Jing Wang, Curvature dimension inequalities and subelliptic heat kernel gradient bounds on contact manifolds. Potential Anal. **40**(2), 163–193 (2014)
7. A.L. Bequillard, A sufficient condition for two Markov semigroups to commute. Ann. Probab. **17**(4), 1478–1482 (1989)
8. L. Bérard-Bergery, J.-P. Bourguignon, Laplacians and Riemannian submersions with totally geodesic fibres. Illinois J. Math. **26**(2), 181–200 (1982)
9. D.E. Blair, *Riemannian geometry of contact and symplectic manifolds, volume 203 of Progress in Mathematics*, 2nd edn. (Birkhäuser Boston Inc, Boston, 2010)
10. Michel Bonnefont, The subelliptic heat kernels on $SL(2,\mathbb{R})$ and on its universal covering $\widetilde{SL(2,\mathbb{R})}$: integral representations and some functional inequalities. Potential Anal. **36**(2), 275–300 (2012)
11. M. Cranston, Gradient estimates on manifolds using coupling. J. Funct. Anal. **99**(1), 110–124 (1991)
12. M. Cranston, A. Greven, Coupling and harmonic functions in the case of continuous time Markov processes. Stoch. Process. Appl. **60**(2), 261–286 (1995)
13. M. Cranston, F.-Y. Wang, A condition for the equivalence of coupling and shift coupling. Ann. Probab. **28**(4), 1666–1679 (2000)
14. E.B. Dynkin, Markov processes, in Vols. I, II, volume 122 of Translated with the authorization and assistance of the author by J. Fabius, V. Greenberg, A. Maitra, G. Majone. Die Grundlehren der Mathematischen Wissenschaften, Bände 121 (Academic Press Inc., Publishers, New York, 1965)
15. K.D. Elworthy, *Stochastic Differential Equations on Manifolds*, LMS Lecture Notes Series 70 (Cambridge University Press, Cambridge, 1982)
16. K.D. Elworthy, W.S. Kendall, Factorization of harmonic maps and Brownian motions, in *From local times to global geometry, control and physics* (*Coventry*, 1984/85), Volume 150 of Pitman Research Notes in Mathematics Series (Longman Science Technology, Harlow, 1986), pp. 75–83
17. K.D. Elworthy, Y.L. Jan, X.-M. Li, The geometry of filtering, *Frontiers in Mathematics* (Birkhäuser Verlag, Basel, 2010)
18. Alexandre Freire, On the Martin boundary of Riemannian products. J. Differ. Geom. **33**(1), 215–232 (1991)
19. Robert Hermann, A sufficient condition that a mapping of Riemannian manifolds be a fibre bundle. Proc. Am. Math. Soc. **11**, 236–242 (1960)
20. N. Ikeda, S. Watanabe, in *Stochastic Differential Equations and Diffusion Processes*, 2nd edn. (North-Holland, 1989)
21. M. Kanai, Analytic inequalities, and rough isometries between noncompact Riemannian manifolds, *Curvature and topology of Riemannian manifolds (Katata, 1985)*, volume 1201 of Lecture Notes in Mathematics (Springer, Berlin, 1986), pp. 122–137
22. H. Kunita, On the decomposition of solutions of stochastic differential equations, *Stochastic integrals (Proceedings of Symposium, University Durham, Durham, 1980)*, volume 851 of Lecture Notes in Mathematics (Springer, Berlin, 1981), pp. 213–255

23. M. Liao, Factorization of diffusions on fibre bundles. Trans. Am. Math. Soc. **311**(2), 813–827 (1989)
24. T. Lindvall, L.C.G. Rogers, Coupling of multidimensional diffusions by reflection. Ann. Probab. **14**(3), 860–872 (1986)
25. P. Molino, *Riemannian foliations, Progress in Mathematics* (Birkhäuser, Boston, 1988)
26. B.L. Reinhart, Differential geometry of foliations, *The Fundamental Integrability Problem*, Ergebnisse der Mathematik und ihrer Grenzgebiete, 99 (Springer, Berlin, 1983)
27. D. Revuz, M. Yor, *Continuous martingales and Brownian motion, volume 293 of Grundlehren der Mathematischen Wissenschaften (Fundamental Principles of Mathematical Sciences)*, 3rd edn. (Springer, Berlin, 1999)
28. Hermann Thorisson, Shift-coupling in continuous time. Probab. Theory Relat. F. **99**(4), 477–483 (1994)
29. P. Tondeur, *Foliations on Riemannian manifolds. Universitext*, 9th edn. (Springer, New York, 1988)
30. P. Tondeur, Geometry of foliations, *Monographs in Mathematics, 90* (Birkhuser, Basel, 1997)
31. J. Wang, The sub-elliptic heat kernel on the CR hyperbolic space. arXiv:1204.3642, 2012

Spatial Risk Measures: Local Specification and Boundary Risk

Hans Föllmer and Claudia Klüppelberg

Abstract We study a mathematical consistency problem motivated by the interplay between local and global risk assessment in a large financial network. In analogy to the theory of Gibbs measures in Statistical Mechanics, we focus on the structure of global convex risk measures which are consistent with a given family of local conditional risk measures. Going beyond the locally law-invariant (and hence entropic) case studied in [11], we show that a global risk measure can be characterized by its behavior on a suitable boundary field. In particular, a global risk measure may not be uniquely determined by its local specification, and this can be seen as a source of "systemic risk", in analogy to the appearance of phase transitions in the theory of Gibbs measures. The proof combines the spatial version [10] of Dynkin's method for constructing the entrance boundary of a Markov process with the non-linear extension [14] of backwards martingale convergence.

Keywords Spatial risk measure · Convex risk measure · Phase transition · Systemic risk

1 Introduction

In a large network of financial institutions, the risk at a given node of the network is usually assessed in terms of some monetary risk measure that involves the marginal distribution at that node. But such an approach neglects the interactive effects that are not captured by the family of marginal distributions. This suggests to take a conditional approach, where the risk measure applied at a given node takes into

H. Föllmer (✉)
Institut für Mathematik, Humboldt Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Germany
e-mail: foellmer@math.hu-berlin.de

C. Klüppelberg
Center for Mathematical Sciences, Technische Universität München, Boltzmannstrasse 3, 85748 Garching, Germany
e-mail: cklu@tum.de

D. Crisan et al. (eds.), *Stochastic Analysis and Applications 2014*,
Springer Proceedings in Mathematics & Statistics 100,
DOI 10.1007/978-3-319-11292-3_11

account the situation at the other nodes of the network; see, for example [1]. The question is whether these conditional risk measures can be aggregated in a consistent manner to a global risk measure, and whether the global risk measure is uniquely determined by the local specification.

With this motivation in mind, we are going to focus on some of the purely mathematical problems which arise in such a spatial setting, and which can be viewed as non-linear analogues to some classical problems in the theory of Gibbs measures. In Dobrushin's probabilistic approach to the analysis of phase transitions in Statistical Mechanics, Gibbs measures are specified by a consistent family of local conditional probability distributions; cf. [6] or [18]. In an infinite spatial network, the global Gibbs measure may not be uniquely determined by the local specification. Non-uniqueness is interpreted as a phase transition, and in that case Gibbs measures can be described as mixtures of phases, defined as extreme points in the convex set of all Gibbs measures.

In analogy to Dobrushin's approach, we start with a given family $(\rho_V)_{V\in\mathscr{V}}$ of local conditional risk measures indexed by the class $\mathscr{V}$ of finite subsets of some infinite set of nodes. These conditional risk measures are convex, and they are assumed to be consistent in the usual sense, that is, $\rho_W(-\rho_V) = \rho_W$ if $V \subseteq W$. Our aim is to clarify the structure of the set $\mathscr{R}$ of global convex risk measures which are consistent with this local specification.

To this end, we assume that the local conditional risk measures ρ_V are absolutely continuous with respect to the local conditional probability distributions π_V in the local specification of a Gibbs measure. Under the stronger assumption of local law invariance, the conditional risk of a financial position X would only depend on the distribution of X under the conditional probability measure π_V. As shown in [11], the local risk measures must then be entropic, and the representation of global risk measures can be described in a rather explicit manner.

In this paper we go beyond the special case of local law invariance. But then the main difficulty consists in extending the local specification $(\rho_V)_{V\in\mathscr{V}}$ to a sufficiently regular conditional risk measure with respect to the tail field. We solve this problem by combining two methods. On the one hand, we use the supermartingale properties implied by local consistency, and in particular the non-linear extension of backwards martingale convergence developed in [14]. On the other hand, we use Dynkin's method [8, 9] of constructing the entrance boundary of a Markov process, adapted to our spatial setting as in [10]. In this way, the set of phases can be described as a spatial "boundary", defined by a sub-σ-field $\hat{\mathscr{F}}$ of the tail field. As our main result, we show that a sufficiently regular global risk measure ρ in $\mathscr{R}$ is uniquely determined by its behavior on the boundary field $\hat{\mathscr{F}}$. In particular, we show that we have non-uniqueness of the global risk measure if the underlying probabilistic structure admits a phase transition. From a financial point of view, this can be viewed as one mathematical aspect of the much broader issue of "systemic risk".

The paper is organized as follows. In Sect. 2 we recall some basic facts from the theory of convex risk measures, and in particular the notion of a convex risk kernel introduced in [11]. In Sect. 3 we describe our spatial setting and the local specification of convex risk measures in terms of local risk kernels. The extension of this local specification to a sufficiently regular convex risk kernel with respect to the tail field is done in two steps. In Sect. 4 we use a straightforward definition of a limiting kernel ρ_∞ and show that it has good properties with respect to any given Gibbs measure P. But this kernel does not behave well enough simultaneously for all such Gibbs measures. To overcome this difficulty, we introduce an additional regularization that involves Dynkin's boundary construction. This second step is carried out in Sect. 5, and the resulting risk kernel $\hat{\rho}_\infty$ is shown to be the key to the structure of global risk measures.

2 Preliminaries on Convex Risk Kernels

In this section we recall some basic definitions and facts from the theory of convex risk measures initiated in [2, 16, 17], and also the notion of a convex risk kernel introduced in [11]. For more details see, for example [12, 15].

Let $(\Omega, \mathscr{F})$ be a measurable space, and denote by $M := M_b(\Omega, \mathscr{F})$ the space of all bounded measurable functions on $(\Omega, \mathscr{F})$. A real-valued functional ρ on M is called a *monetary risk measure* if it is *monotone*, i.e., $\rho(X) \geq \rho(Y)$ whenever $X \leq Y$, *cash-invariant*, i.e., $\rho(X+m) = \rho(X) - m$ for constants m, and *normalized*, i.e., $\rho(0) = 0$. If a monetary risk measure ρ is also convex on M, then ρ will be called a *convex risk measure*. A convex risk measure is called *coherent* if it is also positively homogeneous, that is, $\rho(\lambda X) = \lambda \rho(X)$ for any positive constant λ. We denote by $\mathscr{A} := \{X \in M \mid \rho(X) \leq 0\}$ the *acceptance set* of ρ; in the convex case the acceptance set is convex, in the coherent case a convex cone.

Typically, a convex risk measure has a *dual representation*

$$\rho(X) = \sup_{Q \in \mathscr{Q}} \big(E_Q[-X] - \alpha(Q) \big), \tag{1}$$

in terms of some set $\mathscr{Q}$ of probability measures on $(\Omega, \mathscr{F})$ and some penalty function $\alpha : \mathscr{Q} \to [0, \infty]$. In this case, the representation also holds if we choose

$$\alpha(Q) = \sup_{X \in \mathscr{A}} E_Q[-X], \tag{2}$$

and this is the *minimal penalty function* such that (1) holds.

A necessary condition for (1) is the *Fatou property* of ρ, that is,

$$\lim_{k\to\infty} X_k = X \quad \text{pointwise} \implies \rho(X) \leq \liminf_{k\to\infty} \rho(X_k) \tag{3}$$

for any uniformly bounded sequence $(X_k)_{k=1,2,\ldots}$ in M; cf. [15], Lemma 4.21. We say that ρ has the *Lebesgue property* if (3) is replaced by the stronger condition

$$\lim_{k\to\infty} X_k = X \quad \text{pointwise} \implies \rho(X) = \lim_{k\to\infty} \rho(X_k). \tag{4}$$

This condition is sufficient for the dual representation of ρ, and it implies that the supremum in (1) is actually attained; cf. [15, Theorem 4.22, Exercise 4.22].

Now let P be a probability measure on $(\Omega, \mathscr{F})$.

Definition 1 If ρ is a monetary risk measure on M such that $\rho(X) = \rho(Y)$ whenever $X = Y$ P-almost surely, then we say that ρ is *absolutely continuous with respect to* P, and we write

$$\rho \ll P.$$

In this case, ρ can also be considered as a monetary risk measure on the Banach space $L^\infty(\Omega, \mathscr{F}, P)$. Such a risk measure is called *law-invariant with respect to* P if $\rho(X) = \rho(Y)$ whenever X and Y have the same distribution under P.

If $\rho \ll P$ then the Fatou property is both necessary and sufficient for the dual representation (1) of ρ, regarded as a convex risk measure on $L^\infty(\Omega, \mathscr{F}, P)$. In this case we have $Q \ll P$ for any Q such that $\alpha(Q) < \infty$, and so we can restrict $\mathscr{Q}$ to the class of probability measures which are absolutely continuous with respect to P; see Theorem 4.33 in [15]. If ρ satisfies the stronger Lebesgue property, then the supremum in (1) is actually attained by some $Q \ll P$ depending on X; see Corollary 4.35 in [15], and also [5] for a converse result.

Example 1 Let P be a probability measure P on $(\Omega, \mathscr{F})$, and consider the *entropic risk measure* e_β with parameter $\beta \in [0, \infty)$, defined by

$$e_\beta(X) = \frac{1}{\beta} \log E_P[e^{-\beta X}]; \tag{5}$$

for $\beta = 0$, this will be interpreted as the limiting linear case

$$e_0(\beta) := \lim_{\beta\downarrow 0} e_\beta(X) = E_P[-X]. \tag{6}$$

An entropic risk measure is clearly convex and law-invariant. It has the Lebesgue property, and the minimal penalty function in its dual representation (1) is given by

$$\alpha(Q) = \frac{1}{\beta} H(Q|P),$$

where $H(Q|P)$ denotes the *relative entropy* of Q with respect to P; for $\beta = 0$ the penalty function is to be read as 0 if $Q = P$ and as $+\infty$ if not.

Let $\mathscr{F}_0 \subseteq \mathscr{F}$ be a sub-σ-field of $\mathscr{F}$, and denote by M_0 the space of bounded measurable functions on $(\Omega, \mathscr{F}_0)$. Let us first recall the definition of a *stochastic kernel* $\pi(\omega, d\eta)$ from $(\Omega, \mathscr{F}_0)$ to $(\Omega, \mathscr{F})$: For any $\omega \in \Omega$, $\pi(\omega, \cdot)$ is a probability measure on $(\Omega, \mathscr{F})$, and for any $A \in \mathscr{F}$, the function $\pi(\cdot, A)$ on Ω is $\mathscr{F}_0$-measurable. For a probability measure P on $(\Omega, \mathscr{F}_0)$ we denote by $P\pi$ the probability measure on $(\Omega, \mathscr{F})$ defined by $P\pi[A] = \int \pi(\omega, A)P(d\omega)$. The stochastic kernel will be called *regular* if $\pi(\omega, \cdot) = \delta_\omega$ on $\mathscr{F}_0$. For two such kernels $\pi_i (i = 0, 1)$, their composition $\pi_0\pi_1$ is defined as the stochastic kernel given by $\pi_0\pi_1(\omega, A) = \int \pi_1(\eta, A)\pi_0(\omega, d\eta)$.

Let us now extend the classical definition of a stochastic kernel in the following manner.

Definition 2 A *monetary risk kernel from* $(\Omega, \mathscr{F}_0)$ *to* $(\Omega, \mathscr{F})$ is a real-valued function ρ_0 on $\Omega \times M$ such that

(i) for each $\omega \in \Omega$, the functional $\rho_0(\omega, \cdot)$ is a monetary risk measure on M;
(ii) for each $X \in M$, the function $\rho_0(\cdot, X)$ belongs to M_0.

Such a monetary risk kernel ρ_0 will be called *convex* if all risk measures $\rho_0(\omega, \cdot)$ are convex. It will be called *regular* if

$$\rho_0(\omega, f(X_0, X)) = \rho_0(\omega, f(X_0(\omega), X)) \tag{7}$$

for $\omega \in \Omega$, $X_0 \in M_0$, $X \in M$, and for any bounded measurable function f on R^2. We will say that the risk kernel ρ_0 has the *Fatou property*, or the *Lebesgue property*, if condition (3) or condition (4) holds for each risk measure $\rho_0(\omega, \cdot)$.

Note that regularity of a monetary risk kernel ρ_0 from $(\Omega, \mathscr{F}_0)$ to $(\Omega, \mathscr{F})$ implies the following *local property*:

$$\rho_0(\omega, I_{A_0}X + I_{A_0^c}Y) = I_{A_0}(\omega)\rho_0(\omega, X) + I_{A_0^c}(\omega)\rho_0(\omega, Y) \tag{8}$$

for $\omega \in \Omega$, $X, Y \in M$, and any $A_0 \in \mathscr{F}_0$.

The composition $\rho_0(-\rho_1)$ of two monetary risk kernels ρ_0 and ρ_1 is defined as the monetary risk kernel given by

$$(\rho_0(-\rho_1))(\omega, X) := \rho_0(\omega, -\rho_1(\cdot, X)).$$

If ρ_0 and ρ_1 are both convex, then their composition $\rho_0(-\rho_1)$ is again convex.

If ρ_0 is a regular convex risk kernel from $(\Omega, \mathscr{F}_0)$ to $(\Omega, \mathscr{F})$ such that the risk measures $\rho_0(\omega, \cdot)$ satisfy the condition

$$\rho_0(\omega, \cdot) \ll P \qquad P\text{-a.s.}, \tag{9}$$

then it is easy to check that ρ_0 can be regarded as a conditional convex risk measure in the usual sense, as specified by the following definition.

Definition 3 A map ρ_0 from $L^\infty(\Omega, \mathscr{F}, P)$ to $L^\infty(\Omega, \mathscr{F}_0, P)$ is called a *conditional monetary risk measure with respect to* $\mathscr{F}_0$ *and* P, if it satisfies the following three properties for any $X, Y \in L^\infty(\Omega, \mathscr{F}, P)$:

(i) *Monotonicity:* $\rho_0(X) \geq \rho_0(Y)$ P-a.s. whenever $X \leq Y$ P-a.s.;
(ii) *Conditional cash invariance:* $\rho_0(X + m) = \rho_0(X) - m$ P-a.s. for all $m \in L^\infty(\Omega, \mathscr{F}_0, P)$;
(iii) *Normalization:* $\rho_0(0) = 0$ P-a.s.

Such a conditional risk measure ρ_0 is called *convex* if

$$\rho_0(\lambda X + (1-\lambda)Y) \leq \lambda\rho_0(X) + (1-\lambda)\rho_0(Y) \quad P-\text{a.s.}$$

for any $\mathscr{F}_0$-measurable function λ such that $0 \leq \lambda \leq 1$ P-a.s. It is said to have the *Fatou property* if

$$\lim_{k\to\infty} X_k = X \quad P\text{-a.s.} \implies \rho(X) \leq \liminf_{k\to\infty} \rho(X_k) \quad P\text{-a.s.}$$

for any uniformly bounded sequence $(X_k)_{k=1,2,..}$ in $L^\infty(\Omega, \mathscr{F}, P)$; the *Lebesgue property* is defined in the same manner.

Note that the Fatou or the Lebesgue property of the risk measures $\rho_0(\omega, \cdot)$ in (9) implies the corresponding property of ρ_0, regarded as a conditional risk measure with respect to P.

If a convex conditional risk measure ρ_0 with respect to $\mathscr{F}_0$ and P has the Fatou property, then it admits a conditional version of the dual representation (1). Denoting by

$$\mathscr{A}_0 := \{X \in L^\infty(\Omega, \mathscr{F}, P) \mid \rho_0(X) \leq 0 \quad P\text{-a.s.}\}$$

the *acceptance set* of ρ_0, the dual representation takes the form

$$\rho_0(X) = \operatorname{ess\,sup} \big(E_Q[-X \mid \mathscr{F}_0] - \alpha_0(Q)\big), \tag{10}$$

where the essential supremum is taken with respect to P and over all probability measures $Q \ll P$ such that $Q \approx P$ on the σ-field $\mathscr{F}_0$, and where the minimal penalty function is given by

$$\alpha_0(Q) = \operatorname{ess\,sup}_{X\in\mathscr{A}_0} E_Q[-X \mid \mathscr{F}_0], \tag{11}$$

see [7] or [16, Theorem 11.2]. For a general $Q \ll P$, (11) is defined as an essential supremum under Q. But if Q satisfies the additional condition $Q \approx P$ on $\mathscr{F}_0$ as in (10), then it can as well be read as an essential supremum under P.

3 Local Specification of Spatial Risk Measures

Let I be a countable set of *sites*, and let S be some Polish state space with Borel σ-field $\mathscr{S}$. We assume that each site $i \in I$ can be in some state $s \in S$, and we denote by $\Omega = S^I$ the set of possible *configurations* $\omega : I \to S$. For any subset $J \subseteq I$, we denote by ω_J the restriction of ω to J, by $\mathscr{F}_J$ the σ-field on Ω generated by the projection maps $\omega \to \omega(i)$ for any $i \in J$, and we write $\mathscr{F} = \mathscr{F}_I$. A probability measure P on $(\Omega, \mathscr{F})$ is also called a *random field*.

Let $\mathscr{V}$ denote the class of non-empty finite subsets $V \subseteq I$. For a given set $V \in \mathscr{V}$, the σ-field $\mathscr{F}_V$ describes what is observable on V, while $\mathscr{F}_{V^c}$ describes the situation on $V^c := I \setminus V$, also called the *environment* of V.

Definition 4 A collection $(\rho_V)_{V\in\mathscr{V}}$ of regular convex risk kernels ρ_V from $(\Omega, \mathscr{F}_{V^c})$ to $(\Omega, \mathscr{F})$ is called a *local specification of a convex risk measure* if it satisfies the consistency condition

$$\rho_W(-\rho_V) = \rho_W \tag{12}$$

for any $V, W \in \mathscr{V}$ such that $V \subseteq W$, and if each risk kernel ρ_V is regular in the sense of (7) and has the Fatou property.

From now on we fix a local specification $(\rho_V)_{V\in\mathscr{V}}$ of a convex risk measure.

Definition 5 Let $\mathscr{R}$ denote the set of all convex risk measures ρ on M which are consistent with the local specification $(\rho_V)_{V\in\mathscr{V}}$, that is,

$$\rho(-\rho_V) = \rho \quad \text{for any } V \in \mathscr{V}. \tag{13}$$

Our aim is to clarify the structure of the global risk measures in $\mathscr{R}$. At the general level of Definition 4 there is not much to be said. The situation becomes more transparent if we introduce an underlying probabilistic structure, described by the local specification of a random field; cf. [6, 18].

Definition 6 A collection $(\pi_V)_{V\in\mathscr{V}}$ of regular stochastic kernels π_V from $(\Omega, \mathscr{F}_{V^c})$ to $(\Omega, \mathscr{F})$ is called a *local specification of a random field* if it satisfies the consistency condition

$$\pi_W \pi_V = \pi_W \tag{14}$$

for any $V, W \in \mathscr{V}$ such that $V \subseteq W$.

Definition 7 We denote by $\mathscr{P}$ the convex set of all random fields P which are consistent with this local specification in the sense that

$$P\pi_V = P \quad \text{for any } V \in \mathscr{V}. \tag{15}$$

A random field $P \in \mathscr{P}$ is also called a *Gibbs measure*. The case $|\mathscr{P}| > 1$, where the global random field is not uniquely determined by the local specification $(\pi_V)_{V \in \mathscr{V}}$, is often referred to as a *phase transition*.

For any $V \in \mathscr{V}$, the stochastic kernel π_V serves as a conditional probability distribution with respect to $\mathscr{F}_{V^c}$ which is common to all probability measures $P \in \mathscr{P}$, and so we can write

$$E_P[\, f \mid \mathscr{F}_{V^c}](\omega) = \int f(\eta)\pi_V(\omega, d\eta) \tag{16}$$

for any $P \in \mathscr{P}$ and any measurable function $f \geq 0$ on $(\Omega, \mathscr{F})$.

Let us now fix a local specification $(\pi_V)_{V \in \mathscr{V}}$ of a random field such that

$$\mathscr{P} \neq \emptyset. \tag{17}$$

We connect our local specification $(\rho_V)_{V \in \mathscr{V}}$ of a convex risk measure with the local specification $(\pi_V)_{V \in \mathscr{V}}$ by the following assumption:

Assumption 1. For any $\omega \in \Omega$ and any $V \in \mathscr{V}$, the convex risk measure $\rho_V(\omega, \cdot)$ has the following two properties:

(i) $\rho_V(\omega, \cdot) \ll \pi_V(\omega, \cdot)$

(ii) If X is acceptable for $\rho_V(\omega, \cdot)$ then the expected loss under the measure $\pi_V(\omega, \cdot)$ is uniformly bounded from above, i.e., there is a constant $c \geq 0$ such that

$$\rho_V(\omega, X) \leq 0 \Longrightarrow \int (-X)(\eta)\pi_V(\omega, d\eta) \leq c. \tag{18}$$

Remark 1 The local specification $(\rho_V)_{V \in \mathscr{V}}$ is called *law-invariant* if Assumption 1(i) is replaced by the much stronger assumption that each convex risk measure $\rho_V(\omega, \cdot)$ is law-invariant with respect to the probability measure $\pi_V(\omega, \cdot)$. This implies

$$\rho_V(\omega, X) \geq \int (-X)(\eta)\pi_V(\omega, d\eta)$$

for any $X \in M$, and so condition (18) is satisfied with $c = 0$; see Corollary 4.65 in [16]. Actually much more is true: Under mild regularity conditions, local law invariance together with consistency of the family $(\rho_V)_{V \in \mathscr{V}}$ implies that the risk measures $\rho_V(\omega, \cdot)$ must be *entropic*; see [10] and also [19]. More precisely, the risk kernel ρ_V takes the form

$$\rho_V(\omega, X) = \frac{1}{\beta_\infty(\omega)} \log \int e^{-\beta_\infty(\omega) X(\eta)} \pi_V(\omega, d\eta) \tag{19}$$

with $\beta_\infty(\omega) \in [0, \infty)$, as in Example 1. Due to consistency, the parameter $\beta_\infty(\omega)$ does not depend on V, and this implies that the function $\beta_\infty(\cdot)$ is measurable with respect to the tail field $\mathscr{F}_\infty$ introduced in Sect. 4 below; see [10] for more details.

Lemma 1 *For any $P \in \mathscr{P}$, the risk kernel ρ_V can be regarded as a conditional risk measure*

$$\rho_V : L^\infty(\Omega, \mathscr{F}, P) \to L^\infty(\Omega, \mathscr{F}_{V^c}, P),$$

and this conditional risk measure has the Fatou property with respect to P.

Proof Take X and Y in M such that $X = Y$ P-a.s. We have to show that $\rho_V(\cdot, X) = \rho_V(\cdot, Y)$ P-a.s. Indeed, the consistency condition $P = P\pi_V$ implies $\pi_V(\cdot, X) = \pi_V(\cdot, Y)$ P-a.s., hence $\rho_V(\cdot, X) = \rho_V(\cdot, Y)$ P-a.s. due to part i) of our Assumption 1. The Fatou property of the conditional risk measure with respect to P follows from the Fatou property of the risk kernel ρ_V. □

We now take a closer look at our consistency condition (12). For a given probability measure $P \in \mathscr{P}$, this can be read as a consistency condition for two conditional risk measures with respect to P, as shown by Lemma 1. As such, it can be characterized at the level of the corresponding acceptance sets and also at the level of penalty functions; see, for example [4, 13]. For our purposes, however, we will need an alternative characterization in terms of the following supermartingale property; see [13] and [3, Theorem 2].

Proposition 1 *For any $P \in \mathscr{P}$ and any $V, W \in \mathscr{V}$ such that $V \subseteq W$, the consistency condition $\rho_W(-\rho_V) = \rho_W$ yields the supermartingale inequality*

$$\rho_W(X) + \alpha_W(Q) \geq E_Q[\rho_V(X) + \alpha_V(Q) \,|\, \mathscr{F}_{W^c}] \quad P\text{-a.s.} \tag{20}$$

for any $X \in L^\infty(\Omega, \mathscr{F}, P)$ and any probability measure $Q \ll P$.

4 Passing to the Tail Field

Our aim is to clarify the structure of the class $\mathscr{R}$ of global convex risk measures which are consistent with our local specification $(\rho_V)_{V\in\mathscr{V}}$, in analogy to the classical analysis of the class $\mathscr{P}$ of global random fields which are consistent with the local specification $(\pi_V)_{V\in\mathscr{V}}$.

This problem is trivial if I is finite: In this case we have $I \in \mathscr{V}$ and $\mathscr{F}_{I^c} = \{\emptyset, \Omega\}$, and so $\rho_I(\omega, \cdot)$ does not depend on ω. Thus there is exactly one risk measure $\rho \in \mathscr{R}$, namely $\rho = \rho_I$.

From now we assume $|I| = \infty$, and so $(\Omega, \mathscr{F})$ is an infinite product space. Here we will proceed in two steps. In this section we are going to extend the local specification $(\rho_V)_{V\in\mathscr{V}}$ in a consistent manner to a risk kernel ρ_∞ with respect to the *tail field*

$$\mathscr{F}_\infty := \bigcap_{V \in \mathscr{V}} \mathscr{F}_{V^c},$$

and we shall describe the properties of ρ_∞ as a conditional risk measure with respect to any given measure $P \in \mathscr{P}$. The second step will be done in the next section. It involves a regularization of the initial kernel ρ_∞, and this will be the key to the structure of global risk measures.

Let us fix a sequence $(V_n) \subseteq \mathscr{V}$ increasing to I, and let us use the notation

$$\rho_n := \rho_{V_n}, \quad n = 1, 2, \ldots$$

for the corresponding sequence of convex risk kernels. Now consider the risk kernel ρ_∞ defined by

$$\rho_\infty(\omega, X) := \limsup_{n\to\infty} \rho_n(\omega, X) \tag{21}$$

for any $X \in M$ and any $\omega \in \Omega$. We denote by

$$M_\infty := M_b(\Omega, \mathscr{F}_\infty)$$

the space of all bounded measurable functions on $(\Omega, \mathscr{F}_\infty)$. For any $X \in M$, the function $\rho_\infty(\cdot, X)$ belongs to M_∞, since it is bounded by $||X||$ and clearly measurable with respect to the tail field $\mathscr{F}_\infty$.

Lemma 2 *The functional $\rho_\infty : M \to M_\infty$ defined by (21) is a regular convex risk kernel from $(\Omega, \mathscr{F}_\infty)$ to $(\Omega, \mathscr{F})$, and it satisfies the consistency condition*

$$\rho_\infty(-\rho_V) = \rho_\infty \tag{22}$$

for any $V \in \mathscr{V}$.

Proof For any $\omega \in \Omega$, the functional $\rho_\infty(\omega, \cdot)$ on M inherits from the sequence (ρ_n) the properties of a convex risk measure and also the regularity property (7). Moreover, we have

$$\rho_\infty(-\rho_V(X)) = \limsup_{n\to\infty} \rho_n(-\rho_V(X)) = \limsup_{n\to\infty} \rho_n(X) = \rho_\infty(X)$$

for any $V \in \mathscr{V}$, since $\rho_n(-\rho_V(X)) = \rho_n(X)$ as soon as $V \subset V_n$, due to the consistency condition (12). □

For the rest of this section we fix a probability measure $P \in \mathscr{P}$. We are going to show that the limit superior in (21) is P-almost surely a limit, and that ρ_∞ has good properties as a conditional risk measure with respect to P.

Lemma 1 shows that each risk kernel ρ_n can be regarded as a conditional risk measure under P with respect to $\mathscr{F}_{V_n^c}$, and that it has the Fatou property with respect to P. We denote by

$$\mathscr{A}_n(P) := \{X \in L^\infty(\Omega, \mathscr{F}, P) \mid \rho_n(X) \leq 0 \quad P - \text{a.s.}\}$$

its acceptance set and by

$$\alpha_n(Q) = \text{ess sup}_{X \in \mathscr{A}_n(P)} E_Q[-X|\mathscr{F}_{V_n^c}].$$

its penalty function. It follows that ρ_n admits the dual representation

$$\rho_n(X) = \text{ess sup}\big(E_Q[-X|\mathscr{F}_{V_n^c}] - \alpha_n(Q)\big), \tag{23}$$

where the essential supremum is taken over all $Q \ll P$ such that $Q \approx P$ on $\mathscr{F}_{V_n^c}$. Let us also introduce the set

$$\mathscr{Q}(P) := \{Q \in \mathscr{M}_1(P) | Q = P \text{ on } \mathscr{F}_\infty, \quad \sup_{\mathrm{n}} \mathrm{E_Q[_n(Q)]} < \infty\}.$$

As we shall see in the proof of the following Proposition, we have $P \in \mathscr{Q}(P)$, hence $\mathscr{Q}(P) \neq \emptyset$.

Lemma 3 *For any $Q \in \mathscr{Q}(P)$, the limit*

$$\alpha_\infty(Q) = \lim_{n\to\infty} \alpha_n(Q) \tag{24}$$

exists P-a.s. and satisfies

$$E_P[\alpha_\infty(Q)] < \infty. \tag{25}$$

Proof Take $Q \in \mathscr{Q}(P)$. Applying Proposition 1 for $X = 0$, we see that the consistency condition $\rho_{n+1} = \rho_{n+1}(-\rho_n)$ implies the backwards supermartingale inequality

$$\alpha_{n+1}(Q) \geq E_Q[\alpha_n(Q)|\mathscr{F}_{V_n^c}], \qquad n = 1, 2, \ldots$$

with respect to the decreasing σ-fields $(\mathscr{F}_{V_n^c})_{n=1,2\ldots}$. Since $Q \in \mathscr{Q}(P)$, it follows that $(\alpha_n(Q))_{n=1,2\ldots}$ is a non-negative backwards supermartingale under Q which is bounded in $L^1(Q)$. It is thus convergent, Q-a.s. and in $L^1(Q)$, to a finite limit $\alpha_\infty(Q)$ such that

$$E_Q[\alpha_\infty(Q)] = \lim_{n\to\infty} E_Q[\alpha_n(Q)] < \infty.$$

This implies (25) and also the P-almost sure convergence in (24), since $Q = P$ on $\mathscr{F}_\infty$. □

Combining Lemma 3 with the supermartingale inequality (20), we obtain the first part of the following Proposition. The second part will follow by applying the results in [14] on the behavior of consistent conditional risk measures along decreasing σ-fields.

Proposition 2 *We have*

$$\rho_\infty(\cdot, X) = \lim_{n\to\infty} \rho_n(\cdot, X) \quad P\text{-a.s.}$$

for any $X \in M$*, and the kernel* ρ_∞ *defines a conditional convex risk measure*

$$\rho_\infty : L^\infty(\Omega, \mathscr{F}, P) \to L^\infty(\Omega, \mathscr{F}_\infty, P) \tag{26}$$

under P *with respect to the tail-field* $\mathscr{F}_\infty$*. This conditional risk measure has the Fatou property, and its dual representation is given by*

$$\rho_\infty(X) = \operatorname{ess\,sup}_{Q\in\mathscr{Q}_P}(E_Q[-X|\mathscr{F}_{-\infty}] - \alpha_\infty(Q)), \quad X \in M, \tag{27}$$

where $\alpha_\infty(Q)$ *is given by* (24)*. Moreover,* α_∞ *coincides with the minimal penalty function of* ρ_∞*, i.e.,*

$$\alpha_\infty(Q) = \operatorname{ess\,sup}_{X\in\mathscr{A}_\infty(P)} E_Q[-X|\mathscr{F}_\infty] \tag{28}$$

for any $Q \in \mathscr{Q}(P)$*, where*

$$\mathscr{A}_\infty(P) := \{X \in L^\infty(\Omega, \mathscr{F}, P) \,|\, \rho_\infty(X) \le 0 \quad P-\text{a.s.}\}.$$

Proof (1) Take any $X \in M$ and consider the process

$$S_n(P, X) = \rho_n(X) + \alpha_n(P), \qquad n = 1, 2, \ldots$$

This process is bounded from below by $-||X||$, and the consistency condition $\rho_{n+1} = \rho_{n+1}(-\rho_n)$ implies the backward supermartingale inequality

$$S_{n+1}(P, X) \ge E_P[\, S_n(P, X))\,|\mathscr{F}_{V_n^c}];$$

see Proposition 1 for $Q = P$.

(2) Take any $X \in \mathscr{A}_n(P)$. Since $\rho_n(\cdot, X) \le 0$ P-a.s., we have

$$\rho_n(\cdot, X) \le 0 \qquad \pi_n(\omega, \cdot) - \text{a.s.}$$

for P-almost all ω. Using (16) and our assumption (18), this implies

$$E_P[-X\,|\mathscr{F}_{V_n^c}](\omega) = \int (-X)(\eta)\pi_n(\omega, d\eta) \le c$$

for P-almost all ω. In view of (11), this yields the estimate

$$\alpha_n(P) \le c \qquad P\text{-a.s.}.$$

This bound is valid for any n, and so we have $P \in \mathscr{Q}(P)$.
(3) Since $P \in \mathscr{Q}(P)$, the process $(S_n(P, X))_{n=1,2,...}$ is a backwards supermartingale with respect to P and bounded in $L^1(P)$, hence convergent P-a.s. to some finite limit $S_\infty(P, X)$. Combined with Lemma 3, this yields P-almost sure convergence of the sequence

$$\rho_n(X) = S_n(P, X) + \alpha_n(P), \qquad n = 1, 2, \ldots$$

to $\rho_\infty(X)$ and the equality

$$\rho_\infty(X) = S_\infty(P, X) + \alpha_\infty(P) \quad P - \text{a.s..}$$

(4) Since the backwards supermartingale $(\alpha_n(P))_{n=1,2,...}$ is bounded in $L^1(P)$, we can now apply the results of [14] on the limiting behavior of consistent conditional risk measures along decreasing σ-fields under a fixed reference measure P. Lemma 2 in [14] shows that ρ_∞ has the Fatou property under P, and Theorem 4 in [14] yields the dual representation (27) and the identification of α_∞ as the minimal penalty function of ρ_∞. □

5 Dynkin Boundary and Boundary Risk

In this section we are going to modify the risk kernel ρ_∞ in such a way, that the resulting kernel $\hat{\rho}_\infty$ has good properties in terms of the class $\mathscr{P}$ of Gibbs measures. To this end, we use a method developed by E.B. Dynkin [8] for the construction of the entrance boundary of a Markov process, as it was applied in [10] to the integral representation of the class $\mathscr{P}$. This involves an extension of the local specification $(\pi_V)_{V\in\mathscr{V}}$ to a conditional probability distribution π_∞ with respect to the tail field $\mathscr{F}_\infty$ which is common to all probability measures $P \in \mathscr{P}$. The following Proposition summarizes the results of [8–10] which are relevant for our purpose.

Proposition 3 *There exists a stochastic kernel π_∞ from $(\Omega, \mathscr{F}_\infty)$ to $(\Omega, \mathscr{F})$ with the following properties:*

(i) For any $\omega \in \Omega$, the random field $\pi_\infty(\omega, \cdot)$ belongs to $\mathscr{P}$ and is actually an extreme point of the convex set $\mathscr{P}$. In particular we have

$$\pi_\infty \pi_V = \pi_\infty \quad \text{for any } V \in \mathscr{V}. \tag{29}$$

(ii) For any $\omega \in \Omega$, the probability measure $\pi_\infty(\omega, \cdot)$ is ergodic on the tail field, that is, $\pi_\infty(\omega, A) \in \{0, 1\}$ for any $A \in \mathscr{F}_\infty$, and this implies

$$\pi_\infty(\eta, \cdot) = \pi_\infty(\omega, \cdot) \quad \pi_\infty(\omega, \cdot) - \text{a.s..} \tag{30}$$

(iii) The kernel π_∞ serves, simultaneously for all $P \in \mathscr{P}$, as a conditional distribution with respect to the tail field $\mathscr{F}_\infty$, that is,

$$E_P[\, f \mid \mathscr{F}_\infty\,](\omega) = \int f(\eta)\pi_\infty(\omega, d\eta) \tag{31}$$

P-a.s. for any $P \in \mathscr{P}$ and for any measurable function $f \geq 0$ on $(\Omega, \mathscr{F})$.

We endow the set $\mathscr{P}$ with the canonical σ-field $\mathscr{B}$ generated by the maps $P \to P[A] \quad (A \in \mathscr{F})$. Then the kernel π_∞ can be viewed as a measurable map from $(\Omega, \mathscr{F}_\infty)$ to $(\mathscr{P}, \mathscr{B})$. We denote by

$$\hat{\mathscr{F}} := \sigma(\pi_\infty) \subseteq \mathscr{F}_\infty$$

the σ-field on Ω generated by this map, and by

$$\hat{M} := M_b(\Omega, \hat{\mathscr{F}}) \subseteq M_\infty$$

the corresponding space of bounded measurable functions. We will call $(\Omega, \hat{\mathscr{F}})$ the *Dynkin boundary* of the local specification $(\pi_V)_{V \in \mathscr{V}}$, and $\hat{\mathscr{F}}$ will be called the *boundary field*. Thus, any random field $P \in \mathscr{P}$ admits a representation by a probability measure on the Dynkin boundary, namely

$$P = \hat{P}\pi_\infty := \int \pi_\infty(\omega, \cdot)\, \hat{P}\,(d\omega), \tag{32}$$

where $\hat{P}$ denotes the restriction of P to the σ-field $\hat{\mathscr{F}}$. Conversely, any probability measure $\hat{P}$ on $(\Omega, \hat{\mathscr{F}})$ defines via (32) a random field $P \in \mathscr{P}$, due to (29). In this way, we obtain an *integral representation* of the convex set $\mathscr{P}$ that is coupled to the tail field by the kernel π_∞:

$$\mathscr{P} = \{\hat{P}\pi_\infty \mid \hat{P} \ \text{ is a probability measure on } (\Omega, \hat{\mathscr{F}})\}. \tag{33}$$

In particular, a phase transition $|\mathscr{P}| > 1$ occurs if and only if the Dynkin boundary is non-trivial in the sense that the kernel π_∞ really depends on the tail field, that is, not all measures $\pi_\infty(\omega, \cdot)$ coincide, and so $\hat{\mathscr{F}}$ does not reduce to the trivial σ-field $\{\emptyset, \Omega\}$.

Remark 2 The integral representation (32) shows that the set of extreme points of the convex set $\mathscr{P}$ is given by

$$\mathscr{P}_e := \{\pi_\infty(\omega, \cdot) | \omega \in \Omega\}.$$

In particular, $\mathscr{P}_e$ is a measurable subset of $\mathscr{P}$. Denoting by μ_P the image of P under the map $\pi_\infty : \Omega \to \mathscr{P}_e$, the representation (32) takes the form

$$P = \int_{\mathscr{P}_e} Q\,\mu_P(dQ). \tag{34}$$

Conversely, any probability measure μ on $\mathscr{P}_e$ defines via (34) a random field $P \in \mathscr{P}$, and we have $\mu = \mu_P$. Thus we obtain a Choquet type integral representation of the convex set $\mathscr{P}$, that is, any $P \in \mathscr{P}$ is barycenter of a unique probability measure μ_P on the set $\mathscr{P}_e$ of extreme points; see [8–10].

Let us now regularize the kernel ρ_∞ by introducing the risk kernel $\hat{\rho}_\infty = \pi_\infty \rho_\infty$ defined by

$$\hat{\rho}_\infty(\omega, X) = \int \rho_\infty(\eta, X)\pi_\infty(\omega, d\eta) \tag{35}$$

for $\omega \in \Omega$ and $X \in M$. In order to describe its properties, we first take a closer look at the functions in the space $\hat{M}$.

Lemma 4 *For any function $X \in M_\infty$ and any $\omega \in \Omega$, we have*

$$X(\cdot) = \hat{X}(\omega) \quad \pi_\infty(\omega, \cdot) - \text{a.s.}, \tag{36}$$

where $\hat{X}$ denotes the function in $\hat{M}$ defined by

$$\hat{X}(\omega) := \int X(\eta)\pi_\infty(\omega, d\eta). \tag{37}$$

Moreover, X belongs to $\hat{M}$ if and only if X coincides with $\hat{X}$.

Proof Since $\pi_\infty(\omega, \cdot)$ is 0-1 on the tail field $\mathscr{F}_\infty$, the function $X \in M_\infty$ is constant $\pi_\infty(\omega, \cdot)$-a.s., and this implies (36). The function $\hat{X}$ defined by (37) clearly belongs to $\hat{M}$, and so the identity $X = \hat{X}$ yields $X \in \hat{M}$. Conversely, assume that $X \in \hat{M}$, that is, X is $\hat{\mathscr{F}}$-measurable. Since $\hat{\mathscr{F}}$ is generated by the map $\pi_\infty : \Omega \to \mathscr{P}$, there is a measurable function f on $\mathscr{P}$ such that $\hat{X}(\omega) = f(\pi_\infty(\omega, \cdot))$ for all $\omega \in \Omega$. This implies, for any $\omega \in \Omega$,

$$\hat{X}(\omega) = \int f(\pi_\infty(\eta, \cdot))\pi(\omega, d\eta) = f(\pi_\infty(\omega, \cdot)) = X(\omega),$$

since $\pi_\infty(\eta, \cdot) = \pi_\infty(\omega, \cdot)$ for $\pi(\omega, \cdot)$-almost all η, due to (30). □

Proposition 4 *$\hat{\rho}_\infty$ is a regular convex risk kernel from $(\Omega, \hat{\mathscr{F}})$ to $(\Omega, \mathscr{F})$, and it satisfies the consistency condition*

$$\hat{\rho}_\infty(-\rho_V) = \hat{\rho}_\infty \tag{38}$$

for any $V \in \mathscr{V}$. For fixed $\omega \in \Omega$, we have

$$\hat{\rho}_\infty(\omega, \cdot) \ll \pi_\infty(\omega, \cdot), \tag{39}$$

and the convex risk measure $\hat{\rho}_\infty(\omega, \cdot)$ *has the Fatou property with respect to the probability measure* $\pi_\infty(\omega, \cdot)$.

Proof For any $X \in M$, the function $\hat{\rho}_\infty(\cdot, X)$ is clearly $\hat{\mathscr{F}}$-measurable. For fixed $\omega \in \Omega$, the functional $\hat{\rho}_\infty(\omega, \cdot)$ on M inherits from ρ_∞ the properties of a convex risk measure and also the consistency condition:

$$\begin{aligned}\hat{\rho}_\infty(\omega, -\rho_V(X)) &= \int \rho_\infty(\eta, -\rho_V(X))\pi_\infty(\omega, d\eta)\\ &= \int \rho_\infty(\eta, X)\pi_\infty(\omega, d\eta)\\ &= \hat{\rho}_\infty(\omega, X).\end{aligned}$$

Thus, $\hat{\rho}_\infty$ is a convex risk kernel from $(\Omega, \hat{\mathscr{F}})$ to $(\Omega, \mathscr{F})$ such that $\hat{\rho}_\infty(\omega, \cdot) \in \mathscr{R}$ for any $\omega \in \Omega$. To check its regularity, take $\hat{X} \in \hat{M}$, $X \in M$, and any bounded measurable function f on R^2. Since ρ_∞ is regular by Lemma 2, and since $\hat{X}(\eta) = \hat{X}(\omega)$ for $\pi_\infty(\omega, \cdot)$-almost all η by (36), we obtain

$$\begin{aligned}\hat{\rho}_\infty(\omega, f(\hat{X}, X)) &= \int \rho_\infty(\eta, f(\hat{X}, X))\pi_\infty(\omega, d\eta)\\ &= \int \rho_\infty(\eta, f(\hat{X}(\eta), X))\pi_\infty(\omega, d\eta)\\ &= \int \rho_\infty(\eta, f(\hat{X}(\omega), X))\pi_\infty(\omega, d\eta)\\ &= \hat{\rho}_\infty(\omega, f(\hat{X}(\omega), X)).\end{aligned}$$

It remains to verify the Fatou property of $\hat{\rho}_\infty(\omega, \cdot)$ with respect to the measure $P := \pi_\infty(\omega, \cdot)$. Take any uniformly bounded sequence $(X_k)_{k=1,2,\ldots}$ in M such that X_k converges P-a.s. to some $X \in M$. Since $P \in \mathscr{P}$, Proposition 2 implies

$$\rho_\infty(\cdot, X) \le \liminf_{k\to\infty} \rho_\infty(\cdot, X_k) \quad P\text{-a.s.}.$$

Applying Fatou's lemma, we obtain

$$\begin{aligned}\hat{\rho}_\infty(\omega, X) &= E_P[\rho_\infty(\cdot, X)]\\ &\le E_P[\liminf_{k\to\infty} \rho_\infty(\cdot, X_k)]\\ &\le \liminf_{k\to\infty} E_P[\rho_\infty(\cdot, X_k)]\\ &= \liminf_{k\to\infty} \hat{\rho}_\infty(\omega, X_k).\end{aligned}$$

In the special case $X_k \equiv Y$ we see that $\hat{\rho}_\infty(\omega, X) \leq \hat{\rho}_\infty(\omega, Y)$ whenever $X \leq Y$ $\pi_\infty(\omega, \cdot)$-a.s., and this implies $\hat{\rho}_\infty(\omega, \cdot) \ll \pi_\infty(\omega, \cdot)$. $\square$

Definition 8 Let us say that a monetary risk measure ρ on M has the *Lebesgue property with respect to the class* $\mathscr{P}$ if $\lim_{k\to\infty} \rho(X_k) = \rho(X)$ whenever $(X_k)_{k=1,2,...}$ is a uniformly bounded sequence in M such that

$$\lim_{k\to\infty} X_k = X \quad \mathscr{P} - \text{almost surely},$$

that is, the convergence takes place P-a.s. for any $P \in \mathscr{P}$. We denote by $\mathscr{R}_L$ the class of all risk measures $\rho \in \mathscr{R}$ which have the Lebesgue property with respect to $\mathscr{P}$.

Remark 3 For a monetary risk measure $\hat{\rho}$ on $\hat{M}$, the Lebesgue property with respect to $\mathscr{P}$ is equivalent to the Lebesgue property with respect to pointwise convergence, that is, $\lim_{k\to\infty} \hat{\rho}(\hat{X}_k) = \hat{\rho}(\hat{X})$ whenever $(\hat{X}_k)_{k=1,2,...}$ is a uniformly bounded sequence in $\hat{M}$ such that $\lim_{k\to\infty} \hat{X}_k(\omega) = \hat{X}(\omega)$ for any $\omega \in \Omega$. Indeed, if $\lim_{n\to\infty} \hat{X}_n = \hat{X}$ $\mathscr{P}$-a.s. then the sequence converges $\pi_\infty(\omega, \cdot)$-a.s. for each $\omega \in \Omega$, and this amounts to pointwise convergence on Ω, due to Lemma 4.

The following theorem shows that any risk measure $\rho \in \mathscr{R}_L$ is uniquely determined by its behavior on the Dynkin boundary, that is, by its restriction $\hat{\rho}$ to the space $\hat{M}$.

Theorem 1 *Any risk measure $\rho \in \mathscr{R}_L$ has the form*

$$\rho = \hat{\rho}(-\hat{\rho}_\infty), \tag{40}$$

where $\hat{\rho}$ denotes the restriction of ρ to $\hat{M}$.

Proof Take $\rho \in \mathscr{R}_L$ and any $X \in M$. Since $\rho \in \mathscr{R}$, we have

$$\rho(-\rho_n(X)) = \rho(X)$$

for any $n \geq 1$. The sequence $(\rho_n(X))_{n=1,2,...}$ is uniformly bounded by $||X||$, and Proposition 2 shows that

$$\lim_{n\to\infty} \rho_n(\cdot, X) = \rho_\infty(\cdot, X) \quad \mathscr{P} - \text{almost surely}.$$

Now note that, for any $\omega \in \Omega$, the equality

$$\rho_\infty(\cdot, X) = \int \rho_\infty(\eta, X)\pi_\infty(\omega, d\eta) = \hat{\rho}_\infty(\omega, X) = \hat{\rho}_\infty(\cdot, X)$$

holds $\pi_\infty(\omega, \cdot)$-almost surely, due to Lemma 4. In view of the integral representation (32), this implies $\rho_\infty(\cdot, X) = \hat{\rho}_\infty(\cdot, X)$ P-a.s. for any $P \in \mathscr{P}$, and so we get

$$\lim_{n\to\infty} \rho_n(\cdot, X) = \hat{\rho}_\infty(\cdot, X) \quad \mathscr{P} - \text{almost surely.}$$

Applying the Lebesgue property of ρ with respect to $\mathscr{P}$, we obtain

$$\rho(X) = \lim_{n\to\infty} \rho(-\rho_n(\cdot, X)) = \rho(-\hat{\rho}_\infty(\cdot, X)) = \hat{\rho}(-\hat{\rho}_\infty(\cdot, X)),$$

and this proves the representation (40). □

Remark 4 If a risk measure $\rho \in \mathscr{R}$ has the Fatou property with respect to $\mathscr{P}$ but not the Lebesgue property, then the preceding proof yields the inequality $\rho \le \hat{\rho}(-\hat{\rho}_\infty)$.

Now suppose that the risk kernel $\hat{\rho}_\infty$ is such that each risk measure $\hat{\rho}_\infty(\omega, \cdot)$ has not only the Fatou property but also the Lebesgue property with respect to the measure $\pi_\infty(\omega, \cdot)$; this condition is clearly satisfied in the entropic case of Remark 1. In such a situation we have $\mathscr{R}_L \neq \emptyset$, and there is a one-to-one correspondence between the class $\mathscr{R}_L$ and the class $\hat{\mathscr{R}}_L$ of all convex risk measures $\hat{\rho}$ on $\hat{M}$ that have the Lebesgue property with respect to pointwise convergence:

Corollary 1 *If each risk measure $\hat{\rho}_\infty(\omega, \cdot)$ has the Lebesgue property with respect to the measure $\pi_\infty(\omega, \cdot)$, then we have*

$$\mathscr{R}_L = \{\rho(-\hat{\rho}_\infty) | \hat{\rho} \in \hat{\mathscr{R}}_L\}, \tag{41}$$

and in particular $\mathscr{R}_L \neq \emptyset$.

Proof The inclusion "$\subseteq$" follows from the preceding theorem. Conversely, if $\hat{\rho} \in \hat{\mathscr{R}}_L$ then $\rho := \hat{\rho}(-\hat{\rho}_\infty)$ clearly defines a convex risk measure on M which belongs to the class $\mathscr{R}$. To see that ρ has the Lebesgue property with respect to $\mathscr{P}$ and thus belongs to $\mathscr{R}_L$, take a uniformly bounded sequence (X_n) in M such that $X_n \to X$ $\mathscr{P}$-a.s. In particular, the convergence holds $\pi_\infty(\omega, \cdot)$-a.s. for any $\omega \in \Omega$, and this implies $\lim_{n\to\infty} \hat{\rho}_\infty(\omega, X_n) = \hat{\rho}_\infty(\omega, X)$. Thus we have pointwise convergence of the uniformly bounded sequence $(\hat{\rho}_\infty(\cdot, X_n))_{n=1,2,\ldots}$ in $\hat{M}$. Since $\hat{\rho}$ belongs to $\hat{\mathscr{R}}_L$, we get

$$\lim_{n\to\infty} \rho(X_n) = \lim_{n\to\infty} \hat{\rho}(-\hat{\rho}_\infty(\cdot, X_n)) = \hat{\rho}(-\hat{\rho}_\infty(\cdot, X)) = \rho(X).$$

This proves the converse inclusion "$\supseteq$". In particular we have $\mathscr{R}_L \neq \emptyset$, since $\hat{\mathscr{R}}_L \neq \emptyset$. Indeed, any probability measure $\hat{P}$ on the Dynkin boundary induces via

$$\hat{\rho}(X) = \int (-X) d\hat{P} \tag{42}$$

a convex risk measures $\hat{\rho} \in \hat{\mathscr{R}}_L$. □

Corollary 2 *A risk measure $\rho \in \mathscr{R}_L$ is uniquely determined by the local specification $(\rho_V)_{V\in\mathscr{V}}$ if and only if the local specification $(\pi_V)_{V\in\mathscr{V}}$ admits no phase transition, i.e.,*

$$|\mathscr{R}_L| = 1 \iff |\mathscr{P}| = 1. \tag{43}$$

Proof If $|\mathscr{P}| = 1$ then $\hat{\mathscr{F}}$ is trivial, $\hat{M}$ can be identified with R^1, and there is only one monetary risk measure on $\hat{M}$ given by $\hat{\rho}(m) = -m$. Thus (41) implies $|\mathscr{R}_L| = 1$. Conversely, if $|\mathscr{P}| > 1$ then we can choose $\omega_1, \omega_2 \in \Omega$ such that $\pi_\infty(\omega_1, \cdot) \neq \pi_\infty(\omega_2, \cdot)$. Taking

$$A = \{\omega \in \Omega \mid \pi_\infty(\omega_1, \cdot) = \pi_\infty(\omega_2, \cdot)\} \in \hat{\mathscr{F}},$$

we obtain $\pi_\infty(\omega_1, A) = 1$ and $\pi_\infty(\omega_2, A) = 0$ due to (30). But $\hat{\rho}_\infty(\omega_i, \cdot) \ll \pi_\infty(\omega_i, \cdot)$ for $i = 1, 2$ by Proposition 4, and so we get $\hat{\rho}_\infty(\omega_1, -I_A) = 1$ and $\hat{\rho}_\infty(\omega_2, -I_A) = 0$. This shows that the two risk measures $\hat{\rho}_i := \hat{\rho}_i(\omega, \cdot) \in \mathscr{R}_L$ do not coincide, and so we have $|\mathscr{R}_L| > 1$. □

The absence of a phase transition at the underlying probabilistic level implies $|\mathscr{R}_L| = 1$, but not $|\mathscr{R}| = 1$, as illustrated by the following remark on the entropic case.

Remark 5 Let us return to the special case of local law invariance in Remark 1, where the local risk measures $\rho_V(\omega, \cdot)$ are of the entropic form (19) with some parameter $\beta_\infty(\omega)$ which depends on the tail field $\mathscr{F}_\infty$. For fixed $\omega \in \Omega$, the measure $\pi_\infty(\omega, \cdot)$ is ergodic on $\mathscr{F}_\infty$, and so we have

$$\beta_\infty(\eta) = \hat{\beta}(\omega) := \int \beta_\infty(\zeta)\pi_\infty(\omega, d\zeta)$$

for $\pi_\infty(\omega, \cdot)$-almost all $\eta \in \Omega$. Thus the risk kernel $\hat{\rho}_\infty = \pi_\infty \rho_\infty$ in (35) takes the form

$$\hat{\rho}_\infty(\omega, X) = \frac{1}{\hat{\beta}(\omega)} \log \int e^{-\hat{\beta}(\omega) X(\eta)} \pi_\infty(\omega, d\eta). \tag{44}$$

Clearly, the convex risk measure $\hat{\rho}_\infty(\omega, \cdot)$ has not only the Fatou property but also the Lebesgue property with respect to the probability measure $\pi_\infty(\omega, \cdot)$. Thus we can apply Corollaries 1 and 2.

In the absence of a phase transition we have $\mathscr{P} = \{P\}$ for a single random field P. In this case, the $\hat{\mathscr{F}}$-measurable function $\hat{\beta}$ reduces to the constant

$$\beta := \int \beta_\infty(\omega) P(d\omega) \in [0, \infty),$$

and the unique risk measure ρ in $\mathscr{R}_L$ is given by the entropic risk measure (5) with respect to P and β. In particular we obtain $\rho(X) = E_P[-X]$ for any function $X \in M_\infty$, since $X(\cdot) = E_P[X]$ P-almost surely, due to the ergodicity of P on $\mathscr{F}_\infty$. On the other hand, the convex risk measures $\rho_\infty(\omega, \cdot)$ in (21) all belong to $\mathscr{R}$

due to (22), and they are different from ρ since regularity of the kernel ρ_∞ implies $\rho_\infty(\omega, X) = -X(\omega)$ for any $X \in M_\infty$.

Acknowledgments We began to look at spatial risk measures while the first author was visiting the Departments of Mathematics of LMU and TU in München. Special thanks are due to Codina Cotar who participated in these initial discussions. Further results were obtained while the first author was visiting the FIM at ETH Zürich, the Hausdorff Institute of Mathematics at the University of Bonn. and the Department of Mathematics of the National University of Singapore. The support of these institutions is gratefully acknowledged.

References

1. T. Adrian, M.K. Brunnermeier, CoVaR. Preprint Federal Reserve and Princeton University (2011)
2. P. Artzner, F. Delbaen, J.-M. Eber, D. Heath, Coherent measures of risk. Math. Financ. **9**, 203–228 (1999)
3. B. Acciaio, I. Penner, Dynamic risk measures, in *Advanced Mathematical Methods for Finance*, ed. by G.D. Nunno, B. Oksendal (Springer, Berlin, 2011), pp. 1–34
4. J. Bion-Nadal, Dynamic risk measures: time-consistency and risk measures from BMO martingales. Financ. Stoch. **12**(2), 219–244 (2008)
5. F. Delbaen, Differentiability properties of utility functions, in *Optimality and Risk—Modern Trends in Mathematical Finance*, ed. by F. Delbaen, M. Rásonyi, C. Stricker (Springer, Berlin, 2009), pp. 39–48
6. R.L. Dobrushin, Description of a random field by means of conditional probabilities and conditions of its regularity. Theor. Prob. Appl. **13**, 197–224 (1968)
7. K. Detlefsen, G. Scandolo, Conditional and dynamic convex risk measures. Financ. Stoch. **9**(4), 539–561 (2005)
8. E.B. Dynkin, Integral representation of excessive measures and excessive functions. Russ. Math. Surv. **27**, 43–84 (1972)
9. E.B. Dynkin, Entrance and exit spaces for a Markov process. Actes Congres. Int. Math. **2**, 507–512 (1971)
10. H. Föllmer, Phase Transition and Martin Boundary. Séminaire de Probabilités (Strasbourg) **9**, 305–317 (1975)
11. H. Föllmer, Spatial risk measures and their local specification: the locally law-invariant case. Stat. Risk Model. **31**(1), 79–101 (2014)
12. H. Föllmer, T. Knispel, Convex risk measures: basic facts, law-invariance and beyond, asymptotics for large portfolios. *Handbook of the Fundamentals of Financial Decision Making, Part II*, in L.C. MacLean, W.T. Ziemba eds. 507–554 (World Scientific, 2013)
13. H. Föllmer, I. Penner, Convex risk measures and the dynamics of their penalty functions. Stat. Decis. **24**(1), 61–96 (2006)
14. H. Föllmer, I. Penner, Consistent risk measures and a non-linear extension of backwards martingale convergence, Preprint (2014)
15. H. Föllmer, A. Schied, *Stochastic Finance-An Introduction in Discrete Time*, 3rd edn., Graduate Textbook Series (Walter de Gruyter and Co., Berlin, 2011)
16. H. Föllmer, A. Schied, Convex measures of risk and trading constraints. Financ. Stoch. **6**(4), 429–447 (2002)
17. M. Frittelli, E.R. Gianin, Putting order in risk measures. J. Bank. Financ. **26**, 1473–1486 (2002)
18. H.-O. Georgii, *Gibbs Measures and Phase Transitions*, 2nd ext. ed. De Gruyter Studies in Mathematics 9 (2011)
19. M. Kupper, W. Schachermayer, Representation results for law invariant time consistent functions. Math. Financ. Econ. **2**(3), 189–210 (2009)

On Villat's Kernels and BMD Schwarz Kernels in Komatu-Loewner Equations

Masatoshi Fukushima and Hiroshi Kaneko

Abstract The classical Loewner differential equation for simply connected domains is attracting new attention since Oded Schramm launched in 2000 the stochastic Loewner evolution (SLE) based on it. The Loewner equation itself has been extended to various canonical domains of multiple connectivity after the works by Y. Komatu in 1943 and 1950, but the Komatu-Loewner (K-L) equations have been derived rigorously only in the left derivative sense. In a recent work, Z.-Q. Chen, M. Fukushima and S. Rhode prove that the K-L equation for the standard slit domain is a genuine ODE by using a probabilistic method together with a PDE method, and that the right hand side of the equation admits an expression in terms of the complex Poisson kernel of the Brownian motion with darning (BMD). In the present paper, K-L equations for the annulus and circularly slit annili are investigated. For the annulus, we establish a K-L equation as a genuine ODE possessing a normalized Villat's kernel on its right hand side by using a variant of the Carathéodory convergence theorem for annuli indicated by Komatu. This method is also used to obtain the same K-L equation in the right derivative sense on annulus for a more general family of growing hulls that satisfies a specific right continuity condition usually adopted in the SLE theory. Villat's kernel is then identified with a BMD Schwarz kernel for the annulus. Finally we derive K-L equations for circularly slit annuli in terms of their normalized BMD Schwarz kernels, but only in the left derivative sense when at least one circular slit is present.

Keywords Komatu-Loewner equations · Annulus · Circularly slit annulus · Villat's kernel · Brownian motion with darning · BMD Schwarz kernels

---

Dedicated to Professor Terry Lyons on the occasion of his 60th birthday.

M. Fukushima (✉)
Branch of Mathematical Science, Osaka University, Toyonaka, Osaka 560-8531, Japan
e-mail: fuku2@mx5.canvas.ne.jp

H. Kaneko
Department of Mathematics, Tokyo University of Science, Shinjuku, Tokyo 162-8601, Japan
e-mail: stochos@rs.kagu.tus.ac.jp

D. Crisan et al. (eds.), *Stochastic Analysis and Applications 2014*,
Springer Proceedings in Mathematics & Statistics 100,
DOI 10.1007/978-3-319-11292-3_12

1 Introduction

The celebrated Loewner differential equation for the planar unit disk has been extended to various canonical domains of multiple connectivity, first by Komatu [14] to the annulus, then by Komatu [17] to the circularly slit annulus, much later by Bauer and Friedrich [2] to the circularly slit disk, and further by Bauer and Friedrich [3] to the circularly slit annulus as well as to the standard slit domain, namely, a domain obtained from the upper half plane by removing a finite number of disjoint line segments parallel to the x-axis. However, the Komatu-Loewner differential equation has been derived only in the left derivative sense. Recall that, even in the case of the classical Loewner equation for a disk, its derivation in the right derivative sense is harder (cf.[1, Sect. 6.2]).

In a recent paper by Z.-Q. Chen et al. [7], the Komatu-Loewner equation (the K-L equation in abbreviation) for the standard slit domain is established to be a genuine differential equation with the kernel appearing on its right hand side being the complex Poisson kernel of the Brownian motion with darning (BMD in abbreviation) on the standard slit domain. In order to obtain the right differentiability in t of the family of conformal mappings $g_t(z)$ involved in the equation, a probabilistic representation of $\Im g_t(z)$ in terms of the BMD as well as a Lipschitz continuity of the BMD complex Poisson kernel under the perturbation of the standard slit domains are utilized.

The purpose of the present paper is to investigate the counterparts of K-L equations for the annulus and circularly slit annuli.

In Sect. 3, we consider an annulus whose outer boundary component is the unit circle and establish the K-L equation for it as the genuine differential equation (3.10) with a normalized Villat's kernel on its right hand. The right differentiability of $g_\cdot(z)$ will be shown by using a variant of Carathéodory kernel convergence theorem for annuli formulated in Appendix. In Komatu [14], K-L equations for the annulus were obtained in terms of the Weierstrass zeta function and Jacobi's elliptic function instead of Villat's function. The stated variant of Carathéodory theorem for annuli was also presented in [14] without proof to ensure the continuity of the modulus of the domain with respect to the parameter of the Jordan arc being removed. But the proof of the stated differentiability was not as rigorous as in the present paper. Villat's kernel was adopted 8 years later by Goluzin [12] to derive a K-L equation in a different setting (for annuli located outside the unit disk).

In Sect. 4, we consider a general family of growing hulls in annulus that satisfies a specific right continuity condition usually adopted in the SLE theory (cf. [18]) and in SKLE as well (cf. [6]). We show that the same method as in Sect. 3 works to derive the associated K-L equation (3.10) in the right derivative sense. Zhan presented in [22, Proposition 2.1] a variant of Corollary 4.2 without proof for his study of an annulus SLE that was defined based on the unnormalized Villat's kernel. One may formulate an annulus SLE based directly on the K-L equation (3.10) or its reparametrization (3.21) driven by the Brownian motion (with constant drifts) on the outer circle of the annulus.

The Brownian motion with darning (BMD) for an $(N+1)$-connected planar domain is defined as follows, A closed connected subset of $\mathbb{C}$ containing at least two points is called a *continuum*. Let E be a domain in $\mathbb{C}$ such that $\mathbb{C}\backslash E$ is an unbounded continuum and $\{A_1, \dots, A_N\}$ be a collection of mutually disjoint compact continua contained in E. We write $E_0 = E\backslash\bigcup_{j=1}^{N} A_j$ and consider the topological space $E^* = E_0 \cup \{a_1^*, \dots, a_N^*\}$ obtained from E by rendering each 'hole' A_j of E into a single point a_j^*. Extend the Lebesgue measure m on E_0 to E^* by setting $m(a_j^*) = 0,\ 1 \le j \le N$. There exists then a unique m-symmetric diffusion process Z^* on E^* admitting no killing at $a_1^*, \dots, a_N^*$ whose part (killed) process Z^0 on E_0 is just the absorbing Brownian motion on E_0 (cf. [5, Sect. 7.7]). We call Z^* the *BMD for* E_0. Informally we may say that Z^* is the diffusion process on E^* obtained from the absorbing Brownian motion on E by rendering each hole A_j into a single point a_j^* (darning).

A simple way to conceive the BMD Z^* is to consider the Dirichlet form $(\mathcal{E}^*, \mathcal{F}^*)$ defined by

$$\mathcal{F}^* = \{u \in H_0^1(E) : \widetilde{u} \quad \text{is constant q.e. on each } A_j\}$$

$$\mathcal{E}^*(u, v) = \frac{1}{2}\int_E \nabla u(x) \cdot \nabla v(x)dx,$$

where $\widetilde{u}$ denotes a quasi-continuous version of u. Then $(\mathcal{E}^*, \mathcal{F}^*)$ turns out to be a regular Dirichlet form on $L^2(E^*, m)$ and the associated diffusion process on E^* is nothing but the BMD for E_0 (cf. [7]).

The notion of a BMD-harmonic function for E_0 is well defined to be a function on E^* satisfying a usual probabilistic averaging property with respect to the BMD Z^* (cf. [7]). Thus a BMD-harmonic function is harmonic on E_0 in the classical sense but it has an additional important property that its period around each hole A_j vanishes, and accordingly it admits a unique harmonic conjugate on E_0 up to the addition of a constant. If ∂E is smooth, every bounded BMD-harmonic function u on E^* with continuous boundary value on ∂E admits an expression $u(z) = \int_{\partial E} K^*(z, \zeta)u(\zeta)ds(\zeta),\ z \in E^*$, in terms of the uniquely determined kernel $K^*(z, \zeta),\ z \in E^*,\ \zeta \in \partial E$, called the *BMD-Poisson kernel*.

Since $K^*(z, \zeta)$ is BMD-harmonic in z for each $\zeta \in \partial E$, it admits an analytic function $\Psi(z, \zeta),\ z \in E_0$, with $\Im\Psi(z, \zeta) = K^*(z, \zeta)$ uniquely up to the addition of a real constant. $\Psi(z, \zeta)$ with the normalization $\lim_{z\to\infty} \Psi(z, \zeta) = 0$ is called a *BMD complex Poisson kernel for* E_0 and it appears on the right hand side of the K-L equation for the standard slit domain (cf. [7]).

There exists also a function $\mathcal{S}(z, \zeta),\ z \in E_0,\ \zeta \in \partial E$, analytic in z with $\Re\mathcal{S}(z, \zeta) = K^*(z, \zeta)$ uniquely up to the addition of an imaginary constant. We call $\mathcal{S}(z, \zeta)$ a *BMD Schwarz kernel for* E_0 because its counterpart for the unit disk is the classical Schwarz kernel $\frac{1}{2\pi}\frac{\zeta+z}{\zeta-z}$. We may expect that the BMD Schwarz kernel would play important roles in the K-L equations for the annulus and circularly slit annuli.

Indeed we shall show in Sect. 5 that, in the case of the annulus $\mathbb{A}_q = \{z \in \mathbb{C} : q < |z| < 1\}$, $1 < q < 1$, ($E = \mathbb{D}$, $A = \{z \in \mathbb{C} : |z| \leq q\}$ and $E_0 = \mathbb{A}_q$ in the preceding notation), Villat's kernel for $\mathbb{A}_q$ coincides with a BMD Schwarz kernel for $\mathbb{A}_q$ up to a constant factor.

In Sect. 6, we shall consider more generally a circularly slit annulus and derive a K-L differential equation possessing a normalized BMD Schwarz kernel on its right hand side by making computations similar to [7]. Such a representation of the equation in terms of a BMD Schwarz kernel was obtained neither in [17] nor in [3]. But, when at least one circular slit is present, the equation will be shown to hold only in the sense of left derivative and the problem to make it a genuine ODE is left open.

In this connection, we mention a recent work by C. Boehm and W. Lauf [4] where a K-L equation for a circularly slit disk is obtained as a genuine ODE by using an extended version of the Carathéodory convergence theorem.

2 Villat's Kernel Representing Analytic Functions on Annulus

Define an annulus by $\mathbb{A}_q = \{z \in \mathbb{C} : q < |z| < 1\}$ for $q \in (0, 1)$. Sometimes $\mathbb{A}_q$ is written as $\mathbb{A}$ by omitting q. Define *Villat's function* by

$$\begin{aligned}\mathcal{K}_q(z) &= \lim_{N\to\infty} \sum_{n=-N}^{N} \frac{1+q^{2n}z}{1-q^{2n}z} \\ &= \frac{1+z}{1-z} + \lim_{N\to\infty} \sum_{n=1}^{N} \left(\frac{1+q^{2n}z}{1-q^{2n}z} + \frac{1+q^{-2n}z}{1-q^{-2n}z} \right), \quad z \in \mathbb{A}_q. \end{aligned} \tag{2.1}$$

It holds that

$$\mathcal{K}_q(z) = \frac{1+z}{1-z} + 2\sum_{n=1}^{\infty} \frac{q^{2n}}{q^{2n}-z} + 2\sum_{n=1}^{\infty} \frac{q^{2n}z}{1-q^{2n}z}, \quad z \in \mathbb{A}_q, \tag{2.2}$$

both sums on the righthand side being convergent. This is because

$$\frac{q^{2n}}{q^{2n}-z} + \frac{q^{2n}z}{1-q^{2n}z} = \frac{1}{1-q^{2n}z} + \frac{z}{q^{2n}-z}, \quad n \geq 1.$$

For $z \in \mathbb{A}_q$ and $\zeta \in \partial\mathbb{A}_q$, define *Villat's kernel* by

$$\mathcal{K}_q(z,\zeta) = \mathcal{K}_q(z/\zeta) = \frac{\zeta+z}{\zeta-z} + 2\sum_{n=1}^{\infty} \left(\frac{q^{2n}\zeta}{q^{2n}\zeta-z} + \frac{q^{2n}z}{\zeta-q^{2n}z} \right). \tag{2.3}$$

The following representation by Villat's kernel of any analytic function on $\mathbb{A}$ that is continuous on $\overline{\mathbb{A}}$ has been known:

Theorem 2.1 *If f is analytic on $\mathbb{A}$ and $f \in C(\overline{\mathbb{A}}, \mathbb{C})$, then it holds that*

$$f(z) = \frac{1}{2\pi i} \int\limits_{\partial\mathbb{A}} \Re f(\zeta)\mathcal{K}_q(z,\zeta)\frac{d\zeta}{\zeta} - \frac{1}{2\pi i} \int\limits_{|\zeta|=q} \Re f(\zeta)\frac{d\zeta}{\zeta} + ic, \quad z \in \mathbb{A}, \tag{2.4}$$

where

$$c = \frac{1}{2\pi i} \int\limits_{|\zeta|=q} \Im f(\zeta)\frac{d\zeta}{\zeta}.$$

Furthermore

$$\frac{1}{2\pi i} \int\limits_{\partial\mathbb{A}} \Re f(\zeta)\frac{d\zeta}{\zeta} = 0, \text{ namely, } \int\limits_0^{2\pi} \Re f(e^{i\theta})d\theta = \int\limits_0^{2\pi} \Re f(qe^{i\theta})d\theta. \tag{2.5}$$

This theorem is taken from PhD thesis by Vaitsiakhovich [19] that is quoted in a paper [8] of M.D. Contreras et al. Denote by $\mathcal{L}(z,\zeta)$ the infinite sum in (2.3). For $z \in \mathbb{A}$, $\mathcal{L}(z,\zeta)$ and $\mathcal{L}(1/z,\zeta)$ are both analytic in $\zeta \in \mathbb{A}$ and continuous on $\overline{\mathbb{A}}$, and the expression (2.4) is an easy consequence of the Cauchy theorem and the Cauchy integral formula. Using expression (2.4), we get

$$\lim_{r\uparrow 1} \Re f(re^{i\theta}) = \Re f(e^{i\theta}), \ \lim_{r\downarrow q} \Re f(re^{i\theta}) = \Re f(qe^{i\theta}) + \frac{1}{2\pi i} \int\limits_{\partial\mathbb{A}} \Re f(\zeta)\frac{d\zeta}{\zeta}, \tag{2.6}$$

which yields (2.5).

This theorem goes back to Villat [21]. In page 12–20 of this book, the expression like (2.4) was obtained in terms of the kernel (2.3) by matching the coefficients in the Laurent expansion of f and in Fourier expansion of $\phi\big|_{\partial\mathbb{A}}$. In fact, (2.3) for $|\zeta| = 1$ coincides with $1 + 2S$ for the kernel S in [21]. (2.3) for $|\zeta| = q$ is also related to the kernel T in [21]. The expressions of S and T were then rewritten in [21] to derive the celebrated Villat's formula to represent an analytic function f on $\mathbb{A}$ in terms of the Weierstrass zeta functions. Apparently it was in G.M. Goluzin [12] where the sum (2.2) was first rewritten as a sum (2.1) in the principal value sense.

The next proposition will be utilized in Sects. 3 and 5. We adopt the notations $\mathbb{D} = \{z \in \mathbb{C} : |z| < 1\}$, $\mathbb{D}_q = \{z \in \mathbb{C} : |z| < q\}$.

Proposition 2.2 (i) *Suppose that f is analytic on $\mathbb{A}$, $f \in C(\overline{\mathbb{A}}, \mathbb{C})$ and*

$$\Re f \text{ is equal to a real constant } A \text{ on } \partial\mathbb{D}_q. \tag{2.7}$$

Then

$$\frac{1}{2\pi}\int_0^{2\pi} \Re f(e^{i\theta})d\theta = A, \tag{2.8}$$

and moreover f can be expressed as

$$f(z) = \frac{1}{2\pi}\int_0^{2\pi} \Re f(e^{i\theta})\ \mathcal{K}_q(z, e^{i\theta})d\theta + ic, \quad z \in \mathbb{A}, \tag{2.9}$$

for some real constant c.

(ii) *Conversely, for any $\phi \in C(\partial\mathbb{D}, \mathbb{R})$ and $c \in \mathbb{R}$, define $f(z),\ z \in \mathbb{A}$, by* (2.9) *and A by* (2.8) *with ϕ in place of $\Re f$, respectively. Then*

$$\lim_{r\downarrow q} \Re f(re^{i\eta}) = A \text{ for any } \eta \in [0, 2\pi),\ \lim_{r\uparrow 1} \Re f(re^{i\theta}) = \phi(e^{i\theta}),\ \theta \in [0, 2\pi). \tag{2.10}$$

Proof (i) Condition (2.7) implies (2.8) by Theorem 2.1. Under the condition (2.7), the contribution of the integral on the inner circle $|\zeta| = q$ to the right-hand side of (2.4) is $-\frac{A}{2\pi i}\int_{|\zeta|=q} \mathcal{K}_q(z, \zeta)\frac{d\zeta}{\zeta} - A$, which vanishes because $\frac{1}{2\pi i}\int_{|\zeta|=q} \mathcal{K}_q(z.\zeta)\frac{d\zeta}{\zeta} = -1$ on account of (2.3) and
$\mathrm{Res}_{\{\zeta=0\}}\frac{\zeta+z}{\zeta-z}\cdot\frac{1}{\zeta} = -1,\ \int_{|\zeta|=q}\frac{d\zeta}{\zeta-q^{-2n}z} = 0,\ \mathrm{Res}_{\{\zeta=0\}}\frac{q^{2n}z}{\zeta-q^{2n}z}\cdot\frac{1}{\zeta} = -1,$
$\mathrm{Res}_{\{\zeta=q^{2n}z\}}\frac{q^{2n}z}{\zeta-q^{2n}z}\cdot\frac{1}{\zeta} = 1.$

(ii) By (2.3), we readily have $\lim_{r\downarrow q}\Re\mathcal{K}_q(re^{i\eta}, e^{i\theta}) = 1$ boundedly, yielding the first identity of (2.10). Then f admits the expression (2.4) by the observation made in (i) and so the second identity of (2.10) is nothing but the first one in (2.6). □

The following extension of Proposition 2.2 (i) will be utilized in Sect. 4.

Proposition 2.3 *Suppose that f is analytic and bounded on $\mathbb{A}$, and*

$$\Re f \text{ admits a constant limit } A \text{ at each point of } \partial\mathbb{D}_q. \tag{2.11}$$

Then the limit

$$\phi(e^{i\theta}) = \lim_{r\uparrow 1} \Re f(re^{i\theta}) \tag{2.12}$$

exists for a.e. $\theta \in [0, 2\pi)$ *and*

$$\frac{1}{2\pi}\int_0^{2\pi} \phi(e^{i\theta})d\theta = A. \tag{2.13}$$

Furthermore f can be expressed as

$$f(z) = \frac{1}{2\pi} \int_0^{2\pi} \phi(e^{i\theta})\ \mathcal{K}_q(z, e^{i\theta}) d\theta + ic, \quad z \in \mathbb{A}, \tag{2.14}$$

for some real constant c.

Proof Since $\Re f$ is a bounded harmonic function on $\mathbb{A} = \mathbb{A}_q$, the Fatou theorem (cf. [11]) yields its boundary limit (2.12) on $\partial\mathbb{D}$. On account of the assumption (2.11), f can be extended to be an analytic function on $\{z : q^2 < |z| < 1\}$ denoted by f again across $\partial\mathbb{D}_q$ by the mirror reflection. For any $Q \in (q, 1)$, the function $f_Q(z) = f(Qz)$ is analytic on $\mathbb{A}_q$ continuous on $\overline{\mathbb{A}}_q$ so that (2.4) and (2.5) hold for f_Q. By letting $Q \uparrow 1$, we get (2.13) and also (2.4) with $\Re f\big|_{\partial\mathbb{D}} = \phi$ and $\Re f\big|_{\partial\mathbb{D}_q} = A$, which is reduced to (2.14) as in the proof of Proposition 2.2. □

3 Komatu-Loewner Equation on Annulus in Terms of Villat's Kernel

Fix an annulus $\mathbb{A}_Q$ for $0 < Q < 1$, and a Jordan arc $\gamma = \{\gamma(t) : 0 \le t \le t_\gamma\}$ satisfying $\gamma(0) \in \partial\mathbb{D}$, $\gamma(0, t_\gamma] \subset \mathbb{A}_Q$.

According to [13, Chap. 5, Sect. 1], there exists then a strictly increasing function $\alpha : [0, t_\gamma] \mapsto [Q, Q_\gamma]$ ($\alpha(t_\gamma) = Q_\gamma < 1$) with the following property: if $\alpha(t) = q$, then there is a unique conformal map g_q from $\mathbb{A}_Q \setminus \gamma[0, t]$ onto $\mathbb{A}_q$ with the normalization condition

$$g_q(Q) = q. \tag{3.1}$$

We shall prove the continuity of α eventually, but we do not assume it presently. Nevertheless we can reparametrize the curve γ as $\{\widetilde{\gamma}(q) : \ q \in \mathrm{dom}(\widetilde{\gamma})\}$ by setting $\widetilde{\gamma}(q) = \gamma(\alpha^{-1}(q))$ where $\mathrm{dom}(\widetilde{\gamma}) = \alpha[0, t_\gamma] \subset [Q, Q_\gamma]$.

Take $0 \le t^* < t \le t_\gamma$ and put $q = \alpha(t),\ q^* = \alpha(t^*)$, then $Q \le q^* < q \le Q_\gamma$. Define

$$g_{q^*q} = g_{q^*} \circ g_q^{-1}, \quad S_{q^*q} = g_{q^*}\gamma[t^*t]. \tag{3.2}$$

g_{q^*q} is a conformal map from $\mathbb{A}_q$ onto $B_{q^*q} = \mathbb{A}_{q^*} \backslash S_{q^*q}$ such that

$$g_{q^*q}(q) = q^*. \tag{3.3}$$

Let

$$\lambda(q) = g_q(\widetilde{\gamma}(q)) \tag{3.4}$$

be the image of the tip of the curve $\gamma[0,t]$ under g_q, which is a unique point on the outer circle of $\mathbb{A}_q$. The pre-image $\delta_{q^*q} = g_{q^*q}^{-1}(S_{q^*q})$ is a subarc $\{e^{i\theta} : \beta_1(t^*,t) < \theta < \beta_2(t^*,t)\}$ of the outer circle of $\mathbb{A}_q$ containing the point $\lambda(q)$.

We consider the function

$$\Phi(w) = \log \frac{g_{q^*q}(w)}{w}, \ w \in \mathbb{A}_q, \quad \Phi(q) = \log \frac{q^*}{q}, \tag{3.5}$$

which is a well defined analytic function on $\mathbb{A}_q$, continuously extendable to $\overline{\mathbb{A}}_q$ with

$$\Re\Phi(w) = \log \frac{q^*}{q} \quad \text{for any} \quad w \in \partial\mathbb{D}_q. \tag{3.6}$$

Since $\Re\Phi(e^{i\theta}) = \log |g_{q^*q}(e^{i\theta})|$, we have by Proposition 2.2 (i),

$$\frac{1}{2\pi}\int_0^{2\pi} \log |g_{q^*q}(e^{i\theta})| d\theta = \log \frac{q^*}{q}. \tag{3.7}$$

and, for some real constant c,

$$\log \frac{g_{q^*q}(w)}{w} = \frac{1}{2\pi}\int_0^{2\pi} \log |g_{q^*q}(e^{i\theta})| \mathcal{K}_q(w, e^{i\theta}) d\theta + ic. \tag{3.8}$$

We now substitute $w = g_q(z), \ z \in \mathbb{A}_Q \setminus \gamma[0,t]$ in (3.8) to get

$$\log \frac{g_{q^*}(z)}{g_q(z)} = \frac{1}{2\pi}\int_0^{2\pi} \log |g_{q^*q}(e^{i\theta})| \mathcal{K}_q(g_q(z), e^{i\theta}) d\theta + ic.$$

We next put $z = Q$ and obtain from the normalization condition (3.5) that

$$\log \frac{q^*}{q} = \frac{1}{2\pi}\int_0^{2\pi} \log |g_{q^*q}(e^{i\theta})| \mathcal{K}_q(q, e^{i\theta}) d\theta + ic,$$

and consequently

$$c = -\frac{1}{2\pi}\int_0^{2\pi} \log |g_{q^*q}(e^{i\theta})| \Im\mathcal{K}_q(q, e^{i\theta}) d\theta.$$

Thus we arrive at

$$\log \frac{g_{q^*}(z)}{g_q(z)} = \frac{1}{2\pi} \int_0^{2\pi} \log |g_{q^*q}(e^{i\theta})| \left[\mathcal{K}_q(g_q(z), e^{i\theta}) - i\Im \mathcal{K}_q(q, e^{i\theta}) \right] d\theta. \quad (3.9)$$

Theorem 3.1 *$q = \alpha(t)$ is a strictly increasing continuous function from $[0, t_\gamma]$ onto $[Q, Q_\gamma]$. $g_q(z),\ z \in \mathbb{A}_Q \backslash \gamma[0, t]$, is continuously differentiable in $q \in [Q, Q_\gamma]$ and satisfies the differential equation*

$$\frac{\partial \log g_q(z)}{\partial \log q} = \mathcal{K}_q(g_q(z), \lambda(q)) - i\Im \mathcal{K}_q(q, \lambda(q)),\ Q \le q \le Q_\gamma,\ g_Q(z) = z. \quad (3.10)$$

Proof (I) We first prove that $\alpha(t),\ t \in [0, t_\gamma)$, is left continuous in t, $g_q(z)$ is left-differentiable in q and the Eq. (3.11) holds in the left-derivative sense.
We maintain the notations in the above. Every point on the outer circle of $\mathbb{A}_q$ off the set δ_{q^*q} is sent by g_{q^*q} to a point on the outer circle of $\mathbb{A}_{q^*}$. Accordingly the domain $[0, 2\pi]$ of the integration in both Eqs. (3.7) and (3.9) can be replaced by a smaller interval $[\beta_1(t^*, t), \beta_2(t^*, t)]$.
We fix t and let $t^* \uparrow t$. Denote by $\gamma^+(t^*),\ \gamma^-(t^*)$ the points of 'both sides of the Jordan arc γ corresponding to $\gamma(t^*)$. Then as $t^* \uparrow t$, $\gamma^+(t^*) \to \gamma(t) = \widetilde{\gamma}(q),\ \gamma^-(t^*) \to \gamma(t) = \widetilde{\gamma}(q)$ so that

$$\begin{cases} \beta_1(t^*, t) = g_q(\gamma^-(t^*)) \uparrow g_q(\widetilde{\gamma}(q)) = \lambda(q), \\ \beta_2(t^*, t) = g_q(\gamma^+(t^*)) \downarrow g_q(\widetilde{\gamma}(q)) = \lambda(q). \end{cases} \quad (3.11)$$

Since the integrand in the left hand side of (3.7) is bounded, we have $q^* \uparrow q$ the left continuity of α. We divide the both hand sides of the Eq. (3.9) by the both hand sides of (3.7) and let $t^* \uparrow t$ to obtain the left-differentiablility of $g_q(z)$ in q together with the Eq. (3.10) holding in the left-derivative sense.

(II) We use the following notations: for $r > 0,\ 0 < s < t < \infty$,

$$\mathbb{D}(z, r) = \{w \in \mathbb{C} : |w - z| < r\}, \quad \mathbb{A}_{s,t} = \{w \in \mathbb{C} : s < |w| < t\}.$$

The mirror reflection with respect to the circle $\partial\mathbb{D}(\mathbf{0}, r)$ will be denoted by Π_r. For $0 \le t^* < t \le t_\gamma,\ q^* = \alpha(t^*),\ q = \alpha(t)$ as before, we consider the inverse conformal map

$$h_{q^*q} = g_{q^*q}^{-1} = g_q \circ g_{q^*}^{-1} : \ \mathbb{A}_{q^*} \setminus S_{q^*q} \mapsto \mathbb{A}_q.$$

h_{q^*q} satisfies $h_{q^*q}(q^*) = q$ and it sends the inner circle $\partial\mathbb{D}(\mathbf{0}, q^*)$ of $\mathbb{A}_{q^*}$ onto the inner circle $\partial\mathbb{D}(\mathbf{0}, q)$ of $\mathbb{A}_q$. It further sends $\partial\mathbb{D} \setminus \{\lambda(q^*)\}$ onto $\partial\mathbb{D} \setminus \delta_{q^*q}$. Hence we can extend h_{q^*q} by the mirror reflection Π_{q^*} to a univalent function (denoted by h_{q^*q} again) on

$$\mathbb{A}_{q^{*2}} \setminus (S_{q^*q} \cup \Pi_{q^*} S_{q^*q}) \ (\supset \mathbb{A}_{q^*} \setminus S_{q^*q}).$$

Furthermore, by means of the mirror reflection Π_1, we can extend h_{q^*q} to a univalent function (denoted by h_{q^*q} again) on

$$\mathbb{A}_{q^{*2},(q^*)^{-2}} \setminus (\overline{S}_{q^*q} \cup \Pi_{q^*} S_{q^*q}) \setminus \Pi_1(S_{q^*q} \cup \Pi_{q^*} S_{q^*q}). \tag{3.12}$$

By fixing t^*, we claim that

$$\lim_{t \downarrow t^*} q = q^*, \ \text{namely,} \ \alpha \text{ is right continuous,} \tag{3.13}$$

$$\lim_{t \downarrow t^*} \ h_{q^*q}(z) = z \quad \text{locally uniformly on } \mathbb{A}_{q^{*2},(q^*)^{-2}} \setminus \{\lambda(q^*)\}. \tag{3.14}$$

As $t \downarrow t^*$, the domain of definition of the univalent function h_{q^*q} increases to $\mathbb{A}_{q^{*2},(q^*)^{-2}} \setminus \{\lambda(q^*)\}$. Obviously $\{h_{q^*q} : t \in (t^*, t_\gamma]\}$ is a uniformly bounded family of univalent functions. Take any sequence $\{t_n\}$ decreasing to t^* and write $h_n = h_{q^*q_n}, \ q_n = \alpha(t_n)$. By taking a subsequence if necessary, h_n converges to a function h locally uniformly on $\mathbb{A}_{q^{*2},(q^*)^{-2}} \setminus \{\lambda(q^*)\}$.
To prove the claims (3.13) and (3.14), Let us consider the restriction of h_n to E_n for $E_n = \mathbb{A}_{q^*} \setminus S_{q^*q_n}$, which is denoted by h_n again. Then $\{h_n\}$ satisfies all the conditions (i) $\sim$ (iv) of Corollary 7.2, yielding (3.13) and also (3.14) holding on $\mathbb{A}_{q^*}$. Obviously (3.14) then holds on $\mathbb{A}_{q^{*2},(q^*)^{-2}} \setminus \{\lambda(q^*)\}$ as well. We note that, since $h_{q^*q}(g_{q^*}(z)) = g_q(z)$, (3.14) implies

$$\lim_{t \downarrow t^*} g_q(z) = g_{q^*}(z), \qquad z \in \mathbb{A}_Q \setminus \gamma[0, t^* + \delta], \ \delta > 0. \tag{3.15}$$

(III) The continuity of α has been established by (I) and (3.13). Keeping the notations in (I), we shall prove that

$$\lim_{t \downarrow t^*} \beta_1(t^*, t) = \lambda(q^*), \ \lim_{t \downarrow t^*} \beta_2(t^*, t) = \lambda(q^*), \ \lim_{t \downarrow t^*} \lambda(q) = \lambda(q^*). \tag{3.16}$$

Once (3.16) is established, then we can combine it with (3.15) and the continuity of the Villat's kernel $\mathcal{K}_q$ in q to prove the following readily from (3.7) and (3.9) with the domain of the integration being $[\beta_1(t^*, t), \beta_2(t^*, t)]$ in place of $[0, 2\pi]$: $g_q(z)$ is right differentiable in $q \in [Q, Q_\gamma)$, the equation (3.10) holds in the right-derivative sense and the right hand side of (3.10) is right continuous.
Just as in [10], (3.16) can be obtained from (3.14) in the following way. For any $\epsilon > 0$ with $\epsilon < 1 - q^*$, choose $\delta > 0$ such that

$$\overline{S}_{q^*q} \cup \Pi_1 S_{q^*q} \subset \mathbb{D}(\lambda(q^*), \epsilon) \text{ for any } t \in (t^*, t^* + \delta). \tag{3.17}$$

Let $C = \partial\mathbb{D}(\lambda(q^*), \epsilon)$ and $\chi = h_{q^*q}(C)$. Then $\delta_{q^*q} \subset \text{ins}\ \chi$. By virtue of (3.14), we have for a sufficiently small $\delta > 0$

$$|h_{q^*q}(z) - z| < \epsilon, \quad \text{for any } z \in C \text{ and } t \in (t^*, t^* + \delta), \tag{3.18}$$

which particularly means that diam $\chi < 3\epsilon$. By taking any $z \in C$, we then get for any $t \in (t^*, t^* + \delta)$

$$|\lambda(q^*) - \lambda(q)| \leq |\lambda(q^*) - z| + |z - h_{q^*q}(z)| + |h_{q^*q} - \lambda(q)| < 5\epsilon,$$

$$|\lambda(q^*) - \beta_i(t^*, t)| \leq |\lambda(q^*) - z| + |z - h_{q^*q}(z)| + |h_{q^*q} - \beta_i(t^*, t)| < 5\epsilon,$$

for $i = 1, 2$.

(IV) We finally show that $\lambda(q)$ is left continuous:

$$\lim_{q^* \uparrow q} \lambda(q^*) = \lambda(q), \tag{3.19}$$

which implies the left continuity of the right hand side of the equation (3.10) completing the proof of Theorem 3.1.

It follows from (3.9) that, for $z \in \mathbb{A}_q$,

$$\log \frac{g_{q^*q}(z)}{z} = \frac{1}{2\pi} \int_{\beta_1(t^*,t)}^{\beta_2(t^*,t)} \log |g_{q^*q}(e^{i\theta})| \left[\mathcal{K}_q(z, e^{i\theta}) - i\Im\mathcal{K}_q(q, e^{i\theta}) \right] d\theta.$$

For any $\epsilon > 0$, we can choose $\delta > 0$ such that $\{e^{i\theta} : \beta_1(t^*, t) < \theta < \beta_2(t^*, t)\} \subset \mathbb{D}(\lambda(q), \epsilon)$ for $t^* \in (t - \delta, t)$ by (3.12). For such t^*, we can therefore see from the expression (2.3) of the Villat's kernel $\mathcal{K}_q(z, \zeta)$ that the integrand in the right hand side of the above identity is bounded uniformly in $z \in \overline{\mathbb{A}}_q \setminus \mathbb{D}(\lambda(q), \epsilon)$ and in $q^* = \alpha(t^*)$. Thus we deduce from (3.11)

$$\lim_{q^* \uparrow q} g_{q^*q}(z) = z, \quad \text{locally uniformly in } z \in \overline{\mathbb{A}}_q \setminus \{\lambda(q)\}. \tag{3.20}$$

By the mirror reflection Π_1, we further extend g_{q^*q} to $\mathbb{A}_{q,q^{-1}} \setminus \delta_{t^*t}$ across $\partial\mathbb{D}(\mathbf{0}, 1)$. Then (3.20) is still valid locally uniformly in $z \in \mathbb{A}_{q,q^{-1}} \setminus \{\lambda(q)\}$ and we can repeat the same argument as in (III) for g_{q^*q} in place of h_{q^*q} to obtain (3.19). □

Remark 3.2 For the function $g_q \circ g_{Q_\gamma}^{-1}$ in place of g_q in the above, Komatu [14, 16] derived the Eq. (3.10) in terms of the Weierstrass zeta function as well as Jacobi's elliptic function in place of the present Villat's function. A variant of the Carathéodory kernel convergence theorem for annuli as Theorem 7.1 of the present paper was also stated there without proof, that implicitly implied the continuity of the correspondence $\alpha : t \mapsto q$ (as is shown in step (II) in the above proof). But the proof of the right differentiability of $g_q \circ g_{Q_\gamma}^{-1}$ in q was not given as rigorously as in steps (II), (III) of the present one. Goluzin [12] obtained a counterpart of Theorem 3.1 in terms of Villat's kernel under a different setting for annuli located outside the unit disk $\mathbb{D}$. □

Remark 3.3 Since α is shown to be continuous, the Jordan arc γ can be reparametrized in terms of q as $\{\gamma(q) : Q \le q \le Q_\gamma\}$ by redefining $\gamma(\alpha^{-1}(q))$ as $\gamma(q)$ so that g_q is a conformal map from $\mathbb{A}_Q \setminus \gamma[0, q]$ onto $\mathbb{A}_q$ with the normalization (3.1). $g_q(z)$ satisfies the ODE (3.10) for $z \in \mathbb{A}_Q \setminus \gamma[0, q]$.

It is sometimes convenient to reparametrize the curve γ further in terms of the modulus p of the annulus $\mathbb{A}_q$: $p = -\log q$, $q = e^{-p}$. Denote by P, P_γ the modulus of $\mathbb{A}_Q$, $\mathbb{A}_{Q_\gamma}$, respectively. Villat's kernel is denoted in terms of p as $S_p(z, \zeta) = \mathcal{K}_{e^{-p}}(z, \zeta)$. We further change the parameter q to s in a way that $q = e^{-P}e^s$, $0 \le s \le s_\gamma = P_\gamma - P$. Since the module of $\mathbb{A}_q$ equals $P - s$, (3.10) reads for $z \in \mathbb{A}_q \setminus \gamma[0, s]$ and $s \in [0, s_\gamma]$

$$\frac{\partial \log g_s(z)}{\partial s} = S_{P-s}(g_s(z), \lambda(s)) - i\Im S_{P-s}(e^{s-P}, \lambda(s)), \quad g_0(z) = z, \tag{3.21}$$

for the conformal mapping g_s from $\mathbb{A}_Q \setminus \gamma[0, s]$ onto $\mathbb{A}_{Qe^s}$ with $g_s(Q) = Qe^s$. Here $\lambda(s) = g_s(\gamma(s))$. Zhan defined in [22] an annulus SLE based on the equation (3.21) with the second normalization term of its right hand side being dropped however. One may formulate an annulus SLE based directly on (3.10) or (3.21) driven by the Brownian motion (with constant drifts) on the outer circle of $\mathbb{A}_Q$ by making analogous considerations to the case of standard slit domains in [6]. □

4 K-L Equation on Annulus for Right Continuous Growing Hulls

We consider an annulus $\mathbb{A}_Q$ for a fixed $Q \in (0, 1)$. A closed subset F of $\mathbb{A}_Q$ is called a *hull* in $\mathbb{A}_Q$ if the set $\mathbb{A}_Q \setminus F$ is doubly connected possessing $\partial\mathbb{D}_Q$ as one of its boundary components. A strictly increasing family $\{F_t : 0 < t \le T\}$ of hulls in $\mathbb{A}_Q$ is said to be a *family of growing hulls* in $\mathbb{A}_Q$. A typical example of a family of growing hulls in $\mathbb{A}_Q$ is $\{F_t = \gamma(0, t]; t \in (0, t_\gamma]\}$ for a Jordan arc γ considered in the preceding section.

Let $\{F_t; 0 < t \le T\}$ be a family of growing hulls in $\mathbb{A}_Q$. We define $F_0 = \emptyset$ by convention. According to [13, Chap.5, Sect. 1] again, there exists then a strictly increasing function $\alpha : [0, T] \mapsto [Q, Q_T](\beta(T) = Q_T < 1)$ with the following property: if $\alpha(t) = q$, then there is a unique conformal map g_q from $\mathbb{A}_Q \setminus F_t$ onto $\mathbb{A}_q$ with the normalization condition

$$g_q(Q) = q. \tag{4.1}$$

Needless to say, the function α is determined depending on $\{F_t\}$ and it is different in general from α in the preceding section.

Take $0 \le t^* < t \le t_\gamma$ and put $q = \alpha(t),\ q^* = \alpha(t^*)$, then $Q \le q^* < q \le Q_\gamma$. Define

$$g_{q^*q} = g_{q^*} \circ g_q^{-1}, \quad S_{q^*q} = g_{q^*}(F_t \setminus F_{t^*}). \tag{4.2}$$

g_{q^*q} is a conformal map from $\mathbb{A}_q$ onto $\mathbb{A}_{q^*} \backslash S_{q^*q}$ such that

$$g_{q^*q}(q) = q^*. \tag{4.3}$$

We also consider the inverse map $h_{q^*q} = g_{q^*q}^{-1} (= g_q \circ g_{q^*}^{-1})$. h_{q^*q} is a conformal map from $\mathbb{A}_{q^*} \backslash S_{q^*q}$ onto A_q sending the inner circle of $\mathbb{A}_{q^*}$ onto the inner circle of $\mathbb{A}_q$ homeomorphically.

Denote by $\delta_{q^*q} (\subset \mathbb{C})$ the set of accumulation points of $h_{q^*q}(z)$ as z approaches to S_{q^*q}. δ_{q^*q} is then a closed subset of the outer circle of $\mathbb{A}_q$ so that we can write $\delta_{q^*q} = \{e^{i\theta} : \theta \in \ell_{q^*q}\}$ for a closed subset ℓ_{q^*q} of $[0, 2\pi)$. Observe that any point on the outer circle of $\mathbb{A}_{q^*}$ off the closure of S_{q^*q} is a simple boundary point of $\mathbb{A}_{q^*} \backslash S_{q^*q}$ in the sense of [9]. In view of [9, Theorem 15.3.6], the map h_{q^*q} extends to a continuous one-to-one map (denoted by h_{q^*q} again) from $\overline{\mathbb{A}}_{q^*} \backslash \overline{S}_{q^*q}$ into $\overline{\mathbb{A}}_q$.

We show that

$$h_{q^*q}(\overline{\mathbb{A}}_{q^*} \setminus K) = \overline{\mathbb{A}}_q \setminus \delta_{q^*q} \quad \text{for} \quad K = \overline{S}_{q^*q}. \tag{4.4}$$

Denote the outer circle of $\mathbb{A}_{q^*}$ (resp. $\mathbb{A}_q$) by C^* (resp. C). For any $z \in C^* \backslash K$, take a crosscut γ of $\mathbb{A}_{q^*}$ separating z and K. Then $h_{q^*q}(\gamma)$ separates $h_{q^*q}(z) \in C$ from δ_{q^*q} so that we have the inclusion $\subset$ in (4.4). Next, take any sequence $w_n \in \mathbb{A}_q$ converging to $w \in C \backslash \delta_{q^*q}$. Then $z_n = g_{q*q}(w_n) \in \mathbb{A}_{q^*} \backslash S_{q^*q}$ converges to a point $z \in C^* \cup K$ by taking a suitable subsequence if necessary. If $z \in K$, then $w_n = h_{q^*q}(z_n)$ accumulates to δ_{q^*q} that is absurd. Hence $z \in C^* \backslash K$. Since z is simple, w_n converges to a point $w' \in C$ that must equal w by the assumption, yielding (4.4).

Analogously to Sect. 3, we consider the function

$$\Phi(w) = \log \frac{g_{q^*q}(w)}{w}, \ w \in \mathbb{A}_q, \quad \Phi(q) = \log \frac{q^*}{q}, \tag{4.5}$$

which is a well defined bounded analytic function on $\mathbb{A}_q$ with

$$\Re\Phi(w) = \log \frac{q^*}{q} \quad \text{for any} \quad w \in \partial\mathbb{D}_q. \tag{4.6}$$

Hence, by virtue of Proposition 2.3, the limit

$$\phi(e^{i\theta}) = \lim_{r \uparrow 1} \log |g_{q^*q}(re^{i\theta})| \tag{4.7}$$

exists for a.e. $\theta \in [0, 2\pi)$, and the identities (2.13) with $A = \log \frac{q^*}{q}$ and (2.14) hold true. But, by the observation made above, $\lim_{r\uparrow 1} |g_{q^*q}(re^{i\theta})| = 1$ for any $\theta \in [0, 2\pi)\setminus \ell_{q^*q}$ so that the domain of integration $[0, 2\pi)$ in those identities can be replaced by ℓ_{q^*q}, yielding as in Sect. 3

$$\frac{1}{2\pi} \int_{\ell_{q^*q}} \phi(e^{i\theta}) d\theta = \log \frac{q^*}{q}. \tag{4.8}$$

$$\log \frac{g_{q^*}(z)}{g_q(z)} = \frac{1}{2\pi} \int_{\ell_{q^*q}} \phi(e^{i\theta}) \left[\mathcal{K}_q(g_q(z), e^{i\theta}) - i\Im\mathcal{K}_q(q, e^{i\theta})\right] d\theta. \tag{4.9}$$

We now state a specific right continuity condition on a family of growing hulls. Let $\{F_t; t \in (0, T]\}$ be a family of growing hulls in the annulus $\mathbb{A}_Q$. We keep the related notations introduced above. Let $q^* = \alpha(t^*)$ for $t^* \in [0, T)$. The family is called *right continuous at* t^* *with limit* $\lambda(q^*)$ if there exists a point $\lambda(q^*)$ on the outer boundary of $\mathbb{A}_{q^*}$ such that

$$\bigcap_{t>t^*} \overline{S}_{q^*q} = \lambda(q^*), \tag{4.10}$$

for S_{q^*q} defined by (4.2). This condition is obviously satisfied when the hulls are generated by a Jordan arc γ, in which case $\lambda(q^*) = g_{q^*}(\gamma(t^*))$. But such a condition is also satisfied by more general families of growing hulls arising in SLE (cf. [18]) and in SKLE (cf. [6]) as well.

Theorem 4.1 *Let* $\{F_t; t \in (0, T]\}$ *be a family of growing hulls in the annulus* $\mathbb{A}_Q$ *that is right continuous at* $t^* \in [0, T)$ *with limit* $\lambda(q^*)$. *Then* $q = \alpha(t)$ *is right continuous at* $t = t^*$, $g_q(z)$ *is right differentiable at* $q = q^*$ *and*

$$\left.\frac{\partial^+ \log g_q(z)}{\partial \log q}\right|_{q=q^*} = \mathcal{K}_{q^*}(g_{q^*}(z), \lambda(q^*)) - i\Im\mathcal{K}_{q^*}(q^*, \lambda(q^*)), \tag{4.11}$$

for $z \in \mathbb{A}_Q\setminus F_{t^*+\delta},\ \delta > 0$, *where the left hand side denotes the right derivative.*

Proof It suffices to repeat the steps (II) and (III) in the proof of Theorem 3.1 almost word for word.

Indeed we have verified in the above that the conformal map h_{q^*q} extends to a continuous one-to-one map from $\overline{\mathbb{A}}_{q^*}\setminus\overline{S}_{q^*q}$ onto $\overline{\mathbb{A}}_q\setminus\delta_{q^*q}$. Accordingly, using the mirror reflections Π_{q^*} and Π_1, it can be further extended to a conformal map from the region specified by (3.12) that increases to $\mathbb{A}_{q^{*2},(q^*)^{-2}}\setminus\{\lambda(q^*)\}$ as $t \downarrow t^*$ owing to the current condition (4.10). The functions h_n and regions E_n defined in the paragraph above (3.15) satisfy all the conditions (i)–(iv) of Corollary 7.2 again owing to condition (4.10). Hence we get the right continuity (3.13) of α and a local uniform convergence (3.14) of h_{q^*q} together with the right continuity (3.15) of $g_\cdot(z)$.

For any $\epsilon > 0$ with $\epsilon < 1 - q^*$, we can choose $\delta > 0$ such that (3.17) is valid due to condition (4.10). Let $C = \partial\mathbb{D}(\lambda(q^*), \epsilon)$ and $\chi = h_{q^*q}(C)$. By virtue of (3.14), we have for a sufficiently small $\delta > 0$ the property (3.18) which particularly means that diam $\chi < 3\epsilon$. Since $\delta_{q^*q} \subset$ ins χ, we get for every $\zeta \in \delta_{q^*q}$

$$|\lambda(q^*) - \zeta| < 5\epsilon, \quad \text{for any } t \in (t^*, t^* + \delta). \tag{4.12}$$

By taking the continuity of Villat's kernel $\mathcal{K}_q(z, \zeta)$ and (3.15) into account, we can now deduce the desired conclusion of Theorem 4.1 from (4.8), (4.9) and (4.12). □

Corollary 4.2 *Let* $\{F_t; t \in [0, T]\}$, $F_0 = \emptyset$ *be a family of growing hulls in the annulus* A_Q *satisfying the following conditions:*

(1) α *is continuous on* $[0, T]$ *so that* $\alpha[0, T] = [Q, Q_T]$.
(2) *There exists a continuous map* λ *from* $[Q, Q_T]$ *to* $\partial\mathbb{D}$ *and* F_t *is right continuous at each* $t \in [0, T]$ *with limit* $\lambda(q)$ *for* $q = \alpha(t)$.
(3) $g_q(z)$ *is continuous in* $q \in [Q, Q_T]$ *for each* $z \in \mathbb{A}_Q \backslash F_T$.

Then $g_q(z)$, $z \in \mathbb{A}_Q \backslash F_T$, *is continuously differentiable in* $q \in [Q, Q_T]$ *and satisfies the differential equation*

$$\frac{\partial \log g_q(z)}{\partial \log q} = \mathcal{K}_q(g_q(z), \lambda(q)) - i\Im\mathcal{K}_q(q, \lambda(q)), \quad g_Q(z) = z. \tag{4.13}$$

In fact, under the stated conditions, (4.13) holds in the right derivative sense by virtue of Theorem 4.1. As the right hand side of (4.13) is continuous in q, it becomes a genuine ODE.

5 Villat's Kernel Is a BMD Schwarz Kernel

The Schwarz kernel on a planar domain is by definition an analytic function with its real part being the Poisson kernel to represent harmonic functions by their values on the boundary. But we need to specify which class of harmonic functions and which part of the boundary are involved. We consider a BMD Schwarz kernel $\mathcal{S}(z, \zeta)$ defined in Introduction. $\Re\mathcal{S}(z, \zeta)$, $z \in \mathbb{A}_q$, $\zeta \in \partial\mathbb{D}$, for the annulus $\mathbb{A}_q$ thus represents BMD harmonic functions for $\mathbb{A}_q$ by their boundary values on $\partial\mathbb{D}$. We now deduce from Proposition 2.2 (ii) that the Villat's kernel $\mathcal{K}_q(z, \zeta)$ for $z \in \mathbb{A}_q$, $\zeta \in \partial\mathbb{D}$, is equal to a BMD Schwarz kernel $\mathcal{S}(z, \zeta)$ for $\mathbb{A}_q$ up to a constant factor.

The BMD on $\mathbb{A}_q$ is the diffusion process on $\mathbb{A}_q \cup \{a^*\}$ obtained from the absorbing Brownian motion on $\mathbb{D}$ by rendering the inner concentric disk $\mathbb{D}_q = \{z : |z| < q\}$ into a single point a^*. The BMD-Poisson kernel $K^*(z, e^{i\theta})$, $z \in \mathbb{A}_q$, $0 \le \theta < 2\pi$, to represent BMD-harmonic functions by their values on $\partial\mathbb{D}$ admits the same expression as (5.2) of [7]:

$$K^*(z, e^{i\theta}) = -\frac{1}{2}\frac{d}{dr}G^0(z, re^{i\theta})\big|_{r=1} - \varphi(z)p^{-1}\frac{d}{dr}\varphi(re^{i\theta})\big|_{r=1},$$

where G^0 is the Green function (the 0-order resolvent density) of the ABM on $\mathbb{A}_q$, φ is the hitting probability of $\mathbb{D}_q$ for the ABM on $\mathbb{D}$, and p is the period of φ around $\mathbb{D}_q$. Due to the rotational symmetry, the second term of the right hand side is independent of θ, and $K^*(z, \zeta)$ is a harmonic function in $z \in \mathbb{A}_q$ taking a constant $1/(2\pi)$ on $\partial\mathbb{D}_q$ for each $\theta \in [0, 2\pi)$.

Consider any non-negative continuous function ϕ on $[0, 2\pi)$ with $\int_0^{2\pi}\phi(\theta)d\theta = 1$ and let $u(z) = \int_0^{2\pi} K^*(z, e^{i\theta})\phi(\theta)d\theta,\ z \in \mathbb{A}_q$. Then u is harmonic on $\mathbb{A}_q$, taking a constant $1/(2\pi)$ on $\partial\mathbb{D}_q$ and taking the value $\phi(\theta)$ at each $e^{i\theta} \in \partial\mathbb{D}$. By virtue of Proposition 2.2 (ii), $f(z) = \frac{1}{2\pi}\int_0^{2\pi}\phi(\theta)\mathcal{K}_q(z, e^{i\theta})d\theta$ is an analytic function on $\mathbb{A}_q$ whose real part $\Re f(z)$ possesses the same boundary value on $\partial\mathbb{A}_q$ as u. Therefore $\Re f(z) = u(z),\ z \in \mathbb{A}_q$. By making $\phi \to \delta_{\theta_0}$ for a fixed $\theta_0 \in [0, 2\pi)$, we conclude that $\frac{1}{2\pi}\Re\mathcal{K}_q(z, e^{i\theta_0}) = K^*(z, e^{i\theta_0})$, that is to say, $\frac{1}{2\pi}\mathcal{K}_q(z, e^{i\theta_0}),\ 0 \le \theta_0 < 2\pi$, is nothing but a BMD-Schwarz kernel for the annulus $\mathbb{A}_q$.

6 K-L Equation on Circularly Slit Annulus in Terms of BMD Schwarz Kernel

A domain D of the form $D = \mathbb{A}_q \setminus \bigcup_{j=1}^{N-1} C_j$ is called a *circularly slit annulus* if $\mathbb{A}_q = \{z \in \mathbb{C} : q \le |z| < 1\}$ is an annulus for some $q \in (0, 1)$ and C_j are mutually disjoint concentric circular slits contained in $\mathbb{A}_q$. We denote by $\mathcal{D}$ the collection of all circularly slit annuli. The Komatu-Loewner equation for $\mathcal{D}$ has been formulated by Komatu [17] and Bauer-Friedrich [3]. In this section, we make their descriptions of the equation more precise in terms of a normalized BMD Schwarz kernel introduced below.

We fix $D = \mathbb{A}_Q \setminus \bigcup_{j=1}^{N-1} C_j \in \mathcal{D}$ and consider a Jordan arc $\gamma : [0, t_\gamma] \mapsto \overline{D}$ with $\gamma(0) = \partial\mathbb{D},\ \gamma(0, t_\gamma] \subset D$. According to [17], we can then find a strictly increasing function $\alpha :\ [0, t_\gamma] \mapsto [Q, Q_\gamma],\ (\alpha(t_\gamma) = Q_\gamma)$ such that, for $q = \alpha(t)$, there exists a unique conformal map

$$g_q :\ D \setminus \gamma[0, t]\ \mapsto\ D_q = \mathbb{A}_q \setminus \bigcup_{j=1}^{N-1} C_j(q) \in \mathcal{D}, \quad \text{with} \quad g_q(Q) = q.$$

α may not be continuous as in the annulus case of Sect. 3. Nevertheless we can reparametrize the curve γ as $\{\widetilde{\gamma}(q) : q \in \mathrm{dom}(\widetilde{\gamma})\}$ by setting $\widetilde{\gamma}(q) = \gamma(\alpha^{-1}(q))$, where $\mathrm{dom}(\widetilde{\gamma}) = \alpha[0, t_\gamma] \subset [Q, Q_\gamma]$.

For $D = \mathbb{A}_q \setminus \bigcup_{j=1}^{N-1} C_j \in \mathcal{D}$, let $K_D^*(z, \zeta),\ z \in D,\ \zeta \in \partial\mathbb{D}$, be the BMD Poisson kernel for D. A *BMD Schwarz kernel* $\mathcal{S}_D(z, \zeta)$ for D is by definition a function analytic in $z \in D$ satisfying $\Re\mathcal{S}_D(z, \zeta) = K_D^*(z, \zeta)$. For each $\zeta \in \partial\mathbb{D}$,

$\mathcal{S}_D(\cdot,\zeta)$ exists uniquely up to an imaginary additive constant owing to the zero period property of a BMD harmonic function (cf. [7]). Let us denote by $\widehat{\mathcal{S}}_D(z,\zeta)$ the BMD Schwarz kernel subjected to a normalization

$$\Im\widehat{\mathcal{S}}_D(q,\zeta)=0, \quad \text{for any} \quad \zeta\in\partial\mathbb{D}. \tag{6.1}$$

Any BMD Schwarz kernel $\mathcal{S}_D(z,\zeta)$ gives rise to a normalized one by

$$\widehat{\mathcal{S}}_D(z,\zeta)=\mathcal{S}_D(z,\zeta)-i\Im\mathcal{S}_D(q,\zeta), \quad z\in D, \quad \zeta\in\partial\mathbb{D}. \tag{6.2}$$

If D is just an annulus $\mathbb{A}_q$ with no circular slit, then we see by virtue of the result in the preceding section that its normalized BMD Schwarz kernel equals the *normalized Villat's kernel* multiplied by $\frac{1}{2\pi}$:

$$\widehat{\mathcal{S}}_{\mathbb{D}}(z,\zeta)=\frac{1}{2\pi}[\mathcal{K}_q(z,\zeta)-i\Im\mathcal{K}_q(q,\zeta)]. \tag{6.3}$$

Take $0\le t^*<t\le t_\gamma$ and put $q=\alpha(q),\ q^*=\alpha(t^*)$. Then $Q\le q^*<q\le Q_\gamma$. Define $g_{q^*q}=g_{q^*}\circ g_q^{-1}$ which maps D_q conformally onto $D_{q^*}\backslash S_{q^*q}$ and satisfies

$$g_{q^*q}(q)=q^*, \tag{6.4}$$

where $S_{q^*q}=g_{q^*}\gamma[t^*,t]$. Let

$$\lambda(q)=g_q(\widetilde{\gamma}(q)) \tag{6.5}$$

that is located on an outer circle of D_q. The pre-image $g_{q^*q}^{-1}(S_{q^*q})$ of S_{q^*q} is a subarc $\{e^{i\theta}:\ \beta_1(t^*,t)<\theta<\beta_2(t^*,t)\}$ of the outer circle of D_q containing the point $\lambda(q)$.

Now $\log\left|\dfrac{g_{q^*q}(z)}{z}\right|$, $z\in D_q$, is harmonic on D_q as the imaginary part of the well defined analytic function $\log\dfrac{g_{q^*q}(z)}{z}$ on D_q and takes a constant value on each circular slit $C_j(q)$. Therefore we can verify just as in [7, Sect. 6.3] that

$$\log\left|\frac{g_{q^*q}(z)}{z}\right|=\int\limits_{\partial\mathbb{D}}\left|\log\frac{g_{q^*q}(\zeta)}{\zeta}\right|K_q^*(z,\zeta)s(d\zeta),\ z\in D_q, \tag{6.6}$$

where $K_q^*(z,\zeta)$ is the BMD Poisson kernel for the circularly slit annulus D_q. Hence we get

$$\log\frac{g_{q^*q}(z)}{z}=\int\limits_{\beta_0(t^*,t)}^{\beta_1(t^*,t)}\log|g_{q^*q}(e^{i\varphi})|\widehat{\mathcal{S}}_q(z,e^{i\varphi})d\varphi+ic, \tag{6.7}$$

for the normalized BMD Schwarz kernel $\widehat{\mathcal{S}}_q$ and for some real constant c.

By substituting $z = q$ in (6.7), we obtain from (6.4)

$$\log \frac{q^*}{q} = \int_{\beta_0(t^*,t)}^{\beta_1(t^*,t)} \log |g_{q^*q}(e^{i\varphi})| \widehat{\mathcal{S}}_q(q, e^{i\varphi}) d\varphi + ic,$$

which implies that $c = 0$ on account of (6.1).

On the other hand, the Cauchy integral theorem applied to the analytic function $\log \dfrac{g_{q^*q}(z)}{z}$ on the circularly slit annulus D_q yields just as in [3, Sect. 3.2]

$$\log \frac{q^*}{q} = \int_{\beta_0(t^*,t)}^{\beta_1(t^*,t)} \log |g_{q^*q}(e^{i\varphi})| d\varphi. \tag{6.8}$$

The integrand on the right hand side of (6.8) being uniformly bounded, we get the left continuity of $q = \alpha(t)$ by letting $t^* \uparrow t$ in (6.8).

We next substitute $z = g_q(w)$ into the identity (6.7) with $c = 0$. We then divide the resulting the both hand sides of the resulting identity by those of (6.8) and let $t^* \uparrow t$ in getting the following theorem.

Theorem 6.1 *$q = \alpha(t)$ is left continuous in $t \in (0, t_\gamma]$.*
$g_q(z)$ is left-differentiable in q and it holds for $z \in D \backslash \gamma[0, t]$ that

$$\frac{\partial^- \log g_q(z)}{\partial \log q} = 2\pi \widehat{\mathcal{S}}_q(g_q(z), \lambda(q)), \;\; q \in \alpha(0, t_\gamma] \subset (Q, Q_\gamma], \;\; g_Q(z) = z, \tag{6.9}$$

where the left hand side denotes the left derivative.

Remark 6.2 In the special case that $N = 1$, D is just an annulus $\mathbb{A}_Q$ and the equation (6.9) is reduced to

$$\frac{\partial^- \log g_q(z)}{\partial \log q} = \mathcal{K}_q(g_q(z), \lambda(q)) - i\Im \mathcal{K}_q(q, \lambda(q)). \; q \in \alpha(0, t_\gamma], \;\; g_Q(z) = z, \tag{6.10}$$

by virtue of (6.3), which actually holds in the true derivative sense as has been proved in Theorem 3.1 by making use of the kernel convergence theorem for annuli formulated in Appendix.

In the case where $N > 1$ so that the degree of the multiplicity of the circularly slit annulus D is equal or greater than 3, the problem of proving the equation (6.9) to be a genuine ODE remains open, although Komatu [17] tried to do so by an induction in $N \geq 1$ not quite successfully.

7 Appendix: Carathéodory-Komatu Convergence Theorem for Annuli

As in Sect. 3, we use the notations $\mathbb{D}(z,r) = \{w \in \mathbb{C} : |w-z| < r\}$, $\mathbb{A}(s,t) = \{w \in \mathbb{C} : s < |w| < t\}$ for $r > 0,\ 0 < s < t$.

Consider the following two conditions on a doubly connected domain D in $\mathbb{C}$:

(i) $D \subset \mathbb{A}(1,a)$ for some $a > 1$,
(ii) D admits $\partial\mathbb{D}(0,1)$ as one of the boundary components of D.
We let $\mathcal{D} = \{D : D$ is a doubly connected domain satisfying (i) and (ii)$\}$.

For a sequence $\{D_n\}$ in $\mathcal{D}$, we define its kernel as follows. Suppose that $D_0 \subset \cap_{n=1}^{\infty} D_n$ for some $D_0 \in \mathcal{D}$. Then the *kernel* of $\{D_n\}$ is defined as the maximal doubly connected domain D in $\mathbb{D}(0,1)^c$ such that D satisfies (ii) and any compact subset of D is contained in D_n for sufficiently large n. Otherwise, the kernel is defined to be $\partial\mathbb{D}(0,1)$. A sequence $\{D_n\}$ in $\mathcal{D}$ is said to be *convergent* to D in the sense of *kernel convergence*, if D is the kernel of $\{D_n\}$ and the kernel of any subsequence of $\{D_n\}$ coincides with D. A sequence $\{D_n\}$ in $\mathcal{D}$ is said to be *uniformly bounded* if $D_n \subset \mathbb{A}(1,a),\ n \geq 1$, for some $a > 1$.

It is known that if there exists a conformal map from D onto D' with $D, D' \in \mathcal{D}$, then D and D' admit an identical modulus and the map extends homeomorphically from $\partial\mathbb{D}(\mathbf{0},1) \cup D$ onto $\partial\mathbb{D}(\mathbf{0},1) \cup D'$.

A version of the following theorem was presented in [14, 16] without proof only by mentioning its similarity to a proof of Carathéodory's kernel convergence theorem for a disk. But we give a proof for completeness.

Theorem 7.1 (*Carathéodory-Komatu Convergence Theorem*) *Let $\{D_n\}$ be a uniformly bounded sequence of doubly connected domains in $\mathcal{D}$ and let $\{R_n\}$ be a sequence with $R_n > 1,\ n \geq 1$, such that there is a conformal map F_n from $\mathbb{A}(1,R_n)$ onto D_n satisfying $F_n(1) = 1$ for every n. Then the kernel convergence of $\{D_n\}$ to a doubly connected domain D in $\mathcal{D}$ implies that the sequence $\{R_n\}$ converges to R yielding the modulus of D to be* $\log R$ *and that the sequence $\{F_n\}$ converges locally uniformly to a conformal map F from $\mathbb{A}(1,R)$ onto D.*

Proof The assumption of the uniform boundedness of $\{D_n\}$ and the kernel convergence of $\{D_n\}$ to $D \in \mathcal{D}$ imply that $\partial D_n \subset \overline{\mathbb{A}(1,a)} \setminus \mathbb{A}(1,b),\ n \geq 1$, for some a, b with $1 < b \leq a$. Due to the monotonicity of the moduli (cf. [13, 5,1,Theorem 3]), we then have $b \leq R_n \leq a$.

As $\{F_n\}$ is a normal family, there exist a positive number $R' > 1$ and a subsequence $\{n_k\}$ of $\{n\}$ such that $\lim_{k\to\infty} R_{n_k} = R'$ and $\{F_{n_k}\}$ converges locally uniformly to some analytic function F on $\mathbb{A}(1,R')$, which is non-constant because $\frac{1}{2\pi i}\int_{|z|=(R'+1)/2} d\log F(z)dz = \lim_{k\to\infty} \frac{1}{2\pi i}\int_{|z|=(R'+1)/2} d\log F_{n_k}(z)dz = \lim_{k\to\infty} \frac{1}{2\pi}\int_{|z|=(R'+1)/2} d\arg F_{n_k}(z) = 1$. By virtue of Hurwitz's theorem, we can deduce from the univalence of $\{F_{n_k}\}$ that F is an injective map from $\mathbb{A}(1, R')$ to its image $F(\mathbb{A}(1, R'))$.

It holds that $F(\mathbb{A}(1, R')) \subset D$. In fact, for any $\zeta \in \mathbb{A}(1, R')$, there exists $\delta > 0$ with $\mathbb{D}(\zeta, \delta) \subset \mathbb{A}(1, R')$. Then $\mathbb{D}(\zeta, \delta) \subset \mathbb{A}(1, R_n)$ from some n on. Since the univalence of the function F implies that the coefficient c_1 in the Taylor expansion of $F(z) - F(\zeta) = c_1(z - \zeta) + \cdots$ around ζ does not vanish, we can deduce $|F'_{n_k}(\zeta)| \geq c$ holding for some $c > 0$ and for sufficiently large k from the local uniform convergence of $\{F_{n_k}\}$ to F combined with Cauchy's integral expressions of $F'(\zeta)$ and $F'_{n_k}(\zeta)$. Hence there exists $\rho > 0$ such that $\mathbb{D}(F_{n_k}(\zeta), \rho) \subset F_{n_k}(\mathbb{D}(\zeta, \delta)) \subset D_{n_k}$ from some k on by Koebe 1/4 theorem. Since $\lim_{n_k \to \infty} F_{n_k}(\zeta) = F(\zeta)$, we have $\mathbb{D}(F(\zeta), \rho/2) \subset D_{n_k}$ from some k on, and consequently $F(A(1, R')) \subset D$ because D is also the kernel of D_{n_k}.

Denote by H_n the inverse of F_n. Since the family $\{H_n\}$ is uniformly bounded, we may assume that $\{H_{n_k}\}$ is a locally uniformly convergent sequence by taking a suitable subsequence of $\{n_k\}$ if necessary. Since D is also the kernel of $\{D_{n_k}\}$, we can see that, for any $w \in D$, $w \in D_{n_k}$ for sufficiently large k and $H(w) = \lim_{k \to \infty} H_{n_k}(w)$ is well defined with $1 \leq |H(w)| \leq R'$. Further H is non-constant because of $\frac{1}{2\pi i} \int_{|w|=r} d \log H(w) = \lim_{k \to \infty} \frac{1}{2\pi i} \int_{|w|=r} d \log H_{n_k}(w) = \lim_{k \to \infty} \frac{1}{2\pi} \int_{|w|=r} d \arg H_{n_k}(w) = 1$ for some $r > 1$ satisfying $\partial\mathbb{D}(0, r) \subset \bigcap_{n=1}^{\infty} D_n$.

Therefore, by applying the open mapping theorem to the analytic function H together with the pointwise convergence of $\{H_{n_k}\}$ to H as $k \to \infty$, we see that, for any fixed $w \in D$, there exists a positive number δ such that $H_{n_k}(w) \in \mathbb{D}(H(w), \delta) \subset \mathbb{A}(1, R')$ for sufficiently large k.

If F omits the value w, we have the following contradiction:

$$0 = \frac{1}{2\pi i} \int\limits_{C_{H(w),\delta}} \frac{F'(z)}{F(z) - w} dz = \lim_{k \to \infty} \frac{1}{2\pi i} \int\limits_{C_{H(w),\delta}} \frac{F'_{n_k}(z)}{F_{n_k}(z) - w} dz = 1,$$

where $C_{H(w),\delta} = \partial\mathbb{D}(H(w), \delta)$ with counterclockwise orientation. Accordingly, F takes the value w at some point in $\mathbb{D}(H(w), \delta)$. By combining this with $F(A(1, R')) \subset D$ and the univalence of F, we conclude that F is a conformal map from $\mathbb{A}(1, R')$ onto D.

Owing to the uniqueness of the modulus of the domain D, we have $R = \lim_{k \to \infty} R_{n_k}$ independently of the choice of $\{n_k\}$. Further, $F = \lim_{k \to \infty} F_{n_k}$ gives a conformal map from $\mathbb{A}(1, R)$ onto D. As $F(1) = 1$, F is uniquely determined independently of the choice of $\{n_k\}$ (cf. [13, 5,1, Theorem 2]), yielding the desired conclusion. □

Consider q^* with $0 < q^* < 1$ and a sequence $\{q_n\}$ satisfying $q^* < q_n < 1$ for each n.

Corollary 7.2 *Let $\{h_n\}$ be a sequence of univalent functions satisfying the following conditions*:

(i) *Each h_n is a surjective map from a domain E_n to $\mathbb{A}(q_n, 1)$ with $E_n \subset \mathbb{A}(q^*, 1)$.*
(ii) *$E_n \subset E_{n+1}$ for every n and $\cup_{n=1}^{\infty} E_n = \mathbb{A}(q^*, 1)$.*
(iii) *Each E_n has $\partial\mathbb{D}(\mathbf{0}, q^*)$ as one of its boundary components.*
(iv) *$h_n(q^*) = q_n$ for every n.*

Then $\lim_{n\to\infty} q_n = q^*$ and $\{h_n\}$ converge locally uniformly to the identity map on $\mathbb{A}(q^*, 1)$.

Proof We denote the inverse function of h_n by g_n and define a conformal map F_n from $\mathbb{A}(1, \dfrac{1}{q_n})$ onto D_n satisfying $F_n(1) = 1$ by $F_n(z) = \dfrac{1}{q^*} g_n(q_n z)$ for each n, where $D_n = \{\dfrac{z}{q^*} \in \mathbb{C} : z \in E_n\} \cap \mathbb{A}(1, \dfrac{1}{q^*})$. Then the kernel convergence of the sequence $\{D_n\}$ in $\mathcal{D}$ to $\mathbb{A}(1, \dfrac{1}{q^*}) \in \mathcal{D}$ follows from (ii). Since the modulus of $\mathbb{A}(1, \dfrac{1}{q^*})$ equals q^*, we can apply Theorem 7.1 to deduce that $\lim_{n\to\infty} q_n = q^*$ and that $\{F_n\}$ converges to a conformal map F from $\mathbb{A}(1, \frac{1}{q^*})$ onto itself locally uniformly on $\mathbb{A}(1, \frac{1}{q^*})$. Since $F(1) = 1$, we get $F(z) = z$, $z \in \mathbb{A}(1, \frac{1}{q^*})$, that yields the desired conclusion. □

References

1. L.V. Ahlfors, *Conformal Invariants* (AMS, Providence, 2010)
2. R.O. Bauer, R.M. Friedrich, On radial stochastic Loewner evolution in multiply connected domains. J. Funct. Anal. **237**, 565–588 (2006)
3. R.O. Bauer, R.M. Friedrich, On chordal and bilateral SLE in multiply connected domains. Math. Z. **258**, 241–265 (2008)
4. C. Boehm, W. Lauf, *A Komatu-Loewner Equation for Multiple Slits, To Appear in Computational Methods in Function Theory* (Springer, Berlin, 2014)
5. Z.-Q. Chen, M. Fukushima, *Symmetric Markov Processes, Time Changes, and Boundary Theory* (Princeton University Press, Princeton, 2012)
6. Z.-Q. Chen, M. Fukushima, Stochastic Komatu-Loewner evolutions and a BMD domain constant (Preprint)
7. Z.-Q. Chen, M. Fukushima, S. Rhode, Chordal Komatu-Loewner equation and Brownian motion with darning in multply connected domains, to appear in Trans. Amer. Math. Soc.
8. M.D. Contreras, S. Diaz-Madrigal, P. Gumenyuk, Loewner theory in annulus I: Evolution families and differential equations, to appear in Trans. Amer. Math. Soc. Available on arXiv: 1011.4253
9. J.B. Conway, *Functions of one complex variables II* (Springer, Berlin, 1995)
10. P.L. Duren, *Univalent Functions* (Springer, Berlin, 1983)
11. J.B. Garnett, D.E. Marshall, *Harmonic Measure* (Cambridge University Press, Cambridge, 2005)
12. G.M. Goluzin, On the parametric representation of functions univalent in a ring. Math. Sbornik N.S. **29**(71), 469–476 (1951)

13. G.M. Goluzin, *Geometric Theory of Functions of a Complex Variable* (American Mathematical Society Translations 26, Providence, 1969)
14. Y. Komatu, Untersuchungen über konforme abbildung von zweifach zusammenhängenden gebieten. Proc. Phys.-Math. Soc. Japan **25**, 1–42 (1943)
15. Y. Komatu, Theory of Conformal Mapping, (in Japanese) (Kyouritu Shuppan, vol. 1, 1944)
16. Y. Komatu, Theory of Conformal Mapping, (in Japanese) (Kyouritu Shuppan, vol.2, 1949)
17. Y. Komatu, On conformal slit mapping of multiply-connected domains. Proc. Japan Acad. **26**, 26–31 (1950)
18. G.F. Lawler, Conformally invariant processes in the plane, *Mathematical Surveys and Monographs* (AMS, Providence, 2005)
19. T. Vaitsiakhovich, Boundary value problem for complex partial differential equations in a ring domain, PhD Thesis in Reference [42] of [8]
20. H. Villat, Le problème de Dirichlet dans une aire anulare. Rend. de. Circ. Mat. di Palermo **33**, 134–157 (1912)
21. H. Villat, *Lecons sur l'hydrodynamique*, (Gautiers-Villas, Paris, 1929)
22. D. Zhan, Stochastic Loewner evolution in doubly connected domains. Probab. Theory Relat. Fields **129**, 340–380 (2004)

Skew-Unfolding the Skorokhod Reflection of a Continuous Semimartingale

Tomoyuki Ichiba and Ioannis Karatzas

Abstract The Skorokhod reflection of a continuous semimartingale is unfolded, in a possibly skewed manner, into another continuous semimartingale on an enlarged probability space according to the excursion-theoretic methodology of [14]. This is done in terms of a skew version of the Tanaka equation, whose properties are studied in some detail. The result is used to construct a system of two diffusive particles with rank-based characteristics and skew-elastic collisions. Unfoldings of conventional reflections are also discussed, as are examples involving skew Brownian Motions and skew Bessel processes.

Keywords Skorokhod and conventional reflections · Skew and perturbed Tanaka equations · Skew Brownian and Bessel processes · Pure and Ocone martingales · Local time · Competing particle systems · Asymmetric collisions

AMS 2000 Subject Classifications: Primary 60G42; Secondary 60H10

1 The Result

On a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}),\ \mathbb{F} = \{\mathcal{F}(t)\}_{0 \le t < \infty}$ satisfying the so-called "usual conditions" of right continuity and augmentation by null sets, we consider a real-valued continuous semimartingale $U(\cdot)$ of the form

$$U(t) \;=\; M(t) + A(t)\,, \quad 0 \le t < \infty \tag{1.1}$$

T. Ichiba (✉)
Department of Statistics and Applied Probability, University of California, South Hall, Santa Barbara, CA 93106, USA
e-mail: ichiba@pstat.ucsb.edu

I. Karatzas
Department of Mathematics, Columbia University, New York, NY, USA
e-mail: ik@math.columbia.edu

I. Karatzas
INTECH Investment Management, One Palmer Square, Suite 441, Princeton, NJ 08542, USA
e-mail: ik@enhanced.com

D. Crisan et al. (eds.), *Stochastic Analysis and Applications 2014*,
Springer Proceedings in Mathematics & Statistics 100,
DOI 10.1007/978-3-319-11292-3_13

with $M(\cdot)$ a continuous local martingale and $A(\cdot)$ a process of finite first variation on compact intervals. We assume $M(0) = A(0) = 0$ for concreteness.

There are two ways to "fold", or reflect, this semimartingale about the origin. One is the *conventional reflection*

$$R(t) \;:=\; |U(t)|\,, \quad 0 \le t < \infty\,; \tag{1.2}$$

the other is the *Skorokhod reflection*

$$S(t) \;:=\; U(t) + \max_{0 \le s \le t} \big(-U(s) \big)\,, \quad 0 \le t < \infty\,. \tag{1.3}$$

The following result, inspired by Prokaj [14], shows how the first can be obtained from the second, by suitably unfolding the Skorokhod reflection in a possibly "skewed" manner.

Theorem 1 *Fix a constant* $\alpha \in (0,1)$. *There exists an enlargement* $\big(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{\mathbb{P}}\big)$, $\widetilde{\mathbb{F}} = \{\widetilde{\mathcal{F}}(t)\}_{0 \le t < \infty}$ *of the filtered probability space* $(\Omega, \mathcal{F}, \mathbb{P})$, $\mathbb{F} = \{\mathcal{F}(t)\}_{0 \le t < \infty}$ *with a measure-preserving map* $\pi : \Omega \to \widetilde{\Omega}$, *and on this enlarged space a continuous semimartingale* $X(\cdot)$ *that satisfies*

$$\big|X(\cdot)\big| = S(\cdot)\,, \quad L^X(\cdot) = \alpha\, L^S(\cdot)\,,$$
$$X(\cdot) \;=\; \int_0^{\cdot} \overline{sgn}\big(X(t)\big)\, \mathrm{d}U(t) + \frac{2\,\alpha - 1}{\alpha}\, L^X(\cdot)\,. \tag{1.4}$$

Here and throughout this paper, we use the notation

$$L^U(\cdot) \;:=\; \lim_{\varepsilon \downarrow 0} \frac{1}{2\,\varepsilon} \int_0^{\cdot} 1_{\{0 \le U(t) < \varepsilon\}}\, \mathrm{d}\langle U \rangle(t)\,, \quad \widehat{L}^U(\cdot) \;:=\; \frac{1}{2} \Big(L^U(\cdot) + L^{-U}(\cdot) \Big) \tag{1.5}$$

respectively for the *right* and the *symmetric* local time at the origin of a continuous semimartingale as in (1.1), and the conventions

$$\overline{\mathrm{sgn}}(x) \;:=\; \mathbf{1}_{(0,\infty)}(x) - \mathbf{1}_{(-\infty,0)}(x)\,, \quad \mathrm{sgn}(x) \;:=\; \mathbf{1}_{(0,\infty)}(x) - \mathbf{1}_{(-\infty,0]}(x)\,, \quad x \in \mathbb{R}$$

for the symmetric and the left-continuous versions, respectively, of the signum function. We also denote by $\mathbb{F}^U = \{\mathcal{F}^U(t)\}_{0 \le t < \infty}$ the "natural filtration" of $U(\cdot)$, that is, the smallest filtration that satisfies the usual conditions and with respect to which $U(\cdot)$ is adapted; we set $\mathcal{F}^U(\infty) := \sigma\big(\bigcup_{0 \le t < \infty} \mathcal{F}^U(t)\big)$. Equalities between stochastic processes, such as in (1.4), are to be understood throughout in the almost sure sense.

Theorem 1 constructs a continuous semimartingale $X(\cdot)$ whose conventional reflection coincides with the Skorokhod reflection of the given semimartingale $U(\cdot)$, and which satisfies the stochastic integral equation in (1.4). We think of this equation as a *skew version* of the celebrated Tanaka equation driven by the continuous semimartingale $U(\cdot)$, whose "skew-unfolding" it produces via the parameter α. When there is no skewness, i.e., with $\alpha = 1/2$, the integral equation of (1.4) reduces to the classical Tanaka equation; in this case Theorem 1 is just the main result in the paper [14], which inspired our work.

We shall prove Theorem 1 in Sect. 3, then use it in Sect. 5 to construct a system of two diffusive particles with rank-based characteristics and skew-elastic collisions. Section 4 discusses a similar skew-unfolding of the conventional reflection $R(\cdot) = |U(\cdot)|$ of $U(\cdot)$. In the section that follows we discuss briefly some properties of the *skew Tanaka equation* in (1.4).

2 The Skew Tanaka Equation

A first question that arises regarding the stochastic integral equation in (1.4), is whether it can be written in the more conventional form

$$X(\cdot) = \int_0^{\cdot} \operatorname{sgn}\big(X(t)\big)\, \mathrm{d}U(t) + \frac{2\,\alpha - 1}{\alpha}\, L^X(\cdot)\,, \tag{2.1}$$

in terms of the asymmetric (left-continuous) version of the signum function.

For this, it is necessary and sufficient to have

$$\int_0^{\cdot} \mathbf{1}_{\{X(t)=0\}}\, \mathrm{d}U(t) \equiv 0\,, \quad \text{or equivalently} \quad \int_0^{\cdot} \mathbf{1}_{\{S(t)=0\}}\, \mathrm{d}U(t) \equiv 0 \tag{2.2}$$

in the context of Theorem 1. Now from (1.1), (1.3) it is clear that $M(\cdot)$ is the local martingale part of the continuous semimartingale $S(\cdot)$, so we have $\langle S\rangle(\cdot) = \langle U\rangle(\cdot) = \langle M\rangle(\cdot)$ and

$$\int_0^{\infty} \mathbf{1}_{\{S(t)=0\}}\, \mathrm{d}\langle M\rangle(t) \;=\; 0 \tag{2.3}$$

(e.g., Karatzas and Shreve [9], Exercise 3.7.10). This gives $\int_0^{\cdot} \mathbf{1}_{\{S(t)=0\}}\, \mathrm{d}M(t) \equiv 0$, so (2.2) will follow if and only if

$$\int_0^{\cdot} \mathbf{1}_{\{S(t)=0\}}\, \mathrm{d}A(t) \;\equiv\; 0 \tag{2.4}$$

holds; and on the strength of (2.3), a sufficient condition for (2.4) is that $A(\cdot)$ be absolutely continuous with respect to the quadratic variation process $\langle M\rangle(\cdot)$. We have the following result.

Proposition 1 *For a given continuous semimartingale $U(\cdot)$ of the form (1.1) the stochastic integral equation of (1.4) can be cast equivalently in the form (2.1), if and only if (2.4) holds; and in this case we have the identification $L^S(t) = \max_{0\le s\le t}\big(-U(s)\big)$ and the filtration comparisons*

$$\mathcal{F}^{|X|}(t) = \mathcal{F}^U(t) \subseteq \mathcal{F}^X(t)\,, \quad 0 \le t < \infty\,. \tag{2.5}$$

Whereas, a sufficient condition for (2.4) to hold, is that there exist an $\mathbb{F}-$progressively measurable process $p(\cdot)$, locally integrable with respect to $\langle M\rangle(\cdot)$ and such that

$$A(\cdot) = \int_0^{\cdot} p(t)\,\mathrm{d}\langle M\rangle(t)\,. \tag{2.6}$$

Proof The first and third claims have already been argued. As for the second, we observe that the Itô-Tanaka formula applied to (2.1) gives

$$\begin{aligned} S(\cdot) = \big|X(\cdot)\big| &= \int_0^{\cdot} \mathrm{sgn}\big(X(t)\big)\,\mathrm{d}X(t) + 2L^X(\cdot) = U(\cdot) - \frac{2\alpha-1}{\alpha}\,L^X(\cdot) + 2\,L^X(\cdot) \\ &= U(\cdot) + L^S(\cdot) \end{aligned}$$

on the strength of the second equality in (1.4). It is clear from this expression that the filtration comparison $\mathcal{F}^U(t) \subseteq \mathcal{F}^S(t)$ holds for all $0 \le t < \infty$; whereas the reverse inclusion and the claimed identification are direct consequences of (1.3). $\square$

Remark 1 More generally (that is, in the absence of condition (2.4)), the local time at the origin of the Skorokhod reflection $S(\cdot)$ is $L^S(t) = \max_{0\le s\le t}(-U(s)) + \int_0^t \mathbf{1}_{\{S(u)=0\}}\,\mathrm{d}A(u)\,, \ \ 0 \le t < \infty\,.$

2.1 Uniqueness in Distribution for the Skew Tanaka Equation

A second question that arises regarding the skew-Tanaka equation of (1.4), is whether it can be solved uniquely. It is well-known that we cannot expect pathwise uniqueness or strength to hold for this equation. Such strong existence and uniqueness fail already with $\alpha = 1/2$ and $U(\cdot)$ a standard Brownian motion, in which case we have in (2.5) also the strict inclusion $\mathcal{F}^U(t) \subsetneqq \mathcal{F}^X(t)$ for all $t \in (0,\infty)$ (e.g., [9, Example 5.3.5]). The Skorokhod reflection of $U(\cdot)$ can then be "unfolded" into a Brownian motion $X(\cdot)$, whose filtration is strictly finer than that of the original Brownian motion

$U(\cdot)$: the unfolding cannot be accomplished without the help of some additional randomness.

The issue, therefore, is whether *uniqueness in distribution* holds for the skew-Tanaka equation of (1.4), under appropriate conditions. We shall address this question in the case of a continuous local martingale $U(\cdot)$ with $U(0) = 0$ and $\langle U \rangle(\infty) = \infty$. Let us recall a few notions and facts about such a process, starting with its Dambis-Dubins-Schwarz representation

$$U(t) = B\big(\langle U \rangle(t)\big), \quad 0 \le t < \infty \tag{2.7}$$

(cf. [9, Theorem 3.4.6]); here $B(\theta) = U(Q(\theta)),\ 0 \le \theta < \infty$ is standard Brownian motion, and $Q(\cdot)$ the right-continuous inverse of the continuous, increasing process $\langle U \rangle(\cdot)$.

We say that this $U(\cdot)$ is *pure,* if each $\langle U \rangle(t)$ is $\mathcal{F}^B(\infty)$-measurable; we say that it is an *Ocone martingale,* if the processes $B(\cdot)$ and $\langle U \rangle(\cdot)$ are independent (cf. [4, 12], Appendix). As discussed in [17], a pure Ocone martingale is a Gaussian process.

Proposition 2 *Suppose that $U(\cdot)$ is a continuous local martingale with $U(0) = 0$ and $\langle U \rangle(\infty) = \infty$. Then uniqueness in distribution holds for the skew-Tanaka equation of* (1.4)*, or equivalently of* (2.1)*, provided that either*

(i) *$U(\cdot)$ is pure; or that*
(ii) *the quadratic variation process $\langle U \rangle(\cdot)$ is adapted to a Brownian motion $\Gamma(\cdot) := (\Gamma_1(\cdot), \ldots, \Gamma_n(\cdot))'$, with values in some Euclidean space $\mathbb{R}^n$ and independent of the real-valued Brownian motion $B(\cdot)$ in the representation* (2.7)*.*

Proof Let us consider a continuous local martingale $U(\cdot)$ with $U(0) = 0$, and any continuous semimartingale $X(\cdot)$ that satisfies the stochastic integral equation in (1.4). Then $X(\cdot)$ also satisfies the equation of (2.1), as the condition (2.6) holds in this case trivially with $p(\cdot) \equiv 0$. In fact, the Eq. (2.1) can be written then in the form

$$X(Q(s)) = \int_0^s \mathrm{sgn}\big(X(Q(\theta))\big)\, \mathrm{d}B(\theta) + \frac{2\alpha - 1}{\alpha}\, L^X(Q(s)), \quad 0 \le s < \infty,$$

with $Q(\cdot)$ the right-continuous inverse of the continuous, increasing process $\langle U \rangle(\cdot)$; cf. Proposition 3.4.8 in [9]. Setting

$$\widetilde{X}(s) := X\big(Q(s)\big), \quad \text{it is straightforward to check} \quad L^{\widetilde{X}}(s) = L^X\big(Q(s)\big), \quad 0 \le s < \infty;$$

for this, one uses the representation (1.5) for the local time at the origin, along with the fact that the local martingale part of the continuous seminartingale $X(\cdot)$ in (2.1) has quadratic variation process $\langle U \rangle(\cdot)$. Thus, the time-changed process $\widetilde{X}(\cdot)$ satisfies the stochastic integral equation

$$\widetilde{X}(s) = \int_0^s \operatorname{sgn}\big(\widetilde{X}(\theta)\big)\, \mathrm{d}B(\theta) + \frac{2\,\alpha - 1}{\alpha}\, L^{\widetilde{X}}(s)\,, \quad 0 \le s < \infty\,. \tag{2.8}$$

This can be cast as the Harrison and Shepp equation of [7], namely

$$\widetilde{X}(\cdot) = \widetilde{W}(\cdot) + \frac{2\,\alpha - 1}{\alpha}\, L^{\widetilde{X}}(\cdot) \tag{2.9}$$

for the skew Brownian motion, driven by the standard Brownian motion

$$\widetilde{W}(\cdot) := \int_0^{\cdot} \operatorname{sgn}\big(\widetilde{X}(\theta)\big)\, \mathrm{d}B(\theta)\,. \tag{2.10}$$

It is well-known from the theory of [7] that the Eq. (2.9) has a pathwise unique, strong solution; in fact, the skew Brownian motion $\widetilde{X}(\cdot)$ and the Brownian motion $\widetilde{W}(\cdot)$ generate the same filtration. Since

$$X(t) = \widetilde{X}\big(\langle U\rangle(t)\big)\,, \quad 0 \le t < \infty \tag{2.11}$$

holds with $\widetilde{X}(\cdot)$ adapted to $\mathbb{F}^{\widetilde{W}}$, the distribution of $X(\cdot)$ is uniquely determined whenever

$$\text{the Brownian motion } \widetilde{W}(\cdot) \text{ of (2.10) is independent of the process } \langle U\rangle(\cdot)\,, \tag{2.12}$$

or whenever

$$\langle U\rangle(t) \ \text{ is } \ \mathcal{F}^{\widetilde{W}}(\infty)\text{-measurable, for every } \ t \in [0, \infty)\,. \tag{2.13}$$

But (2.13) holds when $U(\cdot)$ is pure (case (i) of the Proposition); this is because from (2.10) we have $B(\cdot) = \int_0^{\cdot} \operatorname{sgn}\big(\widetilde{X}(\theta)\big)\, \mathrm{d}\widetilde{W}(\theta)$, therefore $\mathcal{F}^B(t) \subseteq \mathcal{F}^{\widetilde{W}}(t)$ for all $t \in [0, \infty)$ and thus $\mathcal{F}^B(\infty) \subseteq \mathcal{F}^{\widetilde{W}}(\infty)$.

On the other hand, (2.12) holds under the condition of case (ii) in the Proposition, as $\langle U\rangle(\cdot)$ is then adapted to the filtration generated by the n-dimensional Brownian motion $\Gamma(\cdot)$; this, in turn, is independent of $\widetilde{W}(\cdot)$ on the strength of the P. Lévy Theorem (e.g., Karatzas and Shreve [9], Theorem 3.3.16), since

$$\langle \widetilde{W}, \Gamma_j\rangle(\cdot) = \int_0^{\cdot} \operatorname{sgn}\big(\widetilde{X}(\theta)\big)\, \mathrm{d}\langle B, \Gamma_j\rangle(\theta) \equiv 0\,, \quad \forall \ j = 1, \ldots, n\,.$$

The proof of the proposition is complete. □

Remark 2 It would be interesting to obtain sufficient conditions for either (2.12) or (2.13) to hold, which are weaker than those of Proposition 2. As Example 2 shows, however—and contrary to our own initial guess—we cannot expect the conclusions of Proposition 2 to remain true for general Ocone martingales.

Example 1 From Brownian Motion to Skew Brownian Motion: Suppose that $U(\cdot)$ is standard, real valued Brownian motion. Then the conditions of Propositions 1 and 2 are satisfied rather trivially; uniqueness in distribution holds for the skew-Tanaka equation of (2.1) (equivalently, of (1.4)); and every continuous semimartingale $X(\cdot)$ that satisfies (2.1) is of the form

$$X(\cdot) \;=\; W(\cdot) + \frac{2\,\alpha - 1}{\alpha}\, L^X(\cdot) \quad \text{with} \quad W(\cdot) \;:=\; \int_0^{\cdot} \operatorname{sgn}\big(X(t)\big)\, \mathrm{d}U(t)\,,$$

or equivalently

$$X(\cdot) \;=\; W(\cdot) + 2\,\big(2\,\alpha - 1\big)\, \widehat{L}^X(\cdot)$$

in terms of the symmetric local time as in (1.5). Of course $W(\cdot)$ is standard Brownian motion by the P. Lévy theorem, and the Harrison-Shepp theory [7] once again characterizes $X(\cdot)$ as skew Brownian motion with parameter α. The processes $W(\cdot)$ and $X(\cdot)$ generate the same filtration, which is strictly finer than the filtration generated by the original Brownian motion $U(\cdot) = \int_0^{\cdot} \operatorname{sgn}\big(X(t)\big)\, \mathrm{d}W(t)$.

Example 2 Failure of Uniqueness in Distribution for General Ocone Martingales: We adapt to our setting a construction of Dubins-Emery-Yor from [4, p. 131]. We start with a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P})$, $\mathbb{F}^B = \{\mathcal{F}^B(t)\}_{0 \le t < \infty}$ where $B(\cdot)$ is standard Brownian motion with $B(0) = 0$, and define the adapted, continuous and strictly increasing process

$$A(t) \;:=\; t \cdot \mathbf{1}_{\{t \le 1\}} + \Big\{1 + \big(u \cdot \mathbf{1}_{\{B(1) > 0\}} + v \cdot \mathbf{1}_{\{B(1) \le 0\}}\big)(t-1)\Big\} \cdot \mathbf{1}_{\{t > 1\}}, \qquad 0 \le t < \infty \tag{2.14}$$

where $u > 0$ and $v > 0$ are given real numbers with $u \neq v$, as well as the processes

$$X(\cdot) \;:=\; B(A(\cdot))\,, \quad \Xi(\cdot) \;:=\; -X(\cdot)\,. \tag{2.15}$$

The Lévy transform

$$\beta(\cdot) \;:=\; \int_0^{\cdot} \operatorname{sgn}(B(t))\, \mathrm{d}B(t)$$

of $B(\cdot)$ is a standard Brownian motion adapted to the filtration $\mathbb{F}^{|B|} = \{\mathcal{F}^{|B|}(t)\}_{0 \le t < \infty}$, which is strictly coarser than $\mathbb{F}^B$; in particular, it can be seen that $\beta(\cdot)$ is independent of $\mathrm{sgn}(B(1)) = 2\,\mathbf{1}_{\{B(1)>0\}} - 1$, and thus of the process $A(\cdot)$ as well.

On the other hand, the process $X(\cdot)$ is a martingale of its natural filtration $\mathbb{F}^X = \{\mathcal{F}^B\big(A(t)\big)\}_{0\le t<\infty}$; therefore, so is its "mirror image" $\Xi(\cdot)$, and more importantly its Lévy transform

$$U(\cdot) \;:=\; \int_0^{\cdot} \mathrm{sgn}\big(X(t)\big)\,\mathrm{d}X(t) \;=\; \beta\big(A(\cdot)\big) \quad \text{with} \quad \langle U\rangle(\cdot) \;=\; A(\cdot)\,,$$

which is thus seen to be an Ocone martingale. Now clearly, both $X(\cdot)$ and $\Xi(\cdot)$ satisfy the Eq. (2.1) with $\alpha = 1/2$ driven by $U(\cdot)$, so pathwise uniqueness fails for this equation. We also note that the conditions of Proposition 2 fail too in this case.

We claim that uniqueness in distribution fails as well. In a manner similar to the treatment in [4], we shall argue that the distributions of $X(\cdot)$ and $\Xi(\cdot)$ at time $t = 2$ are different. Now if the random variables

$$X(2) \;=\; B(1+u)\cdot \mathbf{1}_{\{B(1)>0\}} + B(1+v)\cdot \mathbf{1}_{\{B(1)\le 0\}} \quad \text{and} \quad \Xi(2) \;=\; -X(2)$$

had the same probability distributions, that is, if the distribution of the random variable $X(2)$ were symmetric about the origin, we would have $\mathbb{E}[(X(2))^3] = 0$. However, let us note the decomposition

$$X(2) \;=\; B(1) + \big(B(1+u) - B(1)\big)\cdot \mathbf{1}_{\{B(1)>0\}} + \big(B(1+v) - B(1)\big)\cdot \mathbf{1}_{\{B(1)\le 0\}}\,,$$

which gives

$$\begin{aligned}
\mathbb{E}\big[(X(2))^3\big] &= 3\,\mathbb{E}\Big[B(1)\,\big(B(1+u) - B(1)\big)^2\,\mathbf{1}_{\{B(1)>0\}}\Big] \\
&\quad + 3\,\mathbb{E}\Big[B(1)\,\big(B(1+v) - B(1)\big)^2\,\mathbf{1}_{\{B(1)\le 0\}}\Big] \\
&= 3\,\mathbb{E}\Big[\big(B(1)\big)^+\Big]\,\big(u - v\big) \;\neq\; 0\,.
\end{aligned}$$

This contradiction establishes the claim.

2.2 The Perturbed Skew-Tanaka Equation Is Strongly Solvable

The addition of some independent noise can restore pathwise uniqueness, thus also strength, to weak solutions of the stochastic equation in (1.4) or (2.1). In the spirit of [15] or [6], we have the following result.

Proposition 3 *Suppose that the continuous semimartingale $U(\cdot)$ as in (1.1) satisfies the conditions of Proposition 1, where now the $\mathbb{F}$-progressively measurable process*

$p(\cdot)$ of (2.6) is locally square-integrable with respect to $\langle M\rangle(\cdot)$; and that

$$V(\cdot) \,=\, N(\cdot) + \Delta(\cdot)$$

is another continuous semimartingale, with continuous local martingale part $N(\cdot)$ and finite variation part $\Delta(\cdot)$ which satisfy $N(0) = \Delta(0) = 0$ and

$$\langle M, N\rangle(\cdot) \equiv 0\,, \quad \langle M\rangle(\cdot) \,=\, \int_0^{\cdot} q(t)\,\mathrm{d}\langle N\rangle(t)$$

for some $\mathbb{F}$-progressively measurable process $q(\cdot)$ with values in a compact interval $[0, b]$.

Then pathwise uniqueness holds for the perturbed skew-Tanaka equation

$$X(\cdot) \,=\, \int_0^{\cdot} sgn\big(X(t)\big)\,\mathrm{d}U(t) + V(\cdot) + \frac{2\,\alpha - 1}{\alpha}\,L^X(\cdot)\,, \tag{2.16}$$

provided that either

(i) $\alpha = 1/2$*, or that*
(ii) *$U(\cdot)$ and $V(\cdot)$ are independent, standard Brownian motions. In this case a weak solution to (2.16) exists, and is thus strong.*

The claim of case (i) is proved in Theorem 8.1 of [6], and the claim of case (ii) in an Appendix, Sect. 6. In case (ii) of Proposition 3 the Eq. (2.16) can be written equivalently as

$$X(\cdot) \,=\, \int_0^{\cdot} \mathbf{1}_{\{X(t)>0\}}\,\mathrm{d}W_+(t) \,+\, \int_0^{\cdot} \mathbf{1}_{\{X(t)<0\}}\,\mathrm{d}W_-(t) \,+\, \frac{2\,\alpha - 1}{\alpha}\,L^X(\cdot)\,.$$

Here $W_\pm(\cdot) := V(\cdot) \pm U(\cdot)$ are independent Brownian motions with local variance 2; one of them governs the motion of $X(\cdot)$ during its positive excursions, the other during the negative ones, whereas these excursions get skewed when $\alpha \neq 1/2$.

3 Proof of Theorem 1

We shall follow very closely the methodology of [14], with some necessary modifications related to the skewness. The enlargement of the filtered probability space $(\Omega, \mathcal{F}, \mathbb{P})$, $\mathbb{F} = \{\mathcal{F}(t)\}_{0\le t<\infty}$ is done in terms of a sequence $\{\xi_k\}_{k\in\mathbb{N}}$ of independent random variables with common Bernoulli distribution

$$\mathbb{P}\big(\xi_1 = +1\big) = \alpha\,, \quad \mathbb{P}\big(\xi_1 = -1\big) = 1 - \alpha \tag{3.1}$$

(thus with expectation $\mathbb{E}(\xi_1) = 2\alpha - 1$), which is independent of $\mathcal{F}(\infty) = \sigma\big(\bigcup_{0 \le t < \infty} \mathcal{F}(t)\big)$. On the enlarged probability space $\big(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{\mathbb{P}}\big)$ we have all the objects of the original space, so we keep the same notation for them. We denote by

$$\mathfrak{Z} := \big\{ t \ge 0 : S(t) = 0 \big\} \tag{3.2}$$

the zero set of the Skorokhod reflection $S(\cdot)$ in (1.3), and enumerate as $\{\mathcal{C}_k\}_{k \in \mathbb{N}}$ the disjoint components of $[0, \infty) \setminus \mathfrak{Z}$, that is, the countably-many excursion intervals of the process $S(\cdot)$ away from the origin. This we do in a measurable manner, so that

$$\big\{ t \in \mathcal{C}_k \big\} \in \mathcal{F}(\infty), \quad \forall\ t \ge 0,\ k \in \mathbb{N}\,.$$

In order to simplify notation, we set

$$\mathcal{C}_0 := \mathfrak{Z}\,, \quad \xi_0 := 0\,. \tag{3.3}$$

We define now

$$Z(t) := \sum_{k \in \mathbb{N}_0} \xi_k \, \mathbf{1}_{\mathcal{C}_k}(t)\,, \quad \widetilde{\mathcal{F}}(t) := \mathcal{F}(t) \vee \mathcal{F}^Z(t) \tag{3.4}$$

for all $t \in [0, \infty)$; this gives the enlarged filtration $\widetilde{\mathbb{F}} = \big\{ \widetilde{\mathcal{F}}(t) \big\}_{0 \le t < \infty}$. We posit the following two claims.

Proposition 4 *The process $M(\cdot)$ of (1.1) is a continuous local martingale of the enlarged filtration $\widetilde{\mathbb{F}}$. Consequently, both $U(\cdot)$ and $S(\cdot)$ are continuous $\widetilde{\mathbb{F}}$-semimartingales.*

Proposition 5 *In the notation of (1.3) and (3.4), we have*

$$Z(\cdot)\, S(\cdot) = \int_0^{\cdot} Z(t)\, \mathrm{d}S(t) + \big(2\alpha - 1\big)\, L^S(\cdot)\,. \tag{3.5}$$

Taking the claims of these two propositions at face-value for a moment, we can proceed with the proof of Theorem 1 as follows. We define the process

$$X(\cdot) := Z(\cdot)\, S(\cdot) \tag{3.6}$$

and note

$$Z(\cdot) = \overline{\mathrm{sgn}}\big(X(\cdot)\big)\,, \quad \big|X(\cdot)\big| = S(\cdot) \tag{3.7}$$

thanks to (3.3) and (3.4), as well as

$$X(\cdot) - \int_0^{\cdot} \overline{\mathrm{sgn}}\big(X(t)\big)\, \mathrm{d}S(t) \;=\; Z(\cdot)\, S(\cdot) - \int_0^{\cdot} Z(t)\, \mathrm{d}S(t) \;=\; \big(2\alpha - 1\big)\, L^S(\cdot) \quad (3.8)$$

thanks to (3.6), (3.5). In particular, $X(\cdot)$ is an $\widetilde{\mathbb{F}}$-semimartingale, and we note the property

$$2\, L^X(\cdot) - L^S(\cdot) \;=\; 2\, L^X(\cdot) - L^{|X|}(\cdot) \;=\; \int_0^{\cdot} \mathbf{1}_{\{X(t)=0\}}\, \mathrm{d}X(t)$$

of its local time at the origin (cf. Sect. 2.1 in [8]). In conjunction with (3.8) and the fact that $X(\cdot)$, $S(\cdot)$, and $Z(\cdot)$ all have the same zero set $\mathfrak{Z}$ as in (3.2), (3.3), we get from this last equation

$$\begin{aligned} 2\, L^X(\cdot) - L^S(\cdot) &= \int_0^{\cdot} \mathbf{1}_{\{X(t)=0\}} \Big[\, \overline{\mathrm{sgn}}\big(X(t)\big)\, \mathrm{d}S(t) + \big(2\alpha - 1\big) L^S(t)\, \Big] \\ &= \big(2\alpha - 1\big) L^S(\cdot), \end{aligned} \quad (3.9)$$

thus

$$L^X(\cdot) \;=\; \alpha\, L^S(\cdot)\,, \quad (3.10)$$

establishing the second equality in (1.4). Back in (3.8), this leads to

$$X(\cdot) \;=\; \int_0^{\cdot} \overline{\mathrm{sgn}}\big(X(t)\big) \big[\, \mathrm{d}U(t) + \mathrm{d}C(t) \big] + \big(2\alpha - 1\big)\, L^S(\cdot)\,, \quad (3.11)$$

where $C(\cdot)$ is the continuous, adapted and increasing process

$$C(t) \;:=\; S(t) - U(t) \;=\; \max_{0 \le s \le t} \big(- U(s) \big)\,, \quad 0 \le t < \infty\,.$$

From the theory of the Skorokhod reflection problem we know that this process $C(\cdot)$ is flat off the set $\{t \geq 0 : S(t) = 0\} = \mathfrak{Z}$, so the skew-Tanaka equation of (1.4) follows now from (3.11), (3.10).

The proof of Theorem 1 is complete. □

Proof of Proposition 4: By localization if necessary, it suffices to show that if $M(\cdot)$ is an $\mathbb{F}$-martingale, then it is also an $\widetilde{\mathbb{F}}$-martingale; that is, for any given $0 < \theta < t < \infty$ and $A \in \widetilde{\mathcal{F}}(\theta)$ we have

$$\mathbb{E}\Big[\, \big(M(t) - M(\theta)\big)\, \mathbf{1}_A \,\Big] \;=\; 0\,. \quad (3.12)$$

It is clear from (3.4) that we need to consider only sets of the form $A = B \cap D$, where $B \in \mathcal{F}(\theta)$ and

$$D = \bigcap_{j=1}^{n} \big\{ Z(t_j) = \varepsilon_j \big\} = \bigcap_{j=1}^{n} \big\{ \xi_{\kappa(t_j)} = \varepsilon_j \big\} \tag{3.13}$$

for $n \in \mathbb{N}$, $0 < t_1 < t_2 < \cdots < t_n < \theta < t$ and $\varepsilon \in \{-1, 0, 1\}$. Here we have denoted by $\kappa(u)$ the (random) index of the excursion interval $\mathcal{C}_k$ to which a given $u \in [0, \infty)$ belongs.

For such choices, and because

$$\mathbb{E}\big[\big(M(t) - M(\theta) \big) \mathbf{1}_A \big] = \mathbb{E}\big[\big(M(t) - M(\theta) \big) \mathbf{1}_B \cdot \mathbb{E}\big(\mathbf{1}_D \,|\, \mathcal{F}(\infty) \big) \big],$$

we see that, in order to prove (3.12), it is enough to argue that

$$\mathbb{E}\big(\mathbf{1}_D \,|\, \mathcal{F}(\infty) \big) \quad \text{is} \;\; \mathcal{F}(\theta) - \text{measurable.} \tag{3.14}$$

But the random variables $\kappa(t_j)$ in (3.13) are measurable with respect to $\mathcal{F}(\infty)$, whereas the random variables ξ_1, $\xi_2, \ldots$ are independent of this σ-algebra. Therefore, we have

$$\begin{aligned} \mathbb{E}\big(\mathbf{1}_D \,|\, \mathcal{F}(\infty) \big) &= \mathbb{P}\left[\bigcap_{j=1}^{n} \big\{ \xi_{\kappa(t_j)} = \varepsilon_j \big\} \,\Big|\, \mathcal{F}(\infty) \right] \\ &= \mathbb{P}\big(\xi_{k_1} = \varepsilon_1, \ldots, \xi_{k_n} = \varepsilon_n \big) \Big|_{k_1 = \kappa(t_1), \ldots, k_n = \kappa(t_n)} . \end{aligned} \tag{3.15}$$

For given indices $(k_1, \ldots, k_n)$ and $(\varepsilon_1, \ldots, \varepsilon_n)$, let us denote by m the number of distinct non-zero indices in $(k_1, \ldots, k_n)$, by λ the number from among those distinct indices of the corresponding ε_j's that are equal to 1, and observe

$$\begin{aligned} \mathbb{P}\big(\xi_{k_1} = \varepsilon_1, \ldots, \xi_{k_n} = \varepsilon_n \big) &= 0\,, \;\text{ if } \; (\varepsilon_1, \ldots, \varepsilon_n) \;\text{ contradicts }\; (k_1, \ldots, k_n)\,; \\ &= \alpha^{\lambda} \big(1 - \alpha \big)^{m - \lambda}, \quad \text{otherwise}\,. \end{aligned} \tag{3.16}$$

Here "$(\varepsilon_1, \ldots, \varepsilon_n)$ contradicts $(k_1, \ldots, k_n)$" means that we have either

- $k_i = k_j$ but $\varepsilon_i \neq \varepsilon_j$ for some $i \neq j$; or
- $k_i = 0$ but $\varepsilon_i \neq 0$, for some i ; or
- $k_i \neq 0$ but $\varepsilon_i = 0$, for some i .

We note now that when $k_1 = \kappa(t_1)$, $\ldots$, $k_n = \kappa(t_n)$, the value of m (that is, the number of excursion intervals in $[0, s] \setminus \mathfrak{Z}$ that contain some t_i), the value of λ (i.e., the number of such excursion intervals that are positive) and the statement "$(\varepsilon_1, \ldots, \varepsilon_n)$ contradicts $(k_1, \ldots, k_n)$", can all be determined on the basis of the trajectory $S(u)$, $0 \le u \le \theta$; that is, the quantity on the right-hand side of (3.15) is $\mathcal{F}^S(\theta)$-measurable. As a consequence, the property (3.14) holds. $\square$

Proof of Proposition 5: For any $\varepsilon \in (0,1)$ we define recursively, starting with $\tau_0^\varepsilon := 0$, a sequence of stopping times

$$\tau_{2\ell+1}^\varepsilon := \inf\left\{t > \tau_{2\ell}^\varepsilon \,:\, S(t) > \varepsilon\right\}, \quad \tau_{2\ell+2}^\varepsilon := \inf\left\{t > \tau_{2\ell+1}^\varepsilon \,:\, S(t) = 0\right\}$$

for $\ell \in \mathbb{N}_0$. We use this sequence to approximate the process $Z(\cdot)$ of (3.4) by

$$Z^\varepsilon(t) := \sum_{\ell \in \mathbb{N}_0} Z(t)\, \mathbf{1}_{(\tau_{2\ell+1}^\varepsilon, \tau_{2\ell+2}^\varepsilon]}(t), \quad 0 \le t < \infty.$$

Let us note that the resulting process $Z^\varepsilon(\cdot)$ is constant on each of the indicated intervals; that the sequence of stopping times just defined does not accumulate on any bounded time-interval, on account of the fact that $S(\cdot)$ has continuous paths; and that the process $Z^\varepsilon(\cdot)$ is of finite first variation over compact intervals. We deduce

$$Z^\varepsilon(T)\, S(T) = \int_0^T Z^\varepsilon(t)\, \mathrm{d}S(t) + \int_0^T S(t)\, \mathrm{d}Z^\varepsilon(t), \quad 0 \le T < \infty. \tag{3.17}$$

The piecewise-constant process $Z^\varepsilon(\cdot)$ tends to $Z(\cdot)$ pointwise as $\varepsilon \downarrow 0$, and we have

$$\lim_{\varepsilon \downarrow 0} \int_0^T Z^\varepsilon(t)\, \mathrm{d}S(t) = \int_0^T Z(t)\, \mathrm{d}S(t), \quad \text{in probability} \tag{3.18}$$

for any given $T \in [0,\infty)$; all the while, $|Z^\varepsilon(\cdot)| \le 1$. On the other hand, the second integral in (3.17) can be written as

$$\int_0^T S(t)\, \mathrm{d}Z^\varepsilon(t) = \sum_{\{\ell \,:\, \tau_{2\ell+1}^\varepsilon < T\}} S\big(\tau_{2\ell+1}^\varepsilon\big)\, Z\big(\tau_{2\ell+1}^\varepsilon\big) = \varepsilon \sum_{\{\ell \,:\, \tau_{2\ell+1}^\varepsilon < T\}} Z\big(\tau_{2\ell+1}^\varepsilon\big)$$

$$= \varepsilon \sum_{j=1}^{N(T,\varepsilon)} \xi_{\ell_j} = \varepsilon\, N(T,\varepsilon) \cdot \frac{1}{N(T,\varepsilon)} \sum_{j=1}^{N(T,\varepsilon)} \xi_{\ell_j},$$

where $\{\xi_{\ell_j}\}_{j=1}^{N(T,\varepsilon)}$ is an enumeration of the values $Z\big(\tau_{2\ell+1}^\varepsilon\big)$ and

$$N(T,\varepsilon) := \#\left\{\ell \,:\, \tau_{2\ell+1}^\varepsilon < T\right\}$$

is the number of downcrossings of the interval $(0, \varepsilon)$ that the process $S(\cdot)$ has completed by time T. From [16, Theorem VI.1.10], we have the representation of local time $\lim_{\varepsilon \downarrow 0} \varepsilon N(T, \varepsilon) = L^S(T)$; whereas the strong law of large numbers gives

$$\lim_{\varepsilon \downarrow 0} \frac{1}{N(T, \varepsilon)} \sum_{j=1}^{N(T,\varepsilon)} \xi_{\ell_j} = \mathbb{E}\big(\xi_1\big) .$$

Back into (3.17) and with the help of (3.18), these considerations give

$$Z(T)\, S(T) = \int_0^T Z(t)\, \mathrm{d}S(t) + \mathbb{E}\big(\xi_1\big) \cdot L^S(T) , \quad 0 \le T < \infty ,$$

that is, (3.5). □

4 Conventional Reflection

In a similar manner one can establish the following analogue of Theorem 1, which uses the conventional reflection in place of the Skorokhod reflection.

Theorem 2 *Fix a constant* $\alpha \in (0, 1)$. *There exists an enlargement* $\big(\widehat{\Omega}, \widehat{\mathcal{F}}, \widehat{\mathbb{P}}\big)$, $\widehat{\mathbb{F}} = \{\widehat{\mathcal{F}}(t)\}_{0 \le t < \infty}$ *of the filtered probability space* $(\Omega, \mathcal{F}, \mathbb{P})$, $\mathbb{F} = \{\mathcal{F}(t)\}_{0 \le t < \infty}$, *with a measure-preserving map* $\pi : \Omega \to \widehat{\Omega}$, *and on this enlarged space a continuous semimartingale* $\widehat{X}(\cdot)$ *that satisfies*

$$\big|\widehat{X}(\cdot)\big| = \big|U(\cdot)\big| , \quad L^{\widehat{X}}(\cdot) = \alpha\, L^{|U|}(\cdot) ,$$
$$\widehat{X}(\cdot) = \int_0^{\cdot} \overline{sgn}\big(\widehat{X}(t)\big)\, \mathrm{d}\widehat{U}(t) + \frac{2\alpha - 1}{\alpha}\, L^{\widehat{X}}(\cdot) . \tag{4.1}$$

Here

$$\widehat{U}(\cdot) := \int_0^{\cdot} \overline{sgn}\big(U(t)\big)\, \mathrm{d}U(t) \tag{4.2}$$

is the Lévy transform of the semimartingale $U(\cdot)$, *and the classical reflection* $R(\cdot) = |U(\cdot)|$ *of* $U(\cdot)$ *coincides with the Skorokhod reflection of the process* $\widehat{U}(\cdot)$ *in* (4.2), *namely*

$$\widehat{S}(t) := \widehat{U}(t) + \max_{0 \le s \le t} \big(- \widehat{U}(s) \big) , \qquad 0 \le t < \infty .$$

Indeed, most of the argument of the proof in Sect. 3 goes through verbatim, with $S(\cdot)$, $X(\cdot)$ replaced here by $R(\cdot)$, $\widehat{X}(\cdot)$, up to and including the display (3.10). But now we have

$$R(\cdot) = |U(\cdot)| = \int_0^{\cdot} \overline{\mathrm{sgn}}\big(U(t)\big)\, \mathrm{d}U(t) + L^{|U|}(\cdot) \;=\; \widehat{U}(\cdot) + L^R(\cdot) \tag{4.3}$$

from the Itô-Tanaka formula, so (3.11) is replaced by

$$\widehat{X}(\cdot) \;=\; \int_0^{\cdot} \overline{\mathrm{sgn}}\big(\widehat{X}(t)\big)\, \big[\mathrm{d}\widehat{U}(t) + \mathrm{d}L^R(t)\big] + \big(2\alpha - 1\big) L^R(\cdot)\,.$$

The property $L^{\widehat{X}}(\cdot) = \alpha\, L^R(\cdot)$ is established exactly as in (3.10), so the stochastic integral equation in (4.1) follows from this last display. On the other hand, since the local time $L^R(\cdot)$ grows only on the set $\{t \ge 0 : R(t) = 0\} = \{t \ge 0 : \widehat{X}(t) = 0\}$, the equality of the first and last terms in (4.3) identifies $R(\cdot)$ as the Skorokhod reflection $\widehat{S}(\cdot)$ of the Lévy transform $\widehat{U}(\cdot)$, as claimed in the last sentence of Theorem 2. It is well-known (see, for instance [2]) that the processes $|U(\cdot)|$ and $\widehat{U}(\cdot)$ generate the same filtration.

Remark 3 Let us note that the stochastic integral equation in (4.1) can always be written in the more conventional form

$$\widehat{X}(\cdot) \;=\; \int_0^{\cdot} \mathrm{sgn}\big(\widehat{X}(t)\big)\, \mathrm{d}\widehat{U}(t) + \frac{2\,\alpha - 1}{\alpha}\, L^{\widehat{X}}(\cdot)\,, \tag{4.4}$$

without any additional conditions on $U(\cdot)$. This is because the analogue $\int_0^{\cdot} \mathbf{1}_{\{\widehat{X}(t)=0\}}\, \mathrm{d}\widehat{U}(t) \equiv 0$ of the property in (2.2) is now satisfied trivially, on account of (4.2).

Example 3 From One Skew Brownian Motion to Another: Suppose that $U(\cdot)$ is a skew Brownian motion with parameter $\gamma \in (0, 1)$, i.e.,

$$U(\cdot) \;=\; B(\cdot) + \frac{2\,\gamma - 1}{\gamma}\, L^U(\cdot)$$

for some standard, real-valued Brownian motion $B(\cdot)$. We have in this case $\int_0^{\infty} \mathbf{1}_{\{U(t)=0\}}\, \mathrm{d}t = 0$ as well as the local time property

$$2\, L^U(\cdot) - L^{|U|}(\cdot) = \int_0^{\cdot} \mathbf{1}_{\{U(t)=0\}}\, \mathrm{d}U(t) = \frac{2\,\gamma - 1}{\gamma}\, L^U(\cdot)\,,$$

thus $L^U(\cdot) = \gamma L^{|U|}(\cdot)$ and therefore $R(\cdot) = \big|U(\cdot)\big| = \int_0^\cdot \overline{\mathrm{sgn}}\big(U(t)\big)\,\mathrm{d}U(t) + L^{|U|}(\cdot) = W(\cdot) + L^{|U|}(\cdot)$. Here we have denoted the Lévy transform of (4.2) as

$$W(\cdot) := \widehat{U}(\cdot) = \int_0^\cdot \overline{\mathrm{sgn}}\big(U(t)\big)\left(\mathrm{d}B(t) + \frac{2\gamma-1}{\gamma}\mathrm{d}L^U(t)\right) = \int_0^\cdot \mathrm{sgn}\big(U(t)\big)\mathrm{d}B(t),$$

and observed that it is another standard Brownian motion. Thus, the stochastic integral equation of (4.4) becomes

$$\widehat{X}(\cdot) = \int_0^\cdot \mathrm{sgn}\big(\widehat{X}(t)\big)\,\mathrm{d}W(t) + \frac{2\,\alpha-1}{\alpha}\,L^{\widehat{X}}(\cdot) = \widehat{W}(\cdot) + \frac{2\,\alpha-1}{\alpha}\,L^{\widehat{X}}(\cdot)$$

with $\widehat{W}(\cdot) = \int_0^\cdot \mathrm{sgn}\big(\widehat{X}(t)\big)\,\mathrm{d}W(t)$ yet another standard Brownian motion.

The Harrison and Shepp [7] theory characterizes now $\widehat{X}(\cdot)$ as skew Brownian motion with skewness parameter α. The processes $\widehat{X}(\cdot)$ and $\widehat{W}(\cdot)$ generate the same filtration, as do the processes

$$\widehat{U}(\cdot) = \int_0^\cdot \mathrm{sgn}\big(\widehat{X}(t)\big)\,\mathrm{d}\widehat{W}(t) = W(\cdot) \quad \text{and} \quad R(\cdot) = |U(\cdot)|\,;$$

and the first filtration is finer than the second.

4.1 Skew Bessel Processes

In this subsection suppose that $U^2(\cdot)$ is a squared Bessel process with dimension $\delta \in (1,2)$, i.e., $U^2(\cdot)$ is the unique strong solution of the equation

$$U^2(t) = \delta\, t + 2\int_0^t \sqrt{U^2(t)}\,\mathrm{d}B(t)\,, \quad 0 \le t < \infty$$

for some standard, real-valued Brownian motion $B(\cdot)$. When $\delta \in (1,2)$, the square root $R(\cdot) := |U(\cdot)| \ge 0$ of this process is a semimartingale that keeps visiting the origin almost surely, and can be decomposed as

$$R(\cdot) = \int_0^\cdot \frac{\delta-1}{2\,R(t)} \cdot \mathbf{1}_{\{R(t)\neq 0\}}\mathrm{d}t + B(\cdot) \quad \text{with} \quad L^R(\cdot) \equiv 0\,, \quad \int_0^\cdot \mathbf{1}_{\{R(t)=0\}}\mathrm{d}t \equiv 0\,. \tag{4.5}$$

For the study of the stochastic differential equation (4.5) with $\delta \in (1, 2)$ see, for example, [3].

Given $\alpha \in (0, 1)$, following again the argument of the proof in Sect. 3 verbatim, with $S(\cdot)$, $X(\cdot)$ replaced respectively by $R(\cdot)$, $\widehat{X}(\cdot)$, we unfold the nonnegative Bessel process $R(\cdot)$ to obtain

$$\widehat{X}(\cdot) = Z(\cdot)R(\cdot) = \int_0^{\cdot} Z(t)\mathrm{d}R(t) + (2\alpha - 1)L^R(\cdot)$$
$$= \int_0^{\cdot} \frac{\delta - 1}{2\,\widehat{X}(t)} \cdot \mathbf{1}_{\{\widehat{X}(t)\neq 0\}}\mathrm{d}t + \widehat{\boldsymbol{\beta}}(\cdot)\,, \tag{4.6}$$

with $Z(\cdot) = \mathrm{sgn}(\widehat{X}(\cdot))$ and with $\widehat{\boldsymbol{\beta}}(\cdot) := \int_0^{\cdot} Z(t)\mathrm{d}B(t)$ another standard Brownian motion on an extended probability space, as a consequence of Theorem 4.1 and of the properties in (4.5). We note that the semimartingale $\widehat{X}(\cdot)$ does not accumulate local time at the origin, because of $L^R(\cdot) \equiv 0$.

We claim that the process $\widehat{X}(\cdot)$ constructed here in (4.6) is the δ*-dimensional skew Bessel process with skewness parameter* α. This process was introduced and studied in [1].

Indeed, let us consider the functions $g(x) := |x|^{2-\delta}/(2-\delta)$ and $G(x) := \mathrm{sgn}(x) \cdot g(x)$ for $x \in \mathbb{R}$, and examine $g(\widehat{X}(\cdot))$ and $G(\widehat{X}(\cdot))$. This scaling is a right choice to measure the boundary behavior of $\widehat{X}(\cdot)$ around the origin. By substituting $q = 2 - \delta$, $p = (2-\delta)\,/\,(1-\delta)$, $\nu = -1\,/\,2$ in Proposition XI.1.11 of Revuz and Yor [16], we find that there exists a (nonnegative) one-dimensional Bessel process $\boldsymbol{\rho}(\cdot)$ on the same probability space, such that $\boldsymbol{\rho}(0) = (2-\delta)^{\delta-1} g(\widehat{X}(0))$ and

$$g(\widehat{X}(t)) = \frac{1}{2-\delta}\,\big|\widehat{X}(t)\big|^{2-\delta} = \frac{1}{(2-\delta)^{\delta-1}}\,\boldsymbol{\rho}\big(\Lambda(t)\big)\,, \quad 0 \le t < \infty\,,$$

where

$$\Lambda(t) := \inf\{s \ge 0 : K(s) \ge t\}\,, \quad K(s) := \int_0^s \big(\boldsymbol{\rho}(u)\big)^{\frac{2\delta-2}{2-\delta}}\mathrm{d}u\,;$$

that is, $g(\widehat{X}(\cdot))$ is a time-changed, conventionally reflected Brownian motion with the stochastic clock $\Lambda(\cdot)$. Thus the local time of $g(\widehat{X}(\cdot))$ accumulates at the origin with this clock $\Lambda(\cdot)$.

In the same manner as in the construction of $Z(\cdot)R(\cdot)$ in Theorem 4.1, we obtain here

$$G(\widehat{X}(T)) = \mathrm{sgn}(\widehat{X}(T))g(\widehat{X}(T)) = \int_0^T \mathrm{sgn}(\widehat{X}(t))\mathrm{d}\big(g(\widehat{X}(t))\big) + (2\alpha - 1)\,L^{g(\widehat{X})}(T)$$

as well as

$$L^{G(\widehat{X})}(\cdot) - L^{-G(\widehat{X})}(\cdot) = (2\alpha - 1)\big(L^{G(\widehat{X})}(\cdot) + L^{-G(\widehat{X})}(\cdot)\big) \tag{4.7}$$

and

$$(1-\alpha)\, L^{G(\widehat{X})}(\cdot) = \alpha\, L^{-G(\widehat{X})}(\cdot)\,, \quad L^{g(\widehat{X})}(\cdot) = \frac{1}{2}\big(L^{G(\widehat{X})}(\cdot) + L^{-G(\widehat{X})}(\cdot)\big)$$

in the notation of (1.5). From these relationships (4.7), and on the strength of [1, Theorem 2.22], we identify the process of (4.6) as the δ-dimensional skew Bessel process. Here the process $G(\widehat{X}(\cdot))$ and its local time $L^{G(\widehat{X})}(\cdot)$ correspond to $Y(\cdot)$ and $L_m^X(\cdot)$, respectively, in the notation of [1].

For various properties and representations of this process, we refer the study of [1], in particular, Remark 2.26 there.

5 An Application: Two Diffusive Particles with Asymmetric Collisions

In the paper [6], the authors construct a planar continuous semimartingale $\mathcal{X}(\cdot) = (X_1(\cdot), X_2(\cdot))$ with dynamics

$$\begin{aligned} \mathrm{d}X_1(t) = {} & \big(g\mathbf{1}_{\{X_1(t)\le X_2(t)\}} - h\mathbf{1}_{\{X_1(t)>X_2(t)\}}\big)\mathrm{d}t \\ & + \big(\rho\mathbf{1}_{\{X_1(t)>X_2(t)\}} + \sigma\mathbf{1}_{\{X_1(t)\le X_2(t)\}}\big)\mathrm{d}B_1(t)\,, \end{aligned} \tag{5.1}$$

$$\begin{aligned} \mathrm{d}X_2(t) = {} & \big(g\mathbf{1}_{\{X_1(t)> X_2(t)\}} - h\mathbf{1}_{\{X_1(t)\le X_2(t)\}}\big)\mathrm{d}t \\ & + \big(\rho\mathbf{1}_{\{X_1(t)\le X_2(t)\}} + \sigma\mathbf{1}_{\{X_1(t)> X_2(t)\}}\big)\mathrm{d}B_2(t)\,, \end{aligned} \tag{5.2}$$

for arbitrary real constants g, h and $\rho > 0$, $\sigma > 0$ with $\rho^2 + \sigma^2 = 1$. They show that, for an arbitrary initial condition $(X_1(0), X_2(0)) = (x_1, x_2) \in \mathbb{R}^2$ and with $(B_1(\cdot), B_2(\cdot))$ a planar Brownian motion, the system of (5.1), (5.2) has a pathwise unique, strong solution.

This is a model for two "competing" Brownian particles, with diffusive motions whose drift and dispersion characteristics are assigned according to the particle's ranks.

- In another recent paper [5], a planar continuous semimartingale $\widetilde{\mathcal{X}}(\cdot) = (\widetilde{X}_1(\cdot), \widetilde{X}_2(\cdot))$ is constructed according to the dynamics

$$\begin{aligned} \mathrm{d}\widetilde{X}_1(t) = {} & \big(g\mathbf{1}_{\{\widetilde{X}_1(t)\le \widetilde{X}_2(t)\}} - h\mathbf{1}_{\{\widetilde{X}_1(t)>\widetilde{X}_2(t)\}}\big)\mathrm{d}t \\ & + \big(\rho\mathbf{1}_{\{\widetilde{X}_1(t)>\widetilde{X}_2(t)\}} + \sigma\mathbf{1}_{\{\widetilde{X}_1(t)\le \widetilde{X}_2(t)\}}\big)\mathrm{d}\widetilde{B}_1(t) \\ & + \frac{1-\zeta_1}{2}\,\mathrm{d}L^{\widetilde{X}_1-\widetilde{X}_2}(t) + \frac{1-\eta_1}{2}\,\mathrm{d}L^{\widetilde{X}_2-\widetilde{X}_1}(t)\,, \end{aligned} \tag{5.3}$$

$$\begin{aligned}\mathrm{d}\widetilde{X}_2(t) = {} & \big(g\mathbf{1}_{\{\widetilde{X}_1(t)>\widetilde{X}_2(t)\}} - h\mathbf{1}_{\{\widetilde{X}_1(t)\le\widetilde{X}_2(t)\}}\big)\mathrm{d}t \\ & +\big(\rho\mathbf{1}_{\{\widetilde{X}_1(t)\le\widetilde{X}_2(t)\}} + \sigma\mathbf{1}_{\{\widetilde{X}_1(t)>\widetilde{X}_2(t)\}}\big)\mathrm{d}\widetilde{B}_2(t) \\ & + \frac{1-\zeta_2}{2}\,\mathrm{d}L^{\widetilde{X}_1-\widetilde{X}_2}(t) + \frac{1-\eta_2}{2}\,\mathrm{d}L^{\widetilde{X}_2-\widetilde{X}_1}(t)\,,\end{aligned} \tag{5.4}$$

Here again g, h are arbitrary real constants, $\rho > 0$ and $\sigma > 0$ satisfy $\rho^2 + \sigma^2 = 1$, whereas ζ_i, η_i are real constants satisfying

$$0 \le \alpha := \frac{\eta}{\eta+\zeta} \le 1\,, \quad \zeta := 1+\frac{\zeta_1-\zeta_2}{2}\,, \quad \eta := 1-\frac{\eta_1-\eta_2}{2}\,, \quad \zeta+\eta \ne 0\,.$$

This new system is a version of the previous competing Brownian particle system, but now with *elastic and asymmetric collisions* whose effect is modeled by the local time terms $L^{\widetilde{X}_2-\widetilde{X}_1}(\cdot)$ and $L^{\widetilde{X}_2-\widetilde{X}_1}(\cdot)$. Every time the two particles collide, their trajectories feel a "drag" proportional to these local time terms, whose presence makes the analysis of the system (5.3), (5.4) considerable more involved than that of (5.1), (5.2).

It is shown in [5] under the above conditions that, for an arbitrary initial condition $(\widetilde{X}_1(0), \widetilde{X}_2(0)) = (x_1, x_2) \in \mathbb{R}^2$, and with $(\widetilde{B}_1(\cdot), \widetilde{B}_2(\cdot))$ a planar Brownian motion, the system of (5.3), (5.4) has a pathwise unique, strong solution.

- We shall show how to use the unfolding of Theorem 1, in order to construct the planar process $\widetilde{\mathcal{X}}(\cdot) = (\widetilde{X}_1(\cdot), \widetilde{X}_2(\cdot))$ of (5.3), (5.4) with skew-elastic collisions, starting from the planar diffusion $\mathcal{X}(\cdot) = (X_1(\cdot), X_2(\cdot))$ of (5.1), (5.2). For simplicity, we shall take the initial condition $(x_1, x_2) = (0, 0)$ from now on.

Theorem 3 *Suppose we are given a planar continuous semimartingale* $\mathcal{X}(\cdot) = (X_1(\cdot), X_2(\cdot))$ *that satisfies the system of* (5.1), (5.2) *on some filtered probability space* $\big(\Omega, \mathcal{F}, \mathbb{P}\big)$, $\mathbb{F} = \{\mathcal{F}(t)\}_{0\le t<\infty}$ *with a planar Brownian motion* $(B_1(\cdot), B_2(\cdot))$.

There exists then an enlargement $\big(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{\mathbb{P}}\big)$, $\widetilde{\mathbb{F}} = \{\widetilde{\mathcal{F}}(t)\}_{0\le t<\infty}$ *of this filtered probability space, with a planar Brownian motion* $\big(\widetilde{B}_1(\cdot), \widetilde{B}_2(\cdot)\big)$*, and on it a planar continuous semimartingale* $\widetilde{\mathcal{X}}(\cdot) = (\widetilde{X}_1(\cdot), \widetilde{X}_2(\cdot))$ *that satisfies the system of (5.3), (5.4) with skew-elastic collisions, as well as*

$$\big(X_1(t) - X_2(t)\big) + \sup_{0\le s\le t} \big(X_1(s) - X_2(s)\big)^+ = \big|\widetilde{X}_1(t) - \widetilde{X}_2(t)\big|\,, \quad 0 \le t < \infty\,.$$

In other words, the size of the gap between the new processes $\widetilde{X}_1(\cdot)$, $\widetilde{X}_2(\cdot)$ coincides with the Skorokhod reflection of the difference $X_1(\cdot) - X_2(\cdot)$ of the original processes about the origin. We devote the remainder of this section to the proof of this result.

5.1 Reduction to Symmetric Local Times

First, some preparatory steps. We define the averages $\overline{\zeta} := (\zeta_1 + \zeta_2)\,/\,2\,,\ \ \overline{\eta} := (\eta_1 + \eta_2)\,/\,2\,$, and introduce yet another parameter

$$\beta := \alpha \cdot \frac{\zeta_1 + \zeta_2}{2} + (1-\alpha) \cdot \frac{\eta_1 + \eta_2}{2} = \alpha\,\overline{\zeta} + (1-\alpha)\,\overline{\eta}\,. \tag{5.5}$$

For notational simplicity we shall write all the processes related to the skew collisions with a tilde, e.g., $\widetilde{Y}(\cdot) := \widetilde{X}_1(\cdot) - \widetilde{X}_2(\cdot)\,$. From the relation between the *right* local time $L^{\widetilde{Y}}(\cdot)$ and the *symmetric local time* $\widehat{L}^{\widetilde{Y}}(\cdot)$ as in (1.5), we obtain the relations

$$\zeta L^{\widetilde{Y}}(\cdot) = \eta L^{-\widetilde{Y}}(\cdot), \quad L^{\widetilde{Y}}(\cdot) = 2\alpha \widehat{L}^{\widetilde{Y}}(\cdot), \quad L_{-}^{\widetilde{Y}}(\cdot) := L^{-\widetilde{Y}}(\cdot) = 2(1-\alpha)\widehat{L}^{\widetilde{Y}}(\cdot) \tag{5.6}$$

as in [5]. This way, the system (5.3), (5.4) can be re-cast as

$$\mathrm{d}\widetilde{X}_1(t) = \big(g\mathbf{1}_{\{\widetilde{X}_1(t)\le \widetilde{X}_2(t)\}} - h\mathbf{1}_{\{\widetilde{X}_1(t)> \widetilde{X}_2(t)\}}\big)\mathrm{d}t + \big(\rho\mathbf{1}_{\{\widetilde{X}_1(t)> \widetilde{X}_2(t)\}} + \sigma\mathbf{1}_{\{\widetilde{X}_1(t)\le \widetilde{X}_2(t)\}}\big)\mathrm{d}\widetilde{B}_1(t) + (2\alpha - \beta)\mathrm{d}\widehat{L}^{\widetilde{Y}}(t), \tag{5.7}$$

$$\begin{aligned}\mathrm{d}\widetilde{X}_2(t) = {} & \big(g\mathbf{1}_{\{\widetilde{X}_1(t)> \widetilde{X}_2(t)\}} - h\mathbf{1}_{\{\widetilde{X}_1(t)\le \widetilde{X}_2(t)\}}\big)\mathrm{d}t \\ & + \big(\rho\mathbf{1}_{\{\widetilde{X}_1(t)\le \widetilde{X}_2(t)\}} + \sigma\mathbf{1}_{\{\widetilde{X}_1(t)> \widetilde{X}_2(t)\}}\big)\mathrm{d}\widetilde{B}_2(t) \\ & + (2 - 2\alpha - \beta)\mathrm{d}\widehat{L}^{\widetilde{Y}}(t).\end{aligned} \tag{5.8}$$

We shall construct the system (5.7), (5.8) first, and then obtain from it the system (5.3), (5.4).

5.2 Proof of Theorem 3

By applying a Girsanov change of measure twice, we can remove the drifts from both of the systems (5.1), (5.2) and (5.7), (5.8). Then, in the following, let us construct the two-dimensional Brownian motion with rank-based dispersions and skew-elastic collisions

$$\mathrm{d}\widetilde{X}_1(t) = \big(\rho\mathbf{1}_{\{\widetilde{X}_1(t)> \widetilde{X}_2(t)\}} + \sigma\mathbf{1}_{\{\widetilde{X}_1(t)\le \widetilde{X}_2(t)\}}\big)\mathrm{d}\widetilde{B}_1(t) + (2\alpha - \beta)\,\mathrm{d}\widehat{L}^{\widetilde{Y}}(t)\,, \tag{5.9}$$

$$\mathrm{d}\widetilde{X}_2(t) = \big(\rho\mathbf{1}_{\{\widetilde{X}_1(t)\le \widetilde{X}_2(t)\}} + \sigma\mathbf{1}_{\{\widetilde{X}_1(t)> \widetilde{X}_2(t)\}}\big)\mathrm{d}\widetilde{B}_2(t) + (2 - 2\alpha - \beta)\,\mathrm{d}\widehat{L}^{\widetilde{Y}}(t)$$

from the solution $((X_1(\cdot), X_2(\cdot)), (B_1(\cdot), B_2(\cdot)))$ of the system

$$\begin{aligned} dX_1(t) &= \big(\rho \mathbf{1}_{\{X_1(t)>X_2(t)\}} + \sigma \mathbf{1}_{\{X_1(t)\le X_2(t)\}}\big)\, dB_1(t)\,, \\ dX_2(t) &= \big(\rho \mathbf{1}_{\{X_1(t)\le X_2(t)\}} + \sigma \mathbf{1}_{\{X_1(t)>X_2(t)\}}\big)\, dB_2(t)\,, \end{aligned} \tag{5.10}$$

which is known from [6] to be strongly solvable.

Since there is no drift in these last equations, the difference $Y(\cdot) := X_1(\cdot) - X_2(\cdot)$ between the two components of the system (5.10) is given by the real-valued Brownian motion

$$Y(\cdot) = W(\cdot) := \rho W_1(\cdot) + \sigma W_2(\cdot)\,. \tag{5.11}$$

Here

$$W_1(\cdot) := \int_0^{\cdot} \mathbf{1}_{\{X_1(t)>X_2(t)\}} dB_1(t) - \int_0^{\cdot} \mathbf{1}_{\{X_1(t)\le X_2(t)\}} dB_2(t)\,,$$

$$W_2(t) := \int_0^{\cdot} \mathbf{1}_{\{X_1(t)\le X_2(t)\}} dB_1(t) - \int_0^{\cdot} \mathbf{1}_{\{X_1(t)>X_2(t)\}} dB_2(t)$$

are independent Brownian motions. As in [6], let us recall also the Brownian motion

$$V(\cdot) := \rho V_1(\cdot) + \sigma V_2(\cdot)\,,$$

where again

$$V_1(\cdot) := \int_0^{\cdot} \mathbf{1}_{\{X_1(t)>X_2(t)\}} dB_1(t) + \int_0^{\cdot} \mathbf{1}_{\{X_1(t)\le X_2(t)\}} dB_2(t)\,,$$

$$V_2(\cdot) := \int_0^{\cdot} \mathbf{1}_{\{X_1(t)\le X_2(t)\}} dB_1(t) + \int_0^{\cdot} \mathbf{1}_{\{X_1(t)>X_2(t)\}} dB_2(t)$$

are independent Brownian motions. For a given number $\alpha \in (0, 1)$, there exists by Theorem 1 an adapted, continuous process $\widetilde{Y}(\cdot)$ which satisfies

$$Y(t) + \sup_{0\le s\le t} (-Y(s))^+ = \big|\widetilde{Y}(t)\big|\,, \quad 0 \le t < \infty$$

as well as

$$\widetilde{Y}(\cdot) = \int_0^{\cdot} \overline{\mathrm{sgn}}\big(\widetilde{Y}(t)\big)\mathrm{d}Y(t) + \frac{2\alpha - 1}{\alpha} L^{\widetilde{Y}}(\cdot)$$
$$= \int_0^{\cdot} \mathrm{sgn}\big(\widetilde{Y}(t)\big)\mathrm{d}W(t) + 2(2\alpha - 1)\widehat{L}^{\widetilde{Y}}(\cdot), \tag{5.12}$$

where the last equality follows from Proposition 1 and (5.6). Thus, the "unfolded process" $\widetilde{Y}(\cdot)$ is a skew Brownian motion, with skewness parameter α.

Now let us define the new planar Brownian motion $\big(\widetilde{B}_1(\cdot), \widetilde{B}_2(\cdot)\big)$ as

$$\mathrm{d}\widetilde{B}_1(\cdot) := \big(\mathbf{1}_{\{Y(\cdot)>0, \widetilde{Y}(\cdot)>0\}} - \mathbf{1}_{\{Y(\cdot)\le 0, \widetilde{Y}(\cdot)\le 0\}}\big)\mathrm{d}B_1(\cdot)$$
$$+\big(\mathbf{1}_{\{Y(\cdot)>0, \widetilde{Y}(\cdot)\le 0\}} - \mathbf{1}_{\{Y(\cdot)\le 0, \widetilde{Y}(\cdot)>0\}}\big)\mathrm{d}B_2(\cdot)\,,$$

$$\mathrm{d}\widetilde{B}_2(\cdot) := \big(\mathbf{1}_{\{Y(\cdot)>0, \widetilde{Y}(\cdot)\le 0\}} - \mathbf{1}_{\{Y(\cdot)\le 0, \widetilde{Y}(\cdot)>0\}}\big)\mathrm{d}B_1(\cdot)$$
$$+\big(\mathbf{1}_{\{Y(\cdot)>0, \widetilde{Y}(\cdot)>0\}} - \mathbf{1}_{\{Y(\cdot)\le 0, \widetilde{Y}(\cdot)\le 0\}}\big)\mathrm{d}B_2(\cdot)\,,$$

and, with the number $\beta \in \mathbb{R}$ as in (5.5), the processes $\widetilde{\Xi}(\cdot)$, $\big(\widetilde{X}_1(\cdot), \widetilde{X}_2(\cdot)\big)$ and $\big(\widetilde{V}(\cdot), \widetilde{W}(\cdot)\big)$ by

$$\widetilde{\Xi}(\cdot) := \widetilde{V}(\cdot) + 2(1-\beta)\widehat{L}^{\widetilde{Y}}(\cdot), \quad \widetilde{X}_1(\cdot) := \frac{\widetilde{\Xi}(\cdot) + \widetilde{Y}(\cdot)}{2}, \quad \widetilde{X}_2(\cdot) := \frac{\widetilde{\Xi}(\cdot) - \widetilde{Y}(\cdot)}{2}, \tag{5.13}$$

$$\mathrm{d}\widetilde{V}(\cdot) := \big(\rho\mathbf{1}_{\{\widetilde{Y}(\cdot)>0\}} + \sigma\mathbf{1}_{\{\widetilde{Y}(\cdot)\le 0\}}\big)\mathrm{d}\widetilde{B}_1(\cdot) + \big(\rho\mathbf{1}_{\{\widetilde{Y}(\cdot)\le 0\}} + \sigma\mathbf{1}_{\{\widetilde{Y}(\cdot)>0\}}\big)\mathrm{d}\widetilde{B}_2(\cdot)\,,$$
$$\mathrm{d}\widetilde{W}(\cdot) := \big(\rho\mathbf{1}_{\{\widetilde{Y}(\cdot)>0\}} + \sigma\mathbf{1}_{\{\widetilde{Y}(\cdot)\le 0\}}\big)\mathrm{d}\widetilde{B}_1(\cdot) - \big(\rho\mathbf{1}_{\{\widetilde{Y}(\cdot)\le 0\}} + \sigma\mathbf{1}_{\{\widetilde{Y}(\cdot)>0\}}\big)\mathrm{d}\widetilde{B}_2(\cdot)\,.$$

Then by (5.11) and (5.13) we obtain

$$\mathrm{sgn}(\widetilde{Y}(\cdot))\mathrm{d}W(\cdot) = \mathrm{sgn}(\widetilde{Y}(\cdot))\Big[\big(\rho\mathbf{1}_{\{X_1(\cdot)>X_2(\cdot)\}} + \sigma\mathbf{1}_{\{X_1(\cdot)\le X_2(\cdot)\}}\big)\mathrm{d}B_1(\cdot)$$
$$-\ \big(\rho\mathbf{1}_{\{X_1(\cdot)\le X_2(\cdot)\}} + \sigma\mathbf{1}_{\{X_1(\cdot)>X_2(\cdot)\}}\big)\mathrm{d}B_2(\cdot)\Big]$$
$$= \mathrm{sgn}(\widetilde{Y}(\cdot))\Big[\big(\rho\mathbf{1}_{\{Y(\cdot)>0\}} + \sigma\mathbf{1}_{\{Y(\cdot)\le 0\}}\big)\mathrm{d}B_1(\cdot)$$
$$-\ \big(\rho\mathbf{1}_{\{Y(\cdot)\le 0\}} + \sigma\mathbf{1}_{\{Y(\cdot)>0\}}\big)\mathrm{d}B_2(\cdot)\Big],$$

$$\begin{aligned}
\mathrm{d}\widetilde{W}(\cdot) &= \big(\rho\mathbf{1}_{\{\widetilde{Y}(\cdot)>0\}} + \sigma\mathbf{1}_{\{\widetilde{Y}(\cdot)\le 0\}}\big)\mathrm{d}\widetilde{B}_1(\cdot) \\
&\qquad - \big(\rho\mathbf{1}_{\{\widetilde{Y}(\cdot)\le 0\}} + \sigma\mathbf{1}_{\{\widetilde{Y}(\cdot)>0\}}\big)\mathrm{d}\widetilde{B}_2(\cdot) \\
&= \mathbf{1}_{\{\widetilde{Y}(\cdot)>0\}}\big(\rho\mathrm{d}\widetilde{B}_1(\cdot) - \sigma\mathrm{d}\widetilde{B}_2(\cdot)\big) + \mathbf{1}_{\{\widetilde{Y}(\cdot)\le 0\}}\big(\sigma\mathrm{d}\widetilde{B}_1(\cdot) - \rho\mathrm{d}\widetilde{B}_2(\cdot)\big).
\end{aligned}$$

Because of the relationship between $(B_1(\cdot), B_2(\cdot))$ and $(\widetilde{B}_1(\cdot), \widetilde{B}_2(\cdot))$, it can be shown that

$$\mathrm{d}\widetilde{W}(\cdot) \;=\; \mathrm{sgn}\big(\widetilde{Y}(\cdot)\big)\,\mathrm{d}W(\cdot)\,. \tag{5.14}$$

In fact, these identities can be verified formally via the following table:

Signs of $(Y(\cdot), \widetilde{Y}(\cdot))$	$\mathrm{d}\widetilde{B}_1(\cdot)$	$\mathrm{d}\widetilde{B}_2(\cdot)$	$\mathrm{d}\widetilde{W}(\cdot) \;=\; \mathrm{sgn}(\widetilde{Y}(\cdot))\mathrm{d}W(\cdot)$
$(+,+)$	$\mathrm{d}B_1(\cdot)$	$\mathrm{d}B_2(\cdot)$	$\rho\,\mathrm{d}\widetilde{B}_1(\cdot) - \sigma\,\mathrm{d}\widetilde{B}_2(\cdot) \;=\; \rho\,\mathrm{d}B_1(\cdot) - \sigma\,\mathrm{d}B_2(\cdot)$
$(-,+)$	$-\mathrm{d}B_2(\cdot)$	$-\mathrm{d}B_1(\cdot)$	$\rho\,\mathrm{d}\widetilde{B}_1(\cdot) - \sigma\,\mathrm{d}\widetilde{B}_2(\cdot) \;=\; \sigma\,\mathrm{d}B_1(\cdot) - \rho\,\mathrm{d}B_2(\cdot)$
$(+,-)$	$\mathrm{d}B_2(\cdot)$	$\mathrm{d}B_1(\cdot)$	$\sigma\,\mathrm{d}\widetilde{B}_1(\cdot) - \rho\,\mathrm{d}\widetilde{B}_2(\cdot) \;=\; -\rho\,\mathrm{d}B_1(\cdot) + \sigma\,\mathrm{d}B_2(\cdot)$
$(-,-)$	$-\mathrm{d}B_1(\cdot)$	$-\mathrm{d}B_2(\cdot)$	$\sigma\,\mathrm{d}\widetilde{B}_1(\cdot) - \rho\,\mathrm{d}\widetilde{B}_2(\cdot) \;=\; -\sigma\,\mathrm{d}B_1(\cdot) + \rho\,\mathrm{d}B_2(\cdot)$

Substituting this relation (5.14) into (5.12) and recalling (5.13), we obtain

$$\mathrm{d}\big(\widetilde{X}_1(t) - \widetilde{X}_2(t)\big) \;=\; \mathrm{d}\widetilde{Y}(t) \;=\; \mathrm{d}\widetilde{W}(t) + 2(2\alpha - 1)\,\mathrm{d}\widehat{L}^{\widetilde{Y}}(t)\,. \tag{5.15}$$

Moreover, because of the correspondence between $(\widetilde{V}(\cdot), \widetilde{W}(\cdot))$ and $(V(\cdot), W(\cdot))$ and the relation (5.13), we obtain

$$\frac{1}{2}\,\mathrm{d}\big(\widetilde{V}(t) + \widetilde{W}(t)\big) \;=\; \big(\rho\mathbf{1}_{\{\widetilde{Y}(t)>0\}} + \sigma\mathbf{1}_{\{\widetilde{Y}(t)\le 0\}}\big)\,\mathrm{d}\widetilde{B}_1(t)\,, \tag{5.16}$$

$$\frac{1}{2}\,\mathrm{d}\big(\widetilde{V}(t) - \widetilde{W}(t)\big) \;=\; \big(\sigma\mathbf{1}_{\{\widetilde{Y}(t)>0\}} + \rho\mathbf{1}_{\{\widetilde{Y}(t)\le 0\}}\big)\,\mathrm{d}\widetilde{B}_2(t)\,.$$

Therefore, by calculating the coefficients in front of the local time terms and by combining (5.13), (5.15) and (5.16), we can verify that $(\widetilde{X}_1(\cdot), \widetilde{X}_2(\cdot))$ satisfies

$$\begin{aligned}
\mathrm{d}\widetilde{X}_1(t) &= \big(\rho\mathbf{1}_{\{\widetilde{X}_1(t)>\widetilde{X}_2(t)\}} + \sigma\mathbf{1}_{\{\widetilde{X}_1(t)\le\widetilde{X}_2(t)\}}\big)\mathrm{d}\widetilde{B}_1(t) + (2\alpha - \beta)\mathrm{d}\widehat{L}^{\widetilde{Y}}(t)\,, \\
\mathrm{d}\widetilde{X}_2(t) &= \big(\rho\mathbf{1}_{\{\widetilde{X}_1(t)\le\widetilde{X}_2(t)\}} + \sigma\mathbf{1}_{\{\widetilde{X}_1(t)>\widetilde{X}_2(t)\}}\big)\mathrm{d}\widetilde{B}_2(t) + (2 - 2\alpha - \beta)\mathrm{d}\widehat{L}^{\widetilde{Y}}(t)
\end{aligned} \tag{5.17}$$

that is, (5.9) with the new Brownian motion $(\widetilde{B}_1(\cdot), \widetilde{B}_2(\cdot))$.

By the Girsanov theorem, we obtain (5.7), (5.8); whereas the relationship (5.6) between the left local time $L^{-\widetilde{Y}}(\cdot)$ and the right local time $L^{\widetilde{Y}}(\cdot)$ allows us now to recover the dynamics of (5.3), (5.4) from those of (5.1), (5.2). □

6 Appendix: Proof of Proposition 3

Given a planar Brownian motion $(B_1(\cdot), B_2(\cdot))$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and real constants $\alpha \in (0, 1)$, $x_0 \in \mathbb{R}$, we shall construct a process $X(\cdot) := \mathfrak{q}(Y(\cdot))$ from the solution $Y(\cdot)$ of the stochastic differential equation

$$Y(\cdot) = \mathfrak{p}(x_0) + \int_0^{\cdot} \mathfrak{s}\big(Y(t)\big)\, \mathrm{d}\big(B_1(t) + B_2(t)\big)\,, \tag{6.1}$$

where $\mathfrak{p}(\cdot)$, $\mathfrak{q}(\cdot)$ and $\mathfrak{s}(\cdot)$ are defined by

$$\begin{aligned}
\mathfrak{p}(x) &:= (1-\alpha)\, x\, \mathbf{1}_{(0,\infty)}(x) + \alpha\, x\, \mathbf{1}_{(-\infty,0]}(x)\,, \\
\mathfrak{q}(x) &:= \frac{1}{1-\alpha}\, \mathbf{1}_{(0,\infty)}(x) + \frac{1}{\alpha}\, \mathbf{1}_{(-\infty,0)}(x)\,, \\
\mathfrak{s}(x) &:= (1-\alpha)\, \mathbf{1}_{(0,\infty)}(x) + \alpha\, \mathbf{1}_{(-\infty,0]}(x)\,; \quad x \in \mathbb{R}\,.
\end{aligned}$$

From the work of Nakao in [11] we know that the Eq. (6.1) has a pathwise unique, strong solution.

Since $\mathfrak{q}(\mathfrak{p}(x)) = x$, $x \in \mathbb{R}$, by applying the Itô-Tanaka formula to the process $X(\cdot) = \mathfrak{q}(Y(\cdot))$ we identify the dynamics of $X(\cdot)$ as those of the skew Brownian motion [7], namely

$$X(\cdot) = x_0 + \big(B_1(\cdot) + B_2(\cdot)\big) + \frac{2\alpha - 1}{\alpha} L^X(\cdot)\,, \tag{6.2}$$

driven by the Brownian motion $B_1(\cdot) + B_2(\cdot)$. We rewrite this equation in the form

$$X(\cdot) - x_0 - \int_0^{\cdot} \operatorname{sgn}\big(X(t)\big) \mathrm{d}U(t) - V(\cdot) \;=\; \frac{2\alpha - 1}{\alpha} L^X(\cdot) \;=\; 2\big(2\alpha - 1\big)\widehat{L}^X(\cdot)$$

of (2.16), driven by a new planar Brownian motion $(U(\cdot), V(\cdot))$ with components

$$U(\cdot) \;:=\; \int_0^{\cdot} \mathbf{1}_{\{X(t)>0\}} \mathrm{d}B_1(t) - \int_0^{\cdot} \mathbf{1}_{\{X(t)\le 0\}} \mathrm{d}B_2(t)\,, \tag{6.3}$$

$$V(\cdot) \;:=\; \int_0^{\cdot} \mathbf{1}_{\{X(t)\le 0\}} \mathrm{d}B_1(t) + \int_0^{t} \mathbf{1}_{\{X(t)>0\}} \mathrm{d}B_2(t)\,. \tag{6.4}$$

Therefore, the perturbed skew Tanaka equation (2.16) has the weak solution $(X(\cdot), (U(\cdot), V(\cdot)))$ just constructed.

Conversely, suppose we start with an arbitrary weak solution $(X(\cdot), (U(\cdot), V(\cdot)))$ of the equation (2.16), with $(U(\cdot), V(\cdot))$ a planar Brownian motion. Then we can cast this equation in the form (6.2) in terms of the planar Brownian motion $(B_1(\cdot), B_2(\cdot))$ whose components are given by "disentangling" in (6.3), (6.4), namely

$$B_1(\cdot) = \int_0^{\cdot} \mathbf{1}_{\{X(t)>0\}} \mathrm{d}U(t) + \int_0^{\cdot} \mathbf{1}_{\{X(t)\le 0\}} \mathrm{d}V(t)\,,$$

$$B_2(\cdot) = \int_0^{\cdot} \mathbf{1}_{\{X(t)>0\}} \mathrm{d}V(t) - \int_0^{\cdot} \mathbf{1}_{\{X(t)\le 0\}} \mathrm{d}U(t)\,.$$

But this shows that $X(\cdot)$ is skew Brownian motion, so its probability distribution is determined uniquely.

In other words, the equation of (2.16) admits a weak solution, and this solution is unique in the sense of the probability distribution.

• *Now we shall see that we have not just uniqueness in distribution, but also pathwise uniqueness, for the equation* (2.16) *driven by the planar Brownian motion* $(U(\cdot), V(\cdot))$. The argument that follows is based on [10, Lemma 1], and is almost identical to the proof of Theorem 8.1 of [6] except for the evaluation of the additional local times. Note that Le Gall's Lemma 1 in [10] works for general continuous semimartingales.

Suppose that there are two solutions $X_1(\cdot)$ and $X_2(\cdot)$ of (2.16), defined on the same probability space as the driving planar Brownian motion $(U(\cdot), V(\cdot))$. We shall check that their difference $D(\cdot) := X_1(\cdot) - X_2(\cdot)$ satisfies (c.f. (6.4) in [6]) :

$$\mathbb{E}\Big[\int_0^T \frac{\mathrm{d}\langle D\rangle(s)}{D(s)} \mathbf{1}_{\{D(s)>0\}} \Big] < \infty\,, \quad 0 < T < \infty\,, \tag{6.5}$$

where

$$\langle D\rangle(\cdot) = \int_0^{\cdot} \big(\mathrm{sgn}(X_1(t)) - \mathrm{sgn}(X_2(t))\big)^2 \mathrm{d}t \le 2 \int_0^{\cdot} \big|\mathrm{sgn}(X_1(t)) - \mathrm{sgn}(X_2(t))\big|\, \mathrm{d}t\,.$$

We approximate the signum function by a sequence $\{f_k\}_{k\in\mathbb{N}} \subset C^1(\mathbb{R})$ which converges to the function $f_\infty(\cdot) = \mathrm{sgn}(\cdot)$ pointwise and satisfies $\lim_{k\to\infty}\| f_k \|_{TV} = \| f_\infty \|_{TV}$. Now the parametrized process

$$Z^{(u)}(t) := (1-u)X_1(t) + uX_2(t)\,, \quad 0 \le u \le 1\,, \quad 0 \le t < \infty$$

takes the form of

$$Z^{(u)}(\cdot) = x_0 + \int_0^{\cdot} \big((1-u)\mathrm{sgn}(X_1(t)) + u\,\mathrm{sgn}(X_2(t))\big)\,\mathrm{d}U(t) \\ + V(\cdot) + \frac{2\alpha - 1}{\alpha}\big(uL^{X_1}(\cdot) + (1-u)L^{X_2}(\cdot)\big)\,.$$

The local times in the last term do not affect the size of $\langle Z^{(u)}\rangle(\cdot)$, for which we have the estimate $\mathbb{E}\big(\langle Z^{(u)}\rangle(T)\big) \le 2T$. Proceeding as in [6] we obtain for every $\delta > 0$ the bound

$$\mathbb{E}\Big[\int_0^T \frac{|f_k(X_1(s)) - f_k(X_2(s))|}{X_1(s) - X_2(s)}\mathbf{1}_{\{X_1(s)-X_2(s)>\delta\}}\mathrm{d}t\Big] \\ \le c\|f_k\|_{TV} \cdot \sup_{a,u}\mathbb{E}\big(2L^{(u)}(T,a)\big),$$

where $L^{(u)}(T,a)$ is the right local time of the continuous semimartingale $Z^{(u)}(\cdot)$ accumulated at $a \in \mathbb{R}$ and c is a constant chosen independently of k, u, δ. Letting $k \uparrow \infty$ and $\delta \downarrow 0$, we estimate

$$\mathbb{E}\Big[\int_0^T \frac{\mathrm{d}\langle D\rangle(s)}{D(s)}\mathbf{1}_{\{D(s)>0\}}\Big] < 2\,c\,\|f_\infty\|_{TV} \cdot \sup_{a,u}\mathbb{E}\big(2L^{(u)}(T,a)\big)\,.$$

Finally, we estimate $\mathbb{E}(L^{(u)}(T,a))$ using Tanaka's formula

$$|Z^{(u)}(T) - a| = |Z^{(u)}(0) - a| + \int_0^T \mathrm{sgn}\big(Z^{(u)}(t) - a\big)\mathrm{d}Z^{(u)}(t) + 2L^{(u)}(T,a)\,,$$

and a combination of the Cauchy-Schwartz inequality and the Itô's isometry:

$$\begin{aligned}\mathbb{E}\big(2L^{(u)}(T,a)\big) &\le \mathbb{E}|Z^{(u)}(T) - Z^{(u)}(0)| + \big\{\mathbb{E}(\langle Z^{(u)}\rangle(T))\big\}^{1/2} \\ &\quad + \frac{2\alpha-1}{\alpha}\big(u\,\mathbb{E}(L^{X_1}(T)) + (1-u)\,\mathbb{E}(L^{X_2}(T))\big) \\ &\le 2\Big[\big\{\mathbb{E}\big(\langle Z^{(u)}\rangle(T)\big)\big\}^{1/2} \\ &\quad + \frac{2\alpha-1}{\alpha}\big(u\,\mathbb{E}(L^{X_1}(T)) + (1-u)\,\mathbb{E}(L^{X_2}(T))\big)\Big].\end{aligned}$$

The last term $\mathbb{E}(L^{X_i}(T))$ is evaluated by the same procedure: by Tanaka's formula

$$\frac{1}{\alpha} L^{X_i}(T) = |X_i(T)| - |X_1(0)| - \int_0^T \operatorname{sgn}\big(X_i(t)\big)\, \mathrm{d}V(t) - U(T)\,,$$

and hence

$$\mathbb{E}\big(L^{X_i}(T)\big) \le 2\alpha\big\{\mathbb{E}(\langle X_i\rangle(T))\big\}^{1/2} \le 2^{3/2}\,\alpha\, T^{1/2}\,, \quad i = 1, 2\,.$$

Therefore, we obtain (6.5), and by [10, Lemma 1] we verify $L^D(\cdot) = L^{X_1 - X_2}(\cdot) \equiv 0$.

- *Final step*: By exchanging the rôles of $X_1(\cdot)$ and $X_2(\cdot)$, we obtain $L^{-D}(\cdot) = L^{X_2 - X_1}(\cdot) \equiv 0$ as well as $\widehat{L}^D(\cdot) \equiv 0$. Furthermore, by [13, Corollary 2.6], we obtain

$$\widehat{L}^{X_1 \vee X_2}(t) = \int_0^t \mathbf{1}_{\{X_2(s) \le 0\}}\, \mathrm{d}\widehat{L}^{X_1}(s) + \int_0^t \mathbf{1}_{\{X_1(s) < 0\}}\, \mathrm{d}\widehat{L}^{X_2}(s)\,; \quad 0 \le t < \infty\,.$$

Combining these results with Tanaka's formula, we obtain the dynamics of $M(\cdot) := X_1(\cdot) \vee X_2(\cdot)$:

$$\begin{aligned}
\mathrm{d}M(t) &= \mathbf{1}_{\{X_1(t) \ge X_2(t)\}} \mathrm{d}X_1(t) + \mathbf{1}_{\{X_1(t) < X_2(t)\}} \mathrm{d}X_2(t) + \mathrm{d}L^{X_1 - X_2}(t) \\
&= \mathbf{1}_{\{X_1(t) \ge X_2(t)\}} \Big(\operatorname{sgn}(X_1(t))\mathrm{d}U(t) + \mathrm{d}V(t) + 2(2\alpha - 1)\mathrm{d}\widehat{L}^{X_1}(t)\Big) \\
&\quad + \mathbf{1}_{\{X_1(t) < X_2(t)\}} \Big(\operatorname{sgn}(X_2(t))\mathrm{d}U(t) + \mathrm{d}V(t) + 2(2\alpha - 1)\mathrm{d}\widehat{L}^{X_2}(t)\Big) \\
&= \operatorname{sgn}(M(t))\mathrm{d}U(t) + \mathrm{d}V(t) + 2(2\alpha - 1)\mathrm{d}\widehat{L}^M(t)\,; \quad 0 \le t < \infty\,.
\end{aligned}$$

In other words, each of the continuous semimartingales $X_1(\cdot)$, $X_2(\cdot)$ and $M(\cdot) = X_1(\cdot) \vee X_2(\cdot)$ satisfies the equation (2.16); but uniqueness in the sense of the probability distribution holds for this equation, so all three processes have the same distribution. Since $M(\cdot) \ge X_i(\cdot)$, this forces $M(\cdot) = X_i(\cdot)$, $i = 1, 2$, thus pathwise uniqueness. By the theory of Yamada and Watanabe (e.g., [9, Sect. 5.3.D]), the solution to (2.16) is therefore strong. The proof of Proposition 3 is complete. □

Acknowledgments We are indebted to Marcel Nutz and Daniel Ocone for stimulating discussions, to Vilmos Prokaj and Johannes Ruf for their careful reading of the manuscript and for their suggestions, and to the referee for simplifying the last part of our argument in Example 2. We dedicate this paper to Terry Lyons on the occasion of his 60th birthday.

References

1. S. Blei, On symmetric and skew Bessel processes. Stoch. Process. Appl. **122**(9), 3262–3287 (2012)
2. M. Chaleyat-Maurel, M. Yor, Les filtrations de $|X|$ et X^+, lorsque X est une semi-martingale continue. In “Temps Locaux”. Astérisque **52–53**, 193–196 (1978)
3. A.S. Cherny, On the strong and weak solutions of stochastic differential equations governing Bessel processes. Stoch. Stoch. Rep. **70**(3–4), 213–219 (2000)
4. L.E. Dubins, M. Émery, M. Yor, On the Lévy transformation of Brownian motions and continuous martingales, in *Séminaire de Probabilités XXVII*. Lecture Notes in Mathematics 1577 (Springer, New York, 1993), pp. 122–132
5. E.R. Fernholz, T. Ichiba, I. Karatzas, Two Brownian particles with rank-based characteristics and skew-elastic collisions. Stoch. Process. Appl. **123**, 2999–3026 (2013)
6. E.R. Fernholz, T. Ichiba, I. Karatzas, V. Prokaj, A planar diffusion with rank-based characteristics, and perturbed Tanaka equations. Prob. Theory Relat. F. **156**, 343–374 (2013)
7. J.M. Harrison, L.A. Shepp, On skew Brownian motion. Ann. Prob. **9**, 309–313 (1981)
8. T. Ichiba, I. Karatzas, V. Prokaj, Diffusions with rank-based characteristics and values in the nonnegative orthant. Bernoulli **19**, 2455–2493 (2013)
9. I. Karatzas, S.E. Shreve, *Brownian Motion and Stochastic Calculus*, 2nd edn. (Springer, New York, 1991)
10. J.-F. Le Gall, Applications du temps local aux équations diffé rentielles stochastiques unidimensionnelles, *Séminaire de Probabilités XVII*, Lecture Notes in Mathematics 986 (Springer, New York, 1983), pp. 15–31
11. S. Nakao, On pathwise uniqueness of solutions of stochastic differential equations. Osaka J. Math. **9**, 513–518 (1972)
12. D. Ocone, A symmetry characterization of conditionally independent increment martingales. in *Proceedings of the San Felice Workshop on Stochastic Analysis*, ed by D. Nualart, M. Sanz (Birkhäuser-Verlag, Basel and Boston, 1993), pp. 147–167
13. Y. Ouknine, M. Rutkowski, Local times of functions of continuous semimartingales. Stoch. Analy. Appl. **13**, 211–231 (1995)
14. V. Prokaj, Unfolding the Skorokhod reflection of a semimartingale. Stat. Prob. Lett. **79**, 534–536 (2009)
15. V. Prokaj, The solution of the perturbed Tanaka equation is pathwise unique. Ann. Prob. **41**, 2376–2400 (2013)
16. D. Revuz, M. Yor, *Continuous Martingales and Brownian Motion*, 3rd edn. (Springer, New York, 1999)
17. L. Vostrikova, M. Yor, Some invariance properties of the laws of Ocone martingales, *Séminaire de Probabilités XXXIV*, Lecture Notes in Mathematics 1729 (Springer, New York, 2000), pp. 417–431

Normal Approximation on a Finite Wiener Chaos

David Nualart

Abstract The purpose of this note is to survey some recent developments in the applications of Malliavin calculus combined with Stein's method to derive central limit theorems for random variables on a finite sum of Wiener chaos. Starting from the fourth moment theorem by Nualart and Peccati [23], we will discuss several related topics such as conditions for the convergence in total variation, absolute continuity of probability laws and uniform convergence of densities under suitable non degeneracy assumptions. The fact that the random variables belong to a fixed Wiener chaos (or to a finite sum of Wiener chaos) will play a fundamental role in the results.

Keywords Malliavin calculus · Wiener chaos · Total variation distance · Stein's method

2010 Mathematics Subject Classification 60G15 · 60H07 · 60H10 · 65C30

1 Introduction

The Malliavin calculus is a stochastic calculus of variations in a Gaussian space, developed from the probabilistic proof of Hörmander's hypoellipticity theorem by Malliavin [11]. The main application of this calculus is to establish the regularity of the probability distribution of functionals of an underlying Gaussian process. Basic references for the Malliavin calculus and its applications are the monographs by Malliavin [12] and Nualart [21]. Recently, the integration by parts formula of Malliavin calculus, combined with Stein's method for normal approximations, has proved to be a powerful approach to derive quantitative versions of central limit

---

This work is supported by the NSF grant DMS1208625.

D. Nualart (✉)
Department of Mathematics, University of Kansas, 405 Snow Hall, Lawrence, KS, USA
e-mail: nualart@ku.edu

D. Crisan et al. (eds.), *Stochastic Analysis and Applications 2014*,
Springer Proceedings in Mathematics & Statistics 100,
DOI 10.1007/978-3-319-11292-3_14

theorems, in the case of random variables belonging to a fixed Wiener chaos. A basic reference for this methodology is the monograph by Nourdin and Peccati [17].

The aim of this note is to survey some recent developments in this field. After recalling some basic facts on the Malliavin calculus and the application of the Stein's method to derive the fourth moment theorem, we discuss some advances in this field. More precisely, for random variables in a fixed sum of Wiener chaos, we analyze under which conditions the convergence in law implies the convergence in total variation, we discuss criteria for absolute continuity, and we present some recent results on the the uniform convergence of densities.

2 Elements of Malliavin Calculus

2.1 Gaussian Analysis

Suppose that H is a real separable Hilbert space. Let $X = \{X(h), h \in H\}$ be an *isonormal Gaussian process* over H. This means that X is a centered Gaussian family of random variables on some probability space $(\Omega, \mathcal{F}, P)$ with a covariance given by

$$E(X(h)X(g)) = \langle h, g \rangle_H, \quad h, g \in H.$$

We assume that $\mathcal{F}$ is the P-completion of the σ-field generated by X. For every integer $q \geq 1$, we let $\mathcal{H}_q$ be the qth *Wiener chaos* of X defined as the closed linear subspace of $L^2(\Omega)$ generated by the random variables $\{H_q(X(h)), h \in H, \|h\|_H = 1\}$, where H_q is the qth Hermite polynomial

$$H_q(x) = (-1)^q e^{x^2/2} \frac{d^q}{dx^q} \left(e^{-x^2/2} \right).$$

We also denote by $H^{\otimes q}$ and $H^{\odot q}$, respectively, the qth tensor product and the qth symmetric tensor product of H. For any $q \geq 1$, the mapping $I_q(h^{\otimes q}) = q! H_q(X(h))$ provides a linear isometry between $H^{\odot q}$ (equipped with the modified norm $\sqrt{q!}\|\cdot\|_{H^{\otimes q}}$) and $\mathcal{H}_q$ (equipped with the $L^2(\Omega)$ norm). For $q = 0$. we set $\mathcal{H}_0 = \mathbb{R}$, and I_0 is the identity map.

Any square integrable random variable $F \in L^2(\Omega)$ can be decomposed into an orthogonal sum

$$F = E(F) + \sum_{q=1}^{\infty} I_q(f_q), \tag{2.1}$$

where the $f_q \in H^{\odot q}$ are uniquely determined by F. This is called the *Wiener chaos expansion*.

For example, let $B = \{B_t, t \in [0,1]\}$ be a Brownian motion. Then, if we take $H = L^2([0,1])$ and define $X(h) = \int_0^1 h_t dB_t$, the family $\{X(h), h \in H\}$ is an isonormal Gaussian process. In this case, for any $q \geq 2$, $H^{\odot q} = L^2_{sym}([0,1]^q)$ is the space of symmetric and square integrable functions and I_q is the iterated Itô stochastic integral:

$$I_q(h) = q! \int_0^1 \dots \int_0^{t_2} h(t_1, \dots, t_q) dB_{t_1}, \dots, dB_{t_q}, \quad h \in L^2_{sym}([0,1]^q).$$

Let $\{e_i, i \geq 1\}$ be a complete orthonormal system in H. Given $f \in H^{\odot p}$ and $g \in H^{\odot q}$, for every $r = 0, \dots, p \wedge q$, the *contraction* of f and g of order r is the element of $H^{\otimes(p+q-2r)}$ defined by

$$f \otimes_r g = \sum_{i_1,\dots,i_r=1}^{\infty} \langle f, e_{i_1} \otimes \cdots \otimes e_{i_r} \rangle_{\mathfrak{H}^{\otimes r}} \otimes \langle g, e_{i_1} \otimes \cdots \otimes e_{i_r} \rangle_{\mathfrak{H}^{\otimes r}}.$$

Note that, $f \otimes_0 g = f \otimes g$ equals the tensor product of f and g while, for $p = q$, $f \otimes_p g = \langle f, g \rangle_{H^{\otimes p}}$. The contraction $f \otimes_r g$ is not necessarily symmetric, and we denote by $f \widetilde{\otimes}_r g$ its symmetrization.

2.2 Malliavin Calculus

Let us now introduce some elements of the Malliavin calculus of variations with respect to the isonormal Gaussian process X. Let $\mathcal{S}$ be the set of all smooth and cylindrical random variables of the form

$$F = f\left(X(h_1), \dots, X(h_n)\right), \tag{2.2}$$

where $n \geq 1$, $f : \mathbb{R}^n \to \mathbb{R}$ is an infinitely differentiable function which is bounded together with all its partial derivatives, and $h_i \in H$. The *Malliavin derivative* of F is the element of $L^2(\Omega; H)$ defined as

$$DF = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} \left(X(h_1), \dots, X(h_n)\right) h_i.$$

By iteration, one can define the qth derivative $D^q F$ for every $q \geq 2$, which is an element of $L^2(\Omega; H^{\odot q})$. For any integer $q \geq 1$ and any real number $p \geq 1$, $\mathbb{D}^{q,p}$ denotes the closure of $\mathcal{S}$ with respect to the norm $\|\cdot\|_{\mathbb{D}^{q,p}}$, defined by the relation

$$\|F\|_{\mathbb{D}^{q,p}}^p = E\left[|F|^p\right] + \sum_{i=1}^q E\left(\|D^i F\|_{H^{\otimes i}}^p\right).$$

We denote by δ the adjoint of the operator D, also called the *divergence operator*. A random element $u \in L^2(\Omega; H)$ belongs to the domain of δ, denoted by Domδ, if and only if it verifies

$$\left|E\big(\langle DF, u\rangle_H\big)\right| \le c_u \sqrt{E(F^2)},$$

for any $F \in \mathbb{D}^{1,2}$, where c_u is a constant depending only on u. If $u \in$ Domδ, then the random variable $\delta(u)$ is defined by the duality relationship

$$E(F\delta(u)) = E\big(\langle DF, u\rangle_H\big), \tag{2.3}$$

which holds for every $F \in \mathbb{D}^{1,2}$. Formula (2.3) extends to the multiple Skorohod integral δ^q, and we have

$$E\left(F\delta^q(u)\right) = E\left(\left\langle D^q F, u\right\rangle_{H^{\otimes q}}\right), \tag{2.4}$$

for any element u in the domain of δ^q and any random variable $F \in \mathbb{D}^{q,2}$. Moreover, $\delta^q(h) = I_q(h)$ for any $h \in H^{\odot q}$.

We will make use of the following factorization property. For every $F \in \mathbb{D}^{1,2}$ and every $u \in$ domδ such that Fu and $F\delta(u) - \langle DF, u\rangle_H$ are square integrable, one has that $Fu \in$ domδ and

$$\delta(Fu) = F\delta(u) - \langle DF, u\rangle_H. \tag{2.5}$$

A random variable F with the Wiener chaos expansion given in (2.1) belongs to $\mathbb{D}^{1,2}$ if and only if $\sum_{q=1}^\infty qq!\|f_q\|_{H^{\otimes q}}^2 < \infty$, and, in this case, $E\left(\|DF\|_H^2\right) = \sum_{q=1}^\infty qq!\|f_q\|_{H^{\otimes q}}^2$.

The operator L is defined on a random variable F with the Wiener chaos expansion (2.1) as $LF = \sum_{q=1}^\infty (-q) I_q(f_q)$, and is called the *infinitesimal generator of the Ornstein-Uhlenbeck semigroup*. The domain of this operator in $L^2(\Omega)$ is the set

$$\text{Dom}L = \{F \in L^2(\Omega) : \sum_{q=1}^\infty q^2 q!^2 \left\|f_q\right\|_{H^{\otimes q}}^2 < \infty\} = \mathbb{D}^{2,2}.$$

There is an important relationship between the operators D, δ and L (see [21, Proposition 1.4.3]). A random variable F belongs to the domain of L if and only if $F \in$ Dom (δD) (i.e. $F \in \mathbb{D}^{1,2}$ and $DF \in$ Domδ), and in this case

$$\delta DF = -LF. \tag{2.6}$$

The operator L^{-1} defined by $L^{-1}F = \sum_{q=1}^{\infty}(-1/q)I_q(f_q)$ is the pseudo inverse of L.

The following integration by parts formula is the key ingredient in the applications of Malliavin calculus to normal approximation combined with Stein's method.

Theorem 2.1 *Let $F \in \mathbb{D}^{1,2}$ be such that $E[F] = 0$, and let $f : \mathbb{R} \to \mathbb{R}$ be a $\mathcal{C}^1$ function with a bounded derivative. Then*

$$E[Ff(F)] = E[f'(F)\langle DF, -DL^{-1}F\rangle_H]. \tag{2.7}$$

Proof Taking into account that $E[F] = 0$ and using (2.6), we obtain $F = LL^{-1}F = -\delta(DL^{-1}F)$. Then, the result follows from the duality relationship (2.3)

$$\begin{aligned} E[Ff(F)] &= -E[f(F)\delta(DL^{-1}F)] = E[\langle D(f(F)), -DL^{-1}F\rangle_H] \\ &= E[f'(F)\langle DF, -DL^{-1}F\rangle_H]. \end{aligned} \tag{2.8}$$

This completes the proof of the theorem. □

In the particular case where $F \in \mathcal{H}_q$, with $q \geq 1$, then $DL^{-1}F = -\frac{1}{q}DF$ and (2.8) yields

$$E[f(F)F] = \frac{1}{q}E[f'(F)\|DF\|_H^2]. \tag{2.9}$$

3 Stein's Method for Normal Approximation

We denote by ϕ the density of the standard normal distribution $N(0,1)$ on the real line:

$$\phi(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}.$$

From the fact that ϕ satisfies the differential equation $\phi'(x) = -x\phi(x)$, it follows that a real-valued random variable Z has the normal probability distribution $N(0,1)$, if and only if for every differentiable function $f : \mathbb{R} \to \mathbb{R}$ such that $xf(x)$ and $f'(x)$ are integrable with respect to $N(0,1)$,

$$E[Zf(Z)] = E[f'(Z)].$$

Given a general random variable F, if the expectation $E[Ff(F)] - E[f'(F)]$ is close to zero for a large class of smooth functions f, then we should be able to conclude that the law of F is close to $N(0,1)$ in some sense. This is the heuristics of Stein's method (see [27]). To make this argument rigorous, given a measurable function $h : \mathbb{R} \to \mathbb{R}$ such that $E[|h(Z)|] < \infty$, where Z has the distribution $N(0,1)$, we introduce the *Stein's equation*

$$f'(x) - xf(x) = h(x) - E[h(Z)]. \tag{3.1}$$

The function

$$f_h(x) = e^{x^2/2} \int_{-\infty}^{x} [h(y) - E[h(Z)]]e^{-y^2/2} dy \tag{3.2}$$

turns out to be the unique solution to Eq. (3.1) satisfying $\lim_{x\to\pm\infty} e^{-x^2/2} f(x) = 0$. Substituting x by F in Eq. (3.1) and taking the expectation yields

$$E[h(F)] - E[h(Z)] = E[f_h'(F) - F f_h(F)]. \tag{3.3}$$

One can show that if $\|h\|_\infty \le 1$, then $\|f_h\|_\infty \le \sqrt{\pi/2}$ and $\|f_h'\|_\infty \le 2$. So, taking $h = \mathbf{1}_B$, we obtain the following estimate for the total variation distance between the law of a random variable F and the standard normal distribution

$$\begin{aligned} d_{TV}(F, Z) &= \sup_{B\in\mathcal{B}(\mathbb{R})} |P(F \in B) - P(Z \in B)| \\ &\le \sup_{f\in\mathcal{C}_{TV}} |E[f'(F) - Ff(F)]|, \end{aligned} \tag{3.4}$$

where $\mathcal{C}_{TV}$ is the class of functions f, which are piece-wise differentiable and satisfy $\|f\|_\infty \le \sqrt{\pi/2}$ and $\|f'\|_\infty \le 2$.

4 Central Limit Theorem for Multiple Stochastic Integrals

Consider the context of an isonormal Gaussian process X over a Hilbert space H. Suppose that F is a random variable in a Wiener chaos $\mathcal{H}_q$ for some $q \ge 2$ and $E(F^2) = 1$. Then from (3.4) and (2.9) we can write

$$\begin{aligned} d_{TV}(F, Z) &\le \sup_{f\in\mathcal{C}_{TV}} |E[f'(F) - Ff(F)]| \\ &= \sup_{f\in\mathcal{C}_{TV}} \left| E\left[f'(F)\left(1 - \frac{1}{q}\|DF\|_H^2\right)\right]\right|. \end{aligned}$$

Taking into account that $E(\|DF\|_H^2) = q$ and $\|f'\|_\infty \le 2$ for any $f \in \mathcal{C}_{TV}$, we obtain

$$d_{TV}(F, Z) \le \frac{2}{q}\sqrt{\operatorname{Var}\left(\|DF\|_H^2\right)}.$$

Using (2.9) with $f(x) = x^3$ yields

$$E(F^4) = \frac{3}{q} E[F^2 \|DF\|_H^2]. \tag{4.1}$$

Then, applying (4.1) and the product formula together with orthogonality properties of multiple stochastic integrals, one can show that

$$\mathrm{Var}\left(\|DF\|_H^2\right) \leq \frac{(q-1)q}{3}(E(F^4) - 3) \leq (q-1)\mathrm{Var}\left(\|DF\|_H^2\right).$$

We refer to the monograph by Nourdin and Peccati [17] for a detailed account on the application of Stein's method combined with Malliavin calculus to normal approximations. This methodology leads to a simple and quantitative proof of the so called *Fourth moment theorem* (see Nualart and Peccati [23] and Nualart and Ortiz-Latorre [22]), which represents a drastic simplification of the method of moments and cumulants to prove convergence to the normal distribution.

Theorem 4.1 (Fourth Moment Theorem) *Fix $q \geq 2$. Let $F_n = I_q(f_n) \in \mathcal{H}_q$, $n \geq 1$ be a sequence of elements in the qth chaos, such that*

$$\lim_{n \to \infty} E(F_n^2) = 1.$$

The following conditions are equivalent:

(i) $F_n \Rightarrow N(0, 1)$*, as* $n \to \infty$.
(ii) $E(F_n^4) \to 3$*, as* $n \to \infty$.
(iii) *For all* $1 \leq r \leq q-1$, $f_n \otimes_r f_n \to 0$*, as* $n \to \infty$.
(iv) $\|DF_n\|_H^2 \to q$ *in* $L^2(\Omega)$*, as* $n \to \infty$.

In [25] Peccati and Tudor obtained a multidimensional extension of this result, which can also be derived by Stein's method and Malliavin calculus. There has been a large number of applications of this central limit theorem and its generalizations. We refer the reader to the monograph by Nourdin and Peccati [17] for a detailed account of applications and further developments. As an illustration of the power and wide range of applications of this topic, let us mention the following contributions.

1. A central limit theorem for the renormalized self-intersection local time of the d-dimensional fractional Brownian motion with Hurst parameter $H \in \left[\frac{3}{2d}, \frac{3}{4}\right)$ was proved by Hu and Nualart [8].
2. Quantitative Breuer-Major theorems for functionals of Gaussian stationary sequences have been obtained in a series of works (see, for instance, Nourdin and Peccati [15], Nourdin, Peccati and Podolskij [18] and Breton and Nourdin [3]). A typical example where this methodology is successful is the asymptotic behavior of the p-variation of a fractional Brownian motion.
3. Exact Berry-Esséen asymptotics for functionals of Gaussian processes have been obtained by Nourdin and Peccati [16]. The main result of this paper is the following theorem.

Theorem 4.2 *Let $\{F_n, n \geq 1\}$ be a sequence of variables in a fixed Wiener chaos $\mathcal{H}_q$, $q \geq 2$, such that $E(F_n^2) \to 1$. Let $\varphi(n) = \sqrt{E\left[\left(1 - \frac{1}{q}\|DF_n\|_H^2\right)^2\right]}$, and assume that $\lim_{n\to\infty} \varphi(n) = 0$ and there exists $m \geq 1$ such that $\varphi(n) > 0$ for all $n \geq m$. Suppose that the two-dimensional vector $\left(F_n, \frac{1-\|DF_n\|_H^2/q}{\varphi(n)}\right)$ converges in distribution to a centered two-dimensional Gaussian vector (Z_1, Z_2) with $E(Z_1 Z_2) = \rho$. Then,*

$$\varphi(n)^{-1}[P(F_n \leq z) - P(Z \leq z)] \to \frac{\rho}{3}\Phi^{(3)}(z),$$

as $n \to \infty$, where $\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$.

5 Convergence in Law on a Finite Sum of Wiener Chaos

The convergence in law for real-valued random variables is metrizable by the Fortet-Mourier distance:

$$d_{FM}(F, G) = \sup_{\varphi} |E[\varphi(F)] - E[\varphi(G)]|,$$

where the supremum is taken over all functions $\varphi : \mathbb{R} \to \mathbb{R}$ such that $\|\varphi\|_{Lip} \leq 1$ and $\|\varphi\|_\infty \leq 1$. Here $\|\varphi\|_{Lip}$ denotes the Lipschitz norm

$$\|\varphi\|_{Lip} = \sup_{x \neq y} \frac{|\varphi(x) - \varphi(y)|}{|x - y|}.$$

We have seen that in Theorem 4.1 the convergence is in the total variation distance, which is stronger than the Fortet-Mourier distance. Then, a natural question is whether this remains true whenever we have a sequence of random variables belonging to a finite sum of Wiener chaos which converges in law. An affirmative answer to this question, with a quantitative estimate, has been given by Nourdin and Poly [20] using techniques of Malliavin calculus and the following inequality established by Carbery and Wright [6].

Lemma 5.1 *There is a universal constant $C > 0$ such that, for any polynomial $Q : \mathbb{R}^n \to \mathbb{R}$ of degree at most d and $0 < q < \infty$, one has that*

$$E[Q(X_1, \dots, X_n)^{\frac{q}{d}}]^{\frac{1}{q}} P(|Q(X_1, \dots, X_n)| \leq \alpha) \leq Cq\alpha^{\frac{1}{d}}, \tag{5.1}$$

for all $\alpha > 0$, where $X_1, \dots, X_n$ are independent random variables with law $N(0, 1)$.

The next theorem, proved in [20], says that on a finite sum Wiener chaos, the convergence in law to a non-degenerate limit implies convergence in total variation.

Theorem 5.2 *Let $F_n \in \oplus_{k=1}^{p}\mathcal{H}_k$ be a sequence of random variables converging in law to F_∞, and suppose that F_∞ is not identically zero. Then, there is a constant $c > 0$, depending on p, such that*

$$d_{TV}(F_n, F_\infty) \le c d_{FM}(F_n, F_\infty)^{\frac{1}{2p+1}}.$$

A multidimensional extension of this theorem has been proved by Nourdin et al. [14], using an integration-by-parts formula based on the Poisson kernel developed by Bally and Caramelino [1]. In the multidimensional case a lower bound on the expectation of the determinant of the Malliavin matrix of the random vector is required. We recall that given a random vector $(F_1, \dots, F_d)$ whose components belong to $\mathbb{D}^{1,2}$, its Malliavin matrix is the random matrix defined by $\Gamma = (\langle DF_i, DF_j\rangle_H)_{1\le i,j\le d}$.

Theorem 5.3 *Let $F_n = (F_{1,n}, \dots, F_{d,n})$ be a sequence of d-dimensional random vectors such that $F_{i,n} \in \oplus_{k=1}^{p}\mathcal{H}_k$, $i \in \{1, \dots, d\}$, $n \ge 1$. Suppose that F_n converges in law to F_∞, and*

$$E[\det \Gamma_n] \ge \beta > 0, \tag{5.2}$$

where Γ_n is the Malliavin matrix of F_n. Then for any $\gamma < [(d+1)(4d(q-1)+3)+1]^{-1}$, there is a constant $c_\gamma > 0$ such that

$$d_{TV}(F_n, F_\infty) \le c_\gamma d_{FM}(F_n, F_\infty)^\gamma.$$

Sketch of the Proof

(i) By a truncation argument, we can assume that F_n is bounded, that is, $|F_n| \le M$ for some constant $M > 0$. Let $\varphi : \mathbb{R}^d \to \mathbb{R}$ be a measurable function with support in $[-M, M]^d$ such that $\|\varphi\|_\infty \le 1$. We regularize φ with an approximation of the identity of the form $\rho_\alpha(x) = \alpha^{-d}\rho(x/\alpha)$, where $0 < \alpha \le 1$ and ρ is a nonnegative function in $\mathcal{C}_c^\infty$ with integral equal to one. Then,

$$\begin{aligned} |E[\varphi(F_n) - \varphi(F_m)]| &\le |E[\varphi * \rho_\alpha(F_n) - \varphi * \rho_\alpha(F_m)]| \\ &\quad + 2\sup_n |E[\varphi(F_n) - \varphi * \rho_\alpha(F_n)]| \\ &\le \frac{1}{\alpha} d_{FM}(F_n, F_m) + 2R_\alpha, \end{aligned}$$

where

$$R_\alpha = \sup_n |E[\varphi(F_n) - \varphi * \rho_\alpha(F_n)]|.$$

(ii) Let $h_\alpha = \varphi - \varphi * \rho_\alpha$. For any $\epsilon > 0$,

$$|E[h_\alpha(F_n)]| = \left|E\left[h_\alpha(F_n)\left(\frac{\epsilon}{\det\Gamma_n+\epsilon}+\frac{\det\Gamma_n}{\det\Gamma_n+\epsilon}\right)\right]\right|$$
$$\leq 2\epsilon E[(\det\Gamma_n+\epsilon)^{-1}] + \left|E\left[h_\alpha(F_n)\frac{\det\Gamma_n}{\det\Gamma_n+\epsilon}\right]\right|.$$

(iii) For the first term, using the Carbery-Wright inequality (5.1) and the condition $E[\det\Gamma_n] \geq \beta$, we obtain the estimate

$$2\epsilon E[(\det\Gamma_n+\epsilon)^{-1}] \leq c\epsilon^{\frac{1}{2(q-1)d+1}}.$$

(iv) For the second term we use Malliavin calculus and the representation

$$h_\alpha = \sum_{i=1}^d \partial_i h_\alpha * \partial_i Q_d,$$

where Q_d is the Poisson kernel in $\mathbb{R}^d$, to obtain for any $p \geq 1$,

$$\left|E\left[h_\alpha(F_n)\frac{\det\Gamma_n}{\det\Gamma_n+\epsilon}\right]\right| \leq c\epsilon^{-2}\sup_i\left\|\int_{\mathbb{R}^d}\phi(F_n-y)(\partial_i Q_d-\rho_\alpha*\partial_i Q_d)(y)dy\right\|_{L^p(\Omega)},$$

which implies

$$\left|E\left[h_\alpha(F_n)\frac{\det\Gamma_n}{\det\Gamma_n+\epsilon}\right]\right| \leq c\epsilon^{-2}\alpha^{\frac{1}{d+1}}M^{\frac{d}{d+1}}.$$

(v) Therefore

$$|E[\varphi(F_n)-\varphi(F_m)]| \leq \frac{1}{\alpha}d_{FM}(F_n,F_m) + c\epsilon^{\frac{1}{2(q-1)d+1}} + c\epsilon^{-2}\alpha^{\frac{1}{d+1}}M^{\frac{d}{d+1}},$$

and the desired result follows by letting $m \to \infty$ and optimizing in α, ϵ and M. □

Here are some sufficient conditions for condition (5.2) to hold:

(i) If $\Gamma_n \to M_\infty$ in law and $E[\det M_\infty] > 0$, then (5.2) holds.
(ii) If F_∞ is normal $N_d(0, C)$ with $\det(C) > 0$, then (5.2) holds, because $\Gamma_n \to C$ in $L^2(\Omega)$, as it has been proved by Nualart and Ortiz-Latorre [22].
(iii) If F_∞ has independent and non degenerate components, then (5.2) holds.

The convergence in total variation has also been established in the case where the limit is Gaussian, using techniques from information theory in a recent work by Nourdin, Peccati and Swan [19]. The main ingredient in this paper is the so-called Csiszar-Kullback-Pinsker inequality

$$d_{TV}(F, Z) \leq \sqrt{2D(F||Z)},$$

where $D(F||Z) = \int_{\mathbb{R}^d} f(x) \log \frac{f(x)}{\phi_d(x)} dx$ is the relative entropy of the density f of F, with respect to the density ϕ_d of $Z \sim N_d(0, I)$.

Consider a sequence $F_n = (F_{1,n}, \ldots, F_{d,n})$ of d-dimensional random vectors, such that $F_{i,n} \in \mathcal{H}_{q_i}$, with $1 \leq q_1 \leq q_2 \leq \cdots \leq q_d$, $E(F_{i,n}^2) = 1$, $E(F_{i,n}F_{j,n}) = 0$ for $i \neq j$. The main result of [19] says that if

$$\Delta_n := E[\|F_n\|^4 - \|Z\|^4] \to 0,$$

then

$$D(F_n||Z) \leq C\Delta_n |\log \Delta_n|.$$

The proof is based on the Carbery-Wright inequality, the Malliavin calculus and techniques of information theory.

6 Absolute Continuity of Random Vectors on a Finite Chaos

A basic criterion in Malliavin calculus states that if a random vector $F = (F_1, \ldots, F_d)$, whose components belong to $\mathbb{D}^{1,2}$, satisfies $\det \Gamma > 0$ almost surely (where Γ is the Malliavin matrix of F), then the law of F is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^d$. In a recent work, Nourdin, Nualart and Poly (see [14]) proved the stronger result that for random vectors whose components belong to a finite sum of Wiener chaos, $P(\det \Gamma = 0)$ is zero or one; as a consequence, $P(\det \Gamma > 0) = 1$ turns out to be equivalent to the absolute continuity of F. This is also equivalent to the existence a nonzero polynomial H in d variables such that $H(F) = 0$, which was established by Kusuoka [10]. More precisely, the following result was proved in [14]:

Theorem 6.1 *Let* $F = (F_1, \ldots, F_d)$ *be such that* $F_i \in \bigoplus_{k=1}^q \mathcal{H}_k$. *Let* Γ *be the Malliavin matrix of* F. *Then* $P(\det \Gamma = 0)$ *is zero or one and the following assertions are equivalent:*

(a) *The law of* F *is not absolutely continuous with respect to the Lebesgue measure on* $\mathbb{R}^d$.

(b) *There exists* $H \in \mathbb{R}[X_1, \ldots, X_d] \setminus \{0\}$ *of degree at most* dq^{d-1} *such that, almost surely,*

$$H(F_1, \ldots, F_d) = 0.$$

(c) $E[\det \Gamma] = 0$.

In dimension one it is known that $E[\det \Gamma] = 0$ is equivalent to $F = E(F) = 0$. It would be interesting to know what happens in the multidimensional case, that is, under which conditions we have $E[\det \Gamma] = 0$. Let us denote by C the covariance matrix of the random vector F. Clearly, if $\det C = 0$, then the components of F are linearly dependent and the law of F is not absolutely continuous, which implies that

$E[\det \Gamma] = 0$. The converse of this implication is not true if $d \geq 3$. For instance, the vector $(F_1, F_2, F_1 F_2)$, where F_1 and F_2 are two non-zero independent random variables in the first chaos, satisfies $\det \Gamma = 0$ but $\det C \neq 0$. However, in dimension 2 the converse implication is true for random variables in the second chaos. In fact, we have the following result.

Lemma 6.2 *Let* $(F, G) = (I_2(f), I_2(g))$, *and denote by* Γ *and* C *the Malliaivn matrix and the covariance matrix of the vector* (F, G), *respectively. Then,*

$$E[\det \Gamma] \geq 4 \det C.$$

Proof We have $DF = 2I_1(f)$ and $DG = 2I_1(g)$. As a consequence,

$$\begin{aligned} \|DF\|_H^2 &= 4\|f\|^2 + 4I_2(f \otimes_1 f), \\ \|DG\|_H^2 &= 4\|g\|^2 + 4I_2(g \otimes_1 g), \\ \langle DF, DG \rangle_H &= 4I_2(f \otimes_1 g) + 4\langle f, g \rangle. \end{aligned}$$

This implies

$$\begin{aligned} E[\det \Gamma] &= 16 \left(\|f\|^2 \|g\|^2 - \langle f, g \rangle^2 \right) \\ &\quad + 32 \left(\|f \otimes_1 g\|^2 - \|f \widetilde{\otimes}_1 g\|^2 \right) \geq 4 \det C. \end{aligned}$$

□

Here are some immediate consequences of this lemma:

(i) If $E[\det \Gamma] = 0$, then $\det C = 0$ and F and G are linearly dependent.
(ii) If $\det C > 0$, then $E[\det \Gamma] > 0$ and the law of (F, G) is absolutely continuous.
(iii) If $(F_n, G_n) \Rightarrow (F_\infty, G_\infty)$, and $\det C_\infty > 0$, then the convergence is in total variation.

The equivalence between $E[\det \Gamma] = 0$ and $\det C = 0$ in the particular case of a two-dimensional random vector (F, G) whose components are multiple stochastic integrals of the same order n has been recently proved by Nualart and Tudor in [24]. This implies that the random vector (F, G) has an absolutely continuous law with respect to the Lebesgue measure on $\mathbb{R}^2$ if and only if its components are not proportional, as in the Gaussian case. This result was previously established for $n = 2$ in [14], and for $n = 3, 4$ in [28]. The proof starts from the decomposition of the determinant of the Malliavin matrix into a sum of squares:

$$\det \Gamma = \frac{1}{2} \sum_{i,j=1}^{\infty} \left(\langle DF, e_i \rangle_H \langle DG, e_j \rangle_H - \langle DF, e_j \rangle_H \langle DG, e_i \rangle_H \right)^2,$$

where $\{e_i, i \geq 1\}$ is a complete orthonormal system of H. Then, a basic ingredient of the proof in the general case of 2-dimensional vectors in a fixed Wiener chaos of order n, is the inequality

$$\sum_{k=2}^{\left[\frac{n-1}{2}\right]} \frac{n(n-2k)}{k!^2} E[\det \Gamma^{(k)}] + (n-1)^2 E[\det \Gamma^{(1)}] \geq n^2 \det C,$$

where $\Gamma^{(k)}$ is the k-iterated Malliavin matrix of the vector (F, G) defined by

$$\Gamma^{(k)} = \begin{pmatrix} \|D^{(k)}F\|^2_{H^{\otimes k}} & \langle D^{(k)}F, D^{(k)}G\rangle_{H^{\otimes k}} \\ \langle D^{(k)}F, D^{(k)}G\rangle_{H^{\otimes k}} & \|D^{(k)}G\|^2_{H^{\otimes k}} \end{pmatrix},$$

As a consequence, if $\det C > 0$ then that (F, G) has an absolutely continuous law, and the convergence in law of a sequence (F_n, G_n) to a non-degenerate limit would imply the convergence in total variation.

7 Convergence of Densities

Consider the framework of an isonormal Gaussian process X over a Hilbert space H on a probability space $(\Omega, \mathcal{F}, P)$, where $\mathcal{F}$ is the completion of the σ-field generated by X. We have seen that for random variables belonging to a finite sum of Wiener chaos the convergence in law implies convergence in total variation under some suitable non degeneracy assumptions. Notice that the total variation distance between the law of two random variables F and G with densities p_F and p_G coincides with the $L^1(\mathbb{R})$ distance between the densities:

$$d_{TV}(F, G) = \int_{\mathbb{R}} |p_F(x) - p_G(x)|dx.$$

On the other hand, a sufficient condition for a random variable F in $\mathbb{D}^{1,2}$ to have a density is $\|DF\|_H > 0$ almost surely, and the density if smooth if $E(\|DF\|^{-p}) < \infty$ for all $p \geq 1$. Then, we can expect that imposing uniform boundedness of negative moments of the Malliavin norm (or the determinant of the Malliavin matrix) one can deduce uniform convergence of the corresponding densities. Results of this type have been established in the recent work by Hu et al. [7]. In particular, for one-dimensional random variables in a fixed Wiener chaos one can show the following theorem.

Theorem 7.1 *Suppose that F is a random variable in $\mathcal{H}_q$, $q \geq 2$, such that $E(F^2) = 1$ and $E(\|DF\|_H^{-6}) \leq M$. Let p_F be the density of F, and let Z be a $N(0, 1)$ random variable. Then,*

$$\sup_{s \in \mathbb{R}} |p_F(x) - \phi(x)| \leq C_{M,q}\sqrt{E(F^4) - 3},$$

where ϕ is the density of the law $N(0, 1)$.

Sketch of the Proof

The proof given in [7] is based on the techniques of Malliavin calculus together with Stein's method. The main steps of the proof are as follows.

(i) A basic result in Malliavin calculus is the following formula for the density of the random variable F:

$$p_F(x) = E\left[\mathbf{1}_{\{F>x\}}\delta\left(\frac{DF}{\|DF\|_H^2}\right)\right].$$

Using the factorization property (2.5) and the relation (2.6) we can write

$$\begin{aligned} p_F(x) &= E\left[\mathbf{1}_{\{F>x\}}\frac{qF}{\|DF\|_H^2}\right] - E[\mathbf{1}_{\{F>x\}}\langle DF, D(\|DF\|_H^{-2})\rangle_H] \\ &= E[\mathbf{1}_{\{F>x\}}F] + E[q\|DF\|_H^{-2} - 1] - E[\mathbf{1}_{\{F>x\}}\langle DF, D(\|DF\|_H^{-2})\rangle_H]. \end{aligned}$$

(ii) Using the assumption $E(\|DF\|_H^{-6}) \leq M$ together with Wiener chaos expansions one can show that the terms $E[|q\|DF\|_H^{-2} - 1|]$ and $E[|\langle DF, D(\|DF\|_H^{-2})\rangle_H|]$ can be estimated by a constant times $\sqrt{E(F^4) - 3}$.

(iii) Taking into account that $\phi(x) = E[\mathbf{1}_{\{Z>x\}}Z]$, it suffices to estimate the difference

$$E[\mathbf{1}_{\{F>x\}}F] - E[\mathbf{1}_{\{Z>x\}}Z],$$

which can be done by Stein's method and Malliavin calculus. □

Using the notion of *Fisher information*, Nourdin and Nualart [13], provided an alternative proof to Theorem 7.1 under the weaker assumption $E(\|DF\|_H^{-4-\epsilon}) \leq M$ for some $\epsilon > 0$. We recall that the Fisher information $J(F)$ of F is given by $J(F) = E[s_F(F)^2]$ if the random variable $s_F(F)$ is square-integrable and $J(F) = +\infty$ otherwise, where $s_F(F)$ denotes the *score* associated to F, which is the F-measurable random variable uniquely determined by the integration by parts

$$E[\varphi'(F)] = -E[s_F(F)\varphi(F)] \quad \text{for all test function } \varphi : \mathbb{R} \to \mathbb{R}.$$

We have $J(F) \geq 1 = J(Z)$ with equality if and only if F is standard Gaussian. Then, the gap between $J(F)$ and $1 = J(Z)$ is a measure of how the law of F is close to the standard Gaussian distribution $N(0, 1)$. A quantitative version of this statement is given by the Shimizu's inequality [26]:

$$\|p_F - p_N\|_\infty \leq \sqrt{J(F) - 1}, \tag{7.1}$$

which is the main ingredient in the proof of Theorem 7.1 given in [13].

The following extensions of Theorem 7.1 have been also established in [7]:

(i) One can show the uniform approximation of the mth derivative of p_F by the corresponding mth derivative of the Gaussian density $\phi^{(m)}$. For this approximation to hold one needs the stronger assumption $E(\|DF\|_H^{-\beta}) < \infty$ for some $\beta > 6m + 6\left(\lfloor \frac{m}{2} \rfloor \vee 1\right)$.

(ii) Consider a d-dimensional vector F, whose components are in a fixed chaos, and such that $E[(\det \Gamma)^{-p}] < \infty$ for all p, where Γ denotes the Malliavin matrix of F. In this case for any multi-index $\beta = (\beta_1, \ldots, \beta_k)$, $1 \leq \beta_i \leq d$, one can show

$$\sup_{x \in \mathbb{R}^d} \left|\partial_\beta f_F(x) - \partial_\beta \phi_d(x)\right| \leq c\Big(|C - I|^{\frac{1}{2}} + \sum_{j=1}^{d} \sqrt{E[F_j^4] - 3(E[F_j^2])^2}\Big),$$

where C is the covariance matrix of F, ϕ_d is the standard d-dimensional normal density, and $\partial_\beta = \frac{\partial^k}{\partial x_{\beta_1} \ldots \partial x_{\beta_k}}$.

It would be interesting to check whether one can apply Theorem 7.1 to different examples of central limit theorems for random variables in a fixed chaos. The challenge is to verify the condition $E(\|DF\|_H^{-p}) \leq M$ for some $p > 4$. So far the following examples of applications have been developed.

7.1 Example 1

Let $q = 2$. A random variable F in the second Wiener chaos can be always expressed as

$$F = \sum_{i=1}^{\infty} \lambda_i (X(e_i)^2 - 1),$$

where $\{e_i, i \geq 1\}$ is a complete orthonormal system in H and λ_i is a decreasing sequence of positive numbers such that $\sum_{i=1}^{\infty} \lambda_i^2 < \infty$. Then, if $\lambda_N \neq 0$ for some $N > 4$, we obtain (see [7])

$$\sup_{x \in \mathbb{R}} |p_F(x) - \phi(x)| \leq C_{N,\lambda_N} \sqrt{\sum_{i=1}^{\infty} \lambda_i^4}.$$

7.2 Example 2

Our second example is given by the weighted quadratic variation of the fractional Brownian motion. The fractional Brownian motion $B^H = \{B_t^H, t \geq 0\}$ with Hurst parameter $H \in (0, 1)$ is a zero mean Gaussian process with covariance

$$E(B_t^H B_s^H) = \frac{1}{2}\left(t^{2H} + s^{2H} - |t-s|^{2H}\right).$$

Consider the sequence

$$F_n = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} [(n^H \Delta_j B^H)^2 - 1],$$

where $\Delta_j B^H = B^H_{\frac{j}{n}} - B^H_{\frac{j-1}{n}}$. Then, if $H \in (0, \frac{3}{4})$ it is well-known that F_n converges in law to $N(0, \sigma_H^2)$, where $\sigma_H^2 = 2\sum_{z\in\mathbb{Z}} \rho_H(z)^2$ and

$$\rho_H(z) = \frac{1}{2}(|z+1|^{2H} + |z-1|^{2H} - 2|z|^{2H}). \tag{7.2}$$

If we consider the normalized sequence $G_n = \frac{F_n}{\sqrt{\operatorname{Var} F_n}}$, Biermé, Bonami and León [5] obtained the following result:

$$\sqrt{E(G_n^4) - 3} \le c_H \begin{cases} n^{-\frac{1}{2}} & \text{if } H \in (0, \frac{5}{8}), \\ n^{-\frac{1}{2}} (\log n)^{\frac{3}{2}} & \text{if } H = \frac{5}{8}, \\ n^{4H-3} & \text{if } H \in (\frac{5}{8}, \frac{3}{4}). \end{cases}$$

Using the quantitative version of the fourth moment theorem obtained by Nourdin and Peccati by means of Stein's method, this gives the rate of convergence to zero for the total variation distance $d_{TV}(G_n, Z)$, where Z is a standard normal random variable.

In this context, Nourdin and Nualart [13], obtained the following result:

Proposition 7.2 *For any $p \ge 1$ there exists n_0 (depending on p) such that*

$$\sup_{n \ge n_0} E(\|DG_n\|^{-2p}) < \infty.$$

As a consequence, for n large enough

$$\sup_{x\in\mathbb{R}} |p_{G_n}(x) - \phi(x)| \le C\sqrt{E(G_n^4) - 3} \le Cc_H \begin{cases} n^{-\frac{1}{2}} & \text{if } H \in (0, \frac{5}{8}), \\ n^{-\frac{1}{2}} (\log n)^{\frac{3}{2}} & \text{if } H = \frac{5}{8}, \\ n^{4H-3} & \text{if } H \in (\frac{5}{8}, \frac{3}{4}). \end{cases}$$

A similar estimate can be obtained for the uniform norm of all the derivatives of the density. The proof of Proposition 7.2 is inspired in the following basic lemma, which is obtained by a decomposition in blocks and the application of Carbery-Wright inequality.

Lemma 7.3 *Let G_n be a sequence of random variables in $\mathcal{H}_2$. Fix $N > 2p \geq 2$ and suppose that we have a decomposition of the form*

$$\|DG_n\|^2 = \sum_{i=1}^{N} A_{i,n}^2,$$

where the $A_{i,n}$ are random variables in the first Wiener chaos. Let $\mathcal{F}_{i,n} = \sigma\{A_{j,n}, 1 \leq j \leq i\}$. Suppose that for each $i = 1, \ldots, N$

$$\liminf_{n\to\infty} \text{Var}\left(A_{i,n}|\mathcal{F}_{i-1,n}\right) > 0. \tag{7.3}$$

Then,

$$\sup_{n\geq n_0} E(\|DG_n\|^{-2p}) < \infty.$$

Proof We can write

$$\|DG_n\|^{-2p} = \left(\sum_{i=1}^{N} A_{i,n}^2\right)^{-p} \leq \prod_{i=1}^{N} A_{i,n}^{-\frac{2p}{N}}.$$

Then,

$$E\left[A_{N,n}^{-\frac{2p}{N}}|\mathcal{F}_{N-1,n}\right] \leq 1 + \frac{p}{N}\int_0^1 P(A_{N,n}^2 < x|\mathcal{F}_{N-1,n})x^{-\frac{p}{N}-1}dx,$$

and applying the Carbery-Wright's inequality with $d = q = 2$ (5.1) we obtain

$$\begin{aligned} P(A_{N,n}^2 < x|\mathcal{F}_{N-1,n}) &\leq C\sqrt{x}\left[E\left(A_{N,n}^2|\mathcal{F}_{N-1,n}\right)\right]^{-1/2} \\ &\leq C\sqrt{x}\left[\text{Var}\left(A_{N,n}|\mathcal{F}_{N-1,n}\right)\right]^{-1/2}. \end{aligned}$$

We can conclude the proof taking into account that $N > 2p$ and using condition (7.3). □

7.3 *Example 3*

In a recent paper Hu et al. [9] have derived an application of Theorem 7.1 in the framework of the Breuer-Major theorem (see [4] or Theorem 7.2.4 in [17]) using the approach outlined in Lemma 7.3. Fix $q \geq 2$ and consider a sequence of random variables of the form

$$V_n = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} H_q(X_j), \tag{7.4}$$

where $X = \{X_k, k \geq 1\}$ is a centered Gaussian stationary sequence with unit variance. For all $v \geq 0$, we set $\rho(v) = E[X_1 X_{1+v}]$ and $\rho(-v) = \rho(v)$ if $v < 0$. The main result of [9] is the next theorem.

Theorem 7.4 *Let $X = \{X_k, k \geq 1\}$ be stationary Gaussian sequence whose spectral density f_ρ satisfies $f_\rho \in L^{1/2}([-\pi, \pi])$ and $\log(f_\rho) \in L^1([-\pi, \pi])$. Let V_n be the random variable defined by (7.4), and assume that $\sum_{v\in\mathbb{Z}} |\rho(v)|^q < \infty$. Set $\sigma^2 := q! \sum_{v\in\mathbb{Z}} \rho(v)^q < \infty$. Then for any $p \geq 1$, there exists n_0 such that*

$$\sup_{n\geq n_0} \mathbf{E}\left[\|DV_n\|^{-p}\right] < \infty. \tag{7.5}$$

Therefore, if $F_n = V_n/\sqrt{E[V_n^2]}$, we have

$$\sup_{x\in\mathbb{R}} |p_{F_n}(x) - \phi(x)| \leq c\sqrt{E[F_n^4] - 3}.$$

A particular example of stationary sequence satisfying the assumptions of Theorem 7.4 is $\{X_k = B_k^H - B_{k-1}^H, k \geq 1\}$, where $\{B_t^H, t \geq 0\}$ is a fractional Brownian motion of Hurst parameter $H \in (0, 1)$. In this case, the covariance function $\rho = \rho_H$ is given by (7.2) In this case, the spectral density (see e.g. [2], Eq. (2.17)) given by

$$f(\lambda) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \rho(|k|)e^{i\lambda} = 2c_f(1 - \cos(\lambda)) \sum_{j=-\infty}^{\infty} |2\pi j + \lambda|^{-2H-1}, \quad \lambda \in [-\pi, \pi]$$

satisfies the required conditions. As a consequence, we obtain the uniform convergence of densities (and their derivatives) for the sequence of Hermite variations

$$F_n = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} H_q(n^H \Delta_j B^H), \quad q \geq 2.$$

Acknowledgments We would like to thank an anonymous referee for carefully reading this manuscript and providing useful remarks.

References

1. V. Bally, L. Caramellino, Riesz transform and integration by parts formulas for random variables. Stoch. Proc. Appl. **121**, 1332–1355 (2011)
2. J. Beran, *Statistics for Long-Memory Processes* (Chapman and Hall, Boca Raton, 1994)
3. J.C. Breton, I. Nourdin, Error bounds on the non-normal approximation of Hermite power variations of fractional Brownian motion. Electron. Commun. Probab. **13**, 482–493 (2008)

4. P. Breuer, P. Major, Central limit theorems for non-linear functionals of Gaussian fields. J. Mult. Anal. **13**, 425–441 (1983)
5. H. Biermé, A. Bonami, J. León, Central limit theorems and quadratic variations in terms of spectral density. Electron. J. Probab. **16**(13), 362–395 (2011)
6. A. Carbery, J. Wright, Distributional and L^q norm inequalities for polynomials over convex bodies in $\mathbb{R}^n$. Math. Res. Lett. **8**, 233–248 (2001)
7. Y. Hu, F. Lu, D. Nualart, Convergence of densities for some functionals of Gaussian processes. J. Funct. Anal. **266**, 814–875 (2014)
8. Y. Hu, D. Nualart, Renormalized self-intersection local time for fractional Brownian motion. Ann. Probab. **33**(3), 948–983 (2005)
9. Y. Hu, D. Nualart, S. Tindel, F. Xu, Density convergence in the Breuer-Major theorem for nonlinear Gaussian stationary sequences. Bernoulli (To appear)
10. S. Kusuoka, On the absolute continuity of the law of a system of multiple wiener integral. J. Fac. Sci. Univ. Tokyo Sect. IA Math. **30**, 191–197 (1983)
11. P. Malliavin, *Stochastic calculus of variation and hypoelliptic operators*, in *Proceedings of the International Symposium on Stochastic Differential Equations, Kyoto*, 1976, pp. 195–263, (Wiley, New York, 1978)
12. P. Malliavin, *Stochastic Analysis*, Grundlehren der Mathematischen Wissenschaften, 313 (Springer, Berlin, 1997)
13. I. Nourdin, D. Nualart, Fisher information and the fourth moment theorem. Ann. Inst. H. Poincaré Probab. Statist. (To appear)
14. I. Nourdin, D. Nualart, G. Poly, Absolute continuity and convergence of densities for random vectors on Wiener chaos. Electron. J. Probab. **18**, 1–19 (2013)
15. I. Nourdin, G. Peccati, Stein's method on Wiener chaos. Probab. Theor. Relat. Fields **145**(1–2), 75–118 (2009)
16. I. Nourdin, G. Peccati, Stein's method and exact Berry-Esseen asymptotics for functionals of Gaussian fields. Ann. Probab. **37**(6), 2231–2261 (2009)
17. I. Nourdin, G. Peccati, *Normal Approximations with Malliavin Calculus: From Stein's Method to Universality* (Cambridge University Press, Cambridge, 2012)
18. I. Nourdin, G. Peccati, M. Podolskij, Quantitative Breuer-Major theorems. Stochast. Process. Appl. **121**(4), 793–812 (2011)
19. I. Nourdin, G. Peccati, Y. Swan, Entropy and the fourth moment phenomenon. J. Funct. Anal. **266**, 3170–3207 (2014)
20. I. Nourdin, G. Poly, Convergence in total variation on Wiener chaos. Stoch. Proc. Appl. **123**, 651–674 (2013)
21. D. Nualart, *The Malliavin Calculus and Related Topics*, 2nd edn. (Springer, New York, 2006)
22. D. Nualart, S. Ortiz-Latorre, Central limit theorems for multiple stochastic integrals and Malliavin calculus. Stoch. Proc. Appl. **118**, 614–628 (2008)
23. D. Nualart, G. Peccati, Central limit theorems for sequences of multiple stochastic integrals. Ann. Probab. **33**, 177–193 (2005)
24. D. Nualart, C.A. Tudor, The determinant of the iterated Malliavin matrix and the density of a couple of multiple integrals (Preprint)
25. G. Peccati, C. A. Tudor, *Gaussian limits for vector-valued multiple stochastic integrals* in *Séminaire de Probabilités* XXXVIII, 247–262, Lecture Notes in Math., 1857, (Springer, Berlin, 2005)
26. R. Shimizu, *On Fisher's amount of information for location family* in *Proceedings of A Modern Course on Statistical Distributions in Scientific Work*, pp. 305–312, (Springer, Berlin, 1975)
27. Ch. Stein, *Approximate Computation of Expectations*, IMS Lecture Series - Monograph Series, 7 (CA: Institute of Mathematical Statistics, Hayward, 1986)
28. C.A. Tudor, The determinant of the Malliavin matrix and the determinant of the covariance matrix for multiple integrals. ALEA **X**(2), 681–692 (2013)

An Overview of Viscosity Solutions of Path-Dependent PDEs

Zhenjie Ren, Nizar Touzi and Jianfeng Zhang

Abstract This paper provides an overview of the recently developed notion of viscosity solutions of path-dependent partial differential equations. We start by a quick review of the Crandall-Ishii notion of viscosity solutions, so as to motivate the relevance of our definition in the path-dependent case. We focus on the wellposedness theory of such equations. In particular, we provide a simple presentation of the current existence and uniqueness arguments in the semilinear case. We also review the stability property of this notion of solutions, including the adaptation of the Barles-Souganidis monotonic scheme approximation method. Our results rely crucially on the theory of optimal stopping under nonlinear expectation. In the dominated case, we provide a self-contained presentation of all required results. The fully nonlinear case is more involved and is addressed in [12].

Keywords Path-dependent PDEs · Viscosity solutions · Optimal stopping

AMS 2000 subject classifications: 35D40, 35K10, 60H10, 60H30.

Touzi's research is supported by the ERC grant 321111 RoFiRM, the ANR grant ISOTACE, the Chair *Financial Risks* of the *Risk Foundation* sponsored by Société Générale, and the Chair *Finance and Sustainable Development* sponsored by EDF and Calyon.

Z. Ren · N. Touzi (✉)
CMAP, Ecole Polytechnique Paris, Palaiseau 91128, France
e-mail: nizar.touzi@polytechnique.edu

Z. Ren
e-mail: ren@cmap.polytechnique.fr

J. Zhang
Department of Mathematics, University of Southern California,
Los Angeles, CA, USA
e-mail: jianfenz@usc.edu

D. Crisan et al. (eds.), *Stochastic Analysis and Applications 2014*,
Springer Proceedings in Mathematics & Statistics 100,
DOI 10.1007/978-3-319-11292-3_15

1 Introduction

Let $\Omega := \{\omega \in C^0([0,T], \mathbb{R}^d) : \omega_0 = 0\}$ be the canonical space of continuous paths starting from the origin, B the canonical process defined by $B_t(\omega) := \omega_t$, $t \in [0,T]$, and $\mathbb{F} := \{\mathcal{F}_t, t \in [0,T]\}$ the corresponding filtration. Following Dupire [10], we introduce the pseudo-distance

$$d\big((t,\omega),(t',\omega')\big) := |t-t'| + \|\omega_{\wedge t} - \omega'_{\wedge t'}\|_\infty \quad \text{for all} \quad t,\ t' \in [0,T], \quad \omega,\ \omega' \in \Omega. \tag{1.1}$$

Then, any process $u : [0,T] \times \Omega \longrightarrow \mathbb{R}$, continuous with respect to d, is $\mathbb{F}-$progressively measurable, so that $u(t,\omega) = u\big(t, (\omega_s)_{s \le t}\big)$.

The goal of this paper is to provide a wellposedness theory for the path-dependent partial differential equation (PDE):

$$-\partial_t u(t,\omega) - G\big(t,\omega,u(t,\omega),\partial_\omega u(t,\omega),\partial^2_{\omega\omega}u(t,\omega)\big) = 0, \quad t < T, \quad \omega \in \Omega. \tag{1.2}$$

with boundary condition $u(T,\omega) = \xi(\omega)$. Here, $\xi : (\Omega, \mathcal{F}_T) \longrightarrow (\mathbb{R}, \mathcal{B}_\mathbb{R})$ is a bounded uniformly continuous function, and $G : [0,T] \times \Omega \times \mathbb{R} \times \mathbb{R}^d \times \mathbb{S}_d \longrightarrow \mathbb{R}$ is continuous in (t,ω), Lipschitz-continuous in the remaining variables (y,z,γ), and satisfies the ellipticity condition:

$$\gamma \in \mathbb{S}_d \longmapsto G(t,\omega,y,z,\gamma) \text{ is non-decreasing.} \tag{1.3}$$

The unknown process $u(t,\omega)$ is required to be $\mathbb{F}$-progressively measurable, and the derivatives $\partial_t u, \partial_\omega u, \partial^2_{\omega\omega}u$ are $\mathbb{F}$-progressively measurable processes valued in $\mathbb{R}, \mathbb{R}^d, \mathbb{S}_d$, respectively, which will be defined later. Notice in particular that, as $\mathbb{R}^d$- and $\mathbb{S}_d$-valued process, the derivatives $\partial_\omega u, \partial^2_{\omega\omega}u$ do not correspond to some (infinite-dimensional) gradient and Hessian with respect to the path. Consequently, the Eq. (1.2) is a PDE parameterized by the path, and not a general PDE on the paths space. For this reason, the name path-dependent PDE is more relevant than PDE on the paths space.

There are three particular examples of such equations which can be related to the existing probability theory literature, namely

1. When the nonlinearity G is linear:

$$G^{\text{lin}}(.,y,z,\gamma) := \ell - ky + \frac{1}{2}\text{Tr}(\gamma), \tag{1.4}$$

for some functions ℓ, k defined on $[0,T] \times \Omega$ the natural solution of the equation (1.2) is given by any regular version of the conditional expectation

$$u^{\text{lin}}(t,\omega) := \mathbb{E}^{\mathbb{P}_0}\Big[\int_t^T e^{-\int_t^s k_r dr}\ell_s ds + e^{-\int_t^T k_r dr}\xi\Big|\mathcal{F}_t\Big](\omega), \tag{1.5}$$

where $\mathbb{P}_0$ is the Wiener measure. Similar results hold for more general linear equations.

2. When the nonlinearity G is semilinear:

$$G^{\text{s-lin}}(.,y,z,\gamma) := \frac{1}{2}\text{Tr}(\gamma) + F(.,y,z), \tag{1.6}$$

for some function $F : [0,T]\times\Omega\times\mathbb{R}\times\mathbb{R}^d \longrightarrow \mathbb{R}$, the natural solution of the equation (1.2) is given by any regular version of the backward stochastic differenttial equation:

$$u^{\text{s-lin}}(t,\omega) = Y_t(\omega) \quad \text{where } Y_s = \xi + \int_s^T F_r(Y_r,Z_r)dr - \int_s^T Z_r dB_r,\ \mathbb{P}_0\text{-a.s.}$$

3. The theory of second order backward stochastic differential equations introduced in [5, 29] provides a similar representation of the natural solution of the path-dependent PDE (1.2) for a class of fully nonlinearities G.

Another important particular example, which plays the role of a benchmark, is the so-called Markovian case when $\xi(\omega) = h(\omega_T)$, and $G(t,\omega,y,z,\gamma) = g(t,\omega_t,y,z,\gamma)$ for some functions g and h defined on the corresponding finite-dimensional spaces. In this context, we expect that $u(t,\omega) = v(t,\omega_t)$ for some function $v : [0,T]\times\mathbb{R}^d \longrightarrow \mathbb{R}$, and the path-dependent PDE (1.2) reduces to the standard PDE:

$$-\partial_t v(t,x) - g\big(t,x,v(t,x),Dv(t,x),D^2v(t,x)\big) = 0, \quad t<T, \quad x\in\mathbb{R}^d, \tag{1.7}$$

where ∂_t, D, D^2 denotes respectively the standard time derivative, the gradient and the Hessian with respect to the space variable. In this case, it is well-known that the theory of viscosity solutions introduced by Crandall and Lions [7, 8] is a powerful notion of weak solution for which a solid existence and uniqueness theory has been developed, and which proved its relevance for various applications. Viscosity solutions gained importance by the contributions of Barles and Souganidis [1] to the convergence of numerical schemes, and the work of Cafarelli and Cabre [4] which makes a crucial use of viscosity solutions to obtain sharp regularity results.

Our main concern is the adaptation of the notion of viscosity solutions to the context of our path-dependent PDE (1.2). However, the fact that our underlying space, namely $[0,T]\times\Omega$, is not locally compact raises a major difficulty which needs to be addressed. Indeed, the stability and the uniqueness results in the theory of viscosity solutions is based on the existence of a local maximizer for an arbitrary upper semicontinuous function.

In order to by-pass this difficulty, we introduce a convenient modification of the definition. To explain our definition, let us focus on the notion of viscosity subsolution, the case of a viscosity supersolution is symmetric. For a viscosity subsolution u, the standard definition considers as test functions some point (t_0, x_0) all those functions φ which are pointwisely locally tangent from above to u with contact point (t_0, x_0):

$$(\varphi - u)(t_0, x_0) = \min_{O_r(t_0,x_0)} (\varphi - u), \quad \text{for some } r > 0,$$

where $O_r(t_0, x_0)$ denotes the open ball in $\mathbb{R}^{d+1}$ centered at (t_0, x_0), with radius r.

1. For simplicity, we first consider the case of a nonlinearity $G = G^{\text{lin}}$ as in (1.4), or $G = G^{\text{s-lin}}$ as in (1.6), with $F(t, \omega, y, z)$ independent of the z-component. Our definition follows exactly the spirit of viscosity solutions, but replaces the pointwise tangency by the corresponding notion in mean:
$$(\varphi - u)(t_0, \omega_0) = \min_{\tau} \mathbb{E}^{\mathbb{P}}\big[(\varphi - u)_{\tau\wedge \mathrm{H}}|\mathcal{F}_{t_0}\big](\omega_0),$$
$$\text{for some stopping time } \mathrm{H} > t_0,$$
where the min is over all stopping times $\tau \geq t_0$.
2. For a more general nonlinearity G, our definition replaces the expectation operator $\mathbb{E}^{\mathbb{P}}$ by a the sublinear expectation operator $\overline{\mathcal{E}} := \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}^{\mathbb{P}}$ for some convenient family $\mathcal{P}$ of probability measure. We observe that $\mathcal{P}$ can be chosen to be a dominated family of measures in the semilinear case $G = G^{\text{s-lin}}$. However, in the general nonlinear case, the family $\mathcal{P}$ is not dominated.

The main purpose of this paper is to provide an overview of the available results on the wellposedness of the path-dependent PDE under this notion of viscosity solution. In particular, we highlight that our definition induces a richer family of test function in the Markovian case. Consequently,

(i) the existence may be more difficult to achieve under our definition; however, we shall see that the traditional examples from the applications raise no special difficulty from the existence side; in fact, in contrast with the standard notion of viscosity solution, our definition is tight,

(ii) the uniqueness may be easier under our definition because our notion of viscosity solution is constrained by a bigger set of test functions; indeed recently comparison results were obtained in the semilinear case $G = G^{\text{s-lin}}$ with relatively simple arguments avoiding the Crandall-Ishii's lemma of the standard viscosity solution in the Markovian case; in particular, the comparison result for the linear path-dependent PDE $G = G^{\text{lin}}$ follows from the equivalence between our notion of viscosity subsolution and the (pathwise) submartingale property, whose proof is a simple consequence of the theory of optimal stopping.

This paper also pays a special attention to the stability of our notion of viscosity solutions, which is an essential property of standard viscosity solutions in the

Markovian case, and is responsible for the denomination of this notion. We shall present the present state of stability results, together with the corresponding convergence results of numerical schemes à la Barles and Souganidis [1].

2 Standard Viscosity Solution in the Markovian Case

In this short section, we recall the standard definition of viscosity solutions in the Markovian case, and we review the corresponding existence and uniqueness results. In order for our notations to be consistent with the path-dependent case, our functions will be defined on $\mathrm{cl}(Q) = [0, T] \times \mathbb{R}^d$, where $Q := [0, T) \times \mathbb{R}^d$.

2.1 Definitions and Consistency with Classical Solutions

For $(t, x) \in Q$, $u \in \mathrm{USC}(Q)$, and $v \in \mathrm{LSC}(Q)$, we denote:

$$\underline{A}u(t,x) := \big\{\varphi \in C^{1,2}(Q) : (\varphi - u)(t,x) = \min_Q(\varphi - u)\big\}, \tag{2.1}$$

$$\overline{A}v(t,x) := \big\{\varphi \in C^{1,2}(Q) : (\varphi - v)(t,x) = \max_Q(\varphi - v)\big\}. \tag{2.2}$$

Definition 2.1 (i) $u \in \mathrm{USC}(Q)$ is a viscosity subsolution of Eq. (1.7) if:

$$\big\{-\partial_t\varphi - g(., u, D\varphi, D^2\varphi)\big\}(t,x) \le 0 \quad \text{for all } (t,x) \in Q, \quad \varphi \in \underline{A}u(t,x).$$

(ii) $v \in \mathrm{LSC}(Q)$ is a viscosity supersolution of Eq. (1.7) if:

$$\big\{-\partial_t\varphi - g(., u, D\varphi, D^2\varphi)\big\}(t,x) \ge 0 \quad \text{for all } (t,x) \in Q, \quad \varphi \in \overline{A}u(t,x).$$

(iii) A viscosity solution of (1.7) is a viscosity subsolution and supersolution of (1.7).

From the last definition, it is clear that one may add a constant to the test function φ so that the minimum and the maximum values in (2.1), (2.2) are zero. Then, the pictorial representation of a test function $\varphi \in \underline{A}u(t,x)$ is a smooth function tangent from above to u with contact point at (t,x). The symmetric pictorial representation holds for a test function $\varphi \in \overline{A}v(t,x)$. Notice that $\underline{A}v(t,x)$ may be empty, and in this case the subsolution property at (t,x) holds trivially.

We also observe that we may replace the minimum and maximum in (2.1), (2.2) by the corresponding local notions. Moreover, by the continuity of the nonlinearity g, we may also assume the minimum (reps. maximum) or local minimum (resp. local maximum) to be strict, and we may restrict attention to $C^\infty(Q)$ test functions.

The following consistency property is an easy consequence of the ellipticity condition on g. We state it only for subsolution, but the result can be similarly stated for supersolutions.

Proposition 2.2 *Assume* $g(t, x, y, z, \gamma)$ *is non-decreasing in* γ. *Then, for a function* $u \in C^{1,2}(Q)$, *we have*

u is a classical subsolution of (1.7) *iff u is a viscosity subsolution of* (1.7).

2.2 The Heat Equation Example

In this subsection, we consider the equation

$$-Lu(t,x) := -\partial_t u(t,x) - b(t,x)Du(t,x) - \frac{1}{2}\sigma^2(t,x) : D^2u(t,x) = 0, \quad (t,x) \in Q. \qquad (2.3)$$

where the coefficients $b : Q \longrightarrow \mathbb{R}^d$ and $\sigma : Q \longrightarrow \mathbb{S}_d$ are continuous and Lipschitz-continuous in x uniformly in t. The purpose from studying this simple example is to gain some intuition in view of our extension to the path-dependent case.

Under the above conditions on b and σ, we may consider the unique strong solution $\{X_t, t \in [0,T]\}$ of the stochastic differential equation

$$X_t = X_0 + \int_0^t b(s, X_s)ds + \int_0^t \sigma(s, X_s)dB_s, \qquad \mathbb{P}_0\text{−a.s.} \qquad (2.4)$$

for some given initial data X_0. Then, given a boundary condition $u(T, .) = \psi$ for some $\psi : \mathbb{R}^d \longrightarrow \mathbb{R}$, the natural solution of (2.3) is given by:

$$u^0(t,x) := \mathbb{E}^{\mathbb{P}_0}\big[\psi(X_T)|X_t = x\big], \quad (t,x) \in Q.$$

In the remaining of this section, we verify that u^0 is a viscosity solution of the heat Eq. (2.3), and we make crucial observations which open the door for enlarging the set of test functions.

(a) ***Tower property*** The first step is to use the Markov feature of the process X in conjunction with the tower property to deduce that

$$u(t,x) = \mathbb{E}^{\mathbb{P}_0}\big[u(\tau, X_\tau)|X_t = x\big] \quad \text{for all stopping time } \tau \text{ with values in } [t,T]. \qquad (2.5)$$

We shall use this identity with stopping times $\tau = \tau_h := (t+h) \wedge \inf\{s > t : |X_s - x| \geq 1\}$. For the next development, notice that $\tau_h > t$, a.s., (s, X_s) is bounded on $[t, \tau_h]$, and $\tau_h \longrightarrow t$, a.s. when $h \searrow 0$.
Also, we avoid to discuss the regularity issues of the function u^0. For instance, if ψ is Lipschitz-continuous, then u^0 is immediately seen to be Lipschitz-continuous with respect to the x-variable, uniformly in t, and we verify that u^0 is $\frac{1}{2}$–Hölder-continuous with respect to the t variable, uniformly in x, by using the identity (2.5).

(b) u^0 **is a viscosity subsolution** Let $(t, x) \in Q$ and $\varphi \in \underline{A}u(t, x)$ be given. We denote by $\{X_s^{t,x}, s \in [t, T]\}$ the solution of (2.4) started from $X_t^{t,x} = x$. By definition, we have

$$
\begin{aligned}
&(\varphi - u)(t, x) \leq (\varphi - u) \text{ on } Q, \\
&\text{and then } (\varphi - u)(t, x) \leq \mathbb{E}^{\mathbb{P}_0}\big[(\varphi - u)(\tau_h, X_{\tau_h}^{t,x})\big],
\end{aligned} \tag{2.6}
$$

for all $h > 0$. From the last inequality in mean, together with the identity (2.5), we get

$$
\varphi(t, x) \leq \mathbb{E}^{\mathbb{P}_0}\big[\varphi(\tau_h, X_{\tau_h}^{t,x})\big].
$$

Since the test function φ is smooth, it follows from Itô's formula that

$$
-\mathbb{E}^{\mathbb{P}_0}\Big[\int_t^{\tau_h} L\varphi(r, X_r^{t,x})dr\Big] \leq 0.
$$

Dividing by h and sending $h \searrow 0$, we deduce from the mean value theorem together with the dominated convergence theorem that

$$
-L\varphi(t, x) \leq 0.
$$

(c) u^0 **is a viscosity supersolution** For $(t, x) \in Q$ and $\varphi \in \overline{A}u(t, x)$, notice that we have the analogue of (2.6):

$$
\begin{aligned}
&(\varphi - u)(t, x) \geq (\varphi - u) \text{ on } Q, \\
&\text{and then } (\varphi - u)(t, x) \geq \mathbb{E}^{\mathbb{P}_0}\big[(\varphi - u)(\tau_h, X_{\tau_h}^{t,x})\big].
\end{aligned} \tag{2.7}
$$

Following the same line of argument as in (b), it follows that $L\varphi(t, x) \geq 0$, as required.

Crucial observation Notice that only the right-hand sides of (2.6) and (2.7) have been useful to prove that u^0 is a subsolution and supersolution, respectively, of the heat equation (2.3). The right-hand sides of (2.6) and (2.7) express that the test function φ is tangent to u in mean, locally along the trajectory of the underlying process $(s, X_s^{t,x})$. Of course, the set of smooth functions which are tangent from above (reps. from below) in mean is larger than $\underline{A}u$ (resp. $\overline{A}u$). Consequently, we may have used an alternative definition of viscosity solution with a richer family of test functions (defined by the right-hand sides of (2.6) and (2.7)), and still get the same existence result. The benefit from such a stronger definition may be that the uniqueness theory can be simplified by suitable use of the additional test functions.

Remark 2.3 The $C^{1,2}$ smoothness of the test function φ is only needed in order to apply Itô's formula

$$\varphi\big(\tau_h, X^{t,x}_{\tau_h}\big) - \varphi(t,x)$$
$$= \int_t^{\tau_h} L\varphi(r, X^{t,x}_r)dr + \int_t^{\tau_h} D\varphi(r, X^{t,x}_r)\sigma(r, X^{t,x}_r)dB_r, \qquad \mathbb{P}_0 - \text{a.s.}$$

Motivated by this observation, we shall take Itô's formula as a starting point for the definition of smooth processes in the path-dependent case.

2.3 Existence for HJB Equations

In this subsection, we show that the crucial observation from the previous subsection holds in the context of the fully nonlinear Hamilton-Jacobi-Bellman (HJB) equation:

$$-\partial_t u - \sup_{k\in K}\Big\{b(.,k)Du + \frac{1}{2}\sigma^2(.,k) : D^2u\Big\} = 0, \quad (t,x)\in Q. \tag{2.8}$$

Here, for simplicity, we consider the case of a bounded set of controls K. The controlled coefficients $b : Q \times K \longrightarrow \mathbb{R}^d$ and $\sigma : Q \times K \longrightarrow \mathbb{S}_d$ are continuous in t, Lipschitz-continuous in x uniformly in (t, κ). The controls set is denoted by $\mathcal{K}$, and consists of all $\mathbb{F}$-progressively measurable process with values in K. For all control process $\kappa \in \mathcal{K}$, we introduce the controlled process X^κ as the unique strong solution of the SDE

$$X^\kappa_t = X_0 + \int_0^t b(s, X^\kappa_s, \kappa_s)ds + \int_0^t \sigma(s, X^\kappa_s, \kappa_s)dB_s, \quad \mathbb{P}_0\text{-a.s.}$$

and we denote by $X^{\kappa,t,x}$ the solution corresponding to the initial data $X^{\kappa,t,x}_t = x$. The Dynkin operator associated to X^κ is denoted:

$$L^k := \partial_t + b(.,k)D + \frac{1}{2}\sigma^2(.,k) : D^2.$$

Given a boundary condition $u(T,.) = \psi$ for some $\psi : \mathbb{R}^d \longrightarrow \mathbb{R}$, the natural solution of (2.8) is given by:

$$u^1(t,x) = \sup_{\kappa\in\mathcal{K}} \mathbb{E}^{\mathbb{P}_0}\big[\psi(X_T^{\kappa,t,x})\big], \qquad (t,x)\in Q.$$

In the remaining of this section, we verify that u^1 is a viscosity supersolution of the HJB equation (2.8), and we focus on the crucial observation that only the tangency condition in mean is used for this purpose. The subsolution property can be obtained by similar standard arguments, and the reader can verify that only tangency in mean is needed, again.

(a) **Dynamic programming principle** In the present nonlinear case, the substitute for the tower property identity (2.5) is the following dynamic programming identity:

$$u(t,x) = \sup_{\kappa\in\mathcal{K}} \mathbb{E}^{\mathbb{P}_0}\big[u(\tau^\kappa, X_{\tau^\kappa}^{\kappa,t,x})\big]$$
$$\text{for all stopping times } \tau^\kappa \text{ with values in } [t,T]. \tag{2.9}$$

This identity will be used with stopping times $\tau^\kappa = \tau_h^\kappa := (t+h)\wedge\inf\{s>t : |X_s^{\kappa,t,x} - x| \geq 1\}$. For the next development, notice that $\tau_h > t$, a.s., (s, X_s^κ) is bounded on $[t,\tau_h^\kappa]$, and $\tau_h^\kappa \longrightarrow t$, a.s. when $h \searrow 0$.
The proof of (2.9) is a difficult task relying on involved measurable selections techniques, see [29] for the regular case (which does not require measurable selection arguments), [18, 19] for the general irregular case, and [2] for a weak dynamic programming principle which is sufficient for the task of deriving the viscosity property, while by-passing the measurable selection arguments.
We also avoid here to discuss the regularity issues of the function u^1. For instance, if ψ is Lipschitz-continuous, then u^1 is immediately seen to be Lipschitz-continuous with respect to the x-variable, uniformly in t, and we verify that u^1 is $\frac{1}{2}$-Hölder-continuous with respect to the t- variable, uniformly in x, by using the identity (2.9).

(b) u^1 **is a viscosity supersolution** Let $(t,x)\in Q$ and $\varphi\in\overline{A}u(t,x)$ be given. Fix an arbitrary control process $\kappa\in\mathcal{K}$. For the purpose of the present argument, we may take this control porcess to be constant $\kappa_s = k$ for all $s\in[t,T]$. By definition, we have

$$(\varphi-u)(t,x) \geq (\varphi-u) \text{ on } Q,$$
$$\text{and then } (\varphi-u)(t,x) \geq \mathbb{E}^{\mathbb{P}_0}\big[(\varphi-u)(\tau_h^\kappa, X_{\tau_h^\kappa}^{\kappa,t,x})\big], \tag{2.10}$$

for all $h>0$. From the last inequality in mean, together with the identity (2.9), we get

$$\varphi(t,x) \geq \mathbb{E}^{\mathbb{P}_0}\big[\varphi(\tau_h, X_{\tau_h}^{\kappa,t,x})\big].$$

Since the test function φ is smooth, it follows from Itô's formula that

$$-\mathbb{E}^{\mathbb{P}_0}\Big[\int_t^{\tau_h} L\varphi(r, X_r^{\kappa,t,x})\Big] \geq 0.$$

Dividing by h and sending $h \searrow 0$, we deduce from the mean value theorem together with the dominated convergence theorem that

$$-L^k\varphi(t,x) \geq 0.$$

By the arbitrariness of $k \in K$, this proves the required supersolution property.

Crucial observation Here again, only the right-hand side of (2.10) has been useful to prove that u^1 is a supersolution of the HJB equation (2.8). The right-hand side of (2.10) expresses that the test function φ is tangent to u in mean, locally along the trajectory of the underlying process $(s, X_s^{\kappa,t,x})$, for all possible control process $\kappa \in \mathcal{K}$. The latter is a new feature which appears in the present nonlinear case: while the linear case involves the tangency condition under the expectation operator $\mathbb{E}^{\mathbb{P}_0}$, the present nonlinear case requires the use of a sub linear expectation defined by an additional maximization with respect to all possible choices of control process $\kappa \in \mathcal{K}$.

This additional feature however does not alter the observation that the set of smooth functions which are tangent from below in (sublinear) mean is larger than $\overline{A}u$. Consequently, we may have used an alternative definition of viscosity solution with a richer family of test functions (defined by the right-hand side of (2.10)), and still get the same existence result. Similar to the case of the linear heat equation, the benefit from such a stronger definition may be that the uniqueness theory can be simplified by suitable use of the additional test functions.

2.4 Comparison of Viscosity Solutions

The uniqueness result of viscosity solution of second order fully nonlinear elliptic PDEs is usually obtained as a consequence of the comparison result, which corresponds to the maximum principle.

Definition 2.4 We say that the Eq. (1.7) satisfies comparison of bounded solutions if for all bounded viscosity subsolution u, and bounded viscosity supersolution v, we have

$$(u-v)(T,.) \leq 0 \quad \text{on } \mathbb{R}^d \quad \text{implies } u-v \leq 0 \quad \text{on } \mathrm{cl}(Q).$$

Comparison results for viscosity solution are available for a wide class of equations. The most accessible results are for the case of first order equations where the beautiful trick of doubling variables is remarkably efficient.

For second order equations, comparison results are more difficult and require to introduce a convenient regularization, typically by inf-convolution. The most general approach which covers possibly degenerate equations relies crucially on the Crandall-Ishii Lemma which provides the substitute of first and second order conditions at a local maximum point when the objective function is only upper semicontinuous.

In the context of uniformly elliptic equations, the argument of Caffarelli and Cabre [4] avoids the technique of doubling variables, but still relies crucially on the inf-convolution regularization. We refer to Wang [30] for the extension to the uniformly parabolic case which requires a more involved regularization technique.

All available comparison results for second order elliptic and parabolic equations use the restriction of test functions to paraboloids. This leads to the notion of superjets and subjets. For notations consistency, we continue our discussion with the parabolic case.

For $q \in \mathbb{R}$, $p \in \mathbb{R}^d$, and $\gamma \in \mathbb{S}_d$, we introduce the paraboloid function:

$$\phi^{q,\beta,\gamma}(t,x) := qt + p \cdot x + \frac{1}{2}\gamma x \cdot x, \qquad (t,x) \in Q.$$

For $u \in \mathrm{USC}(Q)$, let $(t_0, x_0) \in Q$, $\varphi \in \underline{A}u(t_0, x_0)$, define $q := \partial_t \varphi(t_0, x_0)$, $p := D\varphi(t_0, x_0)$, and $\gamma := D^2\varphi(t_0, x_0)$. Then, it follows from a Taylor expansion that:

$$u(t,x) \le u(t_0,x_0) + \phi^{q,p,\gamma}(t-t_0, x-x_0) + \circ\big(|t-t_0| + |x-x_0|^2\big).$$

Motivated by this observation, we introduce the *superjet* $J^+u(t_0, x_0)$ by

$$\begin{aligned} J^+u(t_0,x_0) := \big\{(q,p,\gamma) \in \mathbb{R} \times \mathbb{R}^d \times \mathbb{S}_d :\ & \text{for all } (t,x) \in Q \\ & u(t,x) \le u(t_0,x_0) + \phi^{q,p,\gamma}(t-t_0, x-x_0) \\ & + \circ\left(|t-t_0| + |x-x_0|^2\right)\big\}. \end{aligned} \tag{2.11}$$

Then, it can be proved that a function $u \in \mathrm{USC}(Q)$ is a viscosity subsolution of the equation (1.7) if and only if

$$F(t,x,u(t,x),p,\gamma) \le 0 \quad \text{for all } (q,p,\gamma) \in J^+u(t,x).$$

The nontrivial implication of the previous statement requires to construct, for every $(q, p, A) \in J^+u(t,x)$, a smooth test function φ such that the difference $(\varphi - u)$ has a local minimum at (t, x).

Similarly, we define the *subjet* $J^-v(t_0, x_0)$ of a function $v \in \mathrm{LSC}(Q)$ at the point $(t_0, x_0) \in Q$ by

$$J^{-}v(t_0,x_0) := \Big\{(q,p,\gamma) \in \mathbb{R}^d \times \mathbb{S}_d : \text{ for all } (t,x) \in Q \\ v(x) \geq v(t_0,x_0) + \phi^{q,p,\gamma}(t-t_0, x-x_0) \\ + \circ\left(|t-t_0| + |x-x_0|^2\right)\Big\}, \tag{2.12}$$

and $v \in \mathrm{LSC}(Q)$ is a viscosity supersolution of the equation (1.7) if and only if

$$F(t,x,v(t,x),p,\gamma) \geq 0 \quad \text{for all } (q,p,\gamma) \in J^{-}u(t,x).$$

By continuity considerations, we can even enlarge the semijets $J^{\pm}$ to the following closure

$$\tilde{J}^{\pm}w(t,x) := \Big\{(q,p,\gamma) \in \mathbb{R}^d \times \mathbb{S}_d : (t_n,x_n,w(t_n,x_n),q_n,p_n,\gamma_n) \\ \longrightarrow (t,x,w(t,x),q,p,\gamma) \text{ for some sequence} \\ (t_n,x_n,q_n,p_n,\gamma_n)_n \subset \mathrm{Graph}(J^{\pm}w)\Big\},$$

where $(t_n,x_n,q_n,p_n,\gamma_n) \in \mathrm{Graph}(J^{\pm}w)$ means that $(q_n,p_n,\gamma_n) \in J^{\pm}w(t_n,x_n)$. The following result is obvious, and provides an equivalent definition of viscosity solutions.

Proposition 2.5 *Let $u \in \mathrm{USC}(Q)$, and $v \in \mathrm{LSC}(Q)$.*

(i) *Assume that g is lower-semicontinuous. Then, u is a viscosity subsolution of (1.7) iff:*

$$-q - g(t,x,u(t,x),p,\gamma) \leq 0 \\ \textit{for all} \quad (t,x) \in Q \quad \textit{and } (q,p,\gamma) \in \tilde{J}^{+}u(t,x).$$

(ii) *Assume that g is upper-semicontinuous. Then, v is a viscosity supersolution of (1.7) iff:*

$$-q - g(t,x,v(t,x),p,\gamma) \geq 0 \textit{ for all } (t,x) \in Q \textit{ and } (q,p,\gamma) \in \tilde{J}^{-}v(t,x).$$

2.5 Stability of Viscosity Solutions

We conclude this section by reviewing the stability property of viscosity solutions. The following result is expressed in the context of our parabolic fully-nonlinear equation. However, the reader can see from its proof that it holds for general degenerate second order elliptic equations. We consider a family of equations parameterized by $\varepsilon > 0$:

$$-\,\partial_t u - g^{\varepsilon}(x,u,Du,D^2u) = 0 \quad \text{on } Q, \tag{2.13}$$

and we consider the convergence problem of a corresponding family of subsolutions $(u^\varepsilon)_{\varepsilon>0}$. The main ingredient for the stability result is the notion of relaxed semi limits introduced by Barles and Perthame [3]:

$$\overline{u}(t,x) := \limsup_{(\varepsilon,t',x')\to(0,t,x)} u^\varepsilon(t',x') \quad \text{and} \quad \overline{g}(\zeta) := \limsup_{(\varepsilon,\zeta')\to(0,\zeta)} g^\varepsilon(\zeta'),$$

where $\zeta = (t,x,y,z,\gamma)$. Notice that the semilimits here are taken both in the variables and the small parameter ε, and are finite whenever the functions of interest are locally bounded in the corresponding variables and the small parameter ε.

Theorem 2.6 *Let $u^\varepsilon \in USC(Q)$ be a viscosity subsolution of (2.13) for all $\varepsilon > 0$. Suppose that the maps $(\varepsilon,x) \longmapsto u_\varepsilon(x)$ and $(\varepsilon,\zeta) \longmapsto g^\varepsilon(\zeta)$ are locally bounded. Then, $\overline{u} \in USC(Q)$ is a viscosity subsolution of the equation*

$$-\partial_t\overline{u} - \overline{g}(x,\overline{u},D\overline{u},D^2\overline{u}) = 0 \quad \text{on } Q, \tag{2.14}$$

A similar statement holds for supersolutions.

Proof The fact that $\overline{u}$ is upper semicontinuous is an easy exercise. Let $\varphi \in \underline{A}\overline{u}(t,x)$. Without loss of generality, we may assume that the test function φ is strictly tangent from above to $\overline{u}$ at the point (t,x), i.e.

$$(\varphi - \overline{u})(t,x) < (\varphi - \overline{u})(t',x') \quad \text{for all } (t',x') \in Q,\ (t',x') \neq (t,x). \tag{2.15}$$

By definition of $\overline{u}$, there is a sequence $(\varepsilon_n,x_n) \in (0,1] \times \mathbb{R}^d$ such that

$$(\varepsilon_n,t_n,x_n) \longrightarrow (0,t,x) \quad \text{and } u^{\varepsilon_n}(t_n,x_n) \longrightarrow \overline{u}(t,x).$$

Let O be an open subset of Q containing (t,x) and $(t_n,x_n)_n$. Let $(\bar{t}_n,\bar{x}_n)$ be a minimizer of $\varphi - u^{\varepsilon_n}$ on $\mathrm{cl}(O)$. We claim that

$$(\bar{t}_n,\bar{x}_n) \longrightarrow (t,x) \quad \text{and } u^{\varepsilon_n}(\bar{t}_n,\bar{x}_n) \longrightarrow \overline{u}(t,x) \text{ as } n\to\infty. \tag{2.16}$$

Before verifying this, let us complete the proof. We first deduce that $(\bar{t}_n,\bar{x}_n)$ is an interior minimizer of the difference $(\varphi - u^{\varepsilon_n})$. Then, it follows from the viscosity subsolution property of u^{ε_n} that:

$$0 \geq \Big\{ -\partial_t\varphi - g^{\varepsilon_n}\Big(.,u^{\varepsilon_n},D\varphi,D^2\varphi\Big)\Big\}(\bar{t}_n,\bar{x}_n).$$

Then, taking limits on both sides, we see that

$$\begin{aligned} 0 &\geq -\partial_t\varphi(t,x) - \limsup_{n\to\infty} g^{\varepsilon_n}\Big(.,u^{\varepsilon_n},D\varphi,D^2\varphi\Big)(\bar{t}_n,\bar{x}_n) \\ &\geq \Big\{ -\partial_t\varphi - \overline{g}\Big(.,\overline{u},D\varphi,D^2\varphi\Big)\Big\}(t,x), \end{aligned}$$

by (2.16) and the definition of $\overline{g}$.

It remains to prove Claim 2.16. Recall that $(\bar{t}_n, \bar{x}_n)_n$ is valued in the compact set $\mathrm{cl}(O)$. Then, there is a subsequence, still named $(\bar{t}_n, \bar{x}_n)_n$, converging to some $(\bar{t}, \bar{x}) \in \mathrm{cl}(O)$. We now prove that $(\bar{t}, \bar{x}) = (t, x)$ and obtain the second claim in 2.16 as a by-product. By the fact that $(\bar{t}_n, \bar{x}_n)$ is a minimizer of $(\varphi - u^{\varepsilon_n})$ on $\mathrm{cl}(O)$, together with the definition of $\overline{u}$, we see that

$$
\begin{aligned}
0 \; = \; (\varphi - \overline{u})(t, x) &= \lim_{n\to\infty} \left(\varphi - u^{\varepsilon_n}\right)(t_n, x_n) \\
&\geq \limsup_{n\to\infty} \left(\varphi - u^{\varepsilon_n}\right)(\bar{t}_n, \bar{x}_n) \\
&\geq \liminf_{n\to\infty} \left(\varphi - u^{\varepsilon_n}\right)(\bar{t}_n, \bar{x}_n) \\
&\geq (\varphi - \overline{u})(\bar{t}, \bar{x}).
\end{aligned}
$$

We now obtain 2.16 from the fact that (t, x) is a strict minimizer of the difference $(\varphi - \overline{u})$. □

3 Viscosity Solution of Path-Dependent PDEs

We now turn to the main purpose of this paper, namely the theory of viscosity solutions for path-dependent PDEs 1.2:

$$
-\partial_t u(t, \omega) - G\big(t, \omega, u(t, \omega), \partial_\omega u(t, \omega), \partial^2_{\omega\omega} u(t, \omega)\big) = 0, \quad t < T, \quad \omega \in \Omega.
$$

where the generator $G : [0, T] \times \Omega \times \mathbb{R} \times \mathbb{R}^d \times \mathbb{S}_d \longrightarrow \mathbb{R}$ is a continuous map satisfying the ellipticity condition 1.3. We recall that $\Omega := \{\omega \in C^0([0, T], \mathbb{R}^d) : \omega_0 = 0\}$ is the underlying canonical space, $B_t(\omega) := \omega_t$, $t \in [0, T]$, is the canonical process, $\mathbb{P}_0$ is the Wiener measure, $\mathbb{F} := \{\mathcal{F}_t, t \in [0, T]\}$ with $\mathcal{F}_t = \sigma(B_s, s \leq t)$ is the natural filtration equipped with the pseudo-distance d defined in 1.1. Moreover, denote

$$
\Theta := [0, T) \times \Omega, \quad \overline{\Theta} := [0, T] \times \Omega,
$$

and $C^0(\overline{\Theta})$ is the set of continuous processes on $\overline{\Theta}$. We note that any $u \in C^0(\overline{\Theta})$ is $\mathbb{F}$-progressively measurable, namely $u(t, \omega) = u\big(t, (\omega_s)_{s\leq t}\big)$.

3.1 Differentiability

Before introducing the notion of viscosity solutions for this path-dependent PDE, we first need to specify the meaning of the time derivatives $\partial_t u(t, \omega)$ and the spatial derivatives $\partial_\omega u(t, \omega)$ and $\partial^2_{\omega\omega} u(t, \omega)$. Once these derivatives are clearly defined, we would have, on one hand a natural definition of classical solutions for the path-dependent PDE, and on the other hand a natural set of smooth functions to serve as test functions for our notion of viscosity solutions.

These path derivatives were first introduced by Dupire [10]. In particular, [10] defines the vertical derivatives (our spatial derivatives) by bumping the path at time t. While such a definition is natural in the larger space of discontinuous paths, our paths space Ω would require an extension of the map u to the set of discontinuous paths. We refer to Cont and Fournié [6] for this approach, where it is proved in particular that such a vertical derivative, if exists, does not depend on the choice of the extension of u to the set of discontinuous paths. Motivated by Remark 2.3, we adopt the following notion of smoothness.

Definition 3.1 [*Smooth processes*] Let $\mathcal{P}$ be a set of probability measures on Ω with B a $\mathbb{P}$-semimartingale for all $\mathbb{P} \in \mathcal{P}$. We say that $u \in C^{1,2}_{\mathcal{P}}(\Theta)$ if $u \in C^0(\overline{\Theta})$ and there exist processes $\alpha, Z, \Gamma \in C^0(\Theta)$ valued in $\mathbb{R}$, $\mathbb{R}^d$ and $\mathbb{S}_d$, respectively, such that:

$$du_t = \alpha_t dt + \frac{1}{2}\Gamma_t : d\langle B\rangle_t + Z_t dB_t, \quad \mathbb{P} - \text{a.s.} \quad \text{for all } \mathbb{P} \in \mathcal{P}.$$

The processes α, Z and Γ are called the time derivative, spacial gradient and spatial Hessian, respectively, and we denote $\partial_t u := \alpha$, $\partial_\omega u_t := Z_t$, $\partial^2_{\omega\omega} u_t := \Gamma_t$.

We observe that any $C^{1,2}$ process in the Dupire sense is in $C^{1,2}_{\mathcal{P}}(\Theta)$. This is an immediate consequence of the functional Itô formula proved in Dupire [10] and [6]. In particular, our notion of smooth processes is weaker than the corresponding one in [10]. We also note that, when $\mathcal{P}$ is rich enough, our path derivatives are unique.

Remark 3.2 The previous definition does not require that $\partial^2_{\omega\omega} u_t$ be the derivative (in some sense) of $\partial_\omega u_t$. This is very well illustrated by the following example communicated by Mete Soner. Let $d = 2$, and $u_t := \int_0^t B^1_s dB^2_s$ which is defined pathwise due to the results of Karandikar [23].

- Clearly $\partial_t u = 0$. Since $du_t = B^1_t dB^2_t$, under any semimartingale measure, we also deduce that $\partial_\omega u_t = (0, B^1_t)^{\mathrm{T}}$, and $\partial^2_{\omega\omega} u_t = 0_{\mathcal{S}_2}$. Hence $u \in C^{1,2}_{\mathcal{P}}(\Theta)$ for any subset $\mathcal{P}$ of the collection of all semimartingale measures for B.
- Let $\partial^{\mathrm{D}}_\omega u_t$ and $\partial^{\mathrm{D}^2}_{\omega\omega} u_t$ denote the vertical first and second derivatives in the Dupire sense. Direct calculation reveals that $\partial^{\mathrm{D}}_\omega u_t = (0, B^1_t)^{\mathrm{T}} = \partial_\omega u_t$. However,

$$\partial^{\mathrm{D}^2}_{\omega\omega} u_t = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix},$$

which is not symmetric !

- However, we need to point out that in this example u does not belong to $C^0(\overline{\Theta})$.
- We complement this example by the following observation from a private communication with Bruno Dupire. By considering the Dupire vertical derivative as originally defined on the set of discontinuous paths, we see by direct calculation that $\partial^{\mathrm{D}^2}_{\omega\omega} u_t = 0_{\mathcal{S}_2} = \partial^2_{\omega\omega} u_t$.

Definition 3.3 [*Classical solution*] Let $\mathcal{P}$ be a set of probability measures on Ω with B a $\mathbb{P}$-semimartingale for all $\mathbb{P} \in \mathcal{P}$.

(i) $u \in C^{1,2}_{\mathcal{P}}(\Theta)$ is a $\mathcal{P}$-classical subsolution of the path-dependent PDE (1.2) if

$$-\partial_t u - G\big(., u, \partial_\omega u, \partial^2_{\omega\omega} u\big) \le 0 \quad \text{on } \Theta.$$

(ii) $v \in C^{1,2}_{\mathcal{P}}(\Theta)$ is a $\mathcal{P}$-classical supersolution of the path-dependent PDE (1.2) if

$$-\partial_t v - G\big(., v, \partial_\omega v, \partial^2_{\omega\omega} v\big) \ge 0 \quad \text{on } \Theta.$$

(iii) A $\mathcal{P}$-classical solution of (1.2) is both classical subsolution and supersolution.

Example 3.4 Let $u(t,\omega) := \mathbb{E}^{\mathbb{P}_0}[\xi|\mathcal{F}_t]$ for some $\xi \in \mathbb{L}^1(\mathbb{P}_0, \mathcal{F}_T)$, and assume $u \in C^{1,2}_{\mathbb{P}_0}(\Theta)$. By definition, this implies that

$$du_t = \big(\partial_t u_t + \frac{1}{2}\partial^2_{\omega\omega} u_t\big)dt + \partial_\omega u_t dB_t, \qquad \mathbb{P}_0 - \text{a.s.}$$

Since the process u is a martingale, it follows that:

$$\partial_t u_t + \frac{1}{2}\partial^2_{\omega\omega} u_t = 0, \quad (t,\omega) \in \Theta.$$

In other words, u is a $\mathbb{P}_0$- classical solution of the path-dependent heat equation.

Example 3.5 For $\xi \in \mathbb{L}^2(\mathbb{P}_0, \mathcal{F}_T)$, consider the backward stochastic differential equation:

$$du_t = -F_t(\omega, u_t, Z_t)dt + Z_t dB_t, \quad u_T = \xi, \ \mathbb{P}_0 - \text{a.s.}$$

where $F : [0,T] \times \Omega \times \mathbb{R} \times \mathbb{R}^d \longrightarrow \mathbb{R}$ is continuous, uniformly Lipschitz in (y,z), with $F(0,0)$ a square integrable process. Assume $u \in C^{1,2}_{\mathbb{P}_0}(\Theta)$. By definition, this implies that

$$du_t = \big(\partial_t u_t + \frac{1}{2}\partial^2_{\omega\omega} u_t\big)dt + \partial_\omega u_t dB_t = -F_t(\omega, u_t, Z_t)dt + Z_t dB_t, \qquad \mathbb{P}_0 - \text{a.s.}$$

Identifying the martingale terms, we see that $\partial_\omega u_t = Z_t$. Next, identifying the drift term, it follows that u is a $\mathbb{P}_0$-classical solution of the path-dependent semilinear PDE:

$$-\partial_t u_t - \frac{1}{2}\partial^2_{\omega\omega} u_t - F_t(\omega, u_t, \partial_\omega u_t) = 0, \quad (t,\omega) \in \Theta.$$

Remark 3.6 (i) In the Markovian case, strong regularity results are induced by the ellipticity of the underlying diffusion coefficient. The simplest example is when the diffusion is the identity matrix. Let $u(t,x) := \mathbb{E}^{\mathbb{P}_0}[h(B_T)|B_t = x]$. Then $u \in C^\infty([0,T)\times\mathbb{R}^d)$.

(ii) The path-dependency induces specific non-smoothness as outlined by the following example. Let $u(t,\omega) := \mathbb{E}^{\mathbb{P}_0}[B_{\frac{T}{2}}|\mathcal{F}_t] = \omega_{t\wedge\frac{T}{2}}$ for all $t\in[0,T]$. Clearly, $\partial_t u_t = 0$, and $du_t = \mathbf{1}_{t\le\frac{T}{2}}dB_t$ implying that $\partial_\omega u_t$ is not continuous. Hence $u \notin C^{1,2}(\Theta)$.

3.2 Viscosity Solutions of Path-Dependent PDEs

3.2.1 Notations

First recall our canonical setting $(\Omega, B, \mathbb{F}, \mathbb{P}_0)$. We denote by $\mathcal{T}$ the set of all $\mathbb{F}$-stopping times, $\mathcal{T}^+ \subset \mathcal{T}$ the collection of all strictly positive stopping times, and $\mathcal{T}^t \subset \mathcal{T}$ the subset of the $\mathbb{F}$-stopping times larger than t.

For $\omega, \omega' \in \Omega$ and $t\in[0,T]$, we define

$$(\omega\otimes_t\omega')_s := \omega_s\mathbf{1}_{\{s<t\}} + (\omega_t + \omega'_{s-t})\mathbf{1}_{\{s\ge t\}}.$$

Let $\xi:\Omega\to\mathbb{R}$ be $\mathcal{F}_T$-measurable random variable. For any $(t,\omega)\in\Theta$, define

$$\xi^{t,\omega}(\omega') := \xi\big(\omega\otimes_t\omega'\big) \quad \text{for all } \omega'\in\Omega.$$

Clearly, $\xi^{t,\omega}$ is $\mathcal{F}_{T-t}$-measurable, and thus $\mathcal{F}_T$-measurable. Similarly, given a process X defined on Ω, we denote:

$$X^{t,\omega}_s(\omega') := X_{t+s}(\omega\otimes_t\omega'), \quad \text{for } s\in[0,T-t].$$

Clearly, if X is $\mathbb{F}$-adapted, then so is $X^{t,\omega}$.

Let $\mathcal{P}$ be a family of probability measures on Ω. We also introduce the sublinear and superlinear expectation operators associated to $\mathcal{P}$:

$$\overline{\mathcal{E}}^{\mathcal{P}} := \sup_{\mathbb{P}\in\mathcal{P}}\mathbb{E}^{\mathbb{P}} \quad \text{and } \underline{\mathcal{E}}^{\mathcal{P}} := \inf_{\mathbb{P}\in\mathcal{P}}\mathbb{E}^{\mathbb{P}}.$$

3.2.2 Definition of Viscosity Solutions

We recall that the nonlinearity G satisfies the ellipticity condition in (1.3). We assume in addition that it is L_0-Lipschitz with respect to the arguments (y,z,γ), uniformly in (t,ω):

$$\big|G(t,\omega,y,z,\gamma) - G(t,\omega,y',z',\gamma')\big| \le L_0\big(|y-y'|+|z-z'|+|\gamma-\gamma'|\big) \tag{3.1}$$

for all $y,y'\in\mathbb{R}$, $z,z'\in\mathbb{R}^d$, $\gamma,\gamma'\in\mathbb{S}_d$, $(t,\omega)\in\Theta$.

Ou Definition 3.1 of smooth processes involves a family of probability measures that we intentionally did not discuss so far. We now introduce a specific family of semimartingale measures which will be needed for our notion of viscosity solutions.

Definition 3.7 By $\mathcal{P}_L$ we denote the collection of all continuous semimartingale measures $\mathbb{P}$ on Ω whose drift and diffusion characteristics are bounded by L and $\sqrt{2L}$, respectively.

We refer to [12] for properties of this class. In our subsequent analysis, the family of probability measures $\mathcal{P}$ is a subset of $\mathcal{P}_L$ for some $L > 0$.

Motivated by the crucial observations of Sects. 2.2 and 2.3, we introduce the sets of test processes: a

$$\underline{\mathcal{A}}^{\mathcal{P}} u_t(\omega) := \Big\{\varphi \in C^{1,2}_{\mathcal{P}}(\Theta) : (\varphi - u^{t,\omega})_0 = \min_{\tau \in \mathcal{T}} \underline{\mathcal{E}}^{\mathcal{P}}\big[(\varphi - u^{t,\omega})_{\tau \wedge \mathrm{H}}\big] \\ \text{for some } \mathrm{H} \in \mathcal{H}^{\mathcal{P}}_{+}\Big\},$$

$$\overline{\mathcal{A}}^{\mathcal{P}} v_t(\omega) := \Big\{\varphi \in C^{1,2}_{\mathcal{P}}(\Theta) : (\varphi - v^{t,\omega})_0 = \max_{\tau \in \mathcal{T}} \overline{\mathcal{E}}^{\mathcal{P}}\big[(\varphi - v^{t,\omega})_{\tau \wedge \mathrm{H}}\big] \\ \text{for some } \mathrm{H} \in \mathcal{H}^{\mathcal{P}}_{+}\Big\},$$

where $\mathcal{H}^{\mathcal{P}}_{+} \subset \mathcal{T}^{+}$ satisfies the following properties, for all $\mathrm{H}, \mathrm{H}' \in \mathcal{H}^{\mathcal{P}}_{+}$:

$$\begin{aligned} &\textbf{H1} \text{ (stability by minimization) } \mathrm{H} \wedge \mathrm{H}' \in \mathcal{H}^{\mathcal{P}}_{+}, \\ &\textbf{H2} \text{ (stability by localization) } \mathrm{H} \wedge \mathrm{H}_\varepsilon \in \mathcal{H}^{\mathcal{P}}_{+}, \\ &\quad \text{where } \mathrm{H}_\varepsilon := \varepsilon \wedge \inf\big\{t > 0 : |B|_t \geq \varepsilon\big\}. \end{aligned} \tag{3.2}$$

Later, we will call H the localizing stopping time (or the localization) of the corresponding test process φ.

Definition 3.8 [*Viscosity solution of path-dependent PDE*] Let $u, v \in C^0(\overline{\Theta})$.

(i) u is a $\mathcal{P}$-viscosity subsolution of (1.2) if:

$$\big\{-\partial_t\varphi - G\big(., u, \partial_\omega\varphi, \partial^2_{\omega\omega}\varphi\big)\big\}(t,\omega) \leq 0 \text{ for all } (t,\omega) \in \Theta,\ \varphi \in \underline{\mathcal{A}}^{\mathcal{P}} u_t(\omega).$$

(ii) v is a $\mathcal{P}$-viscosity supersolution of (1.2) if:

$$\big\{-\partial_t\varphi - G\big(., v, \partial_\omega\varphi, \partial^2_{\omega\omega}\varphi\big)\big\}(t,\omega) \geq 0 \text{ for all } (t,\omega) \in \Theta,\ \varphi \in \overline{\mathcal{A}}^{\mathcal{P}} v_t(\omega).$$

(iii) A $\mathcal{P}$-viscosity solution of (1.2) is both a $\mathcal{P}$-subsolution and a $\mathcal{P}$-supersolution.

Remark 3.9 in the Markovian case, we may as well use the last definition as an alternative to the standard notion of viscosity solutions. Compared to the standard notion reviewed in Sect. 2, we see that any $\phi \in \underline{A}u(t,x)$ induces a process $\varphi(t,\omega) := \phi(t,\omega_t)$ which obviously lies in $\underline{\mathcal{A}}^{\mathcal{P}} u_t(\omega)$. However, even in the Markovian case

$u_t(\omega) = u(t, \omega_t)$, a test process in $\underline{\mathcal{A}}^{\mathcal{P}} u_t(\omega)$ does not necessarily induce a test function in $\underline{A}u(t, \omega_t)$. Thus, our notion of viscosity solution involves more test functions than the standard notion. A viscosity subsolution/supersolution in sense of Definition 3.8 is restricted by a richer family of test functions. Consequently:

- under our definition, we may hope to take advantage of the richer family of test functions in order to obtain an easier uniqueness proof,
- under our definition, the existence problem is more restricted than under the standard theory of viscosity solutions.

Remark 3.10 Due to the stability property of the set $\mathcal{H}_+^{\mathcal{P}}$ by localization, the viscosity property introduced in Definition 3.8 is a local property. Indeed, in order to check the viscosity property of u at (t, ω), it suffices to know the value of $u^{t,\omega}$ on $[0, \mathrm{H}_\varepsilon]$ for an arbitrarily small $\varepsilon > 0$. In particular, since u and φ are locally bounded, there is no integrability issue in the definition of the set of test functions $\underline{\mathcal{A}}^{\mathcal{P}}$ and $\overline{\mathcal{A}}^{\mathcal{P}}$.

3.3 Semijets Definition and Punctual Differentiability

Similar to the standard notion of viscosity solutions in finite-dimensional spaces, we will now prove that we may reduce our Definition 3.8 to paraboloids:

$$\phi_s^{q,p,\gamma}(\omega) := qs + p \cdot \omega_s + \frac{1}{2}\gamma : \omega_s \omega_s^{\mathrm{T}}, \quad s \in [0, T-t], \quad \omega \in \Omega,$$

for some $(q, p, \gamma) \in \mathbb{R} \times \mathbb{R}^d \times \mathbb{S}_d$. We then introduce the corresponding subjet and superjet:

$$\underline{\mathcal{J}}^{\mathcal{P}} u_t(\omega) := \left\{(q, p, \gamma) \in \mathbb{R} \times \mathbb{R}^d \times \mathbb{S}_d : \phi^{q,p,\gamma} \in \underline{\mathcal{A}}^{\mathcal{P}} u_t(\omega)\right\},$$
$$\overline{\mathcal{J}}^{\mathcal{P}} v_t(\omega) := \left\{(q, p, \gamma) \in \mathbb{R} \times \mathbb{R}^d \times \mathbb{S}_d : \phi^{q,p,\gamma} \in \overline{\mathcal{A}}^{\mathcal{P}} v_t(\omega)\right\}.$$

Proposition 3.11 *Let $\mathcal{P} \subset \mathcal{P}_L$ for some $L > 0$. A process $u \in C^0(\overline{\Theta})$ is a $\mathcal{P}$-viscosity subsolution of (1.2) if and only if:*

$$\begin{aligned} & -q - G(t, \omega, u_t(\omega), p, \gamma) \le 0 \\ & \textit{for all } \ (t, \omega) \in \Theta, \ (q, p, \gamma) \in \underline{\mathcal{J}}^{\mathcal{P}} u_t(\omega). \end{aligned} \tag{3.3}$$

The corresponding statement holds for supersolutions.

Proof We focus on the nontrivial direction, assuming that (3.3) holds. For $(t, \omega) \in \Theta$ and $\varphi \in \underline{\mathcal{A}}^{\mathcal{P}} u_t(\omega)$, we have to prove that $-q - G(t, \omega, u_t(\omega), p, \gamma) \le 0$, where

$$q := \partial_t \varphi(t, \omega), \quad p := \partial_\omega \varphi(t, \omega), \quad \gamma := \partial^2_{\omega\omega} \varphi(t, \omega).$$

Without loss of generality, we take $(t, \omega) = (0, 0)$. For $\varepsilon > 0$, we denote $q_\varepsilon := q + \varepsilon(1 + 2L)$, and $\phi := \phi^{q_\varepsilon, p, \gamma}$. By the smoothness of φ, we may find $\delta_\varepsilon > 0$, such that

$$|\partial_t \varphi - q| \le \varepsilon, |\partial_\omega \varphi - p - \gamma\omega_t| \le \varepsilon, \quad \text{and } |\partial^2_{\omega\omega}\varphi - \gamma| \le \varepsilon$$
$$\text{on } \mathcal{Q}_\varepsilon := \{(t, \omega) : t \le \delta_\varepsilon, |\omega|_t \le \delta_\varepsilon\}.$$

Let H be the stopping time corresponding to φ, and set $\mathrm{H}_\varepsilon := \mathrm{H} \wedge \inf\{t > 0 : (t, \omega) \notin \mathcal{Q}_\varepsilon\}$. Then, for all stopping time $\tau \in \mathcal{T}_0$:

$$\begin{aligned}
&(\phi - u)_0 - \underline{\mathcal{E}}^{\mathcal{P}}\big[(\phi - u)_{\tau\wedge \mathrm{H}_\varepsilon}\big] \\
&\le (\varphi - u)_0 - \underline{\mathcal{E}}^{\mathcal{P}}\big[(\varphi - u)_{\tau\wedge \mathrm{H}_\varepsilon}\big] + \overline{\mathcal{E}}^{\mathcal{P}}\big[(\varphi - \varphi_0 - \phi)_{\tau\wedge \mathrm{H}_\varepsilon}\big] \\
&\le \overline{\mathcal{E}}^{\mathcal{P}}\Big[\int_0^{\tau\wedge \mathrm{H}_\varepsilon}(\partial_t\varphi_s - q_\varepsilon)ds + (\partial_\omega\varphi_s - p - \gamma B_s)dB_s + (\partial^2_{\omega\omega}\varphi_s - \gamma)d\langle B\rangle_s\Big].
\end{aligned}$$

Since $\mathcal{P} \subset \mathcal{P}_L$, it follows that the integral term inside the nonlinear expectation $\overline{\mathcal{E}}^{\mathcal{P}}$ is non-positive, implying that $(\phi - u)_0 - \underline{\mathcal{E}}^{\mathcal{P}}\big[(\phi - u)_{\tau\wedge \mathrm{H}_\varepsilon}\big] \le 0$. Consequently $(q_\varepsilon, p, \gamma) \in \underline{\mathcal{J}}^{\mathcal{P}} u_0$ and therefore $-q_\varepsilon - G(t, \omega, u_t(\omega), p, \gamma) \le 0$ by (3.3). The required result follows by sending $\varepsilon \searrow 0$. □

Proposition 3.12 *For $u^i, v^i \in C^0(\bar\Theta)$, $i = 0, 1$, we have*

$$\underline{\mathcal{J}}^{\mathcal{P}} u^0_t(\omega) + \underline{\mathcal{J}}^{\mathcal{P}} u^1_t(\omega) \subset \underline{\mathcal{J}}^{\mathcal{P}}(u^0 + u^1)_t(\omega) \text{ and}$$
$$\overline{\mathcal{J}}^{\mathcal{P}} v^0_t(\omega) + \overline{\mathcal{J}}^{\mathcal{P}} v^1_t(\omega) \subset \overline{\mathcal{J}}^{\mathcal{P}}(v^0 + v^1)_t(\omega)$$

Proof We only report the argument for the subjets. Let $\theta^i = (q^i, p^i, \gamma^i) \in \underline{\mathcal{J}}^{\mathcal{P}} u^i_t(\omega)$, $i = 0, 1$. By definition, this means that the corresponding paraboloids $\phi^{\theta^i} \in \underline{\mathcal{A}}^{\mathcal{P}} u^i_t(\omega)$, i.e. there is $\mathrm{H}^i \in \mathcal{H}^{\mathcal{P}}_+$ such that

$$-u^i_t \le \mathbb{E}^{\mathbb{P}}\big[(\phi^{\theta^i} - (u^i)^{t,\omega})_{\tau\wedge \mathrm{H}^i}\big] \quad \text{for all } \tau \in \mathcal{T} \text{ and } \mathbb{P} \in \mathcal{P}.$$

With $\mathrm{H} := \mathrm{H}^0 \wedge \mathrm{H}^1 \in \mathcal{H}^{\mathbb{P}}$, this implies that

$$-(u^0 + u^1)_t \le \mathbb{E}^{\mathbb{P}}\big[\big(\phi^{\theta^0} + \phi^{\theta^1} - (u^0 + u^1)^{t,\omega}\big)_{\tau\wedge \mathrm{H}}\big] \quad \text{for all } \tau \in \mathcal{T} \text{ and } \mathbb{P} \in \mathcal{P}.$$

Since $\phi^{\theta^0} + \phi^{\theta^1} = \phi^{\theta^0 + \theta^1}$, this shows that $\theta^0 + \theta^1 \in \underline{\mathcal{J}}^{\mathcal{P}}(u^0 + u^1)_t(\omega)$. □

3.4 Punctual Differentiability

The following notion is adapted from Caffarelli and Cabre [4].

Definition 3.13 A process u is $\mathcal{P}$-punctually $C^{1,2}$ at some point $(t,\omega)\in\Theta$ if

$$\mathcal{J}^{\mathcal{P}}u_t(\omega) := \mathrm{cl}\big(\underline{\mathcal{J}}^{\mathcal{P}}u_t(\omega)\big)\cap \mathrm{cl}\big(\overline{\mathcal{J}}^{\mathcal{P}}u_t(\omega)\big) \;\neq\; \emptyset.$$

The next (immediate) result states that the viscosity property reduces to a point-wise property at points of punctual differentiability.

Proposition 3.14 *Assume that the nonlinearity G is continuous in (z,γ), and let $u\in C^0(\overline{\Theta})$ be a $\mathcal{P}$-viscosity solution of (1.2). Then, if u is $\mathcal{P}$-punctually $C^{1,2}$ at some point $(t,\omega)\in\Theta$, we have*

$$-q-G(t,\omega,u(t,\omega),p,\gamma)=0 \quad \textit{for all } (q,p,\gamma)\in\mathcal{J}^{\mathcal{P}}u_t(\omega).$$

For our subsequent analysis, we need the following additivity property of punctual differentiability, which is a direct consequence of Proposition 3.12.

Proposition 3.15 *Let u, v be two processes which are $\mathcal{P}$-punctually $C^{1,2}$ at some point $(t,\omega)\in\Theta$. Then, $u+v$ is $\mathcal{P}$-punctually $C^{1,2}$ at (t,ω), and*

$$\mathcal{J}^{\mathcal{P}}u_t(\omega)+\mathcal{J}^{\mathcal{P}}u_t(\omega)\subset\mathcal{J}^{\mathcal{P}}(u+v)_t(\omega).$$

3.5 Consistency of Path-Dependent Viscosity Solutions

We conclude this definition subsection by proving consistency of our notion of viscosity solution with classical solutions.

Proposition 3.16 *Let G be continuous, elliptic and uniformly L_0-Lipschitz- continuous in (y,z,γ). Let $\mathcal{P}_{L_0}\subset\mathcal{P}\subset\mathcal{P}_L$ for some $L\geq L_0$. Then, for $u\in C^{1,2}_{\mathcal{P}}(\Theta)$, the following are equivalent:*

(i) *u is a $\mathcal{P}$-classical subsolution (reps. supersolution) for some $L>0$,*
(ii) *u is a $\mathcal{P}$-viscosity subsolution (reps. supersolution).*

Proof We only report the proof of the subsolution property. The supersolution property follows by the same line of argument. If u is a $\mathcal{P}$-viscosity subsolution and $u\in C^{1,2}_{\mathcal{P}}(\Theta)$, then it is clear that $u^{t,\omega}\in\underline{\mathcal{A}}^{\mathcal{P}}u_t(\omega)$ for all $t<T$ and $\omega\in\Omega$, and therefore u is a $\mathcal{P}$-classical subsolution.

We next assume that u is a classical subsolution, and we assume to the contrary that $5c := -\partial_t\varphi - G(.,u,\partial_\omega\varphi,\partial^2_{\omega\omega}\varphi)>0$ for some $t<T$, $\omega\in\Omega$, and $\varphi\in\underline{\mathcal{A}}^{\mathcal{P}}u_t(\omega)$. Without loss of generality, we may assume $(t,\omega)=(0,0)$. Let $\bar{\alpha}\in\mathbb{R}^d$, $\bar{\bar{\beta}}\in\mathbb{S}_d$ be

arbitrary constants with $|\bar\alpha| \le L_0$ and $\frac{1}{2}\mathrm{Tr}[\bar\beta^2] \le L_0$, to be fixed later, and denote by $\bar{\mathbb{P}} := \mathbb{P}^{\bar\alpha,\bar\beta}$ the corresponding probability measure in $\mathcal{P}$, and $\bar{\mathcal{L}} := \bar\alpha\cdot\partial_\omega + \frac{1}{2}\bar\beta^2 : \partial^2_{\omega\omega}$. By the continuity of G, and the fact that $u, \varphi \in C^{1,2}$,

$$-\partial_t\varphi - G_0(u_0, \partial_\omega\varphi_0, \partial^2_{\omega\omega}\varphi_0) \ge 4c,\ |\bar{\mathcal{L}}\varphi - \bar{\mathcal{L}}\varphi_0| \le c,$$
$$\text{and } \big|G(u, \partial_\omega u, \partial^2_{\omega\omega}u) - G_0(u_0, \partial_\omega u_0, \partial^2_{\omega\omega}u_0)\big| \le c,\ |\bar{\mathcal{L}}u - \bar{\mathcal{L}}u_0| \le c, \quad \text{on } [0, \mathrm{H}_\varepsilon],$$

for $\varepsilon > 0$ sufficiently small, where $\mathrm{H}_\varepsilon := \varepsilon \wedge \inf\{s > 0 : |\omega_s| \ge \varepsilon\}$. Since u is a $\mathcal{P}$-classical subsolution, we compute for every $\tau \in \mathcal{T}$ that

$$\begin{aligned}
(\varphi - u)_0 - \mathbb{E}^{\bar{\mathbb{P}}}\big[(\varphi - u)_{\tau\wedge\mathrm{H}_\varepsilon}\big] &= \mathbb{E}^{\bar{\mathbb{P}}}\Big[\int_0^{\tau\wedge\mathrm{H}_\varepsilon} d(u-\varphi)_s\Big] \\
&= \mathbb{E}^{\bar{\mathbb{P}}}\Big[\int_0^{\tau\wedge\mathrm{H}_\varepsilon} \big\{\partial_t(u-\varphi)_s + \bar{\mathcal{L}}(u-\varphi)_s\big\}ds\Big] \\
&\ge \mathbb{E}^{\bar{\mathbb{P}}}\Big[\int_0^{\tau\wedge\mathrm{H}_\varepsilon}\big\{G_0(u_0, \partial_\omega\varphi_0, \partial^2_{\omega\omega}\varphi_0) \\
&\qquad -G_0(u_0, \partial_\omega u_0, \partial^2_{\omega\omega}u_0) + \bar{\mathcal{L}}(u-\varphi)_0\big\}ds\Big] \\
&\quad +c\bar{\mathbb{P}}[\tau \wedge \mathrm{H}_\varepsilon].
\end{aligned}$$

By the definition of $\mathcal{P}$, we may find $\bar\alpha$ so that $G_0(u_0, \partial_\omega\varphi_0, \partial^2_{\omega\omega}\varphi_0) - G_0(u_0, \partial_\omega u_0, \partial^2_{\omega\omega}u_0) + \bar{\mathcal{L}}(u-\varphi)_0 = 0$. Then, whenever $\tau > 0$, $\bar{\mathbb{P}}$-a.s., we have $(\varphi - u)_0 > \mathbb{E}^{\bar{\mathbb{P}}}\big[(\varphi - u)_{\tau\wedge\mathrm{H}_\varepsilon}\big]$, contradicting the fact that $\varphi \in \underline{\mathcal{A}}^{\mathcal{P}}u_0$. □

4 Wellposedness of the Path-Dependent Heat Equation

In this section, we consider the heat equation

$$-\partial_t u - \frac{1}{2}\mathrm{Tr}\big[\partial^2_{\omega\omega}u\big] = 0 \tag{4.1}$$

where, for simplicity, the diffusion matrix is taken to be the identity matrix. We recall that $\mathbb{P}_0$ denotes the Wiener measure. In addition to the previous notations, we denote $\mathbb{F}^*$ as the filtration augmented by all $\mathbb{P}_0$-null sets. Also, denote $\mathcal{T}_*$ (resp. $\mathcal{T}^t_*$) as the set of all $\mathbb{F}^*$-stopping times taking values in $[0, T]$ (resp. $[t, T]$). In this section, we take

$$\mathcal{P} := \{\mathbb{P}_0\} \text{ and } \mathcal{H}^{\mathcal{P}}_+ := \mathcal{T}^+.$$

In this section about the heat equation, the relevant space for our comparison result is

$$C^0_{2,\mathbb{P}_0}(\Theta,\mathbb{R}) := \Big\{ u \in C^0(\bar{\Theta},\mathbb{R}) : \mathbb{E}^{\mathbb{P}_0}\Big[\sup_{t+s\le T} \big|u_s^{t,\omega}\big|^2\Big] < \infty \\ \text{for all } (t,\omega)\in\Theta \Big\}.$$

4.1 Facts from Optimal Stopping Theory

Let $X \in C^0_{2,\mathbb{P}_0}(\Theta,\mathbb{R})$. Our main result uses the Snell envelope characterization of the optimal stopping stopping problem:

$$V_0 := \sup_{\tau\in\mathcal{T}_*} \mathbb{E}^{\mathbb{P}_0}[X_\tau],$$

The standard characterization of this problem uses the dynamic formulation of this problem:

$$Y_t^0 := \operatorname*{ess\,sup}_{\tau\in\mathcal{T}_*^t} \mathbb{E}^{\mathbb{P}_0}\big[X_{\tau\wedge T}\big|\mathcal{F}_t\big], \qquad 0\le t\le T,$$

so that $Y_0^0 = V_0$ by the Blumenthal zero-one law. In this context, an optimal stopping rule is well-known to be defined by the first hitting time

$$\tau^* := \inf\big\{t\ge 0 : Y_t^0 = X_t\big\}.$$

In addition to the standard result, we need an additional refinement by introducing the variable:

$$\mathcal{Y}_\tau(\omega) := \sup_{\theta\in\mathcal{T}_*} \mathbb{E}^{\mathbb{P}_0}\big[X_\theta^{\tau(\omega),\omega}\big], \qquad \text{for all } \tau\in\mathcal{T}_*, \omega\in\Omega.$$

Theorem 4.1 *Let $X \in C^0_{2,\mathbb{P}_0}(\Theta,\mathbb{R})$. Then, there exists an $\mathbb{F}$-adapted version Y of Y^0 satisfying:*

$$Y_{\tau\wedge T} = \mathcal{Y}_{\tau\wedge T},\ \mathbb{P}_0\text{-a.s.} \quad \text{for all } \tau\in\mathcal{T}_*.$$

Moreover, Y is a pathwise continuous $\mathbb{P}_0$-supermartingale, $Y_{\wedge\tau^*}$ is a $\mathbb{P}_0$-martingale, and τ^* is an optimal stopping rule.

This result follows from the more general Theorem 5.2 below.

4.2 Existence, Comparison, and Uniqueness

Definition 4.2 An $\mathbb{F}$-progressively measurable process m is a pathwise $\mathbb{P}_0$-submartingale (resp. supermartingale) if, for any $(t,\omega) \in \Theta$, we have

$$m_t(\omega) - \mathbb{E}^{\mathbb{P}_0}\big[m^{t,\omega}_\tau\big] \le 0 \text{ (resp. } \ge 0) \quad \text{for all } \tau \in \mathcal{T}.$$

Lemma 4.3 *Let $u \in C^0_{2,\mathbb{P}_0}(\Theta,\mathbb{R})$, $(t,\omega) \in \Theta$, and $\mathrm{H} \in \mathcal{T}^+$, be such that $u_t(\omega) > \mathbb{E}^{\mathbb{P}}\big[u^{t,\omega}_{\mathrm{H}}\big]$. Then,*

$$\begin{aligned} &0 \in \underline{\mathcal{A}}^{\mathbb{P}_0} u_{t+t^*}(\omega \otimes_t \omega^*) \\ &\quad \textit{for some } (t^*,\omega^*) \ \textit{ with the localization } \mathrm{H}^* := \mathrm{H}^{t^*,\omega^*} - t^* \in \mathcal{T}^+. \end{aligned}$$

Proof Without loss of generality, we may assume that $(t,\omega) = (0,0)$. Consider the optimal stopping problem $V_0 := \sup_{\tau \in \mathcal{T}_*} \mathbb{E}^{\mathbb{P}_0}\big[u_{\tau\wedge\mathrm{H}}\big]$. Set $X_s := u_{s\wedge\mathrm{H}}$ and let Y be the $\mathbb{F}$-adapted Snell envelope as introduced in Theorem 4.1, τ^* the corresponding optimal stopping rule. From the strict inequality $u_0 > \mathbb{E}^{\mathbb{P}_0}\big[u_{\mathrm{H}}\big]$, it follows that $\mathbb{P}_0[\tau^* < \mathrm{H}] > 0$. By Theorem 4.1, we also have $Y_{\tau^*} = \mathcal{Y}_{\tau^*}$, $\mathbb{P}_0$-a.s. We may then find ω^* such that $t^* := \tau^*(\omega^*) < \mathrm{H}(\omega^*)$, and:

$$u_{t^*}(\omega^*) = Y_{t^*}(\omega^*) \ = \ \max_{\tau \in \mathcal{T}} \mathbb{E}^{\mathbb{P}_0}\big[(u_{\mathrm{H}\wedge\cdot})^{t^*,\omega^*}_\tau\big],$$

By definition of $\underline{\mathcal{A}}^{\mathbb{P}_0} u$, this is exactly the required result. □

The main result of this section is the following.

Theorem 4.4 *For a process $u \in C^0_{2,\mathbb{P}_0}(\Theta,\mathbb{R})$, the following are equivalent:*

(i) *u is a pathwise $\mathbb{P}_0$-submartingale (resp. supermartingale),*
(ii) *u is $\mathbb{P}_0$-viscosity subsolution (resp. supersolution) of the path-dependent heat equation* (4.1).

Proof (i) $\Longrightarrow$ (ii): For arbitrary $(t,\omega) \in \Theta$ and $\varphi \in \underline{\mathcal{A}}^{\mathbb{P}_0} u_t(\omega)$, we have for some $\mathrm{H} \in \mathcal{T}_+$:

$$\varphi_0 - u_t(\omega) \le \mathbb{E}^{\mathbb{P}_0}\big[\varphi_{\tau\wedge\mathrm{H}} - u^{t,\omega}_{\tau\wedge\mathrm{H}}\big] \quad \text{for all } \tau \in \mathcal{T}.$$

For all $\varepsilon > 0$, define $\mathrm{H}_\varepsilon(\omega') := \mathrm{H}(\omega') \wedge \inf\{s \ge 0 : |\omega'_s| \ge \varepsilon\}$. Then, since u is a pathwise $\mathbb{P}_0$-submartingale, it follows that

$$0 \ \ge \ u_t(\omega) - \mathbb{E}^{\mathbb{P}_0}\big[u^{t,\omega}_{\mathrm{H}_\varepsilon}\big] \ge \varphi_0 - \mathbb{E}^{\mathbb{P}_0}\big[\varphi_{\mathrm{H}_\varepsilon}\big] \ = \ \mathbb{E}^{\mathbb{P}_0}\Big[\int_0^{\mathrm{H}_\varepsilon} (-\partial_t\varphi - \frac{1}{2}{:}\,\partial^2_{\omega\omega}\varphi)_s ds\Big]$$

by the smoothness of φ. Sending $\varepsilon \searrow 0$, we see that $(-\partial_t\varphi - \frac{1}{2}\sigma^2 : \partial^2_{\omega\omega}\varphi)_0 \le 0$, as required.

(ii) $\Longrightarrow$ (i): Clearly, it is sufficient to prove that the process $\bar{u} := u^{\varepsilon}_t := u_t + \varepsilon t$ is a pathwise $\mathbb{P}_0$-submartingale for all $\varepsilon > 0$, as the required claim will follow by sending ε to zero. By (ii), we deduce immediately that $\bar{u}$ is a $\mathbb{P}_0$-viscosity subsolution of the equation $\varepsilon - \partial_t\bar{u} - \frac{1}{2}\mathrm{Tr}[\partial^2_{\omega\omega}\bar{u}] \le 0$ on Θ. In particular, this implies that

$$0 \notin \underline{\mathcal{A}}^{\mathbb{P}_0}\bar{u}_t(\omega) \quad \text{ for all } (t, \omega) \in \Theta. \tag{4.2}$$

Suppose to the contrary that $\bar{u}$ is not a pathwise $\mathbb{P}_0$-submartingale, i.e. $\bar{u}_t(\omega) > \mathbb{E}^{\mathbb{P}_0}\big[\bar{u}^{t,\omega}_{\mathrm{H}}\big]$ for some $(t, \omega) \in \Theta$ and $\mathrm{H} \in \mathcal{T}_+$. Then, Lemma 4.3 induces a contradiction of (4.2). $\square$

As an immediate consequence of Theorem 4.4, we obtain the wellposedness of the path-dependent heat equation.

Theorem 4.5 [Comparison and existence for the heat equation]

(i) *Let $u, v \in C^0_{2,\mathbb{P}_0}(\Theta, \mathbb{R})$ be $\mathbb{P}_0$-viscosity subsolution and supersolution, respectively, of the path-dependent heat equation (4.1), with $u_T \le v_T$ on Ω. Then $u \le v$ on $[0, T] \times \Omega$.*

(ii) *For an $\mathcal{F}_T$ r.v. ξ such that $u_t(\omega) := \mathbb{E}^{\mathbb{P}_0}[\xi^{t,\omega}] \in C^0_{2,\mathbb{P}_0}(\Theta, \mathbb{R})$, the process u is the unique $\mathbb{P}_0$-viscosity solution of the path-dependent heat Eq. (4.1) with boundary condition $u_T = \xi$ on Ω.*

Proof (i) By Theorem 4.4, we have $u_t(\omega) \le \mathbb{E}^{\mathbb{P}_0}[(u_T)^{t,\omega}]$ and $\mathbb{E}^{\mathbb{P}_0}[(v_T)^{t,\omega}] \ge v_t(\omega)$ for all $(t, \omega) \in \Theta$. Then $u_T \le v_T$ on Ω implies that $u \le v$ on $[0, T] \times \Omega$.

(ii) Uniqueness is a direct consequence of the comparison result of (i). Clearly the process $u_t(\omega) := \mathbb{E}^{\mathbb{P}_0}\big[\xi^{t,\omega}\big]$ is uniformly continuous on $[0, T] \times \Omega$. Since u is a $\mathbb{P}_0$-martingale, it follows from Theorem 4.4 that it is both a viscosity subsolution and supersolution. $\square$

5 Wellposedness of Semilinear Path-Dependent PDEs

In this section, we consider the equation

$$-\partial_t u - \frac{1}{2}\mathrm{Tr}\big[\partial^2_{\omega\omega}u\big] - F(., u, \partial_\omega u) = 0 \quad \text{ on } \Theta. \tag{5.1}$$

The nonlinearity $F : \Theta \times \mathbb{R} \times \mathbb{R}^d \longrightarrow \mathbb{R}$ is assumed to satisfy the following assumptions which consists with the general assumption as (3.1).

Assumption 5.1 The nonlinearity $F : (t, \omega, y, z) \in \Theta \times \mathbb{R} \times \mathbb{R}^d \longmapsto F(t, \omega, y, z) \in \mathbb{R}$ satisfies the following conditions:

(i) F is uniformly continuous in (t, ω),
(ii) F is uniformly L_0-Lipschitz continuous in (y, z), for some $L_0 \geq 0$, i.e.

$$|F(\cdot, y, z) - F(\cdot, y', z')| \leq L_0 \left(|y - y'| + |z - z'|\right) \quad \text{for all} y,$$
$$y' \in \mathbb{R}, \ z, z' \in \mathbb{R}^d.$$

(iii) The process $F(\cdot, 0, 0)$ is bounded.

For all bounded $\mathbb{F}$-progressively measurable process λ, we denote:

$$d\mathbb{P}_\lambda := Z_T^\lambda \cdot d\mathbb{P}_0 \quad \text{on } \mathcal{F}_T, \quad \text{where } Z_T^\lambda := e^{\int_0^T \lambda_t \cdot dB_t - \frac{1}{2}\int_0^T |\lambda_t|^2 dt}.$$

In this section, we take

$$\mathcal{P} := \left\{\mathbb{P}_\lambda : \ \lambda \text{ bounded by } L\right\}, \tag{5.2}$$

where $L \geq L_0$ is arbitrary. Notice that $\mathbb{P}_0$ is a dominating measure for the family $\mathcal{P}$. For simplicity, we say $\mathbb{P}_\lambda \in \mathcal{P}$ by implying that λ is the corresponding bounded process. Similar to the section of the heat equation, we denote $\mathbb{F}^*$ as the filtration augmented by all $\mathbb{P}_0$-null sets. Also, we consider the set of localizing stopping times as:

$$\mathcal{H}_+^{\mathcal{P}} := \mathcal{T}^+.$$

In this section about the semilinear equation, the relevant space for our comparison result is

$$C^0_{2,\mathcal{P}}(\Theta, \mathbb{R}) := \left\{u \in C^0(\Theta, \mathbb{R}) : \overline{\mathcal{E}}^{\mathcal{P}}\Big[\sup_{t+s \leq T} \left|u_s^{t,\omega}\right|^2\Big] < \infty \quad \text{for all } (t, \omega) \in \Theta\right\}.$$

5.1 *Optimal Stopping Under Dominated Nonlinear Expectation*

For $X \in C^0_{2,\mathcal{P}}(\Theta, \mathbb{R})$, we consider the optimal stopping stopping problem under dominated nonlinear expectation:

$$V_0 := \sup_{\tau \in \mathcal{T}_*} \overline{\mathcal{E}}^{\mathcal{P}}[X_\tau].$$

The corresponding dynamic formulation is defined by:

$$Y_t^0 := \operatorname*{ess\,sup}_{\tau \in \mathcal{T}_*^t} \overline{\mathcal{E}}^{\mathcal{P}}\left[X_{\tau \wedge T} \middle| \mathcal{F}_t\right], \qquad 0 \leq t \leq T.$$

with first hitting time:

$$\tau^* := \inf\{t \geq 0 : Y_t^0 = X_t\}.$$

Since the dominating measure $\mathbb{P}$ satisfies the Blumenthal zero-one law, it follows that $Y_0^0 = V_0$. We also introduce the pointwise optimal stopping problem:

$$\mathcal{Y}_t(\omega) := \sup_{\tau \in \mathcal{T}_*} \overline{\mathcal{E}}^{\mathcal{P}}\big[X^{t,\omega}_{\tau\wedge(T-t)}\big], \quad \text{for all } (t,\omega) \in \bar{\Theta}.$$

Theorem 5.2 *Let $X \in C^0_{2,\mathcal{P}}(\Theta,\mathbb{R})$. Then, there exists an $\mathbb{F}$-adapted version Y of Y^0 satisfying:*

(i) *for all $\tau \in \mathcal{T}$, we have $Y_{\tau\wedge T} = \mathcal{Y}_{\tau\wedge T}$, $\mathbb{P}_0$-a.s.*
(ii) *Y is a pathwise continuous $\mathbb{P}$-supermartingale for all $\mathbb{P} \in \mathcal{P}$, and τ^* is an optimal stopping rule,*
(iii) *$Y_t = \operatorname{ess\,sup}_{\tau\in\mathcal{T}_*^t} \mathbb{E}^{\mathbb{P}^*}\big[X_{\tau\wedge T}|\mathcal{F}_t\big]$ for all $t \in [0,T]$, $\mathbb{P}_0$-a.s. for some $\mathbb{P}^* \in \mathcal{P}$, and*

$$Y = Y_0 + M^* - K^* \ \text{ with } M_0^* = K_0^* = 0, \ \text{ and } \int (Y - X)dK^* = 0, \ \ \mathbb{P}_0\text{-a.s.}$$

for some pathwise continuous martingale M^* and predictable nondecreasing process K^*.

This result can be proved by referring to the corresponding literature in the theory of reflected backward stochastic differential equations, see Remark 7.3 in [13]. For the convenience of those readers who are not familiar with this literature, we report in Sect. 8 a proof purely based on arguments from optimal stopping theory.

5.2 Punctual Smoothness of Submartingales

In this subsection, we prove that a process $u \in C^0_{2,\mathcal{P}}(\Theta,\mathbb{R})$ which is $\mathbb{P}$-submartingale for some $\mathbb{P} \in \mathcal{P}$ is punctually $C^{1,2}_{\mathcal{P}}$-Leb$\otimes\mathbb{P}$-a.e. This is our natural extension of the well-known result that any non-decreasing function is differentiable a.e. and our proof builds on the corresponding standard results in analysis that we quickly review. For a function $f : [0,T] \longrightarrow \mathbb{R}$ with finite variation, we use the following notations for the left-semigradients:

$$\underline{\dot{f}}^{\ell}(t) := \liminf_{\varepsilon\uparrow 0} \frac{f(t+\varepsilon)-f(t)}{\varepsilon} \quad \text{and } \overline{\dot{f}}^{\ell}(t) := \limsup_{\varepsilon\uparrow 0} \frac{f(t+\varepsilon)-f(t)}{\varepsilon}.$$

The right-semigradients $\underline{\dot{f}}^{r}$ and $\overline{\dot{f}}^{r}$ are defined similarly by sending $\varepsilon \downarrow 0$. The function f is differentiable at a point t if

$$\dot{f}(t) = \lim_{\varepsilon \to 0} \frac{f(t+\varepsilon) - f(t)}{\varepsilon} \text{ exists, and therefore } \dot{f}(t) = \underline{\dot{f}}^{\ell}(t) = \overline{\dot{f}}^{\ell}(t) = \underline{\dot{f}}^{r}(t) = \overline{\dot{f}}^{r}(t).$$

Our smoothness results uses crucially the two following properties:

$\mathbf{FV_1}$ The set of points of differentiability of f has full Lebesgue measure.
$\mathbf{FV_2}$ If f is absolutely continuous, then $\lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \int_t^{t+\varepsilon} |\dot{f}(s) - \dot{f}(t)| ds = 0$, Leb-a.e. on $[0, T]$.

For a subset $\Theta_0 \subset \bar{\Theta}$, we denote $\mathbb{T}^{\Theta_0} := \{t : (t, \omega) \in \Theta_0 \text{ for some } \omega \in \Omega\}$ and $\Omega_t^{\Theta_0} := \{\omega : (t, \omega) \in \mathbb{T}^{\Theta_0}\}$.

Theorem 5.3 *Let $\mathbb{P}_\theta \in \mathcal{P}$ and $u \in C^0_{2,\mathcal{P}}(\Theta, \mathbb{R})$ be $\mathbb{P}_\theta$-submartingale. Then u is $\mathcal{P}$-punctually $C^{1,2}$ on Θ_0, for some Θ_0 with*

$$Leb\big[\mathbb{T}^{\Theta_0}\big] = T \quad \textit{and } \mathbb{P}_0\big[\Omega_t^{\Theta_0}\big] = 1 \quad \textit{for all } \ t \in \mathbb{T}^{\Theta_0}. \tag{5.3}$$

Sketch of the proof. For a proof in more details, we refer to [28]. We proceed in two steps.
<u>*Step 1*</u>: By the Doob-Meyer decomposition, we have $u = u_0 + M + A$, $\mathbb{P}_0$-a.s. for some $\mathbb{P}_\theta$-martingale M and nondecreasing predictable process A, with $M_0 = A_0 = 0$. Then, process $M^0 := M - \int_0^{\cdot} \theta_s d\langle M, B\rangle_s$ defines a $\mathbb{P}_0$-martingale.

Since all $\mathbb{P}_0$-martingale have the martingale representation, it follows that $t \longmapsto H_t := \langle M, B\rangle_t = \langle M^0, B\rangle_t$ is absolutely continuous on $[0, T]$, $\mathbb{P}_0 -$ a.s., i.e.

$$h_t := \overline{\dot{H}}^{\ell}_t = \dot{H}_t \quad \text{for a.e. } t \in [0, T], \quad \mathbb{P}_0 - \text{a.s.}$$

By the above property $\mathbf{FV_2}$ together with the Fubini theorem, we see that

$$\text{Leb} \otimes \mathbb{P}_0[\Theta_1] = T \ \text{ where } \ \Theta_1 := \Big\{(t, \omega) : \lim_{\varepsilon \to 0} \tfrac{1}{\varepsilon} \int_t^{t+\varepsilon} |h_s - h_t| ds = 0\Big\}. \tag{5.4}$$

Further, applying property $\mathbf{FV_1}$ to the finite variation process $A^\theta := A + \int_0^{\cdot} \theta_s dH_s$, and using again the Fubini theorem, we see that:

$$\text{Leb} \otimes \mathbb{P}_0[\Theta_2] = T \quad \text{where } \Theta_2 := \big\{(t, \omega) : a_t(\omega) := \dot{A}^\theta_t(\omega) \text{ exists}\big\}. \tag{5.5}$$

<u>*Step 2*</u>: In this step, we prove that for $(t, \omega) \in \Theta_0 := \Theta_1 \cap \Theta_2$.

$$(q^\varepsilon, p, 0) \in \underline{\mathcal{J}}^{\mathcal{P}} u_t(\omega), \quad \text{where } q^\varepsilon := a_t(\omega) - \varepsilon(1 + L), \ \ p := h_t(\omega). \tag{5.6}$$

We define

$$\mathrm{H}(\omega') := \inf\Big\{s > 0 :\ (A^\theta)_s^{t,\omega}(\omega') - A_t^\theta(\omega) \le (a_t(\omega) - \varepsilon)s \quad \text{or} \quad \int_0^s |h_r^{t,\omega}(\omega') - h_t(\omega)|dr \ge \varepsilon s\Big\}.$$

Since $\omega \in \Theta_0$, we have $\mathrm{H} \in \mathcal{T}^+$. Also, note that $M_s^{\lambda,t} := M_s^{t,\omega} - M_t(\omega) - \int_0^s (\theta - \lambda)_r h_r dr$ defines a $\mathbb{P}_\lambda$-martingale. Further, rewriting the Doob-Meyer decomposition, we have

$$u_s^{t,\omega} = u_t(\omega) + (A^\theta)_s^{t,\omega} - A_t^\theta(\omega) + M_s^{\lambda,t} - \int_0^s \lambda_r h_r dr,\ \mathbb{P}_\lambda\text{-a.s.}$$

So, for all $\tau \in \mathcal{T}$, $\mathbb{P}_\lambda \in \mathcal{P}$:

$$\begin{aligned}
&\mathbb{E}^{\mathbb{P}_\lambda}\big[\big(\phi^{q^\varepsilon,p,0} - u^{t,\omega}\big)_{(\tau\wedge\mathrm{H})^{t,\omega}}\big] \\
&= -u_t(\omega) + \mathbb{E}^{\mathbb{P}_\lambda}\Big[(a_t(\omega) - \varepsilon)(\tau \wedge \mathrm{H}) - A^\theta_{\tau\wedge\mathrm{H}} + A_t^\theta(\omega) - \varepsilon L(\tau \wedge \mathrm{H}) \\
&\qquad\qquad + \int_0^{\tau\wedge\mathrm{H}} (h_s - h_t)\lambda_s ds\Big] \\
&\le -u_t(\omega),
\end{aligned}$$

by the definition of H. Then (5.6) holds.
Step 3: From the previous step, it follows that $\big(a_t(\omega), h_t(\omega), 0\big) \in \mathrm{cl}\big(\underline{\mathcal{J}}^{\mathcal{P}} u_t(\omega)\big)$. By a similar argument, we may show that $\big(a_t(\omega), h_t(\omega), 0\big) \in \mathrm{cl}\big(\overline{\mathcal{J}}^{\mathcal{P}} u_t(\omega)\big)$. Consequently, $\big(a_t(\omega), h_t(\omega), 0\big) \in \mathcal{J}^{\mathcal{P}} u_t(\omega)$, and u is punctually $C^{1,2}_{\mathcal{P}}$. □

5.3 Comparison

In this subsection, we are going to show the comparison principle for the semilinear path-dependent equation.

Theorem 5.4 *Let Assumption 5.1 hold true. Let $u, v \in C^0_{2,\mathcal{P}}(\Theta, \mathbb{R})$ be $\mathcal{P}$-viscosity subsolution and supersolution, respectively, of the Eq. (5.1). Assume further that $u_T \le v_T$ on Ω. Then $u \le v$ on $\bar{\Theta}$.*

To show Theorem 5.4, we need some preparation. The following lemma is the analog of Lemma 4.3 in the context of the semilinear path-dependent PDEs. We omit the proof, since it is similar to that of Lemma 4.3.

Lemma 5.5 *Let $u \in C^0_{2,\mathcal{P}}(\Theta, \mathbb{R})$, $(t,\omega) \in \Theta$, and $\mathrm{H} \in \mathcal{T}^+$, be such that $u_t(\omega) > \overline{\mathcal{E}}^{\mathcal{P}}\big[u^{t,\omega}_{\mathrm{H}}\big]$. Then,*

$$0 \in \underline{\mathcal{A}}^{\mathcal{P}} u_{t+t^*}(\omega \otimes_t \omega^*) \quad \textit{for some } (t^*, \omega^*) \textit{ with the localization}$$
$$\mathrm{H}^* := \mathrm{H}^{t^*,\omega^*} - t^* \in \mathcal{T}^+.$$

The next main ingredient is the partial comparison result.

Proposition 5.6 *In the setting of Theorem 5.4, assume in addition that $v \in C^{1,2}_{\mathcal{P}}(\Theta)$. Then $u \le v$ on Θ.*

Proof First, by possibly transforming the problem to the comparison of $\tilde{u}_t := e^{\lambda t} u_t$ and $\tilde{v}_t := e^{\lambda t} v_t$, it follows from the Lipschitz property of the nonlinearity F in y that we may assume without loss of generality that F is decreasing in y.

Suppose to the contrary that $c := (u-v)_t(\omega) > 0$ at some point $(t,\omega) \in \Theta$. Let $c_0 := \frac{c}{2T}$, and define $f_s := (u-v)^+_s + c_0(s-t)$, $s \in [t,T]$. Since $(u-v)_T \le 0$, it follows that $f_t(\omega) > \overline{\mathcal{E}}^{\mathcal{P}}[f^{t,\omega}_{T-t}]$. By Lemma 5.5, we may find a point (t^*, ω^*) such that $t^* \in [t,T)$ and $0 \in \underline{\mathcal{A}}^{\mathcal{P}} f_{t^*}(\omega^*)$. In particular, this implies that

$$-(u-v)^+_{t^*}(\omega^*) - c_0(t^*-t) \le \underline{\mathcal{E}}^{\mathcal{P}}\Big[-\big((u-v)^+_T\big)^{t^*,\omega^*} - c_0(T-t)\Big] = -c_0(T-t),$$

so that $(u-v)^+_{t^*}(\omega^*) \ge c_0(T-t^*) > 0$. Then, since $(u-v)^+ \ge u-v$, we deduce from $0 \in \underline{\mathcal{A}}^{\mathcal{P}} f_{t^*}(\omega^*)$ that

$$(\varphi - u)_{t^*}(\omega^*) \le \underline{\mathcal{E}}^{\mathcal{P}}\Big[(\varphi-u)^{t^*,\omega^*}_{\tau\wedge T}\Big] \quad \text{for all } \tau \in \mathcal{T},$$
$$\text{where} \quad \varphi_s(\omega) := v_s(\omega) - c_0(s-t).$$

Since $v \in C^{1,2}_{\mathcal{P}}(\Theta)$, this means that $\varphi^{t^*,\omega^*} \in \underline{\mathcal{A}}^{\mathcal{P}} u_{t^*}(\omega^*)$. Then, by the viscosity subsolution property of u, and the classical supersolution property of v, we deduce that

$$\begin{aligned}
0 &\ge \Big\{-\partial_t\varphi - \frac{1}{2}\mathrm{Tr}\big[\partial^2_{\omega\omega}\varphi\big] - F(.,u,\partial_\omega\varphi)\Big\}(t^*,\omega^*)\\
&= c_0 + \Big\{-\partial_t v - \frac{1}{2}\mathrm{Tr}\big[\partial^2_{\omega\omega} v\big] - F(.,u,\partial_\omega v)\Big\}(t^*,\omega^*)\\
&\ge c_0 + \{F(.,v,\partial_\omega v) - F(.,u,\partial_\omega v)\}(t^*,\omega^*) \ \ge\ c_0,
\end{aligned}$$

where the last inequality follows from the non-increase of F in y and the fact that $u_{t^*}(\omega^*) \ge v_{t^*}(\omega^*)$. Since $c_0 > 0$, this is the required contradiction. □

Definition 5.7 An $\mathbb{F}$-progressively measurable process m is an $\overline{\mathcal{E}}$-submartingale (resp. $\underline{\mathcal{E}}$-supermartingale), if, for any $(t,\omega) \in \Theta$, we have

$$u_t(\omega) \le \overline{\mathcal{E}}[u^{t,\omega}_Z] \text{ (resp. } \ge \underline{\mathcal{E}}[u^{t,\omega}_Z]) \quad \text{for all } \tau \in \mathcal{T}.$$

Lemma 5.8 *Under Assumption 5.1, there is a constant C such that*

(i) *the process $\left\{u_t + \int_0^t |u_s|ds + Ct, t \in [0,T]\right\}$ is a pathwise $\mathbb{P}^u$ submartingale, for some $\mathbb{P}^u \in \mathcal{P}$,*
(ii) *the process $\left\{v_t - \int_0^t |v_s|ds - Ct, t \in [0,T]\right\}$ is a pathwise $\mathbb{P}^v$ supermartingale, for some $\mathbb{P}^v \in \mathcal{P}$,*
(iii) *u and v are $\mathcal{P}$-punctually $C^{1,2}$ on Θ^u and Θ^v, respectively, where Θ^u and Θ^v satisfy (5.3).*

Proof Assertion (iii) is a direct consequence of (i) and (ii) together with Theorem 5.3. By Assumption 5.1, we may find a constant C such that:

$$|F(t,\omega,y,z)| \le C - 1 + L_0(|y| + |z|)$$

Then, it is easy to verify that $\bar{u}_t := u_t + Ct$ and $\bar{v}_t := v_t - Ct$ are $\mathcal{P}$-viscosity subsolution and supersolution, respectively of:

$$\begin{aligned} &-\mathcal{L}\bar{u} - L_0(|\bar{u} - Ct| + |\partial_\omega \bar{u}|) + 1 \le 0 \quad \text{and} \\ &-\mathcal{L}\bar{v} + L_0(|\bar{v} + Ct| + |\partial_\omega \bar{v}|) - 1 \ge 0 \quad \text{on } [0,T) \times \Omega. \end{aligned} \tag{5.7}$$

In the rest of this proof, we shall show that $\bar{u}$ and $\bar{v}$ are $\overline{\mathcal{E}}^{\mathcal{P}}$-submartingale and $\underline{\mathcal{E}}^{\mathcal{P}}$-supermartingale, respectively. In addition, we prove in Appendix (Proposition 9.2) that a continuous $\overline{\mathcal{E}}^{\mathcal{P}}$-submartingale is a $\mathbb{P}$-submartingale for some $\mathbb{P} \in \mathcal{P}$. This leads to the desired result.

We only prove that $\bar{u}$ is $\overline{\mathcal{E}}^{\mathcal{P}}$-submartingale, as the corresponding statement for $\bar{v}$ follows from the same line of argument.

Suppose to the contrary that $\bar{u}_t(\omega) > \overline{\mathcal{E}}^{\mathcal{P}}[\bar{u}_{\mathrm{H}}^{t,\omega}]$ for some $(t,\omega) \in [0,T) \times \Omega$ and some stopping time $\mathrm{H} \in \mathcal{T}^+$. Then, it follows from Lemma 5.5 that there exist t^* and ω^* such that $0 \in \underline{\mathcal{A}}^{\mathcal{P}}\bar{u}_{t^*}(\omega^*)$, i.e. there exists $\mathrm{H}' \in \mathcal{T}^+$ such that

$$-\bar{u}_{t^*}(\omega^*) \ge \overline{\mathcal{E}}^{\mathcal{P}}\Big[-\bar{u}^{t^*,\omega^*}_{\tau\wedge \mathrm{H}'} - L_0 \int_0^{\tau\wedge\mathrm{H}'} |u_s^{t^*,\omega^*}|ds \Big].$$

As a result, function $\varphi_t := -L_0 \int_0^t |u_s^{t^*,\omega^*}|ds$ is in $\underline{\mathcal{A}}^{\mathcal{P}}u_{t^*}(\omega^*)$. Since $\bar{u}$ is a $\mathcal{P}$-viscosity subsolution of the left equation of (5.7), this leads to

$$L_0|u_{t^*}(\omega^*)| - L_0|u_{t^*}(\omega^*)| + 1 \le 0,$$

which is the required contradiction, thus completing the proof of (i). □

We are now ready for the key-result for the proof of the comparison result. We observe that this statement is an adaptation of the approach of Caffarelli and Cabre [4] to the comparison in the context of the standard theory of viscosity solutions in finite dimensional spaces.

Proposition 5.9 *Let Assumption 5.1 hold, and consider the L in the definition of $\mathcal{P}$ (recall that $L \geq L_0$). Let $u, v \in C^0_{2,\mathcal{P}}(\Theta, \mathbb{R})$ be $\mathcal{P}$-viscosity subsolution and supersolution, respectively, of the path-dependent PDE (5.1). Then, $w := u - v$ is a $\mathcal{P}$-viscosity subsolution of*

$$- \mathcal{L}w(t,\omega) - L|w_t(\omega)| - L|\partial_\omega w_t(\omega)| \leq 0. \tag{5.8}$$

Sketch of Proof Without loss of generality, we only check the viscosity property at $(t, \omega) = (0, 0)$. For an arbitrary $(a, \beta, 0) \in \underline{\mathcal{J}}^{\mathcal{P}}(u - v)_0$, we have to prove that

$$- a - L\,|(u - v)_0| - L|\beta| \leq 0. \tag{5.9}$$

1. Denote as usual by $\phi^{a,\beta} = \phi^{a,\beta,0}$ the corresponding paraboloid process. By definition, there exists $\mathrm{H} \in \mathcal{T}^+$ such that

$$c_0 \ := \ -(u - v)_0 = \min_{\tau \in \mathcal{T}} \underline{\mathcal{E}}^{\mathcal{P}}\big[(\phi^{a,\beta} - u + v)_{\tau \wedge \mathrm{H}}\big].$$

For $\delta > 0$, $r > 0$, and $\mathrm{H}_r := \mathrm{H} \wedge \inf\{t : |\omega_t| \geq r\}$, define the Snell envelop:

$$\hat{m}_t := \operatorname*{ess\,inf}_{\tau \in \mathcal{T}^t_*, \mathbb{P} \in \mathcal{P}} \mathbb{E}^{\mathbb{P}}\left[m_{\tau \wedge \mathrm{H}_r} | \mathcal{F}_t\right], \quad t \in [0, T], \quad \text{where } \ m := \phi^{a+\delta,\beta} - u + v.$$

Clearly,

$$m_0 = c_0, \ \ \underline{\mathcal{E}}^{\mathcal{P}}\left[m_{\mathrm{H}_r}\right] > c_0, \hat{m}_0 \leq m_0, \text{and } \hat{m}_{\mathrm{H}_r} = m_{\mathrm{H}_r}, \quad \mathbb{P}_0\text{-a.s.} \tag{5.10}$$

Further, from Theorem 5.2, we have that:

$$\hat{m}_t = \operatorname*{ess\,inf}_{\tau \in \mathcal{T}^t_*} \mathbb{E}^{\mathbb{P}_{\lambda^*}}\left[m_{\tau \wedge \mathrm{H}_r} | \mathcal{F}_t\right], \mathbb{P}_0 - \text{a.s.} \quad \text{for some} \|\lambda^*\| \leq L. \tag{5.11}$$

2. By classical optimal stopping theory, $\hat{m}$ is a $\mathbb{P}_{\lambda^*}$–submartingale with Doob-Meyer decomposition

$$\hat{m} = \hat{m}_0 + \hat{A} + \hat{M}, \quad \text{with} \hat{A} = \int_0^{\cdot} \mathbf{1}_{\{m = \hat{m}\}}(s) d\hat{A}_s, \qquad \mathbb{P}_{\lambda^*}\text{-a.s.}$$

for some $\mathbb{P}_{\lambda^*}$-martingale $\hat{M}$, and some nondecreasing process $\hat{A}$. In addition, we may prove that $\hat{A}$ is absolutely continuous $\mathbb{P}_0$-a.s. (see Step 4 in the proof of Proposition 7.3 in [28]). Then, it follows from (5.10) that:

$$0 < \underline{\mathcal{E}}^{\mathcal{P}}\left[m_{\mathrm{H}_r} - m_0\right] \leq \underline{\mathcal{E}}^{\mathcal{P}}\left[\hat{m}_{\mathrm{H}_r} - \hat{m}_0\right]$$

$$
\begin{aligned}
&\leq \mathbb{E}^{\mathbb{P}_{\lambda^*}}\left[\int_0^{\mathrm{H}_r} \mathbf{1}_{\{m=\hat{m}\}}(t)d\hat{A}_t\right] \\
&= \lim_{M\to\infty} \mathbb{E}^{\mathbb{P}_{\lambda^*}}\left[\int_0^{\mathrm{H}_r} \mathbf{1}_{\{m=\hat{m}\}}(t)\mathbf{1}_{\{|\dot{\hat{A}}_t|\leq M\}}\dot{\hat{A}}_t\, dt\right] \\
&\leq \lim_{M\to\infty} M\, \mathbb{E}^{\mathbb{P}_{\lambda^*}}\left[\int_0^{\mathrm{H}_r} \mathbf{1}_{\{m=\hat{m}\}}(t)dt\right].
\end{aligned}
$$

This implies that $\mathrm{Leb}\otimes\mathbb{P}_0\big[t < \mathrm{H}_r, m = \hat{m}\big] > 0$, so that, with the subsets Θ^u, Θ^v from Proposition 5.8, we have:

$$\mathrm{Leb}\otimes\mathbb{P}_0\big[\{t \in [0, \mathrm{H}_r), m = \hat{m}\} \cap \Theta^u \cap \Theta^v\big] > 0.$$

Further, by taking in account (i) of Theorem 5.2, we may find a point (t^*, ω^*) such that

$$\begin{gathered}
\mathrm{H}_r^{t^*,\omega^*} - t^* \in \mathcal{T}^+,\ m_{t^*}(\omega^*) = \hat{m}_{t^*}(\omega^*) = \inf_{\tau\in\mathcal{T}_*} \underline{\mathcal{E}}^{\mathcal{P}}[m^{t^*,\omega^*}_{\tau\wedge(\mathrm{H}_r^{t^*,\omega^*}-t^*)}], \\
\text{and } u, v \text{ are } \mathcal{P}\text{-punctually } C^{1,2} \text{ at } (t^*, \omega^*).
\end{gathered} \tag{5.12}$$

3. By Proposition 3.15, it follows that m is $\mathcal{P}$-punctually $C^{1,2}$ at (t^*, ω^*), and $(a^m, \beta^m) := (a + \delta - a^u + a^v, \beta - \beta^u + \beta^v) \in \mathcal{J}^{\mathcal{P}} m(t^*, \omega^*)$ for any $(a^u, \beta^u) \in \mathcal{J}^{\mathcal{P}} u(t^*, \omega^*)$ and $(a^v, \beta^v) \in \mathcal{J}^{\mathcal{P}} v(t^*, \omega^*)$. Then, by using the viscosity subsolution property of u together with Proposition 3.14 and the Lipschitz property of F from Assumption 5.1, we see that:

$$
\begin{aligned}
0 &\geq -a^u - F(t^*, \omega^*, u_{t^*}(\omega^*), \beta^u) \\
&= (-a^v + a^m) - a - \delta - F(t^*, \omega^*, (u - v + v)_{t^*}(\omega^*), \beta + \beta^v - \beta^m) \\
&\geq a^m - L|\beta^m| - a - \delta - L\left|(u - v)_{t^*}(\omega^*)\right| - L|\beta| - a^v \\
&\quad - F(t^*, \omega^*, v_{t^*}(\omega^*), \beta^v)
\end{aligned}
$$

We shall prove in Step 5 below that

$$a^m - L|\beta^m| \geq 0. \tag{5.13}$$

Together with the viscosity supersolution property of v, this provides:

$$0 \geq -a - \delta - L\left|(u - v)_{t^*}(\omega^*)\right| - L|\beta|.$$

Since $t^* \to 0$ as $r \to 0$, and $u, v \in C^0$, this provides $-a-\delta-L\,|(u-v)_0| - L|\beta| \le 0$, which implies (5.9) by sending $\delta \to 0$.

4. It remains to prove (5.13). For the sake of simplicity, we set $t^* = 0$. Recall that $(a^m, \beta^m) \in \mathcal{J}^{\mathcal{P}} m_0$ and $m_0 = \hat{m}_0 = \inf_{\tau\in\mathcal{T}} \underline{\mathcal{E}}^{\mathcal{P}}[m_{\tau\wedge \mathrm{H}_r}]$. Suppose to the contrary that $a^m - L|\beta^m| < 0$. Then, there exists $(\hat{a}, \hat{\beta}) \in \underline{\mathcal{J}}^{\mathcal{P}} m_0$ such that $\hat{a} - L|\hat{\beta}| < 0$. By definition of $\underline{\mathcal{J}}^{\mathcal{P}} m_0$, we have

$$m_0 = \sup_{\tau} \overline{\mathcal{E}}^{\mathcal{P}}\left[m_{\tau\wedge\hat{\mathrm{H}}} - \phi^{\hat{a},\hat{\beta}}_{\tau\wedge\hat{\mathrm{H}}}\right] \quad \text{for some} \hat{\mathrm{H}} \in \mathcal{T}^+ \text{ with } \hat{\mathrm{H}} \le \mathrm{H}_r.$$

Then, considering the process $\lambda := -L\mathrm{sgn}(\hat{\beta})$, we see that:

$$\hat{m}_0 = m_0 \ge \mathbb{E}^{\mathbb{P}_\lambda}\big[m_{\hat{\mathrm{H}}} - \phi^{\hat{a},\hat{\beta}}_{\tau\wedge\hat{\mathrm{H}}}\big] = \mathbb{E}^{\mathbb{P}_\lambda}\big[m_{\hat{\mathrm{H}}}\big] - (\hat{a} - L\hat{\beta})\mathbb{E}^{\mathbb{P}_\lambda}[\hat{\mathrm{H}}] > \mathbb{E}^{\mathbb{P}_\lambda}\big[m_{\hat{\mathrm{H}}}\big].$$

Since $\hat{\mathrm{H}} \le \mathrm{H}_r$ and $\mathbb{P}_\lambda \in \mathcal{P}$, this is in contradiction with the definition of $\hat{m}_0$. □

The previous proposition, together with the partial comparison result of Proposition 5.6, lead directly to the comparison result.

Proof of Theorem 5.4 By Proposition 5.9, $u - v \in C^0_{2,\mathcal{P}}(\Theta, \mathbb{R})$ is a $\mathcal{P}$-viscosity subsolution of the path-dependent equation (5.8). Clearly, 0 is a classical supersolution of the same equation. Since $(u-v)_T \le 0$, we conclude from the partial comparison result of Proposition 5.6 that $u - v \le 0$ on Θ. □

5.4 *Existence*

To establish an existence result of $\mathcal{P}$-viscosity solutions of the equation (5.1) under the above Assumption 5.1, we consider a terminal condition defined by an $\mathcal{F}_T$-measurable r.v. ξ. Then, the PPDE 5.1 with terminal condition $u(T,\omega) = \xi(\omega)$ is closely related to the following backward stochastic differential equation (BSDE):

$$Y^0_t = \xi + \int_t^T F(s, B, Y^0_s, Z^0_s)ds - \int_t^T Z^0_s dB_s, \quad 0 \le t \le T, \qquad \mathbb{P}_0\text{-a.s.} \quad (5.14)$$

We refer to the seminal paper by Pardoux and Peng [24] for the wellposedness of such BSDEs. On the other hand, for any $(t,\omega) \in [0,T] \times \Omega$, by [24] the following BSDE on $[t,T]$ has a unique solution:

$$Y^{0,t,\omega}_s = \xi^{t,\omega} + \int_s^T F^{t,\omega}(r, B^t, Y^{0,t,\omega}_r, Z^{0,t,\omega}_r)dr - \int_s^T Z^{0,t,\omega}_r dB^t_r, \quad \mathbb{P}^t_0\text{-a.s.} \quad (5.15)$$

By the Blumenthal 0–1 law, $Y_t^{0,t,\omega}$ is a constant and we thus define

$$u^0(t,\omega) := Y_t^{0,t,\omega}. \tag{5.16}$$

Theorem 5.10 *Let $\xi \in UCB(\Omega)$ be an $\mathcal{F}_T$-measurable r.v. Then, under Assumption 5.1, u^0 is a viscosity solution of PPDE 1.2 with terminal condition $u_T^0 = \xi$.*

Proof Under our assumptions on the nonlinearity F, it follows from the boundedness and uniform continuity of ξ that u^0 is uniformly continuous on $[0,T]\times\Omega$, see [11]. We show that u^0 is a $\mathcal{P}$-viscosity subsolution, the same line of argument allows to prove that u^0 is a $\mathcal{P}$-viscosity subsolution. We proceed by contradiction, assuming that u^0 is not a viscosity subsolution. Then, there exist $(t,\omega)\in[0,T)\times\Omega$ and $\varphi\in\underline{\mathcal{A}}^{\mathcal{P}}u_t^0(\omega)$ such that:

$$2c \;:=\; -\partial_t\varphi_0 - \frac{1}{2}\mathrm{Tr}\left[\partial^2_{\omega\omega}\varphi_0\right] - F_t(\omega, u_t^0(\omega), \partial_\omega\varphi_0) > 0.$$

Without loss of generality, we assume $u_t(\omega) = \varphi_0$, and we set $(t,\omega) = (0,0)$. Denote:

$$\begin{aligned}\phi_s := \partial_t\varphi_s + \tfrac{1}{2}\mathrm{Tr}\left[\partial^2_{\omega\omega}\varphi_s\right] + F_s(\varphi_s, \partial\varphi_s)\ \text{ so that }\ \phi_0 = -2c,\\ \text{and }\ \tilde{Y}_s := \varphi_s,\ \ \tilde{Z}_s := \partial_\omega\varphi_s,\ \ \delta Y_s := \tilde{Y}_s - Y_s,\ \ \delta Z_s := \tilde{Z}_s - Z_s,\ \ , s\in[0,T].\end{aligned}$$

Applying Itô's formula, we have

$$\begin{aligned}d(\delta Y_s) &= \big(\partial_t\varphi_s + \frac{1}{2}\mathrm{Tr}\big[\partial^2_{\omega\omega}\varphi_s\big]\big)ds + \tilde{Z}_s\cdot dB_s + F_s(Y_s,Z_s)ds - Z_s\cdot dB_s\\ &= \big[\phi_s + F_s(Y_s,Z_s) - F_s(\tilde{Y}_s,\tilde{Z}_s)\big]ds + \delta Z_s\cdot dB_s, \qquad \mathbb{P}_0 - \text{a.s.}\end{aligned}$$

Since $\delta Y_0 = 0$, it follows from the L_0—Lipschitz property of F that for all stopping time $\tau\in\mathcal{T}$:

$$0 \ge (\varphi - u)_\tau - \int_0^\tau \big(\phi_s - L_0|\delta Y_s|\big)ds + \int_0^\tau \big(\delta Z_s\cdot dB_s + L_0|\delta Z_s|ds\big), \qquad \mathbb{P}_0 - \text{a.s.}$$

Define $\mathrm{H}_\varepsilon := \varepsilon\wedge\inf\{s>0 : |B_s|\ge\varepsilon\}\wedge\inf\{s>0 : \phi_s - L_0|\delta Y_s| \ge -c\}$, and notice that $\mathrm{H}_\varepsilon > 0$, $\mathbb{P}_0$-a.s. since $\delta Y_0 = 0$. Then,

$$0 \ge (\varphi - u)_{\mathrm{H}_\varepsilon} + c\ \mathrm{H}_\varepsilon + \int_0^{\mathrm{H}_\varepsilon}\big(\delta Z_s\cdot dB_s + L_0|\delta Z_s|ds\big), \qquad \mathbb{P}_0 - \text{a.s.}$$

By the Girsanov theorem, we may find a probability measure $\bar{\mathbb{P}} \in \mathcal{P}^{L_0} \subset \mathcal{P}$ such that $B + L_0 \int_0^{\cdot} \mathrm{sgn}(Z_s)ds$ is a $\bar{\mathbb{P}}$-Brownian motion. Then, it follows from the previous inequality that $\mathbb{E}^{\bar{\mathbb{P}}}\big[(\varphi - u)_{\mathrm{H}}\big] \leq -c\ \mathbb{E}^{\bar{\mathbb{P}}}[\mathrm{H}] < 0$, contradicting the fact that $\varphi \in \underline{\mathcal{A}}^{\mathcal{P}} u_0^0$.

6 Wellposedness of Fully Nonlinear Path-Dependent PDEs

In this section, we outline the main results established in [14] in the context of the fully nonlinear path-dependent PDE:

$$\mathcal{L}u := -\partial_t u - G\big(., u, \partial_\omega u, \partial^2_{\omega\omega} u\big) = 0 \text{ on } [0, T) \times \Omega. \tag{6.1}$$

Assumption 6.1 The nonlinearity G satisfies:

(i) The process $G(., y, z, \gamma)$ is continuous, and $G(., 0, \mathbf{0}, \mathbf{0})$ is bounded.
(ii) G is elliptic, i.e. nondecreasing in γ.
(iii) G is L_0-Lipschitz in (y, z, γ), uniformly in (t, ω).

In the present fully nonlinear context, we shall consider Definition 3.8 of viscosity solutions with the sets of test processes $\underline{\mathcal{A}}$ and $\overline{\mathcal{A}}$ defined by means of

$$\mathcal{P} := \mathcal{P}_L \text{ for some } L \geq L_0, \text{ and } \mathcal{H} : = \{\mathrm{H} = t \wedge \mathrm{H}_O : t \in [0, T], 0 \in O \subset \mathbb{R}^d, \\ \text{bounded convex}\},$$

where $\mathrm{H}_O := \inf\{t > 0 : B_t \notin O\}$. Observe that, unlike the semilinear case, the set $\mathcal{P}_L$ of Definition 3.7 is a non-dominated family of probability measures.

Following the same line of argument as in the semilinear case, it is shown in [13] that the following partial comparison results hold true.

Theorem 6.2 *Let $u, v \in UCB(\Omega)$ be viscosity subsolutions and supersolution, respectively of the Eq. (6.1), with $u_T \leq v_T$ on Ω. Assume further that either one of them is in $C^{1,2}_{\mathcal{P}}(\Theta)$. Then, under Assumption 6.1, $u \leq v$ on Θ.*

We next report the wellposedness result from [14] which requires further conditions on the path-frozen PDE:

$$(\mathrm{E})^{t,\omega}_{\varepsilon}\ g_{t,\omega}\big(s, v(s, x), Dv(s, x), D^2 v(s, x)\big) = 0,\ (s, x) \in Q^{\varepsilon}_t := [t, T] \times B_{\mathbb{R}^d}(\varepsilon),$$

where $B_{\mathbb{R}^d}(\varepsilon)$ is the centered open ball of $\mathbb{R}^d$ with radius ε. We denote the parabolic boundary of the domain Q^{ε}_t by $\partial Q^{\varepsilon}_t := [t, T) \times B_{\mathbb{R}^d}(\varepsilon) \cup \{T\} \times \mathrm{cl}\big[B_{\mathbb{R}^d}(\varepsilon)\big]$.

Assumption 6.3 (i) The process $G(\cdot, y, z, \gamma)$ is uniformly continuous, uniformly in (y, z, γ);

(ii) For all $\varepsilon > 0$, $(t, \omega) \in \Theta$, and $h \in C^0\big(\partial Q_t^\varepsilon\big)$, we have $\overline{v} = \underline{v}$, where:

$$\begin{aligned}\overline{v}(s,x) &:= \inf\Big\{w(s,x) : w \text{ classical supersolution of } (\mathrm{E})_\varepsilon^{t,\omega} \text{ and } w \geq h \text{ on } \partial Q_t^\varepsilon\Big\},\\ \underline{v}(s,x) &:= \sup\Big\{w(s,x) : w \text{ classical subsolution of } (\mathrm{E})_\varepsilon^{t,\omega} \text{ and } w \leq h \text{ on } \partial Q_t^\varepsilon\Big\}.\end{aligned} \tag{6.2}$$

Remark 6.4 The following sufficient condition for the nonlinearity $g := g_{t,\omega}$ to satisfy Assumption 6.3(ii) is reported from Proposition 8.2 of [14]:

(i) The nonlinearity $g\big(s, y, z, \gamma\big)$ is continuous in s, uniformly Lipschitz in (y, z, γ), and non-decreasing in γ,

(ii) The PDE $(\mathrm{E})_\varepsilon^{t,\omega}$ satisfies existence and comparison in the sense of viscosity solutions within the class of bounded functions,

(iii) Either one of the following conditions holds:

(iii-1) g is convex in (y, z, γ), $g_\delta(., \gamma) := \inf_{A\in\mathbb{S}_d, A\geq 0}\big\{g(., \gamma + A) - \mathrm{Tr}[A]\big\} > -\infty$ for $0 \leq \delta \leq c_0$, for some $c_0 > 0$, and $g_\delta \longrightarrow g$ as $\delta \searrow 0$,

(iii-2) g is convex in γ and uniformly elliptic: for some constant $c_0 > 0$,

$$g(., \gamma) - g(., \gamma') \geq c_0 \mathrm{Tr}[\gamma - \gamma'] \text{ for any } \gamma \geq \gamma'.$$

(iii-3) g is uniformly elliptic and $d \leq 2$.

We finally formulate a technical condition on the final condition ξ. We shall denote $\overline{\omega} := \max_{s\leq t} \omega_s$, $\underline{\omega} := \min_{s\leq t} \omega_s$, and $\omega_s^t := \omega_s - \omega_t$ for all $0 \leq t \leq s \leq T$.

Assumption 6.5 $\xi = g\big(\big(\omega_{t_i}, \overline{\omega}_{t_i}, \underline{\omega}_{t_i}\big)_{1\leq i\leq n}, \omega\big)$ for some $0 = t_0 < \cdots < t_n = T$ and some function $g \in \mathrm{UCB}(\mathbb{R}^{3dn} \times \Omega)$ satisfying for all $\theta \in \mathbb{R}^{3dn}$, $i < n$, and $\omega, \omega' \in \Omega$, there exist some $p > 0$ and continuity modulus ρ such that:

$$|g(\theta, \omega) - g(\theta, \omega')| \leq \rho\Big(\big\|(\omega - \omega')\big\|^p_{\mathbb{L}^p([t_i, t_{i+1}])}\Big)$$
$$\text{whenever } \omega_{\wedge t_i} = \omega'_{\wedge t_i} \text{ and } \omega^{t_{i+1}} = \omega'^{t_{i+1}}.$$

We are now able for the wellposedness result proved in [14].

Theorem 6.6 *Let Assumptions 6.1, 6.3, 6.5 hold true.*

(i) *Let $u, v \in UCB(\Omega)$ be $\mathcal{P}$-viscosity subsolution and supersolution, respectively, of PPDE (6.1) with $u_T \leq \xi \leq v_T$. Then $u \leq v$ on Θ.*

(ii) *The PPDE 6.1 with terminal condition ξ has a unique viscosity solution $u \in UCB(\Theta)$.*

7 Stability of Viscosity Solutions of Path-Dependent PDEs

7.1 Stability

We shall establish the stability in the context of fully nonlinear PPDE, and thus we use the setting in Sect. 6. We first report the fully nonlinear analogue of Lemmas 4.3 and 5.5.

Lemma 7.1 *Let $u \in UCB(\Theta)$, $(t,\omega) \in \Theta$, and $\mathrm{H} \in \mathcal{H}$, be such that $u_t(\omega) > \overline{\mathcal{E}}^{\mathcal{P}}\big[u_{\mathrm{H}}^{t,\omega}\big]$. Then,*

$$0 \in \underline{\mathcal{A}}^{\mathcal{P}} u_{t+t^*}(\omega \otimes_t \omega^*) \text{ for some } (t^*,\omega^*) \text{ with the localization } \mathrm{H}^* := \mathrm{H}^{t^*,\omega^*} - t^* \in \mathcal{H}.$$

We remark that in this case $\mathcal{P}$ has no dominating measure, and consequently the Dominated Convergence Theorem fails under $\overline{\mathcal{E}}^{\mathcal{P}}$. The proof of Lemma 7.1 relies on the theory of optimal stopping under nondominated nonlinear expectation. The Snell envelop approach in this context is rather technical, and makes crucially use of the regularity of X and the special structure of H, see [12].

We now present the stability result. Fix $\mathcal{P} = \mathcal{P}_L$ and simplify the notations: $\overline{\mathcal{E}} := \overline{\mathcal{E}}^{\mathcal{P}}$, $\underline{\mathcal{E}} := \underline{\mathcal{E}}^{\mathcal{P}}$.

Theorem 7.2 *Let G, G^ε satisfy Assumption 6.1 with a common $L_0 \le L$, and $u, u^\varepsilon \in UCB(\Omega)$, for each $\varepsilon > 0$. Assume*

(i) *for each $\varepsilon > 0$, u^ε is a viscosity subsolution (resp. supersolution) of PPDE 6.1 with generator G^ε;*
(ii) *as $\varepsilon \to 0$, $(G^\varepsilon, u^\varepsilon)$ converge to (G, u) locally uniformly.*

Then u is a viscosity subsolution (resp. supersolution) of PPDE 6.1 with generator G.

Proof Without loss of generality we shall only prove the viscosity subsolution property at $(0,\mathbf{0})$. Let $\varphi \in \underline{\mathcal{A}}^{\mathcal{P}} u(0,\mathbf{0})$ with corresponding $\mathrm{H} \in \mathcal{H}$, $\delta_0 > 0$ be a constant such that $\mathrm{H}_{\delta_0} \le \mathrm{H}$ and $\lim_{\varepsilon\to 0} \rho(\varepsilon,\delta_0) = 0$, where $\rho(\varepsilon,\delta_0)$ is the bound of $|G^\varepsilon - G| + |u^\varepsilon - u|$ on the δ_0-neighborhood of $(0,0,y_0,z_0,\gamma_0) := (0,0,u_0,\partial_\omega\varphi_0,\partial^2_{\omega\omega}\varphi_0)$.

Now for $0 < \delta \le \delta_0$, denote $\varphi_\delta(t,\omega) := \varphi(t,\omega) + \delta t$. One can easily show that $\underline{\mathcal{E}}_0[\mathrm{H}_\delta] > 0$, see [13]. Then we have

$$\begin{aligned}(\varphi_\delta - u)_0 = (\varphi - u)_0 &\le \underline{\mathcal{E}}\Big[(\varphi - u)_{\mathrm{H}_\delta}\Big] = \underline{\mathcal{E}}\Big[(\varphi_\delta - u)_{\mathrm{H}_\delta} - \delta \mathrm{H}_\delta\Big]\\ &\le \underline{\mathcal{E}}\Big[(\varphi_\delta - u)_{\mathrm{H}_\delta}\Big] - \delta\underline{\mathcal{E}}[\mathrm{H}_\delta] < \underline{\mathcal{E}}\Big[(\varphi_\delta - u)_{\mathrm{H}_\delta}\Big].\end{aligned}$$

By the local uniform convergence of G^ε and u^ε, there exists $\varepsilon_\delta > 0$ small enough such that

$$(\varphi_\delta - u^\varepsilon)_0 < \underline{\mathcal{E}}\Big[(\varphi_\delta - u^\varepsilon)_{\mathrm{H}_\delta}\Big], \quad \forall \varepsilon \le \varepsilon_\delta. \tag{7.1}$$

By Lemma 7.1, we may find a point (t^*, ω^*) such that

$$0 \in \underline{\mathcal{A}}^{\mathcal{P}}(u^\epsilon - \varphi_\delta)_{t^*}(\omega^*) \text{ with the localization } \mathrm{H}^* := \mathrm{H}_\delta^{t^*,\omega^*} - t^* \in \mathcal{H}.$$

It is straightforward to check that $\varphi_\delta^{t^*,\omega^*} \in \underline{\mathcal{A}}^{\mathcal{P}} u^\varepsilon(t^*, \omega^*)$. Since u^ε is a viscosity subsolution of PPDE 6.1 with generator G^ε, we have

$$\begin{aligned} 0 &\ge \Big[-\partial_t \varphi_\delta - G^\varepsilon(\cdot, u^\varepsilon, \partial_\omega \varphi_\delta, \partial^2_{\omega\omega}\varphi_\delta)\Big](t^*, \omega^*) \\ &= \Big[-\partial_t \varphi - \delta - G^\varepsilon(\cdot, u^\varepsilon, \partial_\omega \varphi, \partial^2_{\omega\omega}\varphi)\Big](t^*, \omega^*). \end{aligned} \tag{7.2}$$

Note that $t^* < \mathrm{H}_\delta(\omega^*)$, then $|u^\varepsilon - u|(t^*, \omega^*) \le \rho(\varepsilon, \delta) \le \rho(\varepsilon, \delta_0)$. By local uniform convergence, we may set δ small enough and then ε small enough so that $(\cdot, u^\varepsilon, \partial_\omega\varphi, \partial^2_{\omega\omega}\varphi)(t^*, \omega^*)$ is in the δ_0-neighborhood of $(0, 0, y_0, z_0, \gamma_0)$. Thus, 7.2 and Assumption 6.1 lead to

$$\begin{aligned} 0 &\ge \Big[-\partial_t \varphi - G(\cdot, u^\varepsilon, \partial_\omega \varphi, \partial^2_{\omega\omega}\varphi)\Big](t^*, \omega^*) - \delta - \rho(\varepsilon, \delta_0) \\ &\ge \Big[-\partial_t \varphi - G(\cdot, u, \partial_\omega \varphi, \partial^2_{\omega\omega}\varphi)\Big](t^*, \omega^*) - \delta - \rho(\varepsilon, \delta_0) - C\rho(\varepsilon, \delta) \\ &\ge \mathcal{L}\varphi_0 - \sup_{(t,\omega): t < \mathrm{H}_\delta(\omega)} \Big| G(\cdot, u, \partial_\omega\varphi, \partial^2_{\omega\omega}\varphi)(t, \omega) - G(\cdot, u, \partial_\omega\varphi, \partial^2_{\omega\omega}\varphi)(0, 0)\Big| \\ &\quad -\delta - C\rho(\varepsilon, \delta_0). \end{aligned}$$

Now by first sending $\varepsilon \to 0$ and then $\delta \to 0$ we obtain $\mathcal{L}\varphi_0 \le 0$. Since $\varphi \in \underline{\mathcal{A}}^{\mathcal{P}} u(0, \mathbf{0})$ is arbitrary, we see that u is a viscosity subsolution of PPDE 6.1 with generator G at $(0, \mathbf{0})$ and thus complete the proof. □

7.2 Monotone Scheme for PPDEs

As an important application of the above stability result (in spirit), in this subsection we study discretization schemes for PPDEs. For any $(t, \omega) \in \Theta$ and $h \in (0, T - t)$, we denote $\mathcal{F}^{t,\omega}_{t+h} := \mathcal{F}_{t+h} \cap \{B_{t\wedge\cdot} = \omega_{t\wedge\cdot}\}$. Let $\mathbb{T}^{t,\omega}_h$ be an operator on $\mathbb{L}^0(\mathcal{F}^{t,\omega}_{t+h})$. For $n \ge 1$, denote $h := \frac{T}{n}$, $t_i := ih$, $i = 0, 1, \ldots, n$, and define:

$$\begin{aligned} &u^h(t_n, \omega) := \xi(\omega), \quad u^h(t, \omega) := \mathbb{T}^{t,\omega}_{t_i - t}\big[u^h(t_i, \cdot)\big], \\ &t \in [t_{i-1}, t_i), \quad i = n, \ldots, 1. \end{aligned} \tag{7.3}$$

where we abuse the notation that:

$$\mathbb{T}_h^{t,\omega}[\varphi_s] := \mathbb{T}_h^{t,\omega}[\varphi_{s-t}^{t,\omega}] \text{ for process } \varphi.$$

Assumption 7.3 Assumption 6.1 holds, and

(i) $\xi : \Omega \to \mathbb{R}$ is bounded and uniformly continuous.
(ii) Comparison principle for PPDE 1.2 holds in the class of bounded viscosity solutions.

Assumption 7.4 The descritization operator $\mathbb{T}_h^{t,\omega}$ satisfies the following conditions:

(i) Consistency: for any $(t, \omega) \in \Theta$ and $\varphi \in C^{1,2}(\Theta)$,

$$\lim_{(t',\omega',h,c)\to(t,\mathbf{0},0,0)} \frac{[c+\varphi](t',\omega') - \mathbb{T}_h^{t',\omega\otimes_t\omega'}\big[[c+\varphi](t'+h,\cdot)\big]}{h} = \mathcal{L}\varphi(t,\omega).$$

where $(t', \omega') \in \Theta$, $h \in (0, T-t)$, $c \in \mathbb{R}$.
(ii) Monotonicity: for some constant $L \geq L_0$ and any $\varphi, \psi \in \mathrm{UCB}(\mathcal{F}_{t+h}^t)$,

$$\overline{\mathcal{E}}^{\mathcal{P}}[(\varphi-\psi)^{t,\omega}] \leq 0 \quad \text{implies} \quad \mathbb{T}_h^{t,\omega}[\varphi] \leq \mathbb{T}_h^{t,\omega}[\psi].$$

(iii) Stability: u^h is uniformly bounded and uniformly continuous in ω, uniformly on h. Moreover, there exists a modulus of continuity function ρ, independent of h, such that

$$|u^h(t,\omega) - u^h(t',\omega_{\cdot\wedge t})| \leq \rho\Big((t'-t)\vee h\Big), \quad \text{for any} t < t' \text{and any } \omega \in \Omega.$$

We now report the result from [32], which extends the seminal work Barles and Souganidis [1] to our path dependent case.

Theorem 7.5 *Let Assumptions 7.3 and 7.4 hold. Then PPDE 6.1 with terminal condition $u(T,\cdot) = \xi$ has a unique bounded viscosity solution u, and u_h converges to u locally uniformly as $h \to 0$.*

Proof By the stability, u^h is bounded. Define

$$\underline{u}(t,\omega) := \liminf_{h\to 0} u^h(t,\omega), \quad \overline{u}(t,\omega) := \limsup_{h\to 0} u^h(t,\omega). \tag{7.4}$$

Clearly $\underline{u}(T,\omega) = \xi(\omega) = \overline{u}(T,\omega)$, $\underline{u} \leq \overline{u}$, and $\underline{u}, \overline{u}$ are bounded and uniformly continuous. We shall show that $\underline{u}$ (resp. $\overline{u}$) is a viscosity supersolution (resp. subsolution) of PPDE 6.1. Then by the comparison principle we see that $\overline{u} \leq \underline{u}$ and thus $u := \overline{u} = \underline{u}$ is the unique viscosity solution of PPDE 6.1. The convergence of u^h is obvious now, which, together with the uniform regularity of u^h and u, implies further the locally uniform convergence.

Without loss of generality, we shall only prove by contradiction that $\underline{u}$ satisfies the viscosity supersolution property at $(0, \mathbf{0})$. Assume not, then there exists $\varphi^0 \in \overline{\mathcal{A}}^{\mathcal{P}}\underline{u}(0, \mathbf{0})$ with corresponding $\mathrm{H} \in \mathcal{H}$ such that $-c_0 := \mathcal{L}\varphi^0(0, 0) < 0$. Denote

$$\varphi(t, \omega) := \varphi^0(t, \omega) - \frac{c_0}{2}t. \tag{7.5}$$

Then

$$\mathcal{L}\varphi(0, 0) = -\frac{c_0}{2} < 0. \tag{7.6}$$

Denote $X^0 := \varphi - \underline{u}$, $X^h := \varphi - u^h$, $\mathrm{H}_\varepsilon := \mathrm{H}^0_\varepsilon \wedge \varepsilon^5 := \inf\{t : |B_t| \geq \varepsilon\} \wedge \varepsilon^5$, and $c_\varepsilon := \frac{1}{3}c_0\varepsilon^5$. Note that $\mathrm{H}_\varepsilon \leq \mathrm{H}$ for ε small enough, and by [13] (2.8),

$$\sup_{\mathbb{P}\in\mathcal{P}_L} \mathbb{P}(\mathrm{H}_\varepsilon \neq \varepsilon^5) = \sup_{\mathbb{P}\in\mathcal{P}_L} \mathbb{P}(\mathrm{H}^0_\varepsilon < \varepsilon^5) \leq CL^4\varepsilon^{-4}\varepsilon^{10} \leq C\varepsilon c_\varepsilon. \tag{7.7}$$

Then

$$\overline{\mathcal{E}}[\varepsilon^5 - \mathrm{H}_\varepsilon] \leq \overline{\mathcal{E}}\Big[\varepsilon^5 \mathbf{1}_{\{\mathrm{H}_\varepsilon \neq \varepsilon^5\}}\Big] \leq C\varepsilon c_\varepsilon.$$

Thus, for ε small, it follows from $\varphi^0 \in \overline{\mathcal{A}}^L\underline{u}(0, \mathbf{0})$ that

$$\begin{aligned}
X^0_0 - \overline{\mathcal{E}}[X^0_{\mathrm{H}_\varepsilon}] &= [\varphi^0 - \underline{u}]_0 - \overline{\mathcal{E}}\Big[(\varphi^0 - \underline{u})_{\mathrm{H}_\varepsilon} - \frac{c_0}{2}\mathrm{H}_\varepsilon\Big] \\
&\geq \overline{\mathcal{E}}\Big[(\varphi^0 - \underline{u})_{\mathrm{H}_\varepsilon}\Big] - \overline{\mathcal{E}}\Big[(\varphi^0 - \underline{u})_{\mathrm{H}_\varepsilon} - \frac{c_0}{2}\mathrm{H}_\varepsilon\Big] \\
&\geq \underline{\mathcal{E}}\Big[\frac{c_0}{2}\mathrm{H}_\varepsilon\Big] = \frac{c_0\varepsilon^5}{2} - \frac{c_0}{2}\overline{\mathcal{E}}[\varepsilon^5 - \mathrm{H}_\varepsilon] \\
&\geq \frac{3c_\varepsilon}{2} - C\varepsilon c_\varepsilon \geq c_\varepsilon > 0.
\end{aligned} \tag{7.8}$$

Let $h_k \downarrow 0$ be a sequence such that

$$\lim_{k\to\infty} u^{h_k}_0 = \underline{u}_0, \tag{7.9}$$

and simplify the notations: $u^k := u^{h_k}$, $X^k := X^{h_k}$. Then (7.8) leads to

$$\begin{aligned}
c_\varepsilon &\leq [\varphi_0 - \liminf_{h\to0} u^h_0] - \overline{\mathcal{E}}\Big[\varphi_{\mathrm{H}_\varepsilon} - \liminf_{h\to0} u^h_{\mathrm{H}_\varepsilon}\Big] \\
&\leq [\varphi_0 - \lim_{k\to\infty} u^k_0] - \overline{\mathcal{E}}\Big[\varphi_{\mathrm{H}_\varepsilon} - \liminf_{k\to\infty} u^k_{\mathrm{H}_\varepsilon}\Big].
\end{aligned}$$

Note that X^k is uniformly bounded. Then by (7.7) we have

$$\overline{\mathcal{E}}\Big[|X^k_{\mathrm{H}_\varepsilon} - X^k_{\varepsilon^5}|\Big] \leq C\varepsilon c_\varepsilon.$$

Since u^h is uniformly continuous, applying the monotone convergence theorem under nonlinear expectation $\overline{\mathcal{E}}$, see e.g. [12] Proposition 2.5, we have

$$\begin{aligned} c_\varepsilon &\leq \lim_{k\to\infty}[\varphi_0 - u^k_0] - \overline{\mathcal{E}}\Big[\limsup_{k\to\infty}[\varphi_{\mathrm{H}_\varepsilon} - u^k_{\mathrm{H}_\varepsilon}]\Big] \\ &\leq \lim_{k\to\infty} X^k_0 - \overline{\mathcal{E}}\Big[\limsup_{k\to\infty} X^k_{\varepsilon^5}\Big] + C\varepsilon c_\varepsilon = \lim_{k\to\infty} X^k_0 - \overline{\mathcal{E}}\Big[\lim_{m\to\infty}\sup_{k\geq m} X^k_{\varepsilon^5}\Big] + C\varepsilon c_\varepsilon \\ &= \lim_{k\to\infty} X^k_0 - \lim_{m\to\infty}\overline{\mathcal{E}}\Big[\sup_{k\geq m} X^k_{\varepsilon^5}\Big] + C\varepsilon c_\varepsilon \leq \lim_{k\to\infty} X^k_0 - \limsup_{k\to\infty}\overline{\mathcal{E}}\Big[X^k_{\varepsilon^5}\Big] + C\varepsilon c_\varepsilon \\ &\leq \lim_{k\to\infty} X^k_0 - \limsup_{k\to\infty}\overline{\mathcal{E}}\Big[X^k_{\mathrm{H}_\varepsilon}\Big] + C\varepsilon c_\varepsilon = \liminf_{k\to\infty}\Big[X^k_0 - \overline{\mathcal{E}}[X^k_{\mathrm{H}_\varepsilon}]\Big] + C\varepsilon c_\varepsilon. \end{aligned}$$

Then, for all ε small enough and k large enough,

$$X^k_0 - \overline{\mathcal{E}}[X^k_{\mathrm{H}_\varepsilon}] \geq \frac{c_\varepsilon}{2}. \tag{7.10}$$

Now, applying Lemma 7.1, we obtain that

$$0 \in \underline{\mathcal{A}}^{\mathcal{P}} X^k_{t^k_*}(\omega^*) \text{ for some } (t^k_*, \omega^*) \text{ with the localization } \mathrm{H}^k_\varepsilon := \mathrm{H}^{t^k_*,\omega^*}_\varepsilon - t^k_*.$$

Moreover, in this case, we may prove that

$$\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\Big[\mathrm{H}^k_\epsilon \leq \delta\Big] \leq C\delta^2 \tag{7.11}$$

for all $\delta \leq h_k$ (see [32]). Let $\{t^k_i, i = 0, \ldots, n_k\}$ denote the time partition corresponding to h_k, and assume $t^k_{i-1} \leq t^k_* < t^k_i$. Note that

$$X^k_{t^k_*}(\omega^k) = Y^k_{t^k_*}(\omega^k) \geq \overline{\mathcal{E}}\Big[(X^k)^{t^k_*,\omega^k}_{\tau\wedge\mathrm{H}^k_\varepsilon}\Big], \quad \forall \tau \in \mathcal{T}.$$

Set $\delta_k := t^k_i - t^k_* \leq h_k$ and $\tau := \delta_k$. Combine the above inequality and 7.11 we have

$$[\varphi - u^k](t^k_*, \omega^k) \geq \overline{\mathcal{E}}\Big[(\varphi - u^k)^{t^k_*,\omega^k}_{\delta_k\wedge\mathrm{H}^k_\varepsilon}\Big] \geq \overline{\mathcal{E}}\Big[(\varphi - u^k)^{t^k_*,\omega^k}_{\delta_k}\Big] - C\delta_k^2.$$

This implies

$$\overline{\mathcal{E}}\Big[\Big(\varphi^{t^k_*,\omega^k}_{\delta_k} - [\varphi - u^k](t^k_*, \omega^k) - C\delta_k^2\Big) - (u^k)^{t^k_*,\omega^k}_{\delta_k}\Big] \leq 0.$$

By the monotonicity condition (Assumption 7.4 (ii)) we have

$$u^k(t_*^k, \omega^k) = \mathbb{T}_{\delta_k}^{t_*^k,\omega^k}[u_{t_i^k}^k] \le \mathbb{T}_{\delta_k}^{t_*^k,\omega^k}\Big[\varphi_{t_i^k} - [\varphi - u^k](t_*^k, \omega^k) - C\delta_k^2\Big]. \quad (7.12)$$

We next use the consistency condition (Assumption 7.4 (i)). For $(t, \omega) = (0, \mathbf{0})$, set

$$t' := t_*^k, \quad \omega' := \omega^k, \quad h := \delta_k, \quad c := -[\varphi - u^k](t_*^k, \omega^k) - C\delta_k^2.$$

By first sending $k \to \infty$ and then $\varepsilon \to 0$, we see that

$$d((t_*^k, \omega^k), (0, \mathbf{0})) \le \mathrm{H}_\varepsilon + \sup_{0\le t\le \mathrm{H}_\varepsilon} |\omega_t^k| \le 2\varepsilon \to 0, \quad h \le h_k \to 0,$$

which, together with 7.5, 7.9, and the uniform continuity of φ and u^k, implies

$$|c| \le \Big|[\varphi - u^k](t_*^k, \omega^k) - [\varphi - u^k](0, \mathbf{0})\Big| + |u_0^k - \underline{u}_0| + C\delta_k^2 \to 0.$$

Then, by the consistency condition, we obtain from 7.12 that

$$\begin{aligned} 0 &\le \frac{u^k(t_*^k, \omega^k) - \mathbb{T}_{\delta_k}^{t_*^k,\omega^k}\Big[\varphi_{t_i^k} - [\varphi - u^k](t_*^k, \omega^k) - C\delta_k^2\Big]}{\delta_k} \\ &= \frac{[c + \varphi](t_*^k, \omega^k) - \mathbb{T}_{\delta_k}^{t_*^k,\omega^k}\Big[[c + \varphi]_{t_i^k}\Big]}{\delta_k} + C\delta_k \to \mathcal{L}\varphi(0, \mathbf{0}). \end{aligned}$$

This contradicts with 7.6.

8 Optimal Stopping Under Dominated Nonlinear Expectation

The objective of this section is to provide a self-contained proof of Theorem 5.2. We follow the setting in Sect. 5. In particular, the family $\mathcal{P}$ of equivalent probability measures is defined as in (5.2).

We emphasize that the main results of this section are available in the literature in reflected backward stochastic differential equations, see [16, 21, 27]. We collect them here for the convenience of the readers who might not be familiar with this literature. Our presentation in Sect. 8.3 and 8.2 is inspired from El Karoui [15] and Appendix D of Karatzas and Shreve [22], which are focused on the standard optimal stopping under linear expectation.

8.1 Preliminaries

For ease of notation, we simply write $\overline{\mathcal{E}} := \overline{\mathcal{E}}^{\mathcal{P}}$. We start by the dominated convergence theorem under $\overline{\mathcal{E}}$ which holds by the fact that $\mathcal{P}$ is dominated by $\mathbb{P}_0$.

Lemma 8.1 *Let X_n be a sequence of random variables. Assume that $X_n^{1+\alpha}$ are uniformly integrable under probability $\mathbb{P}_0$ and $X_n \to X$ $\mathbb{P}_0$-a.s. Then, we have $\overline{\mathcal{E}}[|X_n - X|] \to 0$.*

Proof For any $\mathbb{P}_\lambda \in \mathcal{P}$, we have

$$\begin{aligned}\mathbb{E}^{\mathbb{P}_\lambda}[|X_n - X|] &= \mathbb{E}^{\mathbb{P}_0}\Big[e^{\int_0^T \lambda_s dB_s - \frac{1}{2}\int_0^T |\lambda_s|^2 ds}|X_n - X|\Big] \\ &\le \big(\mathbb{E}^{\mathbb{P}_0}[e^{\int_0^T q\lambda_s dB_s - \int_0^T \frac{q}{2}|\lambda_s|^2 ds}]\big)^{\frac{1}{q}}\big(\mathbb{E}^{\mathbb{P}_0}[|X_n - X|^p]\big)^{\frac{1}{p}},\end{aligned}$$

where $p = 1 + \alpha$ and $\frac{1}{p} + \frac{1}{q} = 1$. Since λ is bounded, we know that $\mathbb{E}^{\mathbb{P}_0}\big[e^{\int_0^T q\lambda_s dB_s - \int_0^T \frac{q}{2}|\lambda_s|^2 ds}\big]$ is bounded. Then, by the convergence theorem, we obtain that $\mathbb{E}^{\mathbb{P}_0}[|X_n - X|^{1+\alpha}] \to 0$. The proof is complete. □

Lemma 8.2 *If $X \ge 0$ $\mathbb{P}_0$-a.s. and $\underline{\mathcal{E}}[X] = 0$, then $X = 0$ $\mathbb{P}_0$-a.s.*

Proof Since $\underline{\mathcal{E}}[X] = 0$, for any $\epsilon > 0$ there exists $\mathbb{P}^\epsilon \in \mathcal{P}$ such that $\mathbb{E}^{\mathbb{P}^\epsilon}[X] < \epsilon$. Also, by Cauchy-Schwarz inequality, we have the estimate:

$$\mathbb{E}^{\mathbb{P}_0}[X^{\frac{1}{2}}] = \mathbb{E}^{\mathbb{P}^\epsilon}\Big[e^{-\int_0^T \lambda^\epsilon dB_s - \frac{1}{2}\int_0^T |\lambda^\epsilon|^2 ds} X^{\frac{1}{2}}\Big] \le C\mathbb{E}^{\mathbb{P}^\epsilon}[X]^{\frac{1}{2}} < C\epsilon^{\frac{1}{2}}.$$

Since ϵ is arbitrary, we get $\mathbb{E}^{\mathbb{P}_0}[X^{\frac{1}{2}}] = 0$. So, we conclude that $X = 0$ $\mathbb{P}_0$-a.s. □

Finally, we state the following lemma, which is a direct consequence of Proposition 3.1. in El Karoui et al. [17].

Lemma 8.3 *Let $\xi \in \mathbb{L}^2(\mathbb{P}_0)$, and $v_t := \operatorname{ess\,sup}_{\mathbb{P}\in\mathcal{P}} \mathbb{E}^{\mathbb{P}}[\xi|\mathcal{F}_t]$. Then, $v_t = \mathbb{E}^{\bar{\mathbb{P}}}[\xi|\mathcal{F}_t]$ $\mathbb{P}_0$-a.s. for all $t \in [0, T]$ for some $\bar{\mathbb{P}} \in \mathcal{P}$.*

8.2 RCLL Version of the $\mathbb{F}^*$-Snell Envelop

Throughout this section, we consider a process $X : [0, T] \times \Omega \longrightarrow \mathbb{R}$ satisfying the following condition.

Assumption 8.4 The process X is piecewise pathwise continuous $\mathbb{F}$-adapted on $[0,T]$, and $\sup_{t\in[0,T]}|X_t| \in \mathbb{L}^2(\mathcal{P})$, i.e.

$$\mathbb{E}^{\mathbb{P}}\Big[\sup_{t\in[0,T]}|X_t|^2\Big] < \infty, \quad \text{for all } \mathbb{P}\in\mathcal{P}.$$

Our starting point is the classical Snell envelop process:

$$Y_t := \operatorname*{ess\,sup}_{\tau\in\mathcal{T}_*^t,\mathbb{P}\in\mathcal{P}} \mathbb{E}^{\mathbb{P}}[X_\tau|\mathcal{F}_t], \qquad t\in[0,T].$$

Clearly, Y_t is $\mathcal{F}_t^*$-measurable for all $t\in[0,T]$.

Lemma 8.5 *For any $t\in[0,T)$, $\big\{\mathbb{E}^{\mathbb{P}}[X_\tau|\mathcal{F}_t];(\tau,\mathbb{P})\in\mathcal{T}_*^t\times\mathcal{P}\big\}$ satisfies the lattice property.*

Proof Let $\tau_1,\ \tau_2\in\mathcal{T}_*^t$ and $\mathbb{P}_1,\ \mathbb{P}_2\in\mathcal{P}$. Let $A:=\{\mathbb{E}^{\mathbb{P}_1}[X_{\tau_1}|\mathcal{F}_t]\ge\mathbb{E}^{\mathbb{P}_2}[X_{\tau_2}|\mathcal{F}_t]\}$, and define

$$\bar\tau := \tau_1 1_A + \tau_2 1_{A^c} \text{ and } \bar{\mathbb{P}}(D) := \mathbb{E}^{\mathbb{P}_1}\Big[\mathbb{E}^{\mathbb{P}_1}[1_{A\cap D}|\mathcal{F}_t] + \mathbb{E}^{\mathbb{P}_2}[1_{A^c\cap D}|\mathcal{F}_t]\Big], \ \ D\in\mathcal{F}_T.$$

Clearly, $\bar\tau\in\mathcal{T}_*^t$, $\bar{\mathbb{P}}\in\mathcal{P}$, and we immediately verify that

$$\mathbb{E}^{\bar{\mathbb{P}}}[X_{\bar\tau}|\mathcal{F}_t] \ge \max\{\mathbb{E}^{\mathbb{P}_1}[X_{\tau_1}|\mathcal{F}_t],\mathbb{E}^{\mathbb{P}_2}[X_{\tau_2}|\mathcal{F}_t]\}, \quad \mathbb{P}_0\text{-a.s.}$$ □

We next introduce the concatenation $\mathbb{P}_1\otimes_t\mathbb{P}_2$ of two probability measures $\mathbb{P}_1,\mathbb{P}_2\in\mathcal{P}$ by:

$$(\mathbb{P}_1\otimes_t\mathbb{P}_2)(D) := \mathbb{E}^{\mathbb{P}_1}\Big[\mathbb{E}^{\mathbb{P}_2}[1_D|\mathcal{F}_t]\Big] \quad \text{for all } D\in\mathcal{F}_T,$$

and we observe that $\mathbb{P}_1\otimes_t\mathbb{P}_2\in\mathcal{P}$.

Lemma 8.6 *Y is an $\overline{\mathcal{E}}$-supermartingale with $\sup_{t\in[0,T]}\mathbb{E}^{\mathbb{P}_0}\big[|Y_t|^2\big]<\infty$ and $\overline{\mathcal{E}}[Y_t]=\sup_{\tau\in\mathcal{T}_*^t}\overline{\mathcal{E}}[X_\tau]$ for all $t\in[0,T]$.*

Proof Denote $|X|_T^* := \sup_{t\in[0,T]}|X_t|$. By the definition of Y, we have

$$\sup_{t\in[0,T]}\mathbb{E}^{\mathbb{P}_0}\big[|Y_t|^2\big] \le \mathbb{E}^{\mathbb{P}_0}\Big[\operatorname*{ess\,sup}_{\mathbb{P}\in\mathcal{P}}\mathbb{E}^{\mathbb{P}}\big[(|X|_T^*)^2\big|\mathcal{F}_t\big]\Big] \le \sup_{\mathbb{P}\in\mathcal{P}}\mathbb{E}^{\mathbb{P}}\big[(|X|_T^*)^2\big] < \infty.$$

For arbitrary $\mathbb{P}\in\mathcal{P}$ and $s\le t$, it follows from Lemma 8.5 and the property of the ess sup that:

$$\begin{aligned}\mathbb{E}^{\mathbb{P}}[Y_t|\mathcal{F}_s] &= \operatorname*{ess\,sup}_{\tau\in\mathcal{T}_*^t;\mathbb{P}'\in\mathcal{P}} \mathbb{E}^{\mathbb{P}\otimes\mathbb{P}'}[X_\tau|\mathcal{F}_s] \\ &\le \operatorname*{ess\,sup}_{\tau\in\mathcal{T}_*^t;\mathbb{P}'\in\mathcal{P}} \mathbb{E}^{\mathbb{P}'}[X_\tau|\mathcal{F}_s] \le \operatorname*{ess\,sup}_{\tau\in\mathcal{T}_*^s;\mathbb{P}'\in\mathcal{P}} \mathbb{E}^{\mathbb{P}'}[X_\tau|\mathcal{F}_s] = Y_s, \quad \mathbb{P}_0-\text{a.s.}\end{aligned}$$

which proves that Y is an $\overline{\mathcal{E}}$-supermartingale.

We finally prove the last claim. For all $\tau \in \mathcal{T}_*^t$ and $\mathbb{P} \in \mathcal{P}$, we have $Y_t \geq \mathbb{E}^{\mathbb{P}}[X_\tau|\mathcal{F}_t]$, $\mathbb{P}_0$-a.s. Hence, we obtain for any $\tau \in \mathcal{T}_*^t$ and $\mathbb{P}$, $\mathbb{P}' \in \mathcal{P}$ that $\overline{\mathcal{E}}[Y_t] \geq \mathbb{E}^{\mathbb{P}'}[Y_t] \geq \mathbb{E}^{\mathbb{P}'\otimes\mathbb{P}}[X_\tau]$, and therefore $\overline{\mathcal{E}}[Y_t] \geq \sup_{\tau\in\mathcal{T}_*^t} \overline{\mathcal{E}}[X_\tau]$. On the other hand, it follows from Lemma 8.5 that:

$$\mathbb{E}^{\mathbb{P}}[Y_t] = \sup_{\tau\in\mathcal{T}_*^t,\mathbb{P}'\in\mathcal{P}} \mathbb{E}^{\mathbb{P}\otimes\mathbb{P}'}[X_\tau] \leq \sup_{\tau\in\mathcal{T}_*^t} \overline{\mathcal{E}}[X_\tau] \text{ for all } \mathbb{P} \in \mathcal{P}.$$ □

Proposition 8.7 (Dynamic programming principle) *For all $t \in [0, T)$ and $\theta \in \mathcal{T}_*^t$:*

$$Y_t = \operatorname*{ess\,sup}_{\tau\in\mathcal{T}_*^t,\ \mathbb{P}\in\mathcal{P}} \mathbb{E}^{\mathbb{P}}\Big[X_\tau 1_{\{\tau<\theta\}} + Y_\theta 1_{\{\tau\geq\theta\}}\Big|\mathcal{F}_t\Big], \quad \mathbb{P}_0\text{-}a.s.$$

Proof Since $X \leq Y$, we have for all $\theta \in \mathcal{T}_*^t$

$$\begin{aligned} Y_t &\leq \operatorname*{ess\,sup}_{\tau\in\mathcal{T}_*^t,\mathbb{P}\in\mathcal{P}} \mathbb{E}^{\mathbb{P}}[X_\tau 1_{\{\tau<\theta\}} + Y_\tau 1_{\{\tau\geq\theta\}}|\mathcal{F}_t] \\ &\leq \operatorname*{ess\,sup}_{\tau\in\mathcal{T}_*^t,\mathbb{P}\in\mathcal{P}} \mathbb{E}^{\mathbb{P}}[X_\tau 1_{\{\tau<\theta\}} + Y_\theta 1_{\{\tau\geq\theta\}}|\mathcal{F}_t], \quad \mathbb{P}_0 - \text{a.s.} \end{aligned}$$

where the last inequality is due to the $\overline{\mathcal{E}}$-supermartingale property of Y of Lemma 8.6. On the other hand, since Y is $\overline{\mathcal{E}}$-supermartingale, we have for all $\tau \in \mathcal{T}_*^t$ and $\mathbb{P} \in \mathcal{P}$:

$$\begin{aligned} Y_t \geq \mathbb{E}^{\mathbb{P}}[Y_{\theta\wedge\tau}|\mathcal{F}_t] &= \mathbb{E}^{\mathbb{P}}[Y_\theta 1_{\{\tau\geq\theta\}} + Y_\tau 1_{\{\tau<\theta\}}|\mathcal{F}_t] \\ &\geq \mathbb{E}^{\mathbb{P}}[Y_\theta 1_{\{\tau\geq\theta\}} + X_\tau 1_{\{\tau<\theta\}}|\mathcal{F}_t], \quad \mathbb{P}_0\text{-a.s.} \end{aligned}$$

The proof is completed by taking ess sup over $\tau \in \mathcal{T}_*^t$ and $\mathbb{P} \in \mathcal{P}$. □

Lemma 8.8 *Y has a $\mathbb{P}_0$-a.s. RCLL $\mathbb{F}^*$-adapted version. Moreover, there exists $\bar{\mathbb{P}} \in \mathcal{P}$ such that $\mathbb{E}^{\bar{\mathbb{P}}}[\sup_{t\in[0,T]} |Y_t|^2] < \infty$.*

Proof Step 1. Let $\{t_n\} \subset [0, T]$ be such that $t_n \searrow t$. By Lemma 8.6, we know that $\overline{\mathcal{E}}[Y_{t_n}] = \sup_{\tau\in\mathcal{T}_*^{t_n}} \overline{\mathcal{E}}[X_\tau] \leq \sup_{\tau\in\mathcal{T}_*^t} \overline{\mathcal{E}}[X_\tau] \leq \overline{\mathcal{E}}[Y_t]$. On the other hand, for any $\tau \in \mathcal{T}_*^t$, denoting $\tau_n := \tau \vee t_n$, it follows from the continuity of X and the $\mathbb{P}_0$-uniform integrability of $\{X_{\tau_n}^2, n \geq 1\}$ that $\overline{\mathcal{E}}[X_\tau] = \lim_{n\to\infty} \overline{\mathcal{E}}[X_{\tau_n}] \leq \liminf_{n\to\infty} \overline{\mathcal{E}}[Y_{t_n}]$. Using again Lemma 8.6, we obtain that $\overline{\mathcal{E}}[Y_t] \leq \liminf_{n\to\infty} \overline{\mathcal{E}}[Y_{t_n}]$. Hence,

$$\overline{\mathcal{E}}[Y_t] = \lim_{s\downarrow t} \overline{\mathcal{E}}[Y_s].$$

Step 2. It follows from Lemma 8.6 that Y is a $\mathbb{P}_0$-supermartingale in the right-continuous filtration $\mathbb{F}^*$. By classical martingale theory, we know that for any $t \in [0, T)$,

$$Y_{t+} := \lim_{s \downarrow\downarrow t, s \in \mathbb{Q}} Y_s \text{ exists} \quad \mathbb{P}_0\text{-a.s.}$$

Note that Y_{t+} is $\mathcal{F}_t^*$-measurable. Also, we have the properties that $\{Y_{t+}\}_t$ is RCLL and $Y_{t+} = \mathbb{E}[Y_{t+}|\mathcal{F}_t^*] \leq Y_t$, $\mathbb{P}_0$-a.s.

We now show that $Y_{t+} = Y_t$, $\mathbb{P}_0$-a.s. Suppose to the contrary that $\mathbb{P}_0[Y_{t+} < Y_t] > 0$. Then, we have $\mathbb{E}^{\mathbb{P}_0}\big[\sqrt{Y_t - Y_{t+}}\big] > 0$, implying that $\underline{\mathcal{E}}[Y_t - Y_{t+}] > 0$. Then, there exists $\delta > 0$ such that:

$$\mathbb{E}^{\mathbb{P}}[Y_t - Y_{t+}] \;\geq\; \delta \;>\; 0 \text{ for all } \mathbb{P} \in \mathcal{P}. \tag{8.1}$$

By the definition of Y_{t+} and the fact that $\{Y_t^{\frac{3}{2}}\}$ are uniformly integrable (by Lemma 8.6), we obtain by Lemma 8.1 that $\overline{\mathcal{E}}[Y_t] = \lim_{s \downarrow t} \overline{\mathcal{E}}[Y_s] = \overline{\mathcal{E}}[Y_{t+}]$. This means that for all $\mathbb{P} \in \mathcal{P}$ and $\epsilon > 0$, there exists $\mathbb{P}' \in \mathcal{P}$ such that $\mathbb{E}^{\mathbb{P}}[Y_t] - \epsilon \leq \mathbb{E}^{\mathbb{P}'}[Y_{t+}]$. Together with (8.1), this implies that $\mathbb{E}^{\mathbb{P}}[Y_t] - \epsilon \leq \mathbb{E}^{\mathbb{P}'}[Y_t] - \delta \leq \overline{\mathcal{E}}[Y_t] - \delta$, and therefore $\overline{\mathcal{E}}[Y_t] - \epsilon \leq \overline{\mathcal{E}}[Y_t] - \delta$. By arbitrariness of $\epsilon > 0$, this provides that $\overline{\mathcal{E}}[Y_t] \leq \overline{\mathcal{E}}[Y_t] - \delta$, which is the required contradiction. So, we have proved that Y_{t+} is an $\mathbb{F}^*$-adapted RCLL version of Y_t.

Step 3. With $|X|_T^* := \sup_{t \in [0,T]} |X_t|$, we have:

$$\sup_{t \in [0,T]} |Y_t|^2 \leq \sup_{t \in [0,T]} \operatorname*{ess\,sup}_{\mathbb{P} \in \mathcal{P}} \mathbb{E}^{\mathbb{P}}\big[\big(|X|_T^*\big)^2 |\mathcal{F}_t\big].$$

By Lemma 8.3, there exists $\bar{\mathbb{P}} \in \mathcal{P}$ such that $\mathbb{E}^{\bar{\mathbb{P}}}[X^*|\mathcal{F}_t] = \operatorname{ess\,sup}_{\mathbb{P} \in \mathcal{P}} \mathbb{E}^{\mathbb{P}}[X^*|\mathcal{F}_t]$ for all t, $\mathbb{P}_0$-a.s. Then, it follows from the Doob inequality that:

$$\mathbb{E}^{\bar{\mathbb{P}}}\Big[\sup_{t \in [0,T]} |Y_t|^2\Big] \leq \mathbb{E}^{\bar{\mathbb{P}}}\Big[\sup_{t \in [0,T]} \mathbb{E}^{\bar{\mathbb{P}}}\big[(|X|_T^*)^2 \big| \mathcal{F}_t\big]\Big] \;\leq\; 4\mathbb{E}^{\bar{\mathbb{P}}}\big[(|X|_T^*)^2\big],$$

which provides the desired result by Assumption 8.4. □

8.3 Doob-Meyer Decomposition of the RCLL $\mathbb{F}^*$-Snell Envelop

From now on, we consider Y in its $\mathbb{F}^*$-adapted RCLL version of Lemma 8.8. For a vector $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$, we denote $|x|_1 := \sum_{i=1}^d |x_i|$.

Proposition 8.9 *There exist $H \in \mathbb{H}_{loc}$ and a non-decreasing previsible process K such that*

$$Y_t \;=\; Y_0 + (H \cdot B)_t - L \int_0^t |H_s|_1 ds - K_t, \quad t \in [0, T], \quad \mathbb{P}_0 - a.s.,$$

with $\mathbb{E}^{\mathbb{P}_0}\big[\sup_{t\in[0,T]}|(H\cdot B)_t|\big]<\infty$.

Proof **1.** By Lemma 8.6, Y is a $\mathbb{P}$-supermartingale, with Doob-Meyer decomposition,

$$Y \;=\; Y_0+M^{\mathbb{P}}-A^{\mathbb{P}}, \quad \mathbb{P}_0-\text{a.s.} \quad \text{for all } \mathbb{P}\in\mathcal{P}, \tag{8.2}$$

for some $\mathbb{P}$-martingale $M^{\mathbb{P}}$ and some non-decreasing previsible process $A^{\mathbb{P}}$. By the martingale representation property, $M^{\mathbb{P}_0}=(H\cdot B)$, $\mathbb{P}_0$-a.s. for some $H\in\mathbb{H}_{loc}$. By the Girsanov theorem, the process $\tilde{M}^{\mathbb{P}_\lambda}:=M^{\mathbb{P}_0}-\int_0^\cdot \lambda_s^T H_s ds$ defines a $\mathbb{P}_\lambda$-local martingale. Then, it follows from the uniqueness of the Doob-Meyer decomposition that $\tilde{M}^{\mathbb{P}_\lambda}$ is a $\mathbb{P}_\lambda$-martingale, and

$$\tilde{M}^{\mathbb{P}_\lambda}=M^{\mathbb{P}_\lambda} \quad \text{and} \int_0^{\cdot}\lambda_s^T H_s ds-A^{\mathbb{P}_0}=-A^{\mathbb{P}_\lambda}, \quad \mathbb{P}_0-\text{a.s.}$$

We next introduce the process λ^* with ith component proportional to the sign of the ith component of H, so that $\lambda^* H=L|H|_1$. Note that $\mathbb{P}^{\lambda^*}\in\mathcal{P}$. Then the required decomposition holds with $K:=A^{\mathbb{P}_{\lambda^*}}$.

2. By Itô's formula, we have

$$A_t^{\mathbb{P}}Y_t-\int_0^t Y_s dA_s^{\mathbb{P}}=\int_0^t A_s^{\mathbb{P}}dY_s=\int_0^t A_s^{\mathbb{P}}dM_s^{\mathbb{P}}-\frac{1}{2}(A_t^{\mathbb{P}})^2, \quad \text{for all } \mathbb{P}\in\mathcal{P}.$$

Let $(\tau_n)_n$ be a localizing sequence for the $\mathbb{P}$-local martingale $\int_0^\cdot A_s^{\mathbb{P}}dM_s^{\mathbb{P}}$. Then,

$$\frac{1}{2}\mathbb{E}^{\mathbb{P}}[(A_{\tau_n}^{\mathbb{P}})^2]\le 2\mathbb{E}^{\mathbb{P}}[\sup_{t\in[0,T]}|Y_t|\cdot A_{\tau_n}^{\mathbb{P}}]\le 2\Big(\mathbb{E}^{\mathbb{P}}\big[\sup_{t\in[0,T]}|Y_t|^2\big]\mathbb{E}^{\mathbb{P}}\big[(A_{\tau_n}^{\mathbb{P}})^2\big]\Big)^{\frac{1}{2}}.$$

For $\mathbb{P}=\bar{\mathbb{P}}$ as in Lemma 8.8, we conclude that $\mathbb{E}^{\bar{\mathbb{P}}}[(A_T^{\bar{\mathbb{P}}})^2]<\infty$. Then, one may easily verify that $\mathbb{E}^{\bar{\mathbb{P}}}\big[\sup_{t\in[0,T]}|M^{\bar{\mathbb{P}}}|^2\big]<\infty$, and therefore $\mathbb{E}^{\bar{\mathbb{P}}}[\langle M^{\bar{\mathbb{P}}}\rangle_T]<\infty$ by the Burkholder-Davis-Gundy inequality. Then, it follows from the Cauchy-Schwartz inequality that

$$\mathbb{E}^{\mathbb{P}}[\langle M^{\mathbb{P}}\rangle_T^{\frac{1}{2}}] \;=\; \mathbb{E}^{\mathbb{P}}[\langle M^{\bar{\mathbb{P}}}\rangle_T^{\frac{1}{2}}] \;\le\; C(\mathbb{E}^{\bar{\mathbb{P}}}[\langle M^{\bar{\mathbb{P}}}\rangle_T])^{\frac{1}{2}}<\infty, \text{ for all } \mathbb{P}\in\mathcal{P},$$

and we conclude that $\mathbb{E}^{\mathbb{P}}\big[\sup_{t\in[0,T]}|M_t^{\mathbb{P}}|\big]<\infty$, by the Burkholder-Davis-Gundy inequality. □

We next provide some further properties of the previsible nondecreasing process K, and we derive an optimal stopping rule.

Proposition 8.10 *The processes Y and K are pathwise continuous, $\int_0^T 1_{\{t:X_t<Y_t\}} dK_t = 0$, $\mathbb{P}_0$-a.s. and the $\mathbb{F}^*$-previsible stopping time $\tau^* := \inf\{t : X_t = Y_t\}$ is an optimal stopping rule.*

In order to prove this result, we introduce the stopping times

$$D_t^\varepsilon \;:=\; \inf\{s \geq t : Y_s \leq X_s + \varepsilon\} \quad \text{for all } t \in [0,T), \;\; \varepsilon > 0.$$

By the right-continuity of Y and the continuity of X, it is clear that $D_t^\varepsilon \in \mathcal{T}_*^t$. The following two lemmas prepare for the proof of Proposition 8.10.

Lemma 8.11 *For all $t \in [0,T)$, we have $\overline{\mathcal{E}}\big[Y_{D_t^\varepsilon} - Y_t\big] = 0$.*

Proof Since Y is $\mathcal{P}$-supermartingale and $D_t^\varepsilon \geq t$, we have $\overline{\mathcal{E}}[Y_{D_t^\varepsilon} - Y_t] \leq 0$. On the other hand, by the dynamic programming principle of Proposition 8.7, we have

$$Y_t = \operatorname*{ess\,sup}_{\tau\in\mathcal{T}_*^t,\mathbb{P}\in\mathcal{P}} \mathbb{E}^{\mathbb{P}}\Big[X_\tau 1_{\{\tau<D_t^\varepsilon\}} + Y_{D_t^\varepsilon} 1_{\{\tau\geq D_t^\varepsilon\}}|\mathcal{F}_t\Big], \qquad \mathbb{P}_0 - \text{a.s.}$$

Here, we may prove the lattice property similar to Lemma 8.5, so that

$$\mathbb{E}^{\mathbb{P}}[Y_t] \;=\; \sup_{\tau\in\mathcal{T}_*^t,\mathbb{P}'\in\mathcal{P}} \mathbb{E}^{\mathbb{P}\otimes_t\mathbb{P}'}\big[X_\tau \mathbf{1}_{\{\tau<D_t^\varepsilon\}} + Y_{D_t^\varepsilon}\mathbf{1}_{\{\tau\geq D_t^\varepsilon\}}\big] \quad \text{for all } \mathbb{P} \in \mathcal{P}.$$

Then, there exists $(\tau_n)_n \subset \mathcal{T}_*^t$ such that

$$\begin{aligned}\mathbb{E}^{\mathbb{P}}[Y_t] &\leq \mathbb{E}^{\mathbb{P}\otimes_t\mathbb{P}_n}\big[X_{\tau_n}\mathbf{1}_{\{\tau_n<D_t^\varepsilon\}} + Y_{D_t^\varepsilon}\mathbf{1}_{\{\tau_n\geq D_t^\varepsilon\}}\big] + \frac{1}{n}\\ &\leq \mathbb{E}^{\mathbb{P}\otimes_t\mathbb{P}_n}\big[Y_{\tau_n\wedge D_t^\varepsilon} - \varepsilon\mathbf{1}_{\tau_n<D_t^\varepsilon}\big] + \frac{1}{n} \;\leq\; \mathbb{E}^{\mathbb{P}\otimes_t\mathbb{P}_n}\big[Y_t - \varepsilon\mathbf{1}_{\tau_n<D_t^\varepsilon}\big] + \frac{1}{n},\end{aligned}$$

where the last inequality follows from the $\overline{\mathcal{E}}$-supermartingale property of Y. Note that

$$\mathbb{E}^{\mathbb{P}\otimes_t\mathbb{P}_n}[Y_t] = \mathbb{E}^{\mathbb{P}\otimes\mathbb{P}_n}\Big[\mathbb{E}^{\mathbb{P}\otimes_t\mathbb{P}_n}[Y_t|\mathcal{F}_t]\Big] = \mathbb{E}^{\mathbb{P}}\Big[\mathbb{E}^{\mathbb{P}\otimes_t\mathbb{P}_n}[Y_t|\mathcal{F}_t]\Big] = \mathbb{E}^{\mathbb{P}}[Y_t].$$

Then $\varepsilon(\mathbb{P}\otimes\mathbb{P}_n)[\tau_n < D_t^\varepsilon] \leq \frac{1}{n}$, and it follows from the previous estimate that:

$$\begin{aligned}\mathbb{E}^{\mathbb{P}}[Y_t] &\leq \mathbb{E}^{\mathbb{P}\otimes\mathbb{P}_n}\big[(X_{\tau_n} - Y_{D_t^\varepsilon})\mathbf{1}_{\{\tau_n<D_t^\varepsilon\}} + Y_{D_t^\varepsilon}\big] + \frac{1}{n}\\ &\leq C(\mathbb{P}\otimes\mathbb{P}_n)\big[\tau_n < D_t^\varepsilon\big]^{\frac{1}{2}} + \mathbb{E}^{\mathbb{P}\otimes\mathbb{P}_n}[Y_{D_t^\varepsilon}] + \frac{1}{n} \;\leq\; \frac{C}{\sqrt{n\varepsilon}} + \frac{1}{n} + \mathbb{E}^{\mathbb{P}\otimes\mathbb{P}_n}\big[Y_{D_t^\varepsilon}\big],\end{aligned}$$

by the fact that $\sup_{t\in[0,T]}|X_t|$ and $Y_{D_t^\varepsilon} \in [X_{D_t^\varepsilon}, X_{D_t^\varepsilon} + \varepsilon]$ are both in $\mathbb{L}^2(\mathcal{P})$. Finally, we obtain

$$\overline{\mathcal{E}}^{\mathcal{P}}\big[Y_{D_t^\varepsilon} - Y_t\big] \;\geq\; \mathbb{E}^{\mathbb{P}\otimes\mathbb{P}_n}\big[Y_{D_t^\varepsilon} - Y_t\big] \;\geq\; -\Big(\frac{C}{\sqrt{n\varepsilon}} + \frac{1}{n}\Big) \;\longrightarrow\; 0 \text{ as } n \to \infty. \;\square$$

Lemma 8.12 *The processes* $\{K_t, t \in [0,T]\}$ *and* $\{K_{D_t^\varepsilon}, t \in [0,T]\}_t$ *are indistinguishable.*

Proof By the decomposition of Proposition 8.9, we have

$$Y_{D_t^\varepsilon} - Y_t = + \int_t^{D_t^\epsilon} H_s dB_s - \int_t^{D_t^\epsilon} L|H_s|_1 ds - K_{D_t^\varepsilon} + K_t, \;\; t \in [0,T], \quad \mathbb{P}_0\text{-a.s.}$$

Since $\overline{\mathcal{E}}[Y_{D_t^\epsilon} - Y_t] = 0$ by Lemma 8.11, we may find a sequence $(\mathbb{P}_n)_{n\geq 1} \subset \mathcal{P}$ such that

$$-\frac{1}{n} \;\leq\; \mathbb{E}^{\mathbb{P}_n}\big[Y_{D_t^\varepsilon} - Y_t\big] \;\leq\; -\mathbb{E}^{\mathbb{P}_n}\big[K_{D_t^\varepsilon} - K_t\big] \;\leq\; -\underline{\mathcal{E}}\big[K_{D_t^\varepsilon} - K_t\big].$$

Then, it follows from the non decrease of K that $\underline{\mathcal{E}}\big[K_{D_t^\varepsilon} - K_t\big] = 0$, and therefore

$$K_{D_t^\varepsilon} \;=\; K_t, \quad \mathbb{P}_0\text{-a.s.} \quad \text{for all } t \in [0,T].$$

Consequently, $\mathbb{P}_0[\Omega'] = 1$, where $\Omega' := \big\{K_{D_t^\epsilon} = K_t, \text{ for all } t \in [0,T] \cap \mathbb{Q}\big\}$. Further, for any $t \in [0,T)$, let $\{t_n\}_n \subset \mathbb{Q}$ and $t_n \downarrow t$. Since K is nondecreasing, we see that $K_t \leq K_{D_t^\varepsilon} \leq K_{D_{t_n}^\varepsilon} = K_{t_n}$ on Ω'. Since K inherits the RCLL property of Y, this shows that $K_{D_t^\varepsilon}$ is right continuous on Ω', and implies that $\{K_t\}_t$ and $\{K_{D_t^\epsilon}\}_t$ are indistinguishable.

Proof of Proposition 8.10

(i) We first prove that $\int (Y - X)dK = 0$, $\mathbb{P}_0$-a.s. From Lemma 8.12, we have $\mathbb{P}_0[\Lambda] = 1$, where $\Lambda = \{\omega : K_t(\omega) = K_{D_t^\varepsilon}(\omega) \text{ for all } t \in [0,T]\}$. Next, consider the decomposition of the process Y into a continuous and a purely discontinuous part $Y = Y^c + Y^d$. From the decomposition of Proposition 8.9 and the fact that K is increasing, we deduce that $\mathbb{P}_0[\Lambda'] = 1$, where $\Lambda' := \{\omega : \Delta Y_t^d(\omega) \leq 0 \text{ for all } t \in [0,T]\}$.
Now fix any $\omega \in \Lambda \cap \Lambda'$. For any $t_0 \in \{t : X_t(\omega) < Y_t(\omega)\}$, denote $2c := Y_{t_0}(\omega) - X_{t_0}(\omega) > 0$. Since $Y(\omega)$ is RCLL with negative jumps, and $X(\omega)$ is continuous, there exists δ such that for all $t \in (t_0 - \delta, t_0]$ we have $Y_t(\omega) - X_t(\omega) > c$, and

$$t_0 \text{ is an interior point of } (t_0 - \delta, D^c_{t_0-\delta}(\omega)) \subset \{t : X_t(\omega) < Y_t(\omega)\}.$$

Further, it is easy to prove that $\{t : X_t(\omega) < Y_t(\omega)\}$ can be covered by a countable number of open intervals in the form of $(t_n, D_{t_n}^{\epsilon_n}(\omega))$. Finally, we have

$$0 \le \int_0^T 1_{\{t: X_t(\omega) < Y_t(\omega)\}} dK_t(\omega) \le \sum_{n=1}^{\infty} (K_{D^{\varepsilon_n}_{t_n}(\omega)}(\omega) - K_{t_n}(\omega)) = 0.$$

(ii) We next prove that Y and K are continuous. Consider the decomposition $K = K^c + K^d$ into a continuous and a purely discontinuous part, and let us show that $\mathbb{P}_0\big[K_t^d = 0 \text{ for all } t \in [0, T]\big] = 1$.
Since K is previsible and $\Delta K_t^d = K_t - K_{t-}$, ΔK^d is also previsible. In the following we set $\inf \emptyset = \infty$. By Theorem 12.3 in Chapter 6 of [31] (p. 333), we know that $\tau^\delta = \inf\{t \in (0, T] : \Delta K_t^d > \delta\}$ is a previsible stopping time (defined in Definition 12.1 in Chapter 6 of [31]), for all $\delta > 0$. Then, by Theorem 12.6 in Chapter 6 of [31], τ^δ can be announced by a sequence of stopping time τ_n, i.e. $\tau_n < \tau^\delta$ and $\tau_n \uparrow \tau^\delta$, $\mathbb{P}_0$-a.s. Then, since K_t and $K_{D_t^\epsilon}$ are indistinguishable by Lemma 8.12, it follows from the definition of τ^δ that $K^d_{D^\varepsilon_{\tau_n}} = K^d_{\tau_n} < K^d_{\tau^\delta}$. Then, $\tau_n \le D^\epsilon_{\tau_n} < \tau^\delta$, $\mathbb{P}_0$-a.s. Hence

$$\mathbb{P}_0[\Omega_0] = 1, \text{ where } \Omega_0 := \big\{\tau_n \uparrow \tau^\delta \text{ and } \tau_n \le D^\varepsilon_{\tau_n} < \tau^\delta\big\}.$$

For all $\omega \in \Omega_0$, we can find a sequence t_n such that $D^\epsilon_{\tau_n}(\omega) \le t_n < \tau^\delta(\omega)$ and $Y_{t_n}(\omega) \le X_{t_n}(\omega) + \epsilon$. Sending $n \to \infty$, we get $Y_{\tau^\delta(\omega)-}(\omega) \le X_{\tau^\delta(\omega)}(\omega) + \epsilon$. So, $Y_{\tau^\delta -} \le X_{\tau^\delta} + \epsilon$, $\mathbb{P}_0$-a.s. Choosing $\epsilon < \delta$, we see that, whenever $\tau^\delta \le T$, $Y_{\tau^\delta} \le Y_{\tau^\delta -} - \delta < X_{\tau^\delta}$, which is the required contradiction. Hence $\tau^\delta = \infty$ for all $\delta > 0$, implying that $K^d = 0$, $\mathbb{P}_0$-a.s.

(iii) We now show that τ^* is an optimal stopping rule. The results of (i) and (ii) lead to $K_{\tau^*} = 0$ $\mathbb{P}_0$-a.s. Recall the generalized Doob-Meyer decomposition in Proposition 8.9. Take λ^* such that $\|\lambda^*\| \le L$ and $\lambda^* H = L|H|_1$. Then, by taking expectation under $\mathbb{P}_{\lambda^*}$, we obtain that

$$Y_0 \;=\; \mathbb{E}^{\mathbb{P}_{\lambda^*}}\big[Y_{\tau^*}\big] \;=\; \mathbb{E}^{\mathbb{P}_{\lambda^*}}\big[X_{\tau^*}\big].$$

The last equality is due to the definition of τ^*. Finally, it is clear that $Y_0 = \overline{\mathcal{E}}^{\mathcal{P}}[X_{\tau^*}]$. Hence, τ^* is an optimal stopping rule. □

8.4 Reduction to a Standard Optimal Stopping Problem

As a consequence of the decomposition in Proposition 8.9 together with Lemma 8.12, we obtain the following reduction.

Proposition 8.13 *There exists a probability $\mathbb{P}^* \in \mathcal{P}$ such that*

$$Y_t = \operatorname*{ess\,sup}_{\tau \in \mathcal{T}_*^t} \mathbb{E}^{\mathbb{P}^*}[X_\tau | \mathcal{F}_t], \quad \mathbb{P}_0\text{-}a.s.$$

In particular, there exists a $\mathbb{P}^*$-martingale M^* such that $Y = Y_0 + M^* - K$, $\mathbb{P}_0$-a.s.

Proof First, for any $\tau \in \mathcal{T}_*^t$ and $\mathbb{P} \in \mathcal{P}$, we have $Y_t \geq \mathbb{E}^{\mathbb{P}}[X_\tau|\mathcal{F}_t]$, $\mathbb{P}_0$-a.s. Hence, $Y_t \geq \operatorname{ess\,sup}_{\tau \in \mathcal{T}_*^t} \mathbb{E}^{\mathbb{P}}[X_\tau|\mathcal{F}_t]$, $\mathbb{P}_0$-a.s.

On the other hand, let λ^* be defined by its ith entry $L\,\mathrm{sgn}(H_t)_i$. From Proposition 8.9, we know that $(H \cdot B) - \int_0^\cdot L|H_s|_1 ds$ is a $\mathbb{P}_{\lambda^*}$-martingale. Then, it follows from the decomposition of Proposition 8.9, together with Lemma 8.12, that

$$Y_t = \mathbb{E}^{\mathbb{P}_{\lambda^*}}\left[Y_{D_t^\varepsilon} + K_{D_t^\varepsilon} - K_t|\mathcal{F}_t\right] = \mathbb{E}^{\mathbb{P}_{\lambda^*}}\left[Y_{D_t^\epsilon}|\mathcal{F}_t\right] \leq \mathbb{E}^{\mathbb{P}_{\lambda^*}}\left[X_{D_t^\epsilon}|\mathcal{F}_t\right] + \varepsilon.$$

Since $D^\varepsilon \to D_t := \inf\{s \geq t : Y_{t-} = X_t\}$, as $\varepsilon \to 0$, this implies that

$$Y_t \leq \mathbb{E}^{\mathbb{P}_{\lambda^*}}\left[X_{D_t}|\mathcal{F}_t\right] \leq \operatorname*{ess\,sup}_{\tau \in \mathcal{T}_*^t} \mathbb{E}^{\mathbb{P}_{\lambda^*}}[X_\tau|\mathcal{F}_t], \qquad \mathbb{P}_0\text{-a.s.} \qquad \square$$

8.5 The $\mathbb{F}$-adapted Snell Envelop

Given the continuity of Y in Proposition 8.10, we now reduce to an $\mathbb{F}$-adapted version.

Proposition 8.14 *There is an $\mathbb{F}$-adapted pathwise continuous indistinguishable version of Y.*

Proof Define

$$Y_t^{\mathbb{F}} = \mathbb{E}^{\mathbb{P}_0}[Y_t|\mathcal{F}_t] \text{ for } t \in [0,T] \cap \mathbb{Q}, \text{ and } Y_t^{\mathbb{F}} := \lim_{s \uparrow\uparrow t, s \in \mathbb{Q}} \tilde{Y}_s \text{ for } t \in [0,T] \setminus \mathbb{Q}.$$

The last limit exists by the pathwise continuity of Y, $\mathbb{P}_0$-a.s., see Proposition 8.10. Clearly, $Y_t^{\mathbb{F}}$ is $\mathcal{F}_t$-measurable. Since Y is $\mathbb{F}^*$-adapted, we have $\mathbb{P}_0\big[Y = Y^{\mathbb{F}}$ on $[0,T] \cap \mathbb{Q}\big] = 1$, and by the pathwise continuity of Y, we deduce that $\mathbb{P}_0\big[Y = Y^{\mathbb{F}}\big] = 1$. Hence Y and $Y^{\mathbb{F}}$ are indistinguishable. $\square$

In this section, we consider the process Y in its version of Proposition 8.14, which we call the $\mathbb{F}$-adapted Snell envelop of X. We next define:

$$Z_\tau(\omega) := \sup_{\theta \in \mathcal{T}_*} \overline{\mathcal{E}}[X_\theta^{\tau(\omega),\omega}], \text{ for all } \tau \in \mathcal{T}.$$

Clearly, $Y_0 = Z_0$. The main result of this subsection is the following.

Proposition 8.15 *Let Y be the $\mathbb{F}$-Snell envelop of X. Then, $Y_\tau = Z_\tau$, $\mathbb{P}_0$-a.s. for all $\tau \in \mathcal{T}_*$*

In preparation for the proof of this result, we prove two lemmas.

Lemma 8.16 *Let $\hat{Y}$ be a continuous $\mathbb{F}$-adapted process such that, for some $\mathbb{P}_0$-martingale $\hat{M}$ and nondecreasing process $\hat{K}$:*

(i) $\hat{Y} = Y_0 + \hat{M}_t - \max_{|\lambda|\le L}\langle \hat{M}, \int_0^\cdot \lambda_s dB_s\rangle - \hat{K}$, $\mathbb{P}_0$-*a.s.*
(ii) $\hat{Y} \ge X$, $\mathbb{P}_0$-*a.s.*
(iii) $\int_0^T 1_{\{t: X_t < \hat{Y}_t\}} d\hat{K}_t = 0$, $\mathbb{P}_0$-*a.s.*

Then, $\hat{Y}_t = \text{ess sup}_{\tau\in\mathcal{T}_*^t, \mathbb{P}\in\mathcal{P}} \mathbb{E}^{\mathbb{P}}[X_\tau|\mathcal{F}_t]$, $\mathbb{P}_0$-*a.s.*

Proof By martingale representation and the Property (i), there exists $\hat{H} \in \mathbb{H}_{loc}$ such that $\hat{Y} = \hat{Y}_0 + (\hat{H}\cdot B) - L\int_0^\cdot |H_s|_1 ds - \hat{K}$,$\mathbb{P}_0$-a.s. By Girsanov theorem, $\hat{M}^\lambda := \int_0^\cdot \hat{H}_s dB_s - \int_0^\cdot \lambda_s^T \hat{H}_s ds$ is $\mathbb{P}_\lambda$-local martingale, and it follows from the previous decomposition that there exists increasing process $\hat{K}^\lambda$ such that

$$\hat{Y} \;=\; \hat{Y}_0 + \hat{M}^\lambda - \hat{K}^\lambda, \quad \mathbb{P}_0 - \text{a.s.} \tag{8.3}$$

By the uniqueness of the Doob-Meyer decomposition, we deduce that $\hat{M}^\lambda$ is a $\mathbb{P}_\lambda$-martingale, and it follows from (8.3) and Property (ii) that

$$\hat{Y}_t \;\ge\; \mathbb{E}^{\mathbb{P}_\lambda}\big[\hat{Y}_\tau\big|\mathcal{F}_t\big] \;\ge\; \mathbb{E}^{\mathbb{P}_\lambda}\big[X_\tau\big|\mathcal{F}_t\big] \text{ for all } \tau\in\mathcal{T}_*^t,\; \mathbb{P}_\lambda\in\mathcal{P}.$$

Hence, $\hat{Y}_t \ge \text{ess sup}_{\tau\in\mathcal{T}_*^t,\mathbb{P}\in\mathcal{P}} \mathbb{E}^{\mathbb{P}}[X_\tau|\mathcal{F}_t]$. For the reverse inequality, consider the stopping time $D_t := \inf\{s\ge t : \hat{Y}_s = X_s\} \in \mathcal{T}_*^t$. Let λ^* be the process defined by its ith entry L sgn$(\hat{H}_i)$. Note that $\hat{K}^{\lambda^*} = \hat{K}$ in (8.3), and therefore

$$\hat{Y}_t = \mathbb{E}^{\mathbb{P}_{\lambda^*}}\big[\hat{Y}_{D_t} + \hat{K}_{D_t} - \hat{K}_t\big|\mathcal{F}_t\big].$$

By property (iii) and the definition of D_t, it follows that $\hat{K}_{D_t} = \hat{K}_t$, $\mathbb{P}_0$-a.s., so that

$$\hat{Y}_t \;=\; \mathbb{E}^{\mathbb{P}_{\lambda^*}}\big[\hat{Y}_{D_t}\big|\mathcal{F}_t\big] \;=\; \mathbb{E}^{\mathbb{P}_{\lambda^*}}\big[X_{D_t}\big|\mathcal{F}_t\big] \;\le\; \operatorname*{ess\,sup}_{\tau\in\mathcal{T}_*^t,\mathbb{P}\in\mathcal{P}} \mathbb{E}^{\mathbb{P}}[X_\tau|\mathcal{F}_t].$$ □

Lemma 8.17 *Let M be a pathwise continuous $\mathbb{P}_0$-martingale with $\mathbb{E}^{\mathbb{P}_0}\big[\sup_{t\in[0,T]} |M_t|\big] < \infty$. Then, there exists an $\mathbb{F}$-adapted indistinguishable version $\tilde{M}$ such that:*

$$\mathbb{P}_0\big[\omega : \tilde{M}^{\tau,\omega} \textit{ is a } \mathbb{P}_0 - \textit{martingale}\big] = 1 \textit{ for all } \tau\in\mathcal{T}.$$

Proof 1. Let $\tilde{M}_T := M_T$, and for all $\omega\in\Omega$:

$$\tilde{M}_s(\omega) := \mathbb{E}^{\mathbb{P}_0}\big[M_T^{s,\omega}\big] \text{ for } s\in\mathbb{Q} \text{ and } \tilde{M}_t(\omega) := \limsup_{s\uparrow t, s\in\mathbb{Q}} \tilde{M}_s(\omega) \text{ for } t\in[0,T]\setminus\mathbb{Q}.$$

Clearly, $\tilde{M}$ is $\mathbb{F}$-adapted, and $\mathbb{P}_0\big[M = \tilde{M} \text{ on } \mathbb{Q}\big] = 1$. Since M is continuous, it is easy to verify that $\mathbb{P}_0[M = \tilde{M}] = 1$, i.e. $\tilde{M}$ is an indistinguishable version of M.

2. Denote $|\tilde{M}|^*_t := \sup_{s\le t} |\tilde{M}_s|$, and

$$I_\tau := \Big\{\omega \in \Omega : \tilde{M}_\tau(\omega) = \mathbb{E}^{\mathbb{P}_0}\big[\tilde{M}^{\tau,\omega}_T\big],\ \mathbb{E}^{\mathbb{P}_0}\big[|\tilde{M}|^{*^{\tau,\omega}}_t\big] < \infty, \\ \text{and } \mathbb{P}_0\{\tilde{M}^{\tau,\omega} \text{ is continuous}\} = 1\Big\}.$$

Since $\tilde{M}$ and M are indistinguishable, $\tilde{M}$ is a $\mathbb{P}_0$-martingale and $\mathbb{P}_0[I_\tau] = 1$. For $\eta \in \mathcal{T}$ with $\eta \ge \tau$, we define a sequence of stopping times $\eta_n := \frac{[2^n\eta]+1}{2^n}$. Note that η_n only take rational values. By the tower property and the definition of $\tilde{M}_s$ for $s \in \mathbb{Q}$, we obtain for $\omega \in I_\tau$:

$$\tilde{M}_\tau(\omega) = \mathbb{E}^{\mathbb{P}_0}[\tilde{M}^{\tau,\omega}_T] = \lim_{n\to\infty} \int_\Omega \mathbb{E}^{\mathbb{P}_0}[\tilde{M}^{\eta_n,\omega\otimes_\tau\omega'}_T]\mathbb{P}_0(d\omega') = \lim_{n\to\infty} \mathbb{E}^{\mathbb{P}_0}[\tilde{M}^{\tau,\omega}_{\eta_n}].$$

Since $\mathbb{E}^{\mathbb{P}_0}\big[|\tilde{M}|^{*^{\tau,\omega}}_t\big] < \infty$, it follows that the family $\{\tilde{M}^{\tau,\omega}_{\eta_n}\}_{n\in\mathbb{N}}$ is $\mathbb{P}_0$-uniformly integrable. Then, it follows from the $\mathbb{P}_0$-a.s. pathwise continuity of $\tilde{M}^{\tau,\omega}$ that $\tilde{M}_\tau(\omega) = \lim_{n\to\infty} \mathbb{E}^{\mathbb{P}_0}[\tilde{M}^{\tau,\omega}_{\eta_n}] = \mathbb{E}^{\mathbb{P}_0}[\tilde{M}^{\tau,\omega}_\eta]$. By the arbitrariness of $\eta \in \mathcal{T}$, this proves that $\tilde{M}^{\tau,\omega}$ is a $\mathbb{P}_0$-martingale. □

Proof of Proposition 8.15 Notice that $Y \ge X$, $\mathbb{P}_0$-a.s., and by Propositions 8.9 and 8.10, there exists $H \in \mathbb{H}_{loc}$ and nondecreasing previsible process K such that, with $M := (H \cdot B)$:

$$Y = Y_0 + M - \max_{|\lambda|\le L} \langle M, \int_0^\cdot \lambda_s dB_s\rangle - K \quad \text{and} \quad \int_0^T 1_{\{t:X_t<Y_t\}} dK_t = 0, \qquad \mathbb{P}_0 - \text{a.s.}$$

The process M is a pathwise continuous $\mathbb{P}_0$-martingale with $\mathbb{E}^{\mathbb{P}_0}\big[\sup_{t\in[0,T]} |M_t|\big] < \infty$, by Proposition 8.9. By Lemma 8.17, we may consider M as the version for which $M^{\tau,\omega}$ is a $\mathbb{P}_0$-martingale, for $\mathbb{P}_0$-a.e. $\omega \in \Omega$.

Let $T' := T - \tau(\omega)$, and define $\tilde{M}^{\tau,\omega}_t(\omega') := M^{\tau,\omega}_t(\omega') - M_\tau(\omega)$ for $t \in [0, T']$. Then, $\tilde{M}^{\tau,\omega}$ is $\mathbb{P}_0$-martingale for $\mathbb{P}_0$-a.e. $\omega \in \Omega$. We now observe that $(Y^{\tau,\omega}, \tilde{M}^{\tau,\omega}, K^{\tau,\omega})$ satisfies the following properties for $\mathbb{P}_0$-a.e. $\omega \in \Omega$:

(i) $Y^{\tau,\omega} = Y_{\tau(\omega)}(\omega) + \tilde{M}^{\tau,\omega} - \max_{|\lambda|\le L} \langle \tilde{M}^{\tau,\omega}, \int_0^\cdot \lambda_s dB_s\rangle - K^{\tau,\omega}$, on $[0, T']$, $\mathbb{P}_0$-a.s.
(ii) $Y^{\tau,\omega} \ge X^{\tau,\omega}$ on $[0, T']$, $\mathbb{P}_0$-a.s.
(iii) $\int_0^{T'} 1_{\{t:X^{\tau,\omega}_t<Y^{\tau,\omega}_t\}} dK^{\tau,\omega}_t = 0$, $\mathbb{P}_0$-a.s.

Then, it follows from Lemma 8.16 that $Y_\tau = Z_\tau$, $\mathbb{P}_0$-a.s. □

9 Appendix: On $\overline{\mathcal{E}}$-submartingales

Definition 9.1 An $\mathbb{F}$-progressively measurable process m is an $\overline{\mathcal{E}}$-submartingale (resp. $\underline{\mathcal{E}}$-supermartingale), if, for any $(t, \omega) \in \Theta$, we have

$$u_t(\omega) \le \overline{\mathcal{E}}[u_Z^{t,\omega}] \text{ (resp. } \ge \underline{\mathcal{E}}[u_Z^{t,\omega}]) \quad \text{for all } \tau \in \mathcal{T}.$$

The main result of this section is the following.

Proposition 9.2 *Let $u \in C^0_{2,\mathcal{P}}(\Theta, \mathbb{R})$ be a $\overline{\mathcal{E}}$-submartingale. Then, there exists $\mathbb{P}^* \in \mathcal{P}$ such that u is a $\mathbb{P}^*$-submartingale.*

Proof **1.** We shall prove in Step 2 below that

$$\overline{\mathcal{E}}\big[u_{s-t}^{t,\omega}\big] = \operatorname*{ess\,sup}_{\mathbb{P}\in\mathcal{P}} \mathbb{E}^{\mathbb{P}}[u_s|\mathcal{F}_t] \quad \text{for } \mathbb{P}_0-\text{a.e. } \omega \in \Omega, \quad \text{for all } t < s \ge T-t. \tag{9.1}$$

Let $t_k^n := kT2^{-n}$, $k \ge 0$, and $\mathbb{I}_n := \{T \wedge t_k^n : k \ge 0\}$. Since $\mathcal{P}$ is weakly compact and $u \in C^0_{2,\mathcal{P}}(\Theta, \mathbb{R})$, we deduce from (9.1) that, for all pair (n, k) with $t_k^n \le T$, there exists $\mathbb{P}^{n,k} \in \mathcal{P}$ such that $u_{t_k^n} \le \mathbb{E}^{\mathbb{P}^{n,k}}[u_{t_{k+1}^n}|\mathcal{F}_{t_k^n}]$, $\mathbb{P}^0$-a.s. Defining $\mathbb{P}^n := \mathbb{P}^{n,0} \otimes_{t_1^n} \mathbb{P}^{n,1} \otimes_{t_2^n}, \cdots$, this implies that

$$u_{t_i^n \wedge T} \le \mathbb{E}^{\mathbb{P}^n}[u_{t_j^n \wedge T}|\mathcal{F}_{t_i^n \wedge T}] \quad \mathbb{P}_0-\text{a.s.} \quad \text{for all } 0 \le i \le j \le n.$$

Since $\mathcal{P}$ is weakly compact, $\mathbb{P}^n$ converges weakly to some $\mathbb{P}^* \in \mathcal{P}$, after possibly passing to a subsequence. Observe that, for $m \ge n$, we have $\mathbb{E}^{\mathbb{P}^n}\big[u_{t_j^n \wedge T}|\mathcal{F}_{t_i^n \wedge T}\big] = \mathbb{E}^{\mathbb{P}^m}\big[u_{t_j^n \wedge T}|\mathcal{F}_{t_i^n \wedge T}\big] \longrightarrow \mathbb{E}^{\mathbb{P}^*}\big[u_{t_j^n \wedge T}|\mathcal{F}_{t_i^n \wedge T}\big]$, as $m \to \infty$, by the fact that $u \in C^0_{2,\mathcal{P}}(\Theta, \mathbb{R})$. Hence, $u_t \le \mathbb{E}^{\mathbb{P}^*}[u_s|\mathcal{F}_t]$, $\mathbb{P}_0$-a.s. for all $t \le s \le T-t$ with $s, t \in \mathbb{I}_n$. By the density of $\mathbb{I}_n$ in $[0, T]$, we further conclude that u is a $\mathbb{P}^*$-submartingale.
2. It remains to prove (9.1). For $t \le s$, define a process:

$$v_t^s := \operatorname*{ess\,sup}_{\mathbb{P}\in\mathcal{P}} \mathbb{E}^{\mathbb{P}}[u_s|\mathcal{F}_t] \quad \text{for } t \in [0, T], \quad 0 \le s \le T-t.$$

Similar to Lemma 8.5, we may check that the family $\{\mathbb{E}^{\mathbb{P}}[u_s|\mathcal{F}_t]; \mathbb{P} \in \mathcal{P}\}$ satisfies the lattice property. Then, for $t_1 \le t_2 \le s$, we have for all $\mathbb{P} \in \mathcal{P}$:

$$\mathbb{E}^{\mathbb{P}}[v_{t_2}^s|\mathcal{F}_{t_1}] = \operatorname*{ess\,sup}_{\mathbb{P}'\in\mathcal{P}} \mathbb{E}^{\mathbb{P}\otimes_{t_2}\mathbb{P}'}[u_s|\mathcal{F}_{t_1}] \le \operatorname*{ess\,sup}_{\mathbb{P}'\in\mathcal{P}} \mathbb{E}^{\mathbb{P}\otimes_{t_1}\mathbb{P}'}[u_s|\mathcal{F}_{t_1}] = v_{t_1}^s.$$

proving that v^s is $\mathbb{P}$-supermartingale on $[0, s]$ for all $\mathbb{P} \in \mathcal{P}$. Similar to Lemma 8.8, we may consider v^s in its $\mathbb{F}^*$-adapted RCLL version.

Following the line of argument in the proof of Proposition 8.9, there exists $H^s \in \mathbb{H}_{loc}$ and increasing process K^s such that

$$v^s = v_0^s + (H^s \cdot B) - L \int_0^{\cdot} |H_r^s|_1 dr - K^s, \quad \mathbb{P}^0 - \text{a.s.}$$

We next prove that $K^s \equiv 0$, $\mathbb{P}^0$-a.s. Indeed, assuming to the contrary that $\mathbb{P}^0[K_s^s > 0] > 0$, it follows that $\underline{\mathcal{E}}[K_s^s] > 0$. Following the line of argument in Lemma 8.6, it can be checked that $\overline{\mathcal{E}}[v_t^s] = \overline{\mathcal{E}}[u_s]$ for all $t \leq s$. Then, since $v_s^s = u_s$, it follows from the previous decomposition that

$$\mathbb{E}^{\mathbb{P}}[v_0^s] \;\geq\; \mathbb{E}^{\mathbb{P}}[u_s + K_s^s] \;\geq\; \mathbb{E}^{\mathbb{P}}[u_s] + \underline{\mathcal{E}}[K_s^s] \quad \text{for all } \mathbb{P} \in \mathcal{P},$$

and therefore $\overline{\mathcal{E}}[v_t^s] > \overline{\mathcal{E}}[u_s]$, which is the required contradiction. This reduces the decomposition of v^s to:

$$v^s = v_0^s + (H^s \cdot B) - L \int_0^{\cdot} |H_r^s|_1 dr, \quad \mathbb{P}^0 - \text{a.s.}$$

so that, with λ^s the process with ith entry L sgn(H_i), we obtain $v_t^s = \mathbb{E}^{\mathbb{P}^{\lambda^s}}[u_s|\mathcal{F}_t]$. We finally prove that (9.1) holds true by following the line of argument in the proof of Proposition 8.15. □

References

1. G. Barles, P.E. Souganidis, Convergence of approximation schemes for fully nonlinear second order equation. Asymptot. Anal. **4**, 271–283 (1991)
2. B. Bouchard, N. Touzi, Weak dynamic programming principle for viscosity solutions. SIAM J. Control Optim. **49**(3), 948–962 (2011)
3. G. Barles, B. Perthame, Discontinuous solutions of deterministic optimal stopping time problems. Math. Model. Numer. Anal. **21**, 557–579 (1987)
4. L. Caffarelli, X. Cabre, *Fully nonlinear elliptic equations*, vol. 43 (American Mathematical Society, Rhode Island, 1995)
5. P. Cheridito, H.M. Soner, N. Touzi, N. Victoir, Second order BSDE's and fully nonlinear PDE's. Commun. Pure Appl. Math. **60**, 1081–1110 (2007)
6. R. Cont, D. Fournie, Functional Itô calculus and stochastic integral representation of martingales. Ann. Probab. **41**, 109–133 (2013)
7. M.G. Crandall, P.-L. Lions, Viscosity solutions of Hamilton-Jacobi equations. Trans. AMS **277**, 42 (1984)
8. M.G. Crandall, P.-L. Lions, Lions, Condition dunicité pour les solutions généralisées dee équations de Hamilton-Jacobi du premier order. C. R. Acad. Sci. Paris **292**, 183–186 (1981)
9. M.G. Crandall, H. Ishii, P.L. Lions, User's guide to viscosity solutions of second order partial differential equations. Bull. Am. Math. Soc. (NS) **27**, 1–67 (1992)
10. B. Dupire, Functional Itô calculus, papers.ssrn.com (2009)
11. I. Ekren, C. Keller, N. Touzi, J. Zhang, On viscosity solutions of path dependent PDEs. Ann. Probab. **42**, 204–236 (2014)
12. I. Ekren, N. Touzi, J. Zhang, Optimal stopping under nonlinear expectation, preprint, arXiv:1209.6601

13. I. Ekren, N. Touzi, J. Zhang, Viscosity Solutions of fully nonlinear parabolic path dependent PDEs: Part I, preprint, arXiv:1210.0006
14. I. Ekren, N. Touzi, J. Zhang, Viscosity solutions of fully nonlinear parabolic path dependent PDEs: Part II, preprint, arXiv:1210.0007
15. N. El Karoui, Les aspects probabilistes du contrôle stochastique. *Ecole dEt de Probabilits de Saint-Flour IX-1979*. Lecture Notes in Mathematics, vol. 876 (1981), pp. 73-238
16. N. El Karoui, C. Kapoudjian, E. Pardoux, S. Peng, M.C. Quenez, Reflected solutions of backward SDEs, and related obstacle problems for PDEs. Ann. Probab. **25**(2), 702–737 (1997)
17. N. El Karoui, S. Peng, M.-C. Quenez, Backward stochastic differential equations in finance. Math. Financ. **7**(1), 1–71 (1997)
18. N. El Karoui, X. Tan, Capacities, measurable selection and dynamic programming Part I: abstract framework, preprint, arXiv:1310.3363
19. N. El Karoui, X. Tan, Capacities, measurable selection and dynamic programming Part II: application in stochastic control problems, preprint, arXiv:1310.3364
20. W. Fleming, H.M. Soner, *Controlled Markov Processes and Viscosity Solutions*, 2nd edn. (Springer, New York, 2006)
21. S. Hamadène, Reflected BSDEs with discontinuous barrier and application. Stoch. Stoch. Rep. **74**(34), 571–596 (2002)
22. I. Karatzas, S. Shreve, *Methods of Mathematical Finance* (Springer, New York, 1998)
23. R. Karandikar, On pathwise stochastic integration. Stoch. Process. Appl. **57**, 11–18 (1995)
24. E. Pardoux, S. Peng, Adapted solutions of backward stochastic differential equations. Syst. Control Lett. **14**, 55–61 (1990)
25. S. Peng, G-Brownian motion and dynamic risk measure under volatility uncertainty, preprint, arXiv:0711.2834
26. S. Peng, Backward stochastic differential equation, nonlinear expectation and their applications, in *Proceedings of the International Congress of Mathematicians, Hyderabad* (2010)
27. S. Peng, M. Xu, The smallest g-supermartingale and reflected BSDE with single and double L2 obstacles. Ann. I. H. Poincaré **41**, 605–630 (2005)
28. Z. Ren, N. Touzi, J. Zhang, Comparison of viscosity solutions of semilinear path-dependent partial differential equations, Preprint
29. H.M. Soner, N. Touzi, J. Zhang, Well-posedness of second order backward SDEs. Probab. Theory Relat. Fields **153**, 149–190 (2012)
30. L. Wang, On the regularity of fully nonlinear parabolic equations: II. Commun. Pure Appl. Math. **45**, 141–178 (1992)
31. D. Williams, L.C.G. Rogers, *Diffusions, Markov Processes, and Martingales: Volume 2* (Cambridge University Press, Cambridge, 1983)
32. J. Zhang, J. Zhuo, Monotone schemes for fully nonlinear parabolic path dependent PDEs. J. Financ. Eng. 1, 1450005, p. 23 (2014) doi:10.1142/S2345768614500056

Logarithmic Asymptotics of the Densities of SPDEs Driven by Spatially Correlated Noise

Marta Sanz-Solé and André Süß

Abstract We consider the family of stochastic partial differential equations indexed by a parameter $\varepsilon \in (0, 1]$,

$$Lu^{\varepsilon}(t, x) = \varepsilon\sigma(u^{\varepsilon}(t, x))\dot{F}(t, x) + b(u^{\varepsilon}(t, x)),$$

$(t, x) \in (0, T] \times \mathbb{R}^d$ with suitable initial conditions. In this equation, L is a second-order partial differential operator with constant coefficients, σ and b are smooth functions and $\dot{F}$ is a Gaussian noise, white in time and with a stationary correlation in space. Let $p^{\varepsilon}_{t,x}$ denote the density of the law of $u^{\varepsilon}(t, x)$ at a fixed point $(t, x) \in (0, T] \times \mathbb{R}^d$. We study the existence of $\lim_{\varepsilon \downarrow 0} \varepsilon^2 \log p^{\varepsilon}_{t,x}(y)$ for a fixed $y \in \mathbb{R}$. The results apply to classes of stochastic wave equations with $d \in \{1, 2, 3\}$ and stochastic heat equations with $d \geq 1$.

Keywords Logarithmic estimates of densities (Varadhan estimates) · Stochastic partial differential equations · Stochastic wave equation · Stochastic heat equation · Malliavin calculus · Large deviation principle · Topological support

2010 Mathematics Subject Classification: Primary: 60H07 · 60H15; Secondary: 60Fxx · 60F10

M. Sanz-Solé—Supported by the grant MTM 2012-31192 from the Dirección General de Investigación, Ministerio de Economía y Competitividad, Spain.
A. Süß—Supported by the grant MTM 2012-31192 from the Dirección General de Investigación, Ministerio de Economía y Competitividad, Spain.

M. Sanz-Solé (✉) · A. Süß
Facultat de Matemàtiques, Universitat de Barcelona, Gran Via de les Corts Catalanes 585, 08007 Barcelona, Spain
e-mail: marta.sanz@ub.edu

D. Crisan et al. (eds.), *Stochastic Analysis and Applications 2014*,
Springer Proceedings in Mathematics & Statistics 100,
DOI 10.1007/978-3-319-11292-3_16

1 Introduction

In this article, we consider the family of stochastic partial differential equations (SPDEs) indexed by a parameter $\varepsilon \in (0, 1]$ defined by

$$Lu^{\varepsilon}(t, x) = \varepsilon\sigma(u^{\varepsilon}(t, x))\dot{F}(t, x) + b(u^{\varepsilon}(t, x)), \tag{1.1}$$

$(t, x) \in (0, T] \times \mathbb{R}^d$, $d \geq 1$, with suitable initial conditions. Here L is a second-order partial differential operator, typical examples are the wave and the heat operators; $\sigma, b : \mathbb{R} \to \mathbb{R}^d$ are smooth functions; $\dot{F}$ is a Gaussian noise, white in time and with a stationary correlation in space.

Equation (1.1) describes a nonhomogeneous initial value problem subject to nonlinear small random fluctuations. The results of this paper are a contribution to the study of the behavior of (1.1) as $\varepsilon \downarrow 0$ and therefore, when the random perturbations disappear. More precisely, denote by $p^{\varepsilon}_{t,x}$ the density of the random variable $u^{\varepsilon}(t, x)$ at a given point $(t, x) \in (0, T] \times \mathbb{R}^d$. We will determine the set of points $y \in \mathbb{R}$ for which one can derive upper and lower bounds for $\lim_{\varepsilon\downarrow 0} \varepsilon^2 \log p^{\varepsilon}_{t,x}(y)$. We will identify these bounds and refer to them as *Varadhan estimates* or *logarithmic estimates*. We will consider examples of stochastic wave equations with $d = \{1, 2, 3\}$ and stochastic heat equations with $d \geq 1$.

For solutions to stochastic differential equations driven by a standard Brownian motion, $\{X_t, t \geq 0\}$, this question is equivalent to the analysis of the density of X_t, when $t \downarrow 0$. Under ellipticity conditions and with analytical methods, it has been firstly studied in [34, 35]. Using Malliavin calculus and large deviation estimates, Varadhan's results have been extended in [15, 16] under hypoelliptic assumptions.

The method of [15, 16] has been applied in [17] to establish Varadhan estimates for an example of hyperbolic SPDE: an Itô equation with two-dimensional parameter. In [24, Propositions 4.4.1 and 4.4.2], a general formulation of that method is given, providing a systematic approach to the study of Varadhan estimates for families of Wiener functionals subject to small perturbations of their sample paths. For example, it has been used in [14] to extend the results of [17], and in [21] for a stochastic heat equation with boundary conditions.

Similarly as in [18], the aim of this paper is to study Varadhan estimates for the class of SPDEs defined by (1.1). However, in comparison with this reference, there are two additional substantial contributions in our results. Firstly, the scope of application of the theory presented in this article is larger. Indeed, we are able to deal with cases where the fundamental solution corresponding to the operator L is a measure, like for example, the stochastic wave equation in spatial dimension $d = 3$. Secondly, in [18, Theorem 1.2] it is not clear for what values of $y \in \mathbb{R}$ the claim $\lim_{\varepsilon\downarrow 0} \varepsilon^2 \log p^{\varepsilon}_{t,x}(y) = -I(y)$, where I is the rate function, holds. This statement requires $p^{\varepsilon}_{t,x}(y) > 0$ for ε small enough, but this problem is not discussed in [18]. Also in [18, Proposition 5.1], it is assumed that the interior of the topological support of the law of $u^{\varepsilon}(t, x)$ is described in a way that we do not see justified. In this paper, these issues are rigorously addressed.

We now describe the contents of this article. In Sect. 2, we formulate the basic assumptions used throughout the paper, we give a rigorous formulation of (1.1) and quote two fundamental results concerning the existence of a unique *random field solution* to (1.1), and on the *existence and smoothness* of the density $p^{\varepsilon}_{t,x}$ (see Theorems 2.2, 2.3, respectively). In Theorem 2.4 we state the main result of the paper on the logarithmic estimates.

Section 3 is devoted to the proof of Theorem 2.4. To obtain the upper bound, we check that $u^{\varepsilon}(t,x)$ is Malliavin differentiable of any order, and that the corresponding Malliavin-Watanabe norm is uniformly bounded in (t,x) and ε. We also prove a quantitative result on the dependence on ε of the L^p norm of the inverse of the Malliavin matrix corresponding to $u^{\varepsilon}(t,x)$. Notice that in Theorem 2.4, the upper bound still makes sense if $\lim_{\varepsilon\downarrow 0}\varepsilon^2\log p^{\varepsilon}_{t,x}(y) = -\infty$.

To establish the lower bound, we prove that the mapping $\varepsilon \mapsto u(t,x;\omega+\varepsilon^{-1}h)$, where h is an admissible shift for the space of paths Ω, is differentiable in the $\mathbb{D}^{\infty}$ topology of Malliavin calculus, and that the mapping given in (2.6) is onto. Then, in order to give full meaning to the lower bound (2.4), it is relevant to know for which set of $y \in \mathbb{R}$, $p^{\varepsilon}_{t,x}(y)$ is strictly positive for ε small enough, and whether the function I is finite. In the analysis of these questions, the characterization of the topological support of the law of the random variables $u^{\varepsilon}(t,x)$, $\varepsilon \in (0,1]$ plays a crucial role. Each one of these random variables are a nonlinear functional Φ (not depending on ε) of the driving Gaussian noise εF. Hence, one should expect the support to be independent of ε. We postpone the proof of a characterization of the support of $u^{\varepsilon}(t,x)$, which in particular shows its independence of ε, to Sect. 4.

The regularity (in the Malliavin sense) of $u^{\varepsilon}(t,x)$ established in Lemma 3.3, combined with [24, Propositions 4.1.1 and 4.1.2] imply that the support of $u^{\varepsilon}(t,x)$ is a nonempty closed interval and that $p^{\varepsilon}_{t,x}(y) > 0$ for all y in the interior of that set. We also prove in Proposition 3.9 that, in these points, $I(y) < \infty$, and also that if the function b is bounded then $\{y \in \mathbb{R} : I(y) < \infty\} = \mathbb{R}$ (see Proposition 3.10).

Section 4 is devoted to the characterization of the topological support of the law of $u^{\varepsilon}(t,x)$ (see Theorem 4.1). The relevant reference is [10], where a characterization of the support of the law of a stochastic wave equation in spatial dimension $d = 3$ with vanishing initial conditions in Hölder norm is established. In comparison with that work, here the SPDE is more general but, instead of considering the sample paths of the solution to (2.1), we take its value at a fixed point (t,x). This makes the analysis significantly easier.

In Sect. 5, we give two examples where the main result is applied: a class of stochastic wave equations with $d \in \{1,2,3\}$ and a class of stochastic heat equation with $d \geq 1$. For the former, owing to results on large deviations, we have $I = J$ and therefore the equality between the upper and lower bounds.

Throughout the paper, we have to deal with different types of evolution equations, including some classes of Hilbert space-valued equations. To provide the suitable background, we prove in the Appendix a result on the existence and uniqueness of random field solution for a very general class of SPDEs.

2 Preliminaries and Statement of the Main Result

Let $\mathcal{C}_0^\infty(\mathbb{R}_+ \times \mathbb{R}^d)$ denote the space of infinitely differentiable functions defined on $\mathbb{R}_+ \times \mathbb{R}^d$ with compact support. On a given probability space $(\Omega, \mathscr{F}, \mathbb{P})$, we consider a Gaussian stochastic process $F = (F(\phi);\ \phi \in \mathcal{C}_0^\infty(\mathbb{R}_+ \times \mathbb{R}^d))$ with mean zero and covariance functional

$$J(\phi, \psi) := \mathbb{E}[F(\phi)F(\psi)] = \int_0^\infty \int_{\mathbb{R}^d} \big(\phi(t) \star \tilde{\psi}(t)\big)(x)\Gamma(\mathrm{d}x)\mathrm{d}t,$$

where $\tilde{\psi}(t, x) := \psi(t, -x)$, the symbol "$\star$" denotes the convolution operator on $\mathbb{R}^d$, and Γ is a nonnegative, nonnegative definite, tempered measure on $\mathbb{R}^d$. We know by [32, Chap. 7, Théorème XVIII] that there exists a nonnegative tempered measure μ on $\mathbb{R}^d$ such that $\mathcal{F}\mu = \Gamma$, where $\mathcal{F}$ denotes the Fourier transform operator given by

$$\mathcal{F}\phi(\xi) = \int_{\mathbb{R}^d} \phi(x)\mathrm{e}^{-2\pi\mathrm{i}\langle\xi,x\rangle}\mathrm{d}x.$$

Following [4], the process F can be extended to a worthy martingale measure $M = (M_t(A);\ t \in \mathbb{R}_+, A \in \mathcal{B}_b(\mathbb{R}^d))$ where $\mathcal{B}_b(\mathbb{R}^d)$ denotes the bounded Borel subsets of $\mathbb{R}^d$. This is achieved by approximating indicator functions 1_A, $A \in \mathcal{B}_b(\mathbb{R}_+ \times \mathbb{R}^d)$ by functions in $\mathcal{C}_0^\infty(\mathbb{R}_+ \times \mathbb{R}^d)$, and thus extending the functional $\phi \mapsto F(\phi)$ to an $L^2(\Omega)$-valued measure $A \mapsto F(1_A)$. Then we define

$$M_t(A) := F(1_{[0,t]\times A}),$$

for all $t \in \mathbb{R}_+$ and $A \in \mathcal{B}_b(\mathbb{R}^d)$.

Throughout this article we use the filtration

$$\mathscr{F}_t := \sigma\Big(M_s(A);\ s \in [0, t], A \in \mathcal{B}_b(\mathbb{R}^d)\Big) \vee \mathscr{N},$$

$t \in \mathbb{R}_+$, where $\mathscr{N}$ is the σ-field generated by the $\mathbb{P}$-null sets.

The SPDE (1.1) is expressed in the *mild formulation*, as follows,

$$\begin{aligned} u^\varepsilon(t, x) = w(t, x) &+ \varepsilon \int_0^t \int_{\mathbb{R}^d} \Lambda(t - s, x - z)\sigma(u^\varepsilon(s, z))M(\mathrm{d}s, \mathrm{d}z) \\ &+ \int_0^t \int_{\mathbb{R}^d} \Lambda(t - s, x - z)b(u^\varepsilon(s, z))\mathrm{d}z\mathrm{d}s, \end{aligned} \tag{2.1}$$

$(t,x) \in (0,T] \times \mathbb{R}^d$, where Λ denotes the fundamental solution to the associated PDE, $Lu = 0$, and w is the contribution of the initial conditions. For $\varepsilon = 1$, we will write $u(t,x)$ instead of $u^1(t,x)$.

We will consider the following assumptions:

(A1) The mapping $t \mapsto \Lambda(t)$ is a deterministic function with values in the space of non-negative tempered distributions with rapid decrease such that

$$\int_0^T \int_{\mathbb{R}^d} \big(\Lambda(s) \star \tilde{\Lambda}(s)\big)(x)\Gamma(\mathrm{d}x)\mathrm{d}s = \int_0^T \int_{\mathbb{R}^d} |\mathcal{F}\Lambda(s)(\xi)|^2 \mu(\mathrm{d}\xi)\mathrm{d}s < \infty.$$

Moreover, for all $t \in (0,T]$, $\Lambda(t)$ is a nonnegative measure, and there exists $\delta > 0$ such that

$$\int_0^t \Lambda(s)(\mathbb{R}^d)ds \le Ct^{\delta}. \tag{2.2}$$

(A2) The mapping $(t,x) \mapsto w(t,x)$ is deterministic, continuous and

$$\sup_{(t,x)\in[0,T]\times\mathbb{R}^d} |w(t,x)| < \infty.$$

Remark 2.1 Later on, we will refer to [5] and also to [7] for results on the stochastic integral in (2.1), and on the existence and uniqueness of solution. These are proved assuming that $\sup_{t\in[0,T]} \Lambda(t)(\mathbb{R}^d) < \infty$. It can be easily checked that they also hold assuming (2.2).

Throughout the paper the following notation will be used. Let Λ be as in hypothesis (A1). For any $s \in [0,T]$, set

$$J_1(s) := \int_{\mathbb{R}^d} \big(\Lambda(s) \star \tilde{\Lambda}(s)\big)(z)\Gamma(\mathrm{d}z) = \int_{\mathbb{R}^d} |\mathcal{F}\Lambda(s)(\xi)|^2 \mu(\mathrm{d}\xi),$$

$$J_2(s) := \Lambda(s)(\mathbb{R}^d),$$

$$g_1(t) := \int_0^t J_1(s)\mathrm{d}s.$$

Notice that (A1) implies $g_1(T) < \infty$.

In (2.1), the last integral denotes the convolution $\int_0^t (\Lambda(t-s) \star b(u(s,\cdot))(x)\mathrm{d}s$ defined pathwise. As for the stochastic integral (also termed *stochastic convolution*), we refer to the construction given in [5] (see [7, Section 2.3] for a summary).

Let $\mathcal{S}(\mathbb{R}^d)$ be the set of Schwartz functions and denote by $\mathcal{H}$ the Hilbert space obtained by completion of the set $\mathcal{S}(\mathbb{R}^d)$ with the inner product

$$\langle \phi, \psi \rangle = \int_{\mathbb{R}^d} (\phi \star \psi)(x)\Gamma(\mathrm{d}x) = \int_{\mathbb{R}^d} \mathcal{F}\phi(\xi)\overline{\mathcal{F}\psi(\xi)}\mu(\mathrm{d}\xi).$$

Set $\mathcal{H}_T := L^2([0, T]; \mathcal{H})$. The Gaussian process F can be extended to an isonormal process $F = (F(\phi);\ \phi \in \mathcal{H}_T)$ in the sense of [25, Definition 1.1.1].

It is useful to identify the isonormal process F with a $\mathcal{H}$-valued cylindrical Wiener process. As shown in [4], by an approximation procedure we define $W_t(\phi) = F(1_{[0,t]}\phi)$, $t \in [0, T]$, $\phi \in \mathcal{H}$. Consider a complete orthonormal system (CONS) of $\mathcal{H}$ denoted by $(e_k)_{k\in\mathbb{N}}$. Then,

$$W = \{W^k(t) := W_t(e_k), t \in [0, T], k \in \mathbb{N}\}$$

defines a sequence of independent standard Brownian motions. Conversely, the process $(F(\phi) = \sum_{k\in\mathbb{N}} \int_0^T \langle \phi(t), e_k\rangle_{\mathcal{H}} \mathrm{d}W^k(t), \phi \in \mathcal{H}_T)$ is an isonormal Gaussian process.

As has been established in [7], there is an equivalence between the stochastic integral in the sense of [5] and the stochastic integral with respect to the cylindrical Wiener process W (see e.g. [9]). In particular, the stochastic integral in (2.1) is equal to

$$\sum_{k\in\mathbb{N}} \int_0^t \langle \Lambda(t-s, x-*)\sigma(u^\varepsilon(s, *)), e_k\rangle_{\mathcal{H}} \mathrm{d}W^k(s).$$

Appealing to [7, Theorem 4.3] and to Remark 2.1, for any fixed $\varepsilon \in (0, 1]$, there exists a stochastic process $\{u^\varepsilon(t, x), (t, x) \in [0, T] \times \mathbb{R}^d\}$ such that (2.1) holds for any $(t, x) \in [0, T] \times \mathbb{R}^d$ a.s. This is termed a *random-field solution* to (2.1). More precisely, we have the following result.

Theorem 2.2 *If Hypotheses* (A1) *and* (A2) *are satisfied and σ and b are Lipschitz continuous functions, then* (2.1) *has a unique random-field solution. Among other properties, this solution is L^2-continuous and for any $p \in [1, \infty)$*

$$\sup_{\varepsilon\in(0,1]} \sup_{(t,x)\in[0,T]\times\mathbb{R}^d} \mathbb{E}\Big[|u^\varepsilon(t, x)|^p\Big] < \infty.$$

We are interested in the family of densities of the probability law of the solution $u^\varepsilon(t, x)$, $\varepsilon \in (0, 1]$ at every fixed point $(t, x) \in (0, T] \times \mathbb{R}^d$. For this, we describe the abstract Wiener space that will be used as framework for the application of the Malliavin Calculus (see [25]).

Let $(\bar{\Omega}, \bar{\mathcal{G}}, \bar{\mu})$ be the canonical space of a standard real-valued Brownian motion on $[0, T]$. With the equivalence shown before, we can identify the canonical probability space of the stochastic process F with that corresponding to a sequence of independent standard Brownian motions $(\Omega, \mathcal{G}, \mathbb{P}) = (\bar{\Omega}^{\mathbb{N}}, \bar{\mathcal{G}}^{\otimes\mathbb{N}}, \bar{\mu}^{\otimes\mathbb{N}})$. This will be the underlying probability space in this article.

Consider the Hilbert space H consisting of sequences $(h^k)_{k\in\mathbb{N}}$ of functions $h^k : [0, T] \to \mathbb{R}$ which are absolutely continuous with respect to the Lebesgue measure and such that $\|h\|_H^2 = \sum_{k\in\mathbb{N}} \int_0^T |\dot{h}^k(s)|^2 ds < \infty$, where $\dot{h}^k$ refers to the derivative of h^k defined almost everywhere. There is an isometry $I : H \to \mathcal{H}_T$ defined by $i(h)(t) = \sum_{k\in\mathbb{N}} \dot{h}^k(t) e_k$. In the sequel we will identify the Hilbert spaces H and $\mathcal{H}_T$ and by an abuse of notation, we will write $i(h) = h$. The triple $(\Omega, H, \mathbb{P})$ is the abstract Wiener space that we shall use as framework for the Malliavin calculus.

Let us introduce some additional assumptions:

(A3) There exist positive constants $C, \gamma > 0$ and $t_0 \in (0, T]$ such that for all $t \in [0, t_0]$,

$$Ct^\gamma \le \int_0^t J_1(s) \mathrm{d}s = g_1(t).$$

(A4) The functions σ and b are infinitely differentiable with bounded derivatives of any order greater or equal than one.

(A5) The function σ satisfies $\inf_{x\in\mathbb{R}} |\sigma(x)| = \sigma_0 > 0$.

The following result in [26] establishes the existence and regularity of the densities for the solution to (2.1) at any point $(t, x) \in (0, T] \times \mathbb{R}^d$.

Theorem 2.3 *Fix* $(t, x) \in (0, T] \times \mathbb{R}^d$ *and* $\varepsilon \in (0, 1]$. *Assume* (A1), (A2), (A3), (A4) *and* (A5). *Then the law of* $u^\varepsilon(t, x)$ *is absolutely continuous with respect to the Lebesgue measure on* $\mathbb{R}$ *and its density, denoted by* $p^\varepsilon_{t,x}$, *is an infinitely differentiable function.*

The last relevant assumption is the following.

(A6) For every $(t, x) \in (0, T] \times \mathbb{R}^d$ the family $(u^\varepsilon(t, x))_{\varepsilon\in(0,1]}$ satisfies a large deviation principle on $\mathbb{R}$ with rate function J.

We refer the reader to [11] for notions and results on large deviations.

We are now in a position to formulate the main result of this paper. It is about the behaviour of the density $p^\varepsilon_{t,x}(y)$ at every fixed $(t, x) \in (0, T] \times \mathbb{R}^d$ and $y \in \mathbb{R}$, as $\varepsilon \to 0$. It will be proved by using the method introduced in [15, 16] (see [24] for a general formulation).

Theorem 2.4 (i) *Fix* $(t, x) \in (0, T] \times \mathbb{R}^d$ *and assume* (A1), (A2), (A3), (A4), (A5) *and* (A6). *Then for any* $y \in \mathbb{R}$,

$$\lim_{\varepsilon\downarrow 0} \varepsilon^2 \log p^\varepsilon_{t,x}(y) \le -J(y). \tag{2.3}$$

(ii) *Let* $(t, x) \in (0, T] \times \mathbb{R}^d$. *Assume* (A1), (A2), (A3), (A4) *and* (A5). *Fix* $y \in \mathbb{R}$ *in the interior of the topological support of the law of* $u(t, x)$. *Then,*

$$\lim_{\varepsilon\downarrow 0} \varepsilon^2 \log p^{\varepsilon}_{t,x}(y) \geq -I(y), \tag{2.4}$$

with

$$I(y) = \inf\left\{\frac{1}{2}\|h\|^2_{\mathcal{H}_T};\ h \in \mathcal{H}_T,\, \Phi^h_{t,x} = y\right\}, \tag{2.5}$$

and where $\Phi^h_{t,x} \in \mathbb{R}$ *is defined by*

$$\Phi^h_{t,x} = w(t,x) + \left\langle \Lambda(t-\cdot, x-*)\sigma(\Phi^h_{\cdot,*}), h\right\rangle_{\mathcal{H}_T} + \int_0^t \int_{\mathbb{R}^d} \Lambda(t-s, x-z) b(\Phi^h_{s,z}) \mathrm{d}z \mathrm{d}s. \tag{2.6}$$

We end this section with some important comments on these statements. The existence and uniqueness of a solution to (2.6) follows from Theorem A.1 in the Appendix. Theorem 2.4 makes sense for those $y \in \mathbb{R}$ such that $p^{\varepsilon}_{t,x}(y) > 0$ for all ε sufficiently small, and the lower bound in (2.4) is nontrivial only if $I(y) < +\infty$. In the last part of Sect. 3.2, we show the connection between these properties and the topological support of the law of $u^{\varepsilon}(t,x)$.

Under some additional assumptions, in Sect. 4 we will prove a characterization of the topological support of $u^{\varepsilon}(t,x)$, $\mathcal{S}$, that exhibits its independence on ε. Proposition 3.10 shows that if b is bounded, $\mathcal{S} = \mathbb{R}$.

We prove in Proposition 3.10 that $I(y)$ in (2.5) is finite for any y in the interior of $\mathcal{S}$. This uses the characterization of the support.

Assume that y belongs to the interior of $\mathcal{S}$. Then, [24, Proposition 4.1.2] yields that $p^{\varepsilon}_{t,x}(y) > 0$ for all $\varepsilon \in (0,1]$.

In Sect. 5 we show that for a class of stochastic wave equations, the hypotheses of Theorem 2.4 are satisfied, and J and I in (2.3) and (2.4) respectively, are identical. Hence, for any y in the interior of the support of the law we have

$$\lim_{\varepsilon\downarrow 0} \varepsilon^2 \log p^{\varepsilon}_{t,x}(y) = -I(y),$$

with I defined in (2.5). With some restrictions, Theorem 2.4 also applies to the stochastic heat equation.

Throughout this article we use the notation C for generic constants that may change from one expression to another. As for the notations and notions of Malliavin calculus, we refer to [24, 25].

3 Proof of the Main Result

The two parts of Theorem 2.4 will be established separately, applying the methods introduced in [15, 16] and extended to an abstract setting in [24] (see Propositions 3.1, 3.2).

By $\|\cdot\|_{k,p}$, $k \in \mathbb{N}$, $p \in [1,\infty)$, we will denote the norm in the space $\mathbb{D}^{k,p}$ (the Watanabe-Sobolev spaces), and by $\|\cdot\|_p$, the $L^p(\Omega)$ norm. We say that a random variable $X : \Omega \to \mathbb{R}$ is non-degenerate if $X \in \mathbb{D}^\infty = \bigcap_{k\in\mathbb{N}} \bigcap_{p\in[1,\infty)} \mathbb{D}^{k,p}$ and the random variable $\gamma_X := \|DX\|^2_{\mathcal{H}_T}$ satisfies $\gamma_X^{-1} \in \bigcap_{p\in[1,\infty)} L^p(\Omega)$, were D denotes the Malliavin derivative. The law of a non-degenerate random variable possesses an infinitely differentiable density.

For the upper bound (2.3), we rely on the following result.

Proposition 3.1 ([24, Proposition 4.4.2]) *Let $(F^\varepsilon)_{\varepsilon\in(0,1]}$ be a family of non-degenerate random variables satisfying*

(i) $\sup_{\varepsilon\in(0,1]} \|F^\varepsilon\|_{k,p} < \infty$ *for all $k \in \mathbb{N}$ and $p \in [1,\infty)$;*
(ii) *for any $p \in [1,\infty)$ there exists $N_p \in [1,\infty)$ such that $\|\gamma_{F^\varepsilon}^{-1}\|_p \le \varepsilon^{-N_p}$;*
(iii) *$(F^\varepsilon)_{\varepsilon\in(0,1]}$ obeys a large deviation principle on $\mathbb{R}$ with rate function J.*

Then

$$\limsup_{\varepsilon\downarrow 0} \varepsilon^2 \log p^\varepsilon(y) \le -J(y),$$

where p^ε denotes the density of F^ε.

We denote by $\mathcal{C}^1(\mathcal{H}_T;\mathbb{R})$ the set of all Fréchet differentiable real functions F defined on $\mathcal{H}_T$. For such deterministic functions, we shall use the notation $\bar{D}F$ for its Fréchet derivative and set $\bar{\gamma}_F = \|\bar{D}F\|^2_{\mathcal{H}_T}$.

The lower bound (2.4) will be established using the following Proposition.

Proposition 3.2 ([24, Proposition 4.4.1]) *Let $(F^\varepsilon)_{\varepsilon\in(0,1]}$ be a family of non-degenerate random variables. Let $\psi \in \mathcal{C}^1(\mathcal{H}_T;\mathbb{R})$ be such that for all $h \in \mathcal{H}_T$*

$$\lim_{\varepsilon\downarrow 0} \frac{F^\varepsilon(\omega + \varepsilon^{-1}h) - \psi(h)}{\varepsilon} = \mathcal{N}(h) \tag{3.1}$$

in the $\mathbb{D}^\infty$-topology, where $\mathcal{N}$ is a random variable belonging to the first Wiener chaos with variance $\bar{\gamma}_\psi(h) = \|\bar{D}\psi(h)\|^2_{\mathcal{H}_T}$. Then

$$\begin{aligned}\liminf_{\varepsilon\downarrow 0} \varepsilon^2 \log p^\varepsilon(y) &\ge -\frac{1}{2} d_R^2(y)\\ &:= -\frac{1}{2} \inf\left\{\|h\|^2_{\mathcal{H}_T};\ h \in \mathcal{H}_T, \psi(h) = y, \bar{\gamma}_\psi(h) > 0\right\},\end{aligned}$$

where p^ε denotes the density of F^ε.

Let us point out that in the proof of this proposition it is implicitly assumed that $p^\varepsilon(y) > 0$ for any $0 < \varepsilon < \varepsilon_0$, with ε_0 small enough.

3.1 Upper Bound

The objective of this section is to apply Proposition 3.1 to the family of random variables $F^\varepsilon := u^\varepsilon(t,x)$, $\varepsilon \in (0,1]$, given in (2.1), where $(t,x) \in (0,T] \times \mathbb{R}^d$ is fixed. We will assume that (A6) holds and check that (A1)–(A5) imply the validity of (i) and (ii) in Proposition 3.1. This will prove the statement (i) in Theorem 2.4.

Lemma 3.3 *Under the conditions* (A1)*,* (A2) *and* (A4) *the Assumption (i) in Proposition* 3.1 *holds. More precisely, we have*

$$\sup_{\varepsilon\in(0,1]} \sup_{(t,x)\in[0,T]\times\mathbb{R}^d} \mathbb{E}\Big[\big\| D^k u^\varepsilon(t,x)\big\|^p_{\mathcal{H}_T^{\otimes k}}\Big] < \infty.$$

Proof This follows along the same lines as in [26, Proposition 6.1]. The difference is that here we are considering a family of SPDEs depending on a parameter $\epsilon \in (0,1]$, and obtain that the norm is uniformly bounded in ε. □

For its further use we recall that for every $\varepsilon \in (0,1]$, the Malliavin derivative of the process $\{u^\varepsilon(t,x), (t,x) \in [0,T] \times \mathbb{R}^d\}$ is an $\mathcal{H}_T$-valued stochastic process $\{Du^\varepsilon(t,x), (t,x) \in [0,T] \times \mathbb{R}^d\}$, solution to

$$\begin{aligned} Du^\varepsilon(t,x) &= \varepsilon\Lambda(t-\cdot, x-*)\sigma(u^\varepsilon(\cdot,*)) \\ &\quad + \varepsilon\int_0^t\int_{\mathbb{R}^d} \Lambda(t-s,x-y)\sigma'(u^\varepsilon(s,y))Du^\varepsilon(s,y)M(\mathrm{d}s,\mathrm{d}y) \\ &\quad + \int_0^t\int_{\mathbb{R}^d} \Lambda(t-s,x-y)b'(u^\varepsilon(s,y))Du^\varepsilon(s,y)\mathrm{d}y\mathrm{d}s. \end{aligned} \tag{3.2}$$

For the background on the Hilbert-valued stochastic and pathwise integrals in the preceding equation, we refer the reader to [26] (see also [7]).

Lemma 3.4 *Fix* $(t,x) \in (0,T] \times \mathbb{R}^d$ *and assume* (A1)*,* (A2)*,* (A3)*,* $\sigma, b \in \mathcal{C}^1$ *with bounded derivatives and* (A5)*. Then for every* $p \in [1,\infty)$ *there exists* $C_p > 0$ *such that*

$$\big\|\gamma^{-1}_{u^\varepsilon(t,x)}\big\|_p \le C_p\varepsilon^{-2},$$

for any $\varepsilon \in (0,1]$.

Proof Fix $\varepsilon \in (0,1]$ and $q \ge 2$. We will prove that there exists $\zeta_0 := \zeta_0(q) > 0$ such that for all $\zeta \in (0,\zeta_0)$

$$P^{\varepsilon}(\zeta) := \mathbb{P}\left\{\varepsilon^{-2}\|Du^{\varepsilon}(t,x)\|^2_{\mathcal{H}_T} \le \zeta\right\} \le C\left(\zeta^q + \zeta^{\frac{2q\delta}{\gamma}}\right), \tag{3.3}$$

where C is a constant not depending on ζ. Then, by the formula

$$\mathbb{E}(Y) = \int_0^{\infty} \mathbb{P}\{Y \ge \zeta\}\mathrm{d}\zeta,$$

valid for nonnegative random variables Y, the assertion will follow.

For any $0 \le s < t$, let $\mathcal{H}_{s,t} = L^2([s,t],\mathcal{H})$. Let t_0 be as defined in (A3). We consider $\rho > 0$ satisfying $\rho < t \wedge t_0$. From (3.2) and the triangular inequality, we clearly have

$$\begin{aligned}\|Du^{\varepsilon}(t,x)\|^2_{\mathcal{H}_T} &\ge \|Du^{\varepsilon}(t,x)\|^2_{\mathcal{H}_{t-\rho,t}}\\ &\ge \frac{1}{2}\|\varepsilon\Lambda(t-\cdot,x-*)\sigma(u^{\varepsilon}(\cdot,*))\|^2_{\mathcal{H}_{t-\rho,t}} - \|X^{\varepsilon}(t,x)\|^2_{\mathcal{H}_{t-\rho,t}},\end{aligned}$$

where

$$X^{\varepsilon}(t,x) := Du^{\varepsilon}(t,x) - \varepsilon\Lambda(t-\cdot,x-*)\sigma(u^{\varepsilon}(\cdot,*)). \tag{3.4}$$

The assumption (A5) yields

$$\|\varepsilon\Lambda(t-\cdot,x-*)\sigma(u^{\varepsilon}(\cdot,*))\|^2_{\mathcal{H}_{t-\rho,t}} \ge \varepsilon^2\sigma_0^2 g_1(\rho).$$

Hence,

$$\begin{aligned}\mathbb{P}\left\{\varepsilon^{-2}\|Du^{\varepsilon}(t,x)\|^2_{\mathcal{H}_T} \le \zeta\right\} &\le \mathbb{P}\left\{\varepsilon^{-2}\|X^{\varepsilon}(t,x)\|^2_{\mathcal{H}_{t-\rho,t}} \ge \frac{\sigma_0^2}{2}g_1(\rho) - \zeta\right\}\\ &\le \left(\frac{\sigma_0^2}{2}g_1(\rho) - \zeta\right)^{-q}\varepsilon^{-2q}\mathbb{E}\big[\|X^{\varepsilon}(t,x)\|^{2q}_{\mathcal{H}_{t-\rho,t}}\big],\end{aligned} \tag{3.5}$$

where in the last inequality we have applied Chebyshev's inequality.

Our next objective is to find an upper bound for $\mathbb{E}\big[\|X^{\varepsilon}(t,x)\|^{2q}_{\mathcal{H}_{t-\rho,t}}\big]$. From (3.2) we have

$$\mathbb{E}\big[\|X^{\varepsilon}(t,x)\|^{2q}_{\mathcal{H}_{t-\rho,t}}\big] \le C\big(T_1(t,x;\rho,q) + T_2(t,x;\rho,q)\big),$$

with

$$T_1^{\varepsilon}(t,x;\rho,q) = \mathbb{E}\left[\left\|\varepsilon\int_0^t\int_{\mathbb{R}^d}\Lambda(t-s,x-y)\sigma'(u^{\varepsilon}(s,y))Du^{\varepsilon}(s,y)M(\mathrm{d}s,\mathrm{d}y)\right\|^{2q}_{\mathcal{H}_{t-\rho,t}}\right],$$

$$T_2^\varepsilon(t,x;\rho,q) = \mathbb{E}\left[\left\| \int_0^t \int_{\mathbb{R}^d} \Lambda(t-s,x-y) b'(u^\varepsilon(s,y)) D u^\varepsilon(s,y) \mathrm{d}y \mathrm{d}s \right\|_{\mathcal{H}_{t-\rho,t}}^{2q}\right].$$

The Malliavin derivative $D_{r,*}u^\varepsilon(s,y)$ vanishes if $r \in (s,T]$. Thus, if $r \in [t-\rho,t]$, the domain of integration of the s variable in the terms $T_1^\varepsilon(t,x;\rho,q)$, $T_2^\varepsilon(t,x;\rho,q)$ can be replaced by $[t-\rho,t]$. Moreover, following the proof of [31, Lemma 8.2], we have

$$\sup_{\varepsilon\in(0,1]} \sup_{s\in[0,\rho]} \sup_{y\in\mathbb{R}^d} \mathbb{E}\big[\|D_{t-\cdot,*}u^\varepsilon(t-s,y)\|_{\mathcal{H}_\rho}^{2q}\big] \le C(g_1(\rho))^q. \tag{3.6}$$

By applying Burholder's inequality for Hilbert valued martingales (see for instance [20]), we obtain

$$\begin{aligned} T_1^\varepsilon(t,x;\rho,q) &\le C\varepsilon^{2q}(g_1(\rho))^q \sup_{\varepsilon\in(0,1]} \sup_{s\in[t-\rho,t]} \sup_{y\in\mathbb{R}^d} \mathbb{E}\big[\|Du^\varepsilon(s,y)\|_{\mathcal{H}_{t-\rho,t}}^{2q}\big] \\ &\le C\varepsilon^{2q}[g_1(\rho)]^{2q}, \end{aligned} \tag{3.7}$$

where in the last inequality we have used (3.6).

We proceed now to the study of the term $T_2^\varepsilon(t,x;\rho,q)$. For this, we use (3.4) and Minkovski's inequality for the norm $\|\cdot\|_{\mathcal{H}}$. So we are left with two terms that we study separately. For the first one, we use that $X_{r,*}(s,y)$ vanishes for $r \in (s,T]$, Hölder's inequality with respect to the finite measure $\Lambda(t-s,x-y)dsdy := \Lambda(t-ds,x-dy)$, the boundedness of b' and (2.2). We obtain

$$\begin{aligned} &\mathbb{E}\left[\left\| \int_0^t \int_{\mathbb{R}^d} \Lambda(t-s,x-y) b'(u^\varepsilon(s,y)) X^\varepsilon(s,y) \mathrm{d}y \mathrm{d}s \right\|_{\mathcal{H}_{t-\rho,t}}^{2q}\right] \\ &\le \mathbb{E}\left[\left(\int_{t-\rho}^t \int_{\mathbb{R}^d} \Lambda(t-s,x-y) |b'(u^\varepsilon(s,y))| \|X^\varepsilon(s,y)\|_{\mathcal{H}_{t-\rho,t}} \mathrm{d}y \mathrm{d}s \right)^{2q}\right] \\ &\le C\rho^{\delta(2q-1)} \int_{t-\rho}^t \sup_{y\in\mathbb{R}^d} \mathbb{E}\big[\|X^\varepsilon(s,y)\|_{\mathcal{H}_{t-\rho,t}}^{2q}\big] J_2(t-s) ds \\ &\le C\rho^{\delta(2q-1)} \int_{t-\rho}^t \sup_{y\in\mathbb{R}^d} \mathbb{E}\big[\|X^\varepsilon(s,y)\|_{\mathcal{H}_{s-\rho,s}}^{2q}\big] J_2(t-s) ds, \end{aligned} \tag{3.8}$$

where the last inequality follows from the property $\|\cdot\|_{\mathcal{H}_{t-\rho,t}} \le \|\cdot\|_{\mathcal{H}_{s-\rho,t}}$, for any $0 \le s \le t \le T$.

Next we consider the second contribution from $T_2^{\varepsilon}(t, x, \rho, q)$. The Lipschitz continuity of σ together with Theorem 2.2 imply

$$\sup_{\varepsilon\in(0,1]} \sup_{(t,x)\in[0,T]\times\mathbb{R}^d} \mathbb{E}\left[|\sigma(u^{\varepsilon}(t,x))|^q\right] \le C\left(1+ \sup_{\varepsilon\in(0,1]} \sup_{(t,x)\in[0,T]\times\mathbb{R}^d} \mathbb{E}\left[|u^{\varepsilon}(t,x)|^q\right]\right),$$

for any $q \in [1, \infty)$. This yields

$$\sup_{y\in\mathbb{R}^d} \mathbb{E}\left[\|\Lambda(s-\cdot, y-*)\sigma(u^{\varepsilon}(\cdot,*))\|^{2q}_{\mathcal{H}_{s-\rho,s}}\right] \le C(g_1(\rho))^q.$$

Using this estimate and proceeding in a similar way as in the study of the previous term, we obtain

$$\mathbb{E}\left[\left\|\int_0^t\int_{\mathbb{R}^d} \Lambda(t-s,x-y)b'(u^{\varepsilon}(s,y))\varepsilon\Lambda(s-\cdot,y-*)\sigma(u^{\varepsilon}(t,x))\mathrm{d}y\mathrm{d}s\right\|^{2q}_{\mathcal{H}_{t-\rho,t}}\right]$$
$$\le C\varepsilon^{2q}\rho^{2q\delta}(g_1(\rho))^q. \tag{3.9}$$

With (3.7), (3.8) and (3.9), we have proved

$$\sup_{x\in\mathbb{R}^d} \mathbb{E}\left[\|X^{\varepsilon}(t,x)\|^{2q}_{\mathcal{H}_{t-\rho,t}}\right] \le C\Bigg(\varepsilon^{2q}\left((g_1(\rho))^{2q} + \rho^{2q\delta}(g_1(\rho))^q\right)$$
$$+ \rho^{\delta(2q-1)}\int_{t-\rho}^t \sup_{y\in\mathbb{R}^d} \mathbb{E}\left[\|X^{\varepsilon}(s,y)\|^{2q}_{\mathcal{H}_{s-\rho,s}}\right]J_2(t-s)\mathrm{d}s\Bigg).$$

Applying Gronwall's lemma in [5, Lemma 15] to the function

$$f(t) = \sup_{x\in\mathbb{R}^d} \mathbb{E}\left[\|X^{\varepsilon}(t,x)\|^{2q}_{\mathcal{H}_{t-\rho,t}}\right],$$

we have

$$\sup_{x\in\mathbb{R}^d} \mathbb{E}\left[\|X^{\varepsilon}(t,x)\|^{2q}_{\mathcal{H}_{t-\rho,t}}\right] \le C\varepsilon^{2q}\left((g_1(\rho))^{2q} + \rho^{2q\delta}(g_1(\rho))^q\right).$$

Plugging this estimate in (3.5) we obtain

$$\mathbb{P}\left\{\varepsilon^{-2}\|Du^{\varepsilon}(t,x)\|^2_{\mathcal{H}_T} \le \zeta\right\} \le C\left(\frac{\sigma_0^2}{2}g_1(\rho) - \zeta\right)^{-q}\left((g_1(\rho))^{2q} + \rho^{2q\delta}(g_1(\rho))^q\right).$$

Let $0 < \rho = \rho(\zeta) \le t \wedge t_0$ be such that $g_1(\rho) = \frac{4}{\sigma_0^2}\zeta$, which by (A3) implies $\rho \le C\zeta^{1/\gamma}$. With this choice of ρ, the preceding inequality yields (3.3).

The proof of Theorem 2.4(i) is now complete. □

3.2 Lower Bound

The purpose of this section is to prove that the family of random variables $(F^\varepsilon)_{\varepsilon\in(0,1]} = (u^\varepsilon(t,x))_{\varepsilon\in(0,1]}$, with fixed $(t,x) \in (0,T]\times\mathbb{R}^d$, satisfies the assumptions of Proposition 3.2, with $\psi(h) = \Phi^h_{t,x}$ (see (2.6)), and we will identify the random variable $\mathcal{N}$. We will also prove that for any $h \in \mathcal{H}_T$, $\bar{\gamma}_{\Phi^h_{t,x}} > 0$.

Lemma 3.5 *Assume* (A1), (A2) *and that* $\sigma, b \in \mathcal{C}^1$ *with Lipschitz continuous and bounded derivatives. Then, for all* $(t,x) \in [0,T]\times\mathbb{R}^d$, *the mapping* $\mathcal{H}_T \ni h \mapsto \Phi^h_{t,x}$ *defined in* (2.6) *is Fréchet differentiable.*

Proof Fix $(t,x) \in [0,T]\times\mathbb{R}^d$, $h \in \mathcal{H}_T$. We use Cauchy-Schwarz' inequality, (A2), (A1) and the Lipschitz continuity of σ and b to obtain

$$\begin{aligned}
|\Phi^h_{t,x}|^2 &\le C\Bigg\{1 + \|h\|^2_{\mathcal{H}_T}\|\Lambda(t-\cdot,x-*)\sigma(\Phi^h_{t,x})\|^2_{\mathcal{H}_T} \\
&\qquad + \left|\int_0^t\int_{\mathbb{R}^d}\Lambda(t-s,x-z)b(\Phi^h_{s,z})\mathrm{d}z\mathrm{d}s\right|^2\Bigg\} \\
&\le C(\|h\|^2_{\mathcal{H}_T}+1)\int_0^t\Big(1+\sup_{(r,y)\in[0,s]\times\mathbb{R}^d}|\Phi^h_{r,y}|^2\Big)\big(J_1(t-s)+J_2(t-s)\big)\mathrm{d}s.
\end{aligned}$$

Gronwall's Lemma yields

$$\sup_{(t,x)\in[0,T]\times\mathbb{R}^d}|\Phi^h_{t,x}|^2 \le C(\|h\|^2_{\mathcal{H}_T}+1)\int_0^T\big(J_1(s)+J_2(s)\big)\mathrm{d}s < \infty, \tag{3.10}$$

where the constant C is independent of $h \in \mathcal{H}_T$. Now fix $h_0 \in \mathcal{H}_T$ and note that

$$\begin{aligned}
\Phi^{h+h_0}_{t,x} - \Phi^h_{t,x} =& \big\langle\Lambda(t-\cdot,x-*)\sigma(\Phi^{h+h_0}_{\cdot,*}),h_0\big\rangle_{\mathcal{H}_T} \\
&+\big\langle\Lambda(t-\cdot,x-*)\big(\sigma(\Phi^{h+h_0}_{\cdot,*})-\sigma(\Phi^h_{\cdot,*})\big),h\big\rangle_{\mathcal{H}_T} \\
&+\int_0^t\int_{\mathbb{R}^d}\Lambda(t-s,x-z)\big(b(\Phi^{h+h_0}_{s,z})-b(\Phi^h_{s,z})\big)\mathrm{d}z\mathrm{d}s.
\end{aligned}$$

With the same arguments as for the proof of (3.10), we get that for all $h_0 \in \mathcal{H}_T$

$$|\Phi^{h+h_0}_{t,x} - \Phi^h_{t,x}|^2 \le C\|h_0\|^2_{\mathcal{H}_T}\|\Lambda(t-\cdot, x-*)\sigma(\Phi^{h+h_0}_{\cdot,*})\|^2_{\mathcal{H}_T}$$
$$+ C\big(\|h\|^2_{\mathcal{H}_T}+1\big)\int_0^t \sup_{(r,y)\in[0,s]\times\mathbb{R}^d} |\Phi^{h+h_0}_{r,y} - \Phi^h_{r,y}|^2\big(J_1(t-s)+J_2(t-s)\big)\mathrm{d}s.$$

Due to (3.10) and (A1), the first term is bounded (up to a constant depending on h and h_0) by $\|h_0\|^2_{\mathcal{H}_T}$. Then applying Gronwall's Lemma we obtain

$$\sup_{(t,x)\in[0,T]\times\mathbb{R}^d} |\Phi^{h+h_0}_{t,x} - \Phi^h_{t,x}| \le C_{h,h_0}\|h_0\|_{\mathcal{H}_T}. \tag{3.11}$$

Note that the constant C_{h,h_0} does not blow up as $\|h_0\| \to 0$. With (3.10) and (3.11), we can prove the existence of the Fréchet derivative of the map $h \mapsto \Phi^h_{t,x}$. First, we provide a candidate for it at the point $g \in \mathcal{H}_T$, as follows:

$$\begin{aligned}\bar{D}\Phi^h_{t,x}(g) &= \big\langle \Xi^h(t,x), g\big\rangle_{\mathcal{H}_T}\\ &= \big\langle \Lambda(t-\circ, x-\bullet)\sigma(\Phi^h_{\circ,\bullet}), g\big\rangle_{\mathcal{H}_T}\\ &\quad + \big\langle \Lambda(t-\cdot, x-*)\sigma'(\Phi^h_{\cdot,*})\big\langle \Xi^h(\cdot,*), g\big\rangle_{\mathcal{H}_T}, h\big\rangle_{\mathcal{H}_T}\\ &\quad + \int_0^t\int_{\mathbb{R}^d} \Lambda(t-s, x-z)b'(\Phi^h_{s,z})\big\langle \Xi^h(s,z), g\big\rangle_{\mathcal{H}_T}\mathrm{d}z\mathrm{d}s,\end{aligned} \tag{3.12}$$

where $\Xi^h(t,x)$ is defined by the integral equation on $\mathcal{H}_T$:

$$\begin{aligned}\Xi^h_{\circ,\bullet}(t,x) &= \Lambda(t-\circ, x-\bullet)\sigma(\Phi^h_{\circ,\bullet}) + \big\langle \Lambda(t-\cdot, x-*)\sigma'(\Phi^h_{\cdot,*})\Xi^h_{\circ,\bullet}(\cdot,*), h\big\rangle_{\mathcal{H}_T}\\ &\quad + \int_0^t\int_{\mathbb{R}^d} \Lambda(t-s, x-z)b'(\Phi^h_{s,z})\Xi^h_{\circ,\bullet}(s,z)\mathrm{d}z\mathrm{d}s.\end{aligned} \tag{3.13}$$

According to Theorem A.1, this equation has a unique solution. Note that in the previous two formulas $(\circ, \bullet)$ is the argument in $[0,T]\times\mathbb{R}^d$ which interacts with $g \in \mathcal{H}_T$ (the element at which $\Xi^h(t,x)$ is evaluated) and $(\cdot, *)$ is the argument in $[0,T]\times\mathbb{R}^d$ that interacts with $h \in \mathcal{H}_T$ which is the point where the Fréchet derivative is taken.

From (3.12) and (3.13), we clearly have

$$
\begin{aligned}
&\frac{\Phi^{h+h_0}_{t,x} - \Phi^h_{t,x} - \bar{D}\Phi^h_{t,x}(h_0)}{\|h_0\|_{\mathcal{H}_T}}\\
&\quad = \frac{1}{\|h_0\|_{\mathcal{H}_T}} \big\langle \Lambda(t-\circ, x-\bullet)\big(\sigma(\Phi^{h+h_0}_{\circ,\bullet}) - \sigma(\Phi^h_{\circ,\bullet})\big), h_0\big\rangle_{\mathcal{H}_T}\\
&\qquad + \left\langle \Lambda(t-\cdot, x-*)\left(\frac{\sigma(\Phi^{h+h_0}_{\cdot,*}) - \sigma(\Phi^h_{\cdot,*}) - \sigma'(\Phi^h_{\cdot,*})\bar{D}\Phi^h_{\cdot,*}(h_0)}{\|h_0\|_{\mathcal{H}_T}}\right), h\right\rangle_{\mathcal{H}_T}\\
&\qquad + \int_0^t \int_{\mathbb{R}^d} \Lambda(t-s, x-z) \frac{b(\Phi^{h+h_0}_{s,z}) - b(\Phi^h_{s,z}) - b'(\Phi^h_{s,z})\bar{D}\Phi^h_{t,x}(h_0)}{\|h_0\|_{\mathcal{H}_T}} \mathrm{d}z\mathrm{d}s.
\end{aligned}
\tag{3.14}
$$

Our aim is to have an upper bound for the absolute value of each term on the right-hand side of (3.14). By applying Cauchy-Schwarz' inequality, the fact that σ is Lipschitz continuous, (A1) and (3.11), we have

$$
\begin{aligned}
&\frac{1}{\|h_0\|^2_{\mathcal{H}_T}} \left|\big\langle \Lambda(t-\circ, x-\bullet)\big(\sigma(\Phi^{h+h_0}_{\circ,\bullet}) - \sigma(\Phi^h_{\circ,\bullet})\big), h_0\big\rangle^2_{\mathcal{H}_T}\right|\\
&\qquad \le \left\|\Lambda(t-\cdot, x-*)\big(\sigma(\Phi^{h+h_0}_{\cdot,*}) - \sigma(\Phi^h_{\cdot,*})\big)\right\|^2_{\mathcal{H}_T}\\
&\qquad \le \sup_{(r,y)\in[0,T]\times\mathbb{R}^d} \left|\sigma(\Phi^{h+h_0}_{r,y}) - \sigma(\Phi^h_{r,y})\right|^2 \|\Lambda(t-\cdot, x-*)\|^2_{\mathcal{H}_T}\\
&\qquad \le C_{h,h_0}\|h_0\|^2_{\mathcal{H}_T}.
\end{aligned}
$$

For the second term, we first use Cauchy-Schwarz' inequality and apply the usual procedure involving the Fourier transformation. Then we use the mean-value theorem to obtain

$$
\begin{aligned}
&\left|\left\langle \Lambda(t-\cdot, x-*)\left(\frac{\sigma(\Phi^{h+h_0}_{\cdot,*}) - \sigma(\Phi^h_{\cdot,*}) - \sigma'(\Phi^h_{\cdot,*})\bar{D}\Phi^h_{\cdot,*}(h_0)}{\|h_0\|_{\mathcal{H}_T}}\right), h\right\rangle_{\mathcal{H}_T}\right|^2\\
&\quad \le \|h\|^2_{\mathcal{H}_T} \int_0^t \sup_{(r,y)\in[0,s]\times\mathbb{R}^d} \left|\frac{\sigma(\Phi^{h+h_0}_{r,y}) - \sigma(\Phi^h_{r,y}) - \sigma'(\Phi^h_{r,y})\bar{D}\Phi^h_{r,y}(h_0)}{\|h_0\|_{\mathcal{H}_T}}\right|^2 J_1(t-s)\mathrm{d}s\\
&\quad = \|h\|^2_{\mathcal{H}_T} \int_0^t \sup_{(r,y)\in[0,s]\times\mathbb{R}^d} \left|\frac{\sigma'(\xi^{h,h_0}_{r,y})(\Phi^{h+h_0}_{r,y} - \Phi^h_{r,y}) - \sigma'(\Phi^h_{r,y})\bar{D}\Phi^h_{r,y}(h_0)}{\|h_0\|_{\mathcal{H}_T}}\right|^2 J_1(t-s)\mathrm{d}s\\
&\quad \le \|h\|^2_{\mathcal{H}_T} \int_0^t \sup_{(r,y)\in[0,s]\times\mathbb{R}^d} \left|\big(\sigma'(\xi^{h,h_0}_{r,y}) - \sigma'(\Phi^h_{r,y})\big)\frac{\Phi^{h+h_0}_{r,y} - \Phi^h_{r,y}}{\|h_0\|_{\mathcal{H}_T}}\right|^2 J_1(t-s)\mathrm{d}s\\
&\qquad + \|h\|^2_{\mathcal{H}_T} \int_0^t \sup_{(r,y)\in[0,s]\times\mathbb{R}^d} \left|\sigma'(\Phi^h_{r,y})\frac{\Phi^{h+h_0}_{r,y} - \Phi^h_{r,y} - \bar{D}\Phi^h_{r,y}(h_0)}{\|h_0\|_{\mathcal{H}_T}}\right|^2 J_1(t-s)\mathrm{d}s,
\end{aligned}
$$

where $\xi^{h,h_0}_{r,y}$ is a real number in the convex hull of $\Phi^{h+h_0}_{r,y}$ and $\Phi^h_{r,y}$.

Using (3.11) and the Lipschitz continuity property of σ', along with (A3), the first term on the right-hand side of the last inequality can be bounded from above by

$$C\|h\|^2_{\mathcal{H}_T} \sup_{(t,x)\in[0,T]\times\mathbb{R}^d} \left|\xi^{h,h_0}_{t,x} - \Phi^h_{t,x}\right|^2,$$

and therefore also by $C_{h,h_0}\|h_0\|^2_{\mathcal{H}_T}\|h\|^2_{\mathcal{H}_T}$.

We are assuming that σ' is bounded. Hence, we have proved

$$\begin{aligned}
&\left|\left\langle \Lambda(t-\cdot, x-*)\left(\frac{\sigma(\Phi^{h+h_0}_{\cdot,*}) - \sigma(\Phi^h_{\cdot,*}) - \sigma'(\Phi^h_{\cdot,*})\bar{D}\Phi^h_{\cdot,*}(h_0)}{\|h_0\|_{\mathcal{H}_T}}\right), h\right\rangle_{\mathcal{H}_T}\right|^2\\
&\le C_{h,h_0}\|h_0\|^2_{\mathcal{H}_T} + C_h\int_0^t \sup_{(r,y)\in[0,s]\times\mathbb{R}^d}\left|\frac{\Phi^{h+h_0}_{r,y} - \Phi^h_{r,y} - \bar{D}\Phi^h_{r,y}(h_0)}{\|h_0\|_{\mathcal{H}_T}}\right|^2 J_1(t-s)\mathrm{d}s.
\end{aligned}$$

A similar estimate, with $J_1(t-s)$ replaced by $J_2(t-s)$ holds for the last term on the right-hand side of (3.14).

Summarizing, we have proved that

$$\begin{aligned}
&\left|\frac{\Phi^{h+h_0}_{t,x} - \Phi^h_{t,x} - \bar{D}\Phi^h_{t,x}(h_0)}{\|h_0\|_{\mathcal{H}_T}}\right|^2\\
&\qquad\le C_{h,h_0}\|h_0\|^2_{\mathcal{H}_T} + C_h\int_0^t \sup_{(r,y)\in[0,s]\times\mathbb{R}^d}\left|\frac{\Phi^{h+h_0}_{r,y} - \Phi^h_{r,y} - \bar{D}\Phi^h_{r,y}(h_0)}{\|h_0\|_{\mathcal{H}_T}}\right|^2\\
&\qquad\quad\times\big(J_1(t-s) + J_2(t-s)\big)\mathrm{d}s.
\end{aligned}$$

Since for any $h \in \mathcal{H}_T$, $\sup_{\|h_0\|_{\mathcal{H}_T}\le 1} C_{h,h_0} < \infty$, by Gronwall's Lemma we conclude

$$\lim_{\|h_0\|_{\mathcal{H}_T}\to 0}\left|\frac{\Phi^{h+h_0}_{t,x} - \Phi^h_{t,x} - \bar{D}\Phi^h_{t,x}(h_0)}{\|h_0\|_{\mathcal{H}_T}}\right| = 0.$$

This ends the proof of the Lemma. □

Remark 3.6 Assume (A4). By further differentiating the term $\Xi^h(t,x)$ in (3.13) and repeating the calculation involving the definition of higher-order Fréchet differentiability, it can be shown that $\Phi_{t,x}$ is Fréchet differentiable of any order.

Now we are in position to check (3.1).

Lemma 3.7 *Fix* $(t,x) \in [0,T]\times\mathbb{R}^d$ *and assume* (A1), (A2), (A3) $\sigma, b \in \mathcal{C}^1$ *with Lipschitz continuous and bounded derivatives, and* (A5). *Then, for all* $h \in \mathcal{H}_T$, (3.1) *holds with* $F^\varepsilon = u^\varepsilon(t,x)$, $\psi(h) = \Phi^h_{t,x}$ *and* $\mathcal{N}_{t,x}(h)$ *given by the SPDE*

$$
\begin{aligned}
\mathcal{N}_{t,x}(h) = & \int_0^t \int_{\mathbb{R}^d} \Lambda(t-s, x-y)\sigma(\Phi^h_{s,y}) M(\mathrm{d}s, \mathrm{d}y) \\
& + \langle \Lambda(t-\cdot, x-*)\sigma'(\Phi^h_{\cdot,*})\mathcal{N}_{\cdot,*}(h), h\rangle_{\mathcal{H}_T} \\
& + \int_0^t \int_{\mathbb{R}^d} \Lambda(t-s, x-y) b'(\Phi^h_{s,y})\mathcal{N}_{s,y}(h)\mathrm{d}y\mathrm{d}s. \qquad (3.15)
\end{aligned}
$$

Proof First we note that by Theorem A.1 there exists a unique solution to (3.15). The integrand in the stochastic integral term of this equation is deterministic, consequently, the random variable $\mathcal{N}_{t,x}(h)$ is Gaussian and therefore it belongs to the first Wiener chaos. Its variance is given by $\|D\mathcal{N}_{t,x}(h)\|^2_{\mathcal{H}_T}$, where D denotes the Malliavin derivative.

The Malliavin derivative of $\mathcal{N}_{t,x}(h)$ satisfies the equation

$$
\begin{aligned}
D\mathcal{N}_{t,x}(h) = & \Lambda(t-\cdot, x-*)\sigma(\Phi^h_{\cdot,*}) + \langle \Lambda(t-\cdot, x-*)\sigma'(\Phi^h_{\cdot,*})D\mathcal{N}_{\cdot,*}(h), h\rangle_{\mathcal{H}_T} \\
& + \int_0^t \int_{\mathbb{R}^d} \Lambda(t-s, x-y) b'(\Phi^h_{s,y}) D\mathcal{N}_{s,y}(h)\mathrm{d}y\mathrm{d}s. \qquad (3.16)
\end{aligned}
$$

Comparing this equation with the one for $\bar{D}\Phi^h_{t,x}$ in (3.12), (3.13) and invoking the uniqueness of solution, we see that, for any $h \in \mathcal{H}_T$, the $\mathcal{H}_T$-valued stochastic processes $\{D\mathcal{N}_{t,x}(h), (t,x) \in [0,T] \times \mathbb{R}^d\}$ and $\{\bar{D}\Phi^h_{t,x}, (t,x) \in [0,T] \times \mathbb{R}^d\}$ are indistinguishable. In particular, the variance of $\mathcal{N}_{t,x}(h)$ is $\|\bar{D}\Phi^h_{t,x}\|^2_{\mathcal{H}_T}$.

Set $u^{\varepsilon,h}(t,x) := u(t,x;\omega + \varepsilon^{-1}h)$. According to Lemma A.2, the process $(u^{\varepsilon,h}(t,x), (t,x) \in [0,T] \times \mathbb{R}^d)$ satisfies (A.6). By uniqueness of solution we clearly have $u^{\varepsilon,0}(t,x) = u^{\varepsilon}(t,x)$ and $u^{0,h}(t,x) = \lim_{\varepsilon \downarrow 0} u^{\varepsilon,h}(t,x) = \Phi^h_{t,x}$, for any $(t,x) \in [0,T] \times \mathbb{R}^d$.

Next, we prove in our context the convergence (3.1) in $L^p(\Omega)$ norm, for $p \in [2,\infty)$. Set

$$
Z^{\varepsilon,h}_{t,x} = \varepsilon^{-1}(u^{\varepsilon,h}(t,x) - \Phi^h_{t,x}) - \mathcal{N}_{t,x}(h).
$$

By using the equations satisfied by each one of the terms on the right hand-side of that expression, we see that

$$
\mathbb{E}\big[|Z^{\varepsilon,h}_{t,x}|^p\big] \le C \sum_{i=1}^3 T^{\varepsilon,h,i}_{t,x},
$$

where

$$T^{\varepsilon,h,1}_{t,x} = \mathbb{E}\Big[\Big|\int_0^t\int_{\mathbb{R}^d} \Lambda(t-s,x-z)\big(\sigma(u^{\varepsilon,h}(s,z)) - \sigma(\Phi^h_{s,z})\big)M(\mathrm{d}s,\mathrm{d}z)\Big|^p\Big],$$

$$T^{\varepsilon,h,2}_{t,x} = \mathbb{E}\Big[\Big|\Big\langle \Lambda(t-\cdot,x-*)\Big(\frac{\sigma(u^{\varepsilon,h}(\cdot,*)) - \sigma(\Phi^h_{\cdot,*})}{\varepsilon} - \sigma'(\Phi^h_{\cdot,*})\mathcal{N}_{\cdot,*}(h)\Big),h\Big\rangle_{\mathcal{H}_T}\Big|^p\Big],$$

$$T^{\varepsilon,h,3}_{t,x} = \mathbb{E}\Big[\Big|\int_0^t\int_{\mathbb{R}^d} \Lambda(t-s,x-z)\Big(\frac{b(u^{\varepsilon,h}(s,z)) - b(\Phi^h_{s,z})}{\varepsilon} - b'(\Phi^h_{s,z})\mathcal{N}_{s,z}(h)\Big)\mathrm{d}z\mathrm{d}s\Big|^p\Big].$$

We will prove that each one of these terms tends to zero as $\varepsilon \downarrow 0$.

By the usual estimates on moments of stochastic and pathwise integrals, we have

$$\begin{aligned}
&\sup_{(r,y)\in[0,t]\times\mathbb{R}^d} \mathbb{E}\big[|u^{\varepsilon,h}(r,y) - \Phi^h_{r,y}|^p\big] \\
&\quad\le C\varepsilon^p\Big(1 + \sup_{\varepsilon\in(0,1]}\sup_{(t,x)\in[0,T]\times\mathbb{R}^d} \mathbb{E}\big[|u^{\varepsilon,h}(t,x)|^p\big]\Big) \\
&\qquad + C\int_0^t \sup_{(r,y)\in[0,s]\times\mathbb{R}^d} \mathbb{E}\big[|u^{\varepsilon,h}(r,y) - \Phi^h_{r,y}|^p\big]\big(J_1(t-s) + J_2(t-s)\big)\mathrm{d}s.
\end{aligned}$$

By Gronwall's Lemma this yields

$$\lim_{\varepsilon\downarrow 0}\Big(\sup_{(t,x)\in[0,T]\times\mathbb{R}^d} \mathbb{E}\big[|u^{\varepsilon,h}(t,x) - \Phi^h_{t,x}|^p\big]\Big) = 0. \tag{3.17}$$

Since

$$T^{\varepsilon,h,1}_{t,x} \le C \sup_{(t,x)\in[0,T]\times\mathbb{R}^d} \mathbb{E}\big[|u^{\varepsilon,h}(t,x) - \Phi^h_{t,x}|^p\big],$$

we deduce

$$\lim_{\varepsilon\downarrow 0}\Big(\sup_{(t,x)\in[0,T]\times\mathbb{R}^d} T^{\varepsilon,h,1}_{t,x}\Big) = 0. \tag{3.18}$$

Next, we deal with the term $T^{\varepsilon,h,2}_{t,x}$. Cauchy-Schwarz's inequality and the mean-value theorem applied to σ yield

$$\begin{aligned}
&\mathbb{E}\Big[\Big|\Big\langle \Lambda(t-\cdot,x-*)\Big(\frac{\sigma(u^{\varepsilon,h}(\cdot,*)) - \sigma(\Phi^h_{\cdot,*})}{\varepsilon} - \sigma'(\Phi^h_{\cdot,*})\mathcal{N}_{\cdot,*}(h)\Big),h\Big\rangle_{\mathcal{H}_T}\Big|^p\Big] \\
&\le C\|h\|^p_{\mathcal{H}_T} \\
&\quad\times \int_0^t \sup_{(r,y)\in[0,s]\times\mathbb{R}^d} \mathbb{E}\Big[\Big|\frac{\sigma(u^{\varepsilon,h}(r,y)) - \sigma(\Phi^h_{r,y})}{\varepsilon} - \sigma'(\Phi^h_{r,y})\mathcal{N}_{r,y}(h)\Big|^p\Big]J_1(t-s)\mathrm{d}s
\end{aligned}$$

$$\leq C\left\{\int_0^t \sup_{(r,y)\in[0,s]\times\mathbb{R}^d} \mathbb{E}\left[\left|\sigma'(\xi^{\varepsilon,h}_{r,y})\left(\frac{u^{\varepsilon,h}(r,y)-\Phi^h_{r,y}}{\varepsilon}-\mathcal{N}_{r,y}(h)\right)\right|^p\right]J_1(t-s)\mathrm{d}s\right.$$
$$\left.+\sup_{(t,x)\in[0,T]\times\mathbb{R}^d}\mathbb{E}\left[\left|\left(\sigma'(\Phi^h_{t,x})-\sigma'(\xi^{\varepsilon,h}_{t,x})\right)\mathcal{N}_{t,x}(h)\right|^p\right]\int_0^t J_1(t-s)\mathrm{d}s\right\}, \tag{3.19}$$

where $\xi^{\varepsilon,h}_{r,y}(\omega)$ is a point lying in the open interval determined by $\Phi^h_{r,y}$ and $u^{\varepsilon,h}(r,y;\omega)$.

From Theorem A.1 and the Lipschitz continuity of σ' we have

$$\sup_{(t,x)\in[0,T]\times\mathbb{R}^d}\mathbb{E}\left[\left|\left(\sigma'(\Phi^h_{t,x})-\sigma'(\xi^{\varepsilon,h}_{t,x})\right)\mathcal{N}_{t,x}(h)\right|^p\right]$$
$$\leq \sup_{(t,x)\in[0,T]\times\mathbb{R}^d}\left(\mathbb{E}\left[\left|\sigma'(\Phi^h_{t,x})-\sigma'(\xi^{\varepsilon,h}_{t,x})\right|^{2p}\right]\right)^{\frac{1}{2}}\sup_{(t,x)\in[0,T]\times\mathbb{R}^d}\left(\mathbb{E}\left[\left|\mathcal{N}_{t,x}(h)\right|^{2p}\right]\right)^{\frac{1}{2}}$$
$$\leq C\sup_{(t,x)\in[0,T]\times\mathbb{R}^d}\left(\mathbb{E}\left[\left|\Phi^h_{t,x}-u^{\varepsilon,h}_{t,x}\right|^{2p}\right]\right)^{1/2}.$$

Consequently,

$$T^{\varepsilon,h,2}_{t,x}\leq C\sup_{(t,x)\in[0,T]\times\mathbb{R}^d}\left(\mathbb{E}\left[\left|\Phi^h_{t,x}-u^{\varepsilon,h}_{t,x}\right|^{2p}\right]\right)^{1/2}$$
$$+C\int_0^t\sup_{(r,y)\in[0,s]\times\mathbb{R}^d}\mathbb{E}\left[\left|\frac{u^{\varepsilon,h}(r,y)-\Phi^h_{r,y}}{\varepsilon}-\mathcal{N}_{r,y}(h)\right|^p\right]J_1(t-s)\mathrm{d}s, \tag{3.20}$$

With similar arguments, one can check that

$$T^{\varepsilon,h,3}_{t,x}\leq C\sup_{(t,x)\in[0,T]\times\mathbb{R}^d}\left(\mathbb{E}\left[\left|\Phi^h_{t,x}-u^{\varepsilon,h}_{t,x}\right|^{2p}\right]\right)^{1/2}$$
$$+C\int_0^t\sup_{(r,y)\in[0,s]\times\mathbb{R}^d}\mathbb{E}\left[\left|\frac{u^{\varepsilon,h}(r,y)-\Phi^h_{r,y}}{\varepsilon}-\mathcal{N}_{r,y}(h)\right|^p\right]J_2(t-s)\mathrm{d}s, \tag{3.21}$$

Thus from (3.18), (3.20), (3.21) it follows that

$$\sup_{(r,y)\in[0,t]\times\mathbb{R}^d} \mathbb{E}\big[|Z^{\varepsilon,h}_{r,y}|^p\big]$$
$$\leq C_\varepsilon + C\int_0^t \sup_{(r,y)\in[0,s]\times\mathbb{R}^d} \mathbb{E}\big[|Z^{\varepsilon,h}_{r,y}|^p\big]\big(J_1(t-s)+J_2(t-s)\big)\mathrm{d}s,$$

where C_ε converges to zero as $\varepsilon \downarrow 0$. Applying Gronwall's Lemma we see that $Z^{\varepsilon,h}_{t,x}$ converges to zero in L^p as $\varepsilon \downarrow 0$ for all $h \in \mathcal{H}_T$, uniformly in $(t,x) \in [0,T]\times\mathbb{R}^d$.

The next step consists of proving the convergence to zero of $Z^{\varepsilon,h}_{t,x}$ in the $\mathbb{D}^{1,p}$ norm, for any $p \in [2,\infty)$. Since $\Phi^h_{t,x}$ is deterministic this reduces to show that $\varepsilon^{-1}Du^\varepsilon(t,x;\omega+h) - D\mathcal{N}(h)$ converges to zero as $\varepsilon \downarrow 0$ in $L^p(\Omega;\mathcal{H}_T)$.

By applying the Malliavin derivative operator to Eq. (A.6), one can show that the process $Du^{\varepsilon,h}(t,x)$ satisfies the SPDE

$$\begin{aligned} Du^{\varepsilon,h}(t,x) &= \varepsilon\Lambda(t-\cdot,x-*)\sigma(u^{\varepsilon,h}(\cdot,*)) \\ &\quad + \varepsilon\int_0^t\int_{\mathbb{R}^d} \Lambda(t-s,x-z)\sigma'(u^{\varepsilon,h}(s,z))Du^{\varepsilon,h}(s,z)M(\mathrm{d}s,\mathrm{d}z) \\ &\quad + \langle\Lambda(t-\cdot,x-*)\sigma'(u^{\varepsilon,h}(\cdot,*))Du^{\varepsilon,h}(\cdot,*),h\rangle_{\mathcal{H}_T} \\ &\quad + \int_0^t\int_{\mathbb{R}^d} \Lambda(t-s,x-z)b'(u^{\varepsilon,h}(s,z))Du^{\varepsilon,h}(s,z)\mathrm{d}z\mathrm{d}s. \end{aligned} \tag{3.22}$$

For its further use, we remark that

$$\lim_{\varepsilon\downarrow 0}\left(\sup_{(t,x)\in[0,T]\times\mathbb{R}^d} \mathbb{E}\big[\|Du^{\varepsilon,h}(t,x)\|^p_{\mathcal{H}_T}\big]\right) = 0. \tag{3.23}$$

Indeed, this follows from the estimate

$$\begin{aligned} \mathbb{E}\big[\|Du^{\varepsilon,h}(t,x)\|^p_{\mathcal{H}_T}\big] &\leq C\varepsilon^p\left(1+\sup_{\varepsilon\in(0,1]}\sup_{(t,x)\in[0,T]\times\mathbb{R}^d}\mathbb{E}\big[|u^{\varepsilon,h}(t,x)|^p\big]\right) \\ &\quad + C\int_0^t \sup_{\varepsilon\in(0,1]}\sup_{(r,y)\in[0,s]\times\mathbb{R}^d}\mathbb{E}\big[\|Du^{\varepsilon,h}(r,y)\|^p_{\mathcal{H}_T}\big] \\ &\quad \times \big(J_1(t-s)+J_2(t-s)\big)\mathrm{d}s, \end{aligned}$$

along with Gronwall's lemma.

By (3.16) and (3.22) we easily obtain

$$
\begin{aligned}
&\mathbb{E}\big[\|\varepsilon^{-1}Du^{\varepsilon,h}(t,x)-D\mathcal{N}_{t,x}(h)\|^p_{\mathcal{H}_T}\big]\\
&\quad\le C\mathbb{E}\big[\big\|\Lambda(t-\cdot,x-*)\big(\sigma(u^{\varepsilon,h}(\cdot,*))-\sigma(\Phi^h_{\cdot,*})\big)\big\|^p_{\mathcal{H}_T}\big]\\
&\qquad+C\mathbb{E}\Bigg[\Bigg\|\int_0^t\int_{\mathbb{R}^d}\Lambda(t-s,x-z)\sigma'(u^{\varepsilon,h}(s,z))Du^{\varepsilon,h}(s,z)M(\mathrm{d}s,\mathrm{d}z)\Bigg\|^p_{\mathcal{H}_T}\Bigg]\\
&\qquad+C\mathbb{E}\Bigg[\Bigg\|\big\langle\Lambda(t-\cdot,x-*)\big(\varepsilon^{-1}\sigma'(u^{\varepsilon,h}(\cdot,*))Du^{\varepsilon,h}(\cdot,*)-\sigma'(\Phi^h_{\cdot,*})D\mathcal{N}_{\cdot,*}(h)\big),h\big\rangle_{\mathcal{H}_T}\Bigg\|^p_{\mathcal{H}_T}\Bigg]\\
&\qquad+C\mathbb{E}\Bigg[\Bigg\|\int_0^t\int_{\mathbb{R}^d}\Lambda(t-s,x-z)\big(\varepsilon^{-1}b'(u^{\varepsilon,h}(s,z))Du^{\varepsilon,h}(s,z)-b'(\Phi^h_{s,z})\\
&\qquad\qquad\times D\mathcal{N}_{s,z}(h)\big)\mathrm{d}z\mathrm{d}s\Bigg\|^p_{\mathcal{H}_T}\Bigg].
\end{aligned}
\tag{3.24}
$$

Each term on the right-hand side of (3.24) converges to zero as $\varepsilon\downarrow 0$. Indeed, for the first and second terms, this is a consequence of (3.17) and (3.23), respectively. For the analysis of the last two ones, we use the argument involving the mean-value Theorem as in (3.19). Then Gronwall's Lemma yields the assertion.

In order to finish the proof, we must check the convergence of $Z^{\varepsilon,h}_{t,x}$ in the $\mathbb{D}^{k,p}$ norm, for any $k\ge 2$, $p\in[2,\infty)$. Since $\mathcal{N}_{t,x}(\bar h)$ is a Gaussian random variable, this reduces to show that $\varepsilon^{-1}D^k u^{\varepsilon,h}(t,x)$ converges to zero in $L^p(\Omega;\mathcal{H}_T{}^{\otimes k})$, which is proved recursively on $k\ge 2$. We leave the details to the reader. □

Thanks to Proposition 3.2, the results proved so far establish the lower bound in (2.4) with

$$
I(y)=\inf\left\{\frac{1}{2}\|h\|^2_{\mathcal{H}_T};\ h\in\mathcal{H}_T,\,\Phi^h_{t,x}=y,\,\bar\gamma_{\Phi^h_{t,x}}>0\right\}.
$$

In the next lemma it is shown that under the standing assumptions, the condition $\bar\gamma_\psi(h)>0$ is satisfied. Hence, $I(y)$ is as in (2.5).

Lemma 3.8 *Fix $(t,x)\in(0,T]\times\mathbb{R}^d$ and assume* (A1), (A2), (A3), *$\sigma,b\in\mathcal{C}^1$ with bounded derivatives and* (A5). *Then $\bar\gamma_{\Phi^h_{t,x}}>0$ for all $h\in\mathcal{H}_T$.*

Proof The proof follows the same strategy as in Lemma 3.4, the difference being that here we use deterministic arguments.

Fix $h\in\mathcal{H}_T$ and let $0<\rho<t\wedge t_0$, with t_0 as in (A3). Remember that $\bar\gamma_{\Phi^h_{t,x}}=\|\bar D\Phi^h_{t,x}\|^2_{\mathcal{H}_T}$, where $\bar D$ stands for the Fréchet derivative. Using (3.12) and (3.13) we clearly obtain

$$
\bar\gamma_{\Phi^h_{t,x}}=\|\Xi^h_{\circ,\bullet}(t,x)\|^2_{\mathcal{H}_T}\ge\|\Xi^h_{r,\bullet}(t,x)\|^2_{\mathcal{H}_{t-\rho,t}}\ge\frac{1}{2}A^1_{t,x}(\rho)-A^2_{t,x}(\rho),
$$

with

$$A^1_{t,x}(\rho) = \|\Lambda(t-\circ, x-\bullet)\sigma(\Phi^h_{\circ,\bullet})\|^2_{\mathcal{H}_{t-\rho,t}},$$
$$A^2_{t,x}(\rho) = \|\chi_{\circ,\bullet}(t,x)\|^2_{\mathcal{H}_{t-\rho,t}}$$

and

$$\chi_{\circ,\bullet}(t,x) = \Xi^h_{\circ,\bullet}(t,x) - \Lambda(t-\circ, x-\bullet)\sigma(\Phi^h_{\circ,\bullet}).$$

Using (A3), we have

$$A^1_{t,x}(\rho) \geq \sigma_0^2 g_1(\rho), \tag{3.25}$$

and this bound is uniform in $(t,x) \in [0,T] \times \mathbb{R}^d$ and in $h \in \mathcal{H}_T$.

Our next aim is to prove that there exists $\zeta > 0$ and $0 < \rho < t \wedge t_0$ such that $\frac{\sigma_0^2}{2} g_1(\rho) - A^2_{t,x}(\rho) \geq \zeta$, or equivalently,

$$\left(\frac{\sigma_0^2}{2} g_1(\rho) - \zeta\right)^{-1} A^2_{t,x}(\rho) \leq 1. \tag{3.26}$$

For this, we will find a suitable upper bound for $A^2_{t,x}(\rho)$. By using the definition of $\chi_{\circ,\bullet}(t,x)$ and (3.13), we have

$$A^2_{t,x}(\rho) \leq C\left(N^1_{t,x}(\rho) + N^2_{t,x}(\rho)\right),$$

with

$$N^1_{t,x}(\rho) = \left\|\left\langle \Lambda(t-\cdot, x-*)\sigma'(\Phi^h_{\cdot,*})\,\Xi^h_{\circ,\bullet}(\cdot,*), h\right\rangle_{\mathcal{H}_T}\right\|^2_{\mathcal{H}_{t-\rho,t}},$$
$$N^2_{t,x}(\rho) = \left\|\int_0^t \int_{\mathbb{R}^d} \Lambda(t-s, x-z) b'(\Phi^h_{s,z})\,\Xi^h_{\circ,\bullet}(s,z)\mathrm{d}z\mathrm{d}s\right\|^2_{\mathcal{H}_{t-\rho,t}}.$$

Remember that, similarly as in (3.13), $(\circ,\bullet)$ is the argument in $[t-\rho,t]\times\mathbb{R}^d$ relevant for the $\mathcal{H}_{t-\rho,t}$ norm, while $(\cdot,*)$ interacts with h in the $\mathcal{H}_T$ norm.

From the definition of $\Xi(s,y)$ given in (3.13) and Proposition A.3, we deduce the estimate

$$\sup_{y\in\mathbb{R}^d} \|\Xi^h_{\circ,\bullet}(s,y)\|^2_{\mathcal{H}_{s-\rho,s}} \leq C g_1(\rho).$$

By applying first Schwarz's inequality to the inner product in $\mathcal{H}_T$, the preceding estimate yields

$$N^1_{t,x}(\rho) \leq C(g_1(\rho))^2. \tag{3.27}$$

(See (3.7) for an analogous result).

By similar arguments, we have

$$N^2_{t,x}(\rho) \le C\rho^{2\delta} g_1(\rho). \tag{3.28}$$

(Notice the analogy with (3.8)).

With (3.27), (3.28), we see that the left-hand side of (3.26) is bounded by

$$C\left(\frac{\sigma_0^2}{2} g_1(\rho) - \zeta\right)^{-1} \left(g_1(\rho)^2 + \rho^{2\delta} g_1(\rho)\right).$$

Fix $\rho = \rho(\zeta) > 0$ such that $g_1(\rho) = \frac{4}{\sigma_0^2}\zeta$ which by (A3) implies that $\rho \le C\zeta^{1/\gamma}$. Then the previous expression is bounded by $C(\zeta + \zeta^{\frac{2\delta}{\gamma}})$. Hence, (3.26) holds for a suitable choice of $\zeta > 0$. This ends the proof. □

The results established so far show that Theorem 2.4(ii) holds for the set $y \in \mathbb{R}$ such that $p^{\varepsilon}_{t,x}(y) > 0$ for all ε small enough. The next objective is to analyze when this condition is satisfied and also whether the function I defined in (2.4) is finite. Both questions are related to the characterisation of the topological support of the law of the random variable $u^{\varepsilon}(t,x)$, with $\varepsilon \in (0,1]$ and $(t,x) \in (0,T] \times \mathbb{R}^d$ fixed.

Under suitable conditions, we prove in Theorem 4.1 that the support of $u^{\varepsilon}(t,x)$ does not depend on the parameter ε and is given by

$$\mathcal{S} := \mathrm{supp}(P \circ [u^{\varepsilon}(t,x)]^{-1}) = \overline{\{\Phi^h_{t,x};\ h \in \mathcal{H}_T\}}.$$

In particular, $\mathcal{S}$ is the topological support of the law of the random variable $u^1(t,x)$, that we denote by $u(t,x)$.

Since for any $p \in [2,\infty)$, $u^{\varepsilon}(t,x) \in \mathbb{D}^{1,p}$ (see Lemma 3.3), we can apply Fang's result quoted in [24, Proposition 4.1.1] to deduce that $\mathcal{S}$ is a closed interval. Moreover, applying [24, Proposition 4.1.2], we obtain that for all points in the interior of $\mathcal{S}$, denoted by $\mathring{\mathcal{S}}$, we have $p^{\varepsilon}_{t,x}(y) > 0$. Therefore, $\log p^{\epsilon}_{t,x}(y)$ is well-defined for all $y \in \mathring{\mathcal{S}}$. Notice that $\mathring{\mathcal{S}} \ne \emptyset$.

The next statements provide results on the finiteness of I, defined in (2.5).

Proposition 3.9 *The hypotheses are the same as in Theorem 4.1. In addition, we suppose that $\sigma, b \in \mathcal{C}^1$ are Lipschitz continuous and have bounded derivatives. Then, for all $z \in \mathring{\mathcal{S}}$, $I(z) < \infty$.*

Proof Let $z \in \mathring{\mathcal{S}}$ and $\rho^z := \mathrm{dist}(z, \partial\mathcal{S})$. Define $z_1^* := z - \rho^z/2$ and $z_2^* := z + \rho^z/2$. Since the set $S_1 := \{\Phi^h_{t,x};\ h \in \mathcal{H}_T\}$ is dense in $\mathcal{S}$, there exists $z_1 \in S_1 \cap B_{\rho^z/4}(z_1^*)$ and $z_2 \in S_1 \cap B_{\rho^z/4}(z_2^*)$. By definition of I in (2.5), $I(z_1)$ and $I(z_2)$ are finite. The function $h \mapsto \Phi^h_{t,x}$ is continuous (see Lemma 3.5). Hence, by the intermediate value theorem [23, Theorem 24.3], we conclude that for all $\bar{z} \in (z_1, z_2)$ there exists an $h_{\bar{z}} \in \mathcal{H}_T$ such that $\Phi^{h_{\bar{z}}}_{t,x} = \bar{z}$ and therefore $I(z) < \infty$. □

If the function b is bounded, one can show that $\{z \in \mathbb{R};\ I(z) < \infty\} = \mathbb{R}$. Therefore $\mathrm{supp}(P \circ [u^\varepsilon(t,x)]^{-1}) = \mathbb{R}$, $p^\varepsilon_{t,x}(y) > 0$ for all $y \in \mathbb{R}$ and (2.4) holds for any $y \in \mathbb{R}$. This is a consequence of the following Proposition.

Proposition 3.10 *Assume* (A1), (A2), (A5)*, and that σ and b are Lipschitz continuous. Suppose also that b is bounded. Then $\{z \in \mathbb{R};\ I(z) < \infty\} = \mathbb{R}$.*

Proof Fix $(t,x) \in (0,T] \times \mathbb{R}^d$. Owing to (A1), for every $h \in \mathcal{H}_T$,

$$\left| \int_0^t \int_{\mathbb{R}^d} \Lambda(t-s, x-z) b(\Phi^h_{s,z}) \mathrm{d}z \mathrm{d}s \right| \le |b|_\infty \int_0^t \Lambda(s)(\mathbb{R}^d) \mathrm{d}s =: I_2. \tag{3.29}$$

Note that this bound does not depend on $h \in \mathcal{H}_T$. Moreover, (A5) and (A1) imply

$$I_1 := \sigma_0^2 \|\Lambda(t-\cdot, x-*)\|^2_{\mathcal{H}_T} < \infty. \tag{3.30}$$

Fix $\alpha > 0$, $z \in \mathbb{R}$, $h \in \mathcal{H}_T$, and set

$$h_{z,\alpha}(\cdot,*) := \frac{|z| + \alpha + I_2 + |w(t,x)|}{I_1} \Lambda(t-\cdot, x-*)\sigma(\Phi^h_{\cdot,*}).$$

Using (3.10) one can easily check that $h_{z,\alpha} \in \mathcal{H}_T$. From (3.29), (3.30), along with (A5), we obtain

$$\begin{aligned}
\Phi^{h_{z,\alpha}}_{t,x} &= w(t,x) + \frac{|z| + \alpha + I_2 + |w(t,x)|}{I_1} \\
&\qquad \times \langle \Lambda(t-\cdot, x-*)\sigma(\Phi^{h_{z,\alpha}}_{\cdot,*}), \Lambda(t-\cdot, x-*)\sigma(\Phi^h_{\cdot,*})\rangle_{\mathcal{H}_T} \\
&\quad + \int_0^t \int_{\mathbb{R}^d} \Lambda(t-s, x-z) b(\Phi^{h_{z,\alpha}}_{s,z}) \mathrm{d}z \mathrm{d}s \\
&\ge -|w(t,x)| + \frac{|z| + \alpha + I_2 + |w(t,x)|}{I_1} I_1 - I_2 \\
&> z,
\end{aligned}$$

and similarly,

$$\begin{aligned}
\Phi^{-h_{z,\alpha}}_{t,x} &= w(t,x) - \frac{|z| + \alpha + I_2 + |w(t,x)|}{I_1} \\
&\qquad \times \langle \Lambda(t-\cdot, x-*)\sigma(\Phi^{-h_{z,\alpha}}_{\cdot,*}), \Lambda(t-\cdot, x-*)\sigma(\Phi^h_{\cdot,*})\rangle_{\mathcal{H}_T} \\
&\quad + \int_0^t \int_{\mathbb{R}^d} \Lambda(t-s, x-z) b(\Phi^{-h_{z,\alpha}}_{s,z}) \mathrm{d}z \mathrm{d}t
\end{aligned}$$

$$\leq |w(t,x)| - \frac{|z| + \alpha + I_2 + |w(t,x)|}{I_1} I_1 + I_2$$
$$< z.$$

Thus, for all $z \in \mathbb{R}$ there exists $h_{z,\alpha} \in \mathcal{H}_T$ such that $\Phi_{t,x}^{-h_{z,\alpha}} < z < \Phi_{t,x}^{h_{z,\alpha}}$. By the intermediate value theorem [23, Theorem 24.3] together with Lemma 3.5, there exists some $h_z \in \mathcal{H}_T$ such that $\Phi_{t,x}^{h_z} = z$. This finishes the proof. □

Assume as in the previous proposition that (A1), (A2), (A5) hold and that σ and b are Lipschitz continuous. Suppose that σ is bounded. Then, for all $y \in \mathbb{R}$, $p_{t,x}(y) > 0$ (see [27, Theorem 5.1]) and therefore $\mathcal{S} = \mathbb{R}$.

4 Support Theorem

In this section we prove a characterization of the topological support of the probability law of the random variable $u^\varepsilon(t,x)$, for a fixed $(t,x) \in (0,T] \times \mathbb{R}^d$, defined by (2.1). This is the smallest closed subset $\mathcal{X}^\varepsilon \subseteq \mathbb{R}$ satisfying $\mathbb{P} \circ (u^\varepsilon(t,x))^{-1}(\mathcal{X}^\varepsilon) = 1$. Under stronger assumptions than in Theorem 2.2, we prove in Theorem 4.1 that $\mathcal{X}^\varepsilon = \overline{\{\Phi_{t,x}^h; h \in \mathcal{H}_T\}}$ and therefore also that $\mathcal{X}^\varepsilon$ does not depend on ε. This will be a consequence of two approximation results, as follows.

Fix $\varepsilon \in (0,1]$ and consider a sequence $(v^{n,\varepsilon})_{n\in\mathbb{N}}$ of $\mathcal{H}_T$-valued random variables such that

(C1) $\lim_{n\to\infty} \mathbb{P}\big[\big|u^\varepsilon(t,x) - \Phi_{t,x}^{v^{n,\varepsilon}}\big| > \eta\big] = 0$,
(C2) for any $h \in \mathcal{H}_T$, $\lim_{n\to\infty} \mathbb{P}\big[\big|u(t,x;\omega - v^{n,\varepsilon} + h) - \Phi_{t,x}^h\big| > \eta\big] = 0$,

for any $\eta > 0$.

By Portmanteau's Theorem, (C1) implies that $\mathcal{X}^\varepsilon \subseteq \overline{\{\Phi_{t,x}^h; h \in \mathcal{H}_T\}}$. From (C2) together with Girsanov's theorem, we deduce the converse inclusion $\mathcal{X}^\varepsilon \supseteq \overline{\{\Phi_{t,x}^h; h \in \mathcal{H}_T\}}$. Without loss of generality, we may assume that $\varepsilon = 1$ and write $u(t,x)$ and v^n instead of $u^\varepsilon(t,x)$ and $v^{n,\varepsilon}$, respectively.

It is easy to see that both convergences (C1) and (C2) (with $\varepsilon = 1$) can be formally derived from a single convergence result. Indeed, let $A, B, G, b : \mathbb{R} \to \mathbb{R}$, $w : [0,T] \times \mathbb{R}^d \to \mathbb{R}$ and $h \in \mathcal{H}_T$. We consider the SPDEs

$$\begin{aligned} X(t,x) = w(t,x) &+ \int_0^t \int_{\mathbb{R}^d} \Lambda(t-s, x-y)(A+B)(X(s,y))M(\mathrm{d}s,\mathrm{d}y) \\ &+ \langle \Lambda(t-\cdot, x-*)G(X(\cdot,*)), h\rangle_{\mathcal{H}_T} \\ &+ \int_0^t \int_{\mathbb{R}^d} \Lambda(t-s,x-y)b(X(s,y))\mathrm{d}y\mathrm{d}s, \end{aligned} \tag{4.1}$$

and, for all $n \in \mathbb{N}$

$$\begin{aligned} X_n(t,x) = w(t,x) &+ \int_0^t \int_{\mathbb{R}^d} \Lambda(t-s,x-y)A(X_n(s,y))M(\mathrm{d}s,\mathrm{d}y) \\ &+ \langle \Lambda(t-\cdot,x-*)B(X_n(\cdot,*)), v^n\rangle_{\mathcal{H}_T} \\ &+ \langle \Lambda(t-\cdot,x-*)G(X_n(\cdot,*)), h\rangle_{\mathcal{H}_T} \\ &+ \int_0^t \int_{\mathbb{R}^d} \Lambda(t-s,x-y)b(X_n(s,y))\mathrm{d}y\mathrm{d}s. \end{aligned} \tag{4.2}$$

Suppose we can prove that the sequence $(X_n(t,x))_{n\in\mathbb{N}}$ converges in probability to $X(t,x)$, for fixed $(t,x) \in (0,T]\times\mathbb{R}^d$. Then with the choice $A = G = 0$ and $B = \sigma$, we obtain (C1). By taking $A = G = \sigma$ and $B = -\sigma$, we get (C2).

The sequence $(v^n)_{n\in\mathbb{N}}$ will consist of smooth approximations of the stochastic process F. As has been described in Sect. 2, F can be identified with a sequence of independent standard Brownian motions $W = \{W^k(t), t \in [0,T], k \in \mathbb{N}\}$.

Fix $n \in \mathbb{N}$ and consider the partition of $[0,T]$ determined by $\frac{iT}{2^n}$, $i = 0, 1, \dots, 2^n$. Denote by Δ_i the interval $[\frac{iT}{2^n}, \frac{(i+1)T}{2^n}[$ and by $|\Delta_i| = T2^{-n}$ its length. We write $W_j(\Delta_i)$ for the increment $W_j(\frac{(i+1)T}{2^n}) - W_j(\frac{iT}{2^n})$, $i = 0, \dots, 2^n-1$, $j \in \mathbb{N}$. Define differentiable approximations of $(W_j, j \in \mathbb{N})$ as follows:

$$W^n = \left(W_j^n = \int_0^{\cdot} \dot{W}_j^n(s)ds,\ j \in \mathbb{N} \right),$$

where for $j > n$, $\dot{W}_j^n = 0$, and for $1 \le j \le n$,

$$\dot{W}_j^n(t) = \begin{cases} \sum_{i=0}^{2^n-2} 2^n T^{-1} W_j(\Delta_i) 1_{\Delta_{i+1}}(t) & \text{if} \quad t \in [2^{-n}T, T], \\ 0 & \text{if} \quad t \in [0, 2^{-n}T[. \end{cases}$$

Then, let

$$v^n(t,x) = \sum_{j\in\mathbb{N}} \dot{W}_j^n(t) e_j(x).$$

By Theorem A.1, Eq. (4.1) has a unique random-field solution, and this solution possesses moments of any order, uniformly in (t,x). That theorem cannot be applied to Eq. (4.2), because the $\mathcal{H}_T$-valued random variable v^n does not satisfy (A.3). For this reason (but also for others that will become clear later), we fix a parameter $\theta > 0$ and introduce a localization on Ω defined by

$$L_n(t) = \left\{ \sup_{1\le j\le n} \sup_{0\le i\le \lfloor 2^n t T^{-1}-1\rfloor_+} |W_j(\Delta_i)| \le 2^{n(\theta-1)} \right\},$$

where $n \in \mathbb{N}$ and $t \in [0, T]$. Notice that $L_n(t)$ decreases with t. Similarly as in [22, Lemma 2.1], one can prove that if $\theta > \frac{1}{2}$,

$$\lim_{n\to\infty} \mathbb{P}\left(L_n(t)^c\right) = 0, \ t \in [0, T]. \tag{4.3}$$

It is easy to check that

$$\|v^n(t, *)1_{L_n(t)}\|_{\mathcal{H}} \le Cn2^{n\theta},$$

and for $0 \le t < t' \le T$,

$$\|v^n 1_{L_n(t')} 1_{[t,t']}\|_{\mathcal{H}_T} \le Cn2^{n\theta}|t - t'|^{\frac{1}{2}}. \tag{4.4}$$

On each $L_n(t)$, the assumptions of Theorem A.1 are satisfied. Thus, by localization we can prove the existence of a unique solution to (4.2), and that this solution is bounded in probability.

For the formulation of the main result, it is necessary to introduce an additional assumption:

(A7) As in (A1), the mapping $t \mapsto \Lambda(t)$ is a deterministic function with values in the space of non-negative tempered distributions with rapid decrease, and for any $t \in [0, T]$, $\Lambda(t)$ is a non-negative measure. Moreover, there exist $\eta, \delta > 0$ such that

(i) $\int_0^t \int_{\mathbb{R}^d} |\mathcal{F}\Lambda(s)(\xi)|^2 \mu(\mathrm{d}\xi)\mathrm{d}s \le Ct^\eta$, for any $t \in (0, T]$,
(ii) $\sup_{0\le s\le T} \int_{\mathbb{R}^d} |\mathcal{F}\Lambda(s)(\xi)|^2 \mu(\mathrm{d}\xi) = \sup_{0\le s\le T} J_1(s) < \infty$,
(iii) $\int_0^t \Lambda(s)(\mathbb{R}^d)\mathrm{d}s \le Ct^\delta$, for any $t \in (0, T]$.

Clearly, (A7) is stronger than (A1).

Let $\Gamma(dx) = |x|^{-\beta}dx$, $\beta \in (0, d \wedge 2)$, and therefore $\mu(\mathrm{d}\xi) = |\xi|^{-(d-\beta)}\mathrm{d}\xi$. In Sect. 5, we will see that the fundamental solution to the wave equation with $d = \{1, 2, 3\}$ satisfies (A7).

The main result of this section is the following.

Theorem 4.1 *The hypotheses are* (A2) *and* (A7). *We also suppose that σ and b are Lipschitz continuous functions. Let $u^\varepsilon(t, x)$ be the solution to* (2.1) *at a given point $(t, x) \in (0, T] \times \mathbb{R}^d$. Then the topological support of the probability law $\mathbb{P} \circ (u^\varepsilon(t, x))^{-1}$ is the closure of the set $\{\Phi^h_{t,x}; h \in \mathcal{H}_T\}$, where $\Phi^h_{t,x}$ is defined in* (2.6).

By the preceding discussion, the theorem is a corollary of the next Proposition.

Proposition 4.2 *Assume that $A, B, G, b : \mathbb{R} \to \mathbb{R}$ are Lipschitz continuous functions and that assumptions* (A2) *and* (A7) *are satisfied. Fix $(t, x) \in (0, T] \times \mathbb{R}^d$ and,* *in the definition of* $L_n(t)$*, fix* $\theta > \frac{1}{2}$ *such that* $\left(\frac{\eta}{2} - \theta + \frac{1}{2}\right) \wedge \delta - \theta > 0$*. Then, for any* $p \in [1, \infty)$*,*

$$\lim_{n\to\infty} \left\| (X_n(t,x) - X(t,x)) 1_{L_n(t)} \right\|_p = 0, \tag{4.5}$$

where $\|\cdot\|_p$ *denotes the* $L^p(\Omega)$*-norm.*

Indeed, owing to (4.3), the convergence (4.5) yields $\lim_{n\to\infty} X_n(t,x) = X(t,x)$ in probability.

The rest of the section is devoted to the proof of Proposition 4.2. First we introduce some additional notation.

For any $n \in \mathbb{N}$, $t \in [0, T]$, set

$$\underline{t}_n = \max\left\{k2^{-n}T, k = 0, \ldots, 2^n - 1 : k2^{-n}T < t\right\},$$

and define $t_n = \max\{\underline{t}_n - 2^{-n}T, 0\}$. To strengthen the $\mathcal{F}_t$-measurability properties of $X(t,x)$ and $X_n(t,x)$, we consider stochastic processes defined by a modification of Eqs. (4.1), (4.2), respectively, as follows:

$$\begin{aligned} X(t, t_n, x) = {} & w(t,x) + \int_0^{t_n}\int_{\mathbb{R}^d} \Lambda(t-s, x-y)(A+B)(X(s,y))M(\mathrm{d}s, \mathrm{d}y) \\ & + \langle \Lambda(t-\cdot, x-*)G(X(\cdot,*))1_{[0,t_n]}(\cdot), h\rangle_{\mathcal{H}_T} \\ & + \int_0^{t_n}\int_{\mathbb{R}^d} \Lambda(t-s, x-y) b(X(s,y))\mathrm{d}y\mathrm{d}s, \end{aligned} \tag{4.6}$$

and

$$\begin{aligned} X_n^-(t,x) = {} & w(t,x) + \int_0^{t_n}\int_{\mathbb{R}^d} \Lambda(t-s, x-y)A(X_n(s,y))M(\mathrm{d}s, \mathrm{d}y) \\ & + \langle \Lambda(t-\cdot, x-*)B(X_n(\cdot,*))1_{[0,t_n]}(\cdot), v^n\rangle_{\mathcal{H}_T} \\ & + \langle \Lambda(t-\cdot, x-*)G(X_n(\cdot,*))1_{[0,t_n]}(\cdot), h\rangle_{\mathcal{H}_T} \\ & + \int_0^{t_n}\int_{\mathbb{R}^d} \Lambda(t-s, x-y) b(X_n(s,y))\mathrm{d}y\mathrm{d}s. \end{aligned}$$

In the proof of Proposition 4.2, we will use the following facts: for any $p \in [1, \infty)$ and every integer $n \geq 1$,

(P1)

$$\sup_{(t,x)\in[0,T]\times\mathbb{R}^d} \|X(t,x) - X(t,t_n,x)\|_p \leq C 2^{-n(\frac{\eta}{2}\wedge\delta)}, \tag{4.7}$$

(P2)

$$\sup_{n\in\mathbb{N}} \sup_{(t,x)\in[0,T]\times\mathbb{R}^d} \|X(t,t_n,x)\|_p \le C, \tag{4.8}$$

(P3)

$$\sup_{(t,x)\in[0,T]\times\mathbb{R}^d} \|(X_n(t,x)-X_n^-(t,x))1_{L_n(t)}\|_p \le Cn2^{-n([(\frac{\eta}{2}-\theta+\frac{1}{2})\wedge\delta]}. \tag{4.9}$$

The estimate (4.7) can be easily obtained by adapting the arguments of the proof of [10, Lemma 4.1] and applying the assumption (A7). From (4.7) and Theorem A.1 applied to the Equation (4.1), we obtain (4.8). Finally, (4.9) is proved adapting the arguments of the proof of [10, Lemmas 4.2 and 4.3], and assuming (A7).

Proof of Proposition 4.2.

Using (4.1), (4.2), we write the difference $X_n(t,x)-X(t,x)$ grouped into comparable terms and prove their convergence to zero. The main difficulty lies in the convergence of $\langle\Lambda(t-\cdot,x-*)B(X_n(\cdot,*)),v^n\rangle_{\mathcal{H}_T}$ to $\int_0^t\int_{\mathbb{R}^d}\Lambda(t-s,x-y)B(X(s,y))M(\mathrm{d}s,\mathrm{d}y)$.

We write

$$X(t,x)-X_n(t,x)=\sum_{i=1}^{10}U_n^i(t,x),$$

where

$$U_n^1(t,x)=\int_0^t\int_{\mathbb{R}^d}\Lambda(t-s,x-y)\,[A(X(s,y))-A(X_n(s,y))]\,M(\mathrm{d}s,\mathrm{d}y),$$

$$U_n^2(t,x)=\langle\Lambda(t-\cdot,x-*)[G(X(\cdot,*))-G(X_n(\cdot,*))],h\rangle_{\mathcal{H}_T},$$

$$U_n^3(t,x)=\int_0^t\int_{\mathbb{R}^d}\Lambda(t-s,x-y)[b(X(s,y))-b(X_n(s,y))]\mathrm{d}y\mathrm{d}s,$$

$$U_n^4(t,x)=\int_{t_n}^t\int_{\mathbb{R}^d}\Lambda(t-s,x-y)B(X(s,y))M(ds,dy),$$

$$U_n^5(t,x)=\langle\Lambda(t-\cdot,x-*)[B(X(\cdot,*))-B(X_n(\cdot,*))]1_{[t_n,t]}(\cdot),v^n\rangle_{\mathcal{H}_T},$$

$$U_n^6(t,x)=-\langle\Lambda(t-\cdot,x-*)B(X(\cdot,*))1_{[t_n,t]}(\cdot),v^n\rangle_{\mathcal{H}_T},$$

$$U_n^7(t,x)=\int_0^{t_n}\int_{\mathbb{R}^d}\Lambda(t-s,x-y)[B(X(s,y))-B(X^-(s,y))]M(\mathrm{d}s,\mathrm{d}y),$$

$$U_n^8(t,x)=\int_0^{t_n}\int_{\mathbb{R}^d}\Lambda(t-s,x-y)B(X^-(s,y))M(\mathrm{d}s,\mathrm{d}y)$$

$$
\begin{aligned}
&- \langle \Lambda(t-\cdot, x-*)B(X^-(\cdot,*))1_{[0,t_n]}(\cdot), v^n\rangle_{\mathcal{H}_T},\\
U_n^9(t,x) &= \langle \Lambda(t-\cdot, x-*)[B(X^-(\cdot,*)) - B(X_n^-(\cdot,*))]1_{[0,t_n]}(\cdot), v^n\rangle_{\mathcal{H}_T},\\
U_n^{10}(t,x) &= \langle \Lambda(t-\cdot, x-*)[B(X_n^-(\cdot,*)) - B(X_n(\cdot,*))]1_{[0,t_n]}(\cdot), v^n\rangle_{\mathcal{H}_T}.
\end{aligned}
$$

Here, we have used the abridged notation $X^-(\cdot,*)$ for the stochastic process $X(t,t_n,x)$ defined in (4.6). Notice that although not apparent in this new notation, $X^-(\cdot,*)$ does depend on n.

Fix $p \in [2,\infty[$. Clearly,

$$
\mathbb{E}\left(|X_n(t,x) - X(t,x)|^p\, 1_{L_n(t)}\right) \le C\sum_{i=1}^{10}\mathbb{E}\left(\left|U_n^i(t,x)\right|^p 1_{L_n(t)}\right).
$$

We start by analyzing the contribution of $U_n^i(t,x)$, $i=1,2,3$ on the left-hand side of this expression.

Burkholder's and Hölder's inequalities yield

$$
\mathbb{E}\left(\left|U_n^1(t,x)\right|^p 1_{L_n(t)}\right) \le C\int_0^t \sup_{y\in\mathbb{R}^d}\mathbb{E}\Big(\left|X(s,y) - X_n(s,y)\right|^p 1_{L_n(s)}\Big) J_1(t-s)\mathrm{d}s. \tag{4.10}
$$

Schwarz's inequality implies

$$
\begin{aligned}
&\mathbb{E}\Big[\left|U_n^2(t,x)\right|^p 1_{L_n(t)}\Big]\\
&\quad\le \|h\|_{\mathcal{H}_T}^p\,\mathbb{E}\left(\left\|\Lambda(t-\cdot,x-*)[G(X(\cdot,*)) - G(X_n(\cdot,*))]1_{L_n(t)}\right\|_{\mathcal{H}_T}^2\right)^{p/2}.
\end{aligned}
$$

Then, by using Hölder's inequality we obtain

$$
\mathbb{E}\left(\left|U_n^2(t,x)\right|^p 1_{L_n(t)}\right) \le C\int_0^t \sup_{y\in\mathbb{R}^d}\mathbb{E}\Big(\left|X(s,y) - X_n(s,y)\right|^p 1_{L_n(s)}\Big) J_1(t-s)\mathrm{d}s. \tag{4.11}
$$

We apply Hölder's inequality to $U_n^3(t,x)$ and obtain

$$
\mathbb{E}\left(\left|U_n^3(t,x)\right|^p 1_{L_n(t)}\right) \le C\int_0^t \sup_{y\in\mathbb{R}^d}\mathbb{E}\Big(\left|X(s,y) - X_n(s,y)\right|^p 1_{L_n(s)}\Big) J_2(t-s)\mathrm{d}s. \tag{4.12}
$$

Next we consider the terms $U_n^i(t,x)$ for $i=4,5,6$. Let $i=4$. Hölder's inequalities and assumption (A7) yield

$$
\begin{aligned}
&\mathbb{E}\left(\left|U_n^4(t,x)\right|^p 1_{L_n(t)}\right) \\
&\quad \le C\left(1+\sup_{(t,x)\in[0,T]\times\mathbb{R}^d}\mathbb{E}\big[|X(t,x)|^p\big]\right)\left(\int_{t_n}^{t} J_1(t-s)\mathrm{d}s\right)^{p/2} \\
&\quad \le C2^{-pn\eta/2}.
\end{aligned}
\tag{4.13}
$$

Using Hölder's inequality, (4.4) and Assumption (A7), we have

$$
\begin{aligned}
&\mathbb{E}\left(\left|U_n^5(t,x)\right|^p 1_{L_n(t)}\right) \\
&\quad \le Cn^p 2^{np(\theta-1/2)}\mathbb{E}\big[\|\Lambda(t-\cdot,x-*)[B(X(s,y))-B(X_n(\cdot,*))] \\
&\qquad\qquad 1_{[t_n,t]}(\cdot)1_{L_n(t)}\|^p_{\mathcal{H}_T}\big] \\
&\quad \le Cn^p 2^{np(\theta-1/2)}\left(\int_{t_n}^{t} J_1(t-s)ds\right)^{p/2-1} \\
&\qquad \times\int_{t_n}^{t}\sup_{y\in\mathbb{R}^d}\mathbb{E}\Big(\left|X(s,y)-X_n(s,y)\right|^p 1_{L_n(s)}\Big)J_1(t-s)ds \\
&\quad \le Cn^p 2^{-n[p(\frac{\eta}{2}-\theta+\frac{1}{2})-\eta]}\int_0^t\sup_{y\in\mathbb{R}^d}\mathbb{E}\Big(\left|X(s,y)-X_n(s,y)\right|^p 1_{L_n(s)}\Big)J_1(t-s)ds.
\end{aligned}
\tag{4.14}
$$

Since $\frac{\eta}{2}-\theta+\frac{1}{2}>0$, for $p>\eta(\frac{\eta}{2}-\theta+\frac{1}{2})^{-1}$, we clearly have $p(\frac{\eta}{2}-\theta+\frac{1}{2})-\eta>0$.

For $U_n^6(t,x)$, we proceed in a similar manner as for $U_n^5(t,x)$ applying the fact that $X(t,x)$ has uniformly bounded moments of all orders. We obtain

$$
\begin{aligned}
&\mathbb{E}\left(\left|U_n^6(t,x)\right|^p 1_{L_n(t)}\right) \\
&\quad \le Cn^p 2^{pn(\theta-1/2)}\left(1+\sup_{(t,x)\in[0,T]\times\mathbb{R}^d}\mathbb{E}\big[|X(t,x)|^p\big]\right)\left(\int_{t_n}^{t} J_1(t-s)ds\right)^{p/2} \\
&\quad \le Cn^p 2^{-pn(\eta/2-\theta+1/2)}.
\end{aligned}
\tag{4.15}
$$

Finally, we study $U_n^i(t,x)$, $i=7,8,9,10$. The arguments based on Burkholder's and Hölder's inequalities and (4.7) give

$$\mathbb{E}\left(\left|U_n^7(t,x)\right|^p 1_{L_n(t)}\right) \le C\int_0^t \sup_{y\in\mathbb{R}^d}\mathbb{E}\Big(\left|X^-(s,y)-X(s,y)\right|^p 1_{L_n(s)}\Big)J_1(t-s)\mathrm{d}s$$
$$\le C2^{-np[\frac{\eta}{2}\wedge\delta]}. \tag{4.16}$$

In the following, let τ_n be the operator defined on functions $f:[0,T]\times\mathbb{R}^d\to\mathbb{R}$ by $\tau_n(f)(t,x):=f((t+2^{-n})\wedge T,x)$. Since $t_n<T-2^{-n}$, the restriction '$\wedge T$' is not active on $t\in[0,t_n]$. Let π_n be the projection operator from $\mathcal{H}_T$ onto the Hilbert subspace generated by the set of functions

$$\{2^nT^{-1}1_{\Delta_i}(\cdot)\otimes e_k(*),\, i=0,\dots,2^n-1, k=1,\dots,n\}.$$

Note that $\pi_n\circ\tau_n$ is a uniformly bounded operator in $n\in\mathbb{N}$ and $\pi_n\circ\tau_n$ converges to $I_{\mathcal{H}_T}$ strongly, where $I_{\mathcal{H}_T}$ denotes the identity operator on $\mathcal{H}_T$. Moreover, $\Upsilon_t:=\big(\pi_n\circ\tau_n\big)-I_{\mathcal{H}_T}$ is a contraction operator on $\mathcal{H}_T$.

Since $X_n^-(s,*)$, $X^-(s,*)$ are $\mathcal{F}_{s_n}$-measurable random variables, by using the definition of v^n one checks that

$$U_n^9(t,x)=\int_0^{t_n}\int_{\mathbb{R}^d}(\pi_n\circ\tau_n)\big[\Lambda(t-s,x-y)$$
$$(B(X_n^-(s,y))-B(X^-(s,y)))\big]M(\mathrm{d}s,\mathrm{d}y).$$

Thus, after having applied Burkholder's inequality, we obtain

$$\mathbb{E}\Big[\Big(\left|U_n^9(t,x)\right|^p\Big)1_{L_n(t)}\Big]$$
$$\le C\mathbb{E}\left(\left\|(\pi_n\circ\tau_n)\big[\Lambda(t-\cdot,x-*)(B(X^-)-B(X_n^-))\big](\cdot,*)1_{[0,t_n]}(\cdot)1_{L_n(\cdot)}\right\|_{\mathcal{H}_T}^p\right)$$
$$\le C\mathbb{E}\left(\int_0^{t_n}\left\|\Lambda(t-s,x-*)(B(X^-(s,*))-B(X_n^-(s,*)))1_{L_n(s)}\right\|_{\mathcal{H}}^2\mathrm{d}s\right)^{\frac{p}{2}}.$$

Then similarly as for $U_n^2(t,x)$, we have

$$\mathbb{E}\Big[\Big(\left|U_n^9(t,x)\right|^p\Big)1_{L_n(t)}\Big]$$
$$\le C\int_0^{t_n}\sup_{y\in\mathbb{R}^d}\mathbb{E}\Big(\left|X^-(s,y)-X_n^-(s,y)\right|^p 1_{L_n(s)}\Big)J_1(t-s)\mathrm{d}s.$$

This clearly implies

$$\begin{aligned}
&\mathbb{E}\Big[\Big(\Big|U_n^9(t,x)\Big|^p\Big)1_{L_n(t)}\Big]\\
&\le C\Big(\int_0^{t_n}\sup_{y\in\mathbb{R}^d}\mathbb{E}\Big(\Big|X^-(s,y)-X(s,y)\Big|^p 1_{L_n(s)}\Big)J_1(t-s)\mathrm{d}s\\
&\qquad+\int_0^{t_n}\sup_{y\in\mathbb{R}^d}\mathbb{E}\Big(\Big|X(s,y)-X_n(s,y)\Big|^p 1_{L_n(s)}\Big)J_1(t-s)\mathrm{d}s\\
&\qquad+\int_0^{t_n}\sup_{y\in\mathbb{R}^d}\mathbb{E}\Big(\Big|X_n(s,y)-X_n^-(s,y)\Big|^p 1_{L_n(s)}\Big)J_1(t-s)\mathrm{d}s\Big).
\end{aligned}$$

Recall that $X^-(s,y)=X(s,s_n,y)$. By applying (4.7) and (4.9), we obtain

$$\begin{aligned}
&\mathbb{E}\Big[\Big(\Big|U_n^9(t,x)\Big|^p\Big)1_{L_n(t)}\Big]\\
&\le C\int_0^{t}\sup_{y\in\mathbb{R}^d}\mathbb{E}\Big(\Big|X(s,y)-X_n(s,y)\Big|^p 1_{L_n(s)}\Big)J_1(t-s)\mathrm{d}s\\
&\quad+Cn^p2^{-np[(\frac{\eta}{2}-\theta+\frac{1}{2})\wedge\delta]}.
\end{aligned}\tag{4.17}$$

For the study of $U_n^{10}(t,x)$, we first apply Schwarz's inequality. Then (4.4) and (4.9) yield

$$\mathbb{E}\Big[\Big(\Big|U_n^{10}(t,x)\Big|^p\Big)1_{L_n(t)}\Big]\le Cn^{2p}2^{-np[(\frac{\eta}{2}-\theta+\frac{1}{2})\wedge\delta-\theta]}.\tag{4.18}$$

Finally, we consider $U_n^8(t,x)$. We are assuming that $t>0$. Hence, for n big enough, $t_n-2^{-n}>0$ and $t_n+2^{-n}<t$. Define

$$\begin{aligned}
U_n^{8,1}(t,x)=\int_0^{t_n}\int_{\mathbb{R}^d}\pi_n\Big(&\Lambda(t-\cdot,x-*)B(X^-(\cdot,*))\\
&-\tau_n\big[\Lambda(t-\cdot,x-*)B(X^-(\cdot,*))\Big)(s,y)M(\mathrm{d}s,\mathrm{d}y),\\
U_n^{8,2}(t,x)=\int_0^{t_n}\int_{\mathbb{R}^d}\Big(&\Lambda(t-s,x-y)B(X^-(s,y))\\
&-\pi_n\big[\Lambda(t-\cdot,x-*)B(X^-(\cdot,*))\big]\Big)M(\mathrm{d}s,\mathrm{d}y).
\end{aligned}$$

Clearly,

$$U_n^8(t,x) = U_n^{8,1}(t,x) + U_n^{8,2}(t,x).$$

To facilitate the analysis, we write $U_n^{8,1}(t,x)$ more explicitly, as follows.

$$\begin{aligned} U_n^{8,1}(t,x) = \int_0^{t_n}\int_{\mathbb{R}^d} \Big(&\pi_n\big[\Lambda(t-\cdot, x-*)B(X^-(\cdot,*))\big](s,y) \\ &- \pi_n\big[\Lambda(t-\cdot, x-*)B\big(X^-\big(\cdot,*\big)\big)\big](s+2^{-n},y)\Big)M(\mathrm{d}s,\mathrm{d}y). \end{aligned} \tag{4.19}$$

For the second integral on the right-hand side of (4.19) we perform a change of variable $s+2^{-n} \mapsto s$. Therefore we obtain

$$\mathbb{E}\left(\left|U_n^{8,1}(t,x)\right|^p 1_{L_n(t)}\right) \le C\left(V_n^{8,1}(t,x) + V_n^{8,2}(t,x)\right),$$

where

$$\begin{aligned} V_n^{8,1}(t,x) := \mathbb{E}\Big[\Big|\int_{t_n}^{t_n+2^{-n}}\int_{\mathbb{R}^d} &\pi_n\big[\Lambda(t-\cdot, x-*)B(X^-(\cdot,*))\big](s,y) \\ &\times M(\mathrm{d}s,\mathrm{d}y)\Big|^p 1_{L_n(t)}\Big], \\ V_n^{8,2}(t,x) := \mathbb{E}\Big[\Big|\int_0^{2^{-n}}\int_{\mathbb{R}^d} &\pi_n\big[\Lambda(t-\cdot, x-*)B\big(X^-\big(\cdot,*\big)\big)\big](s,y) \\ &\times M(\mathrm{d}s,\mathrm{d}y)\Big|^p 1_{L_n(t)}\Big]. \end{aligned}$$

By the usual procedure involving Burkholder's and Hölder's inequalities and (4.8) we have

$$\begin{aligned} V_n^{8,1}(t,x) &\le C\Bigg(1 + \sup_{(t,x)\in[0,T]\times\mathbb{R}^d} \mathbb{E}\big[|X^-(t,x)|^p\big]\Bigg)\Bigg(\int_{t_n}^{t_n+2^{-n}} J_1(t-s)\mathrm{d}s\Bigg)^{\frac{p}{2}} \\ &\le C\Bigg(1 + \sup_{(t,x)\in[0,T]\times\mathbb{R}^d} \mathbb{E}\big[|X^-(t,x)|^p\big]\Bigg)\Bigg(\int_0^{2^{-n+1}} J_1(s)ds\Bigg)^{\frac{p}{2}} \\ &\le C2^{-np\frac{\eta}{2}}, \end{aligned} \tag{4.20}$$

where in the last inequality, after a change of variable, we have applied (i) of assumption (A7).

Notice that for $s \in [0, 2^{-n}]$, $X^-(s, y) = X(s, s_n, y) = w(s, y)$. Therefore, condition (ii) in (A7) implies,

$$
\begin{aligned}
V_n^{8,2}(t,x) &\le C\left(1 + \sup_{(t,x)\in[0,T]\times\mathbb{R}^d} |w(t,x)|^p\right)\left(\int_0^{2^{-n}} J_1(t-s)\mathrm{d}s\right)^{\frac{p}{2}} \\
&\le C2^{-n\frac{p}{2}}.
\end{aligned}
\tag{4.21}
$$

Thus, by (4.20) and (4.21) we have proved the convergence

$$
\lim_{n\to\infty} \sup_{(t,x)\in[0,T]\times\mathbb{R}^d} \mathbb{E}\left(\left|U_n^{8,1}(t,x)\right|^p 1_{L_n(t)}\right) = 0. \tag{4.22}
$$

Let us now consider $U_n^{8,2}(t,x)$. After applying Burkholder's inequality we have

$$
\begin{aligned}
&\mathbb{E}\left(\left|U_n^{8,2}(t,x)\right|^p 1_{L_n(t)}\right) \\
&\quad \le C\mathbb{E}\left(\left\|\left(\pi_n - I_{\mathcal{H}_T}\right)\left[\Lambda(t-\cdot, x-*)B(X^-(\cdot,*))\right]1_{L_n(\cdot)}1_{[0,t_n]}(\cdot)\right\|_{\mathcal{H}_T}^p\right).
\end{aligned}
$$

We want to prove that the right-hand side of this inequality tends to zero as $n \to \infty$.

Set

$$
Z_n(t,x) = \left\|\left(\pi_n - I_{\mathcal{H}_T}\right)\left[\Lambda(t-\cdot, x-*)B(X^-(\cdot,*))\right]1_{L_n(\cdot)}1_{[0,t_n]}(\cdot)\right\|_{\mathcal{H}_T}.
$$

Since π_n is a projection on the Hilbert space $\mathcal{H}_T$, the sequence $\{Z_n(t,x), n \ge 1\}$ decreases to zero as $n \to \infty$ and can be bounded from above by $\sup_{n\in\mathbb{N}} \|\Lambda(t-\cdot, x-*)B(X^-(\cdot,*))\|_{\mathcal{H}_T}$. Remember that $X^-(s,y)$ stands for $X(s, s_n, y)$, defined in (4.6), and therefore it depends on n.

Assume that

$$
\mathbb{E}\left(\sup_n \|\Lambda(t-\cdot, x-*)B(X^-(\cdot,*))\|_{\mathcal{H}_T}^p\right) < \infty, \tag{4.23}
$$

for any $p \in [1, \infty)$. Then, by bounded convergence theorem, we can conclude that $\lim_{n\to\infty} \mathbb{E}[(Z_n(t,x))^p] = 0$.

Let us sketch the main arguments for the proof of (4.23). By considering first a convolution in the space variable of $\Lambda(t-\cdot, x-*)B(X^-(\cdot,*))$ with an approximation of the identity, and then passing to the limit, we prove

$$\mathbb{E}\left(\sup_n \|\Lambda(t-\cdot, x-*)B(X^-(\cdot,*))\|^p_{\mathcal{H}_T}\right)$$
$$\leq C\left(1+\sup_{(t,x)\in[0,T]\times\mathbb{R}^d} E\left(\sup_n |X(t,t_n,x)|^p\right)\right)\left(\int_0^t J_1(s)\mathrm{d}s\right)^{\frac{p}{2}} \tag{4.24}$$

(see [26, Proposition 3.3] for the arguments).

From the definition of $X(t,t_n,x)$, we see that for the second and third terms in (4.6), the supremum in n can be easily handled, since they are defined pathwise. For the stochastic integral term, we consider the discrete martingale

$$\left\{\int_0^{t_n}\int_{\mathbb{R}^d} \Lambda(s_0-s,x-y)(A+B)(X(s,y))M(\mathrm{d}s,\mathrm{d}y), \mathcal{F}_{t_n}, n\in\mathbb{N}\right\},$$

where $s_0 \in]0,T]$ is fixed. By applying first Doob's maximal inequality and then Burkholder's inequality, we obtain

$$E\left(\sup_n\left|\int_0^{t_n}\int_{\mathbb{R}^d}\Lambda(s_0-s,x-y)(A+B)(X(s,y))M(\mathrm{d}s,\mathrm{d}y)\right|^p\right)$$
$$\leq CE\left(\left|\int_0^{t}\int_{\mathbb{R}^d}\Lambda(s_0-s,x-y)(A+B)(X(s,y))M(\mathrm{d}s,\mathrm{d}y)\right|^p\right)$$
$$\leq CE\left(\|\Lambda(s_0-\cdot,x-*)(A+B)(X(\cdot,*))\|^{\frac{p}{2}}_{\mathcal{H}_T}\right).$$

Finally, we take $s_0 := t$. Using the property $\sup_{(t,x)\in[0,T]\times\mathbb{R}^d} E\left(|X(t,x)|^p\right)$, we obtain that the expression (4.24) is finite.

Hence, we have proved that

$$\lim_{n\to\infty} E\left(|U_n^{8,2}(t,x)|^p 1_{L_n(t)}\right)=0, \tag{4.25}$$

and from (4.22) and (4.25), we conclude

$$\lim_{n\to\infty}\mathbb{E}(|U_n^8(t,x)|^p 1_{L_n(t)})=0. \tag{4.26}$$

Taking into account (4.10)–(4.18) and (4.26), we see that

$$\mathbb{E}\left(|X(t,x) - X_n(t,x)|^p 1_{L_n(t)}\right)$$
$$\leq C_n + C\int_0^t \sup_{y\in\mathbb{R}^d} \mathbb{E}\left(|X(s,y) - X_n(s,y)|^p 1_{L_n(s)}\right)\left(J_1(t-s) + J_2(t-s)\right)ds,$$

where $(C_n, n \geq 1)$ is a sequence of real numbers satisfying $\lim_{n\to\infty} C_n = 0$. By applying a version of Gronwall's lemma, see [5, Lemma 15], we finish the proof of the Proposition. □

5 Examples

In this section we illustrate Theorem 2.4 with some examples.

Stochastic wave equation

Let F be the Gaussian process introduced in Sect. 2. Consider the family of stochastic wave equations indexed by $\varepsilon \in (0, 1]$,

$$\left(\frac{\partial^2}{\partial t^2} - \Delta_d\right)u^\varepsilon(t,x) = \varepsilon\sigma(u^\varepsilon(t,x))\dot{F}(t,x) + b(u^\varepsilon(t,x)),\ (t,x) \in (0,T] \times \mathbb{R}^d,$$
$$u^\varepsilon(0,x) = u_0(x), \quad \frac{\partial u^\varepsilon}{\partial t}(0,x) = u_1(x), \tag{5.1}$$

where Δ_d stands for the d-dimensional Laplacian, and $d \in \{1, 2, 3\}$.

We write (5.1) in the *mild form* (2.1) with

$$w(t,x) = \int_{\mathbb{R}^d} \Lambda(t, x-y)u_1(y)\mathrm{d}y + \frac{\partial}{\partial t}\int_{\mathbb{R}^d} \Lambda(t, x-y)u_0(y)\mathrm{d}y,$$

where Λ is the fundamental solution to the wave equation.

For any $t \in (0, T]$, $\Lambda(t)$ is a measure with support included in $B_t(0)$ (the closed ball of $\mathbb{R}^d$ centered at zero and with radius t), and $\Lambda(t)(\mathbb{R}^d) = t$. For example, if $d = 3$, it is the uniform surface measure on $\partial B_t(0)$, normalized by the factor $\frac{1}{4\pi t}$. It is also well-known that the Fourier transform of Λ is $\mathcal{F}\Lambda(t)(\xi) = \frac{\sin(2\pi t|\xi|)}{2\pi|\xi|}$ for any d (see e.g. [13, Chap. 5]).

For the sake of illustration, we will assume that the covariance measure of F is given by $\Gamma(dx) = |x|^{-\beta}\mathrm{d}x$, with $\beta \in (0, d \wedge 2)$, although a deeper analysis might allow to go beyond this case (see [6]). Then $\mu(d\xi) := \mathcal{F}^{-1}(\Gamma)(d\xi) = |\xi|^{-(d-\beta)}d\xi$ and

$$\int_{\mathbb{R}^d} |\mathcal{F}\Lambda(s)(\xi)|^2\mu(\mathrm{d}\xi) = \int_{\mathbb{R}^d} \frac{\sin^2(2\pi s|\xi|)}{4\pi^2|\xi|^{d-\beta+2}}\mathrm{d}\xi.$$

With the change of variable $\xi \mapsto (2\pi s)\xi$, we easily obtain that the last integral is equal to $Cs^{2-\beta}$, with $C > 0$. Thus,

$$\int_0^t \int_{\mathbb{R}^d} |\mathcal{F}\Lambda(s)(\xi)|^2 \mu(\mathrm{d}\xi)\mathrm{d}s = Ct^{3-\beta},\ t \in [0, T].$$

Consequently, the assumptions (A3) and (A7) hold with $\gamma = \eta = 3 - \beta$, $\bar{\eta} = 1$, $\delta = 2$ and $t_0 \in (0, T]$.

In this setting, we have the following result on (5.1).

Theorem 5.1 *Assume* (A2), (A4) and (A5). *Then for all* $(t, x) \in (0, T] \times \mathbb{R}^d$,

$$\lim_{\varepsilon \downarrow 0} \varepsilon^2 \log p^{\varepsilon}_{t,x}(y) = -I(y), \tag{5.2}$$

for all y in the interior of the support of $u(t, x)$, where I is defined in (2.5). *If in addition, either b or σ is bounded, then* (5.2) *holds for any $y \in \mathbb{R}$.*

Notice that under the standing hypotheses, Theorems 2.2, 2.3 and 4.1 hold. We refer to [7, Lemma 4.2] for sufficient conditions on the functions u_0, u_1 implying (A2).

Next we comment on the validity of assumption (A6). The sample paths of the random field solution to Eq. (5.1) belong to the space $\mathcal{C}^{\alpha}([0, T] \times \mathbb{R}^d)$ of α-Hölder continuous functions of degree $\alpha \in (0, \frac{2-\beta}{2})$ (see [8, Sect. 2.1] for a summary of results and references). In the present framework, and for spatial dimension $d = 3$, a *large deviation principle* (LDP) for (5.1) in the space $\mathcal{C}^{\alpha}([0, T] \times \mathbb{R}^d)$, with rate function $J \equiv I$, is established in [28, Theorem 1.1] (see also [29]). Its proof is carried out following the variational approach of Budhiraja and Dupuis in [1] (see also [12]). By the classical *contraction principle* of LDP ([11, Theorem 4.2.1]), this implies (A6). The proof in [28, Theorem 1.1] also applies to $d \in \{1, 2\}$. For $d = 2$ and with a different method, Chenal [3] establishes the same LDP. For $d = 1$, the reduced form of the stochastic wave equation driven by space-time white noise is considered in [17], and logarithmic estimates for the density are proved.

Proof of Theorem 5.1. From the preceding discussion, we infer that the random field solution to the stochastic wave Equation (5.1) at a fixed point $(t, x) \in (0, T] \times \mathbb{R}^d$ satisfies the assumptions of Theorem 2.4, and that $J = I$. □

Stochastic heat equation

Consider the family of stochastic heat equations indexed by $\varepsilon \in (0, 1]$,

$$\left(\frac{\partial}{\partial t} - \Delta_d\right) u^{\varepsilon}(t, x) = \varepsilon\sigma(u^{\varepsilon}(t, x))\dot{F}(t, x) + b(u^{\varepsilon}(t, x)),\ (t, x) \in (0, T] \times \mathbb{R}^d,$$
$$u^{\varepsilon}(0, x) = u_0(x), \tag{5.3}$$

where the process F is the same as in the example of the stochastic wave equation, and $d \in \mathbb{N}$.

As in the preceding example, we interpret (5.3) in the *mild form* (2.1) with

$$w(t,x) = \int_{\mathbb{R}^d} \Lambda(t, x-y) u_0(y) \mathrm{d}y,$$

and

$$\Lambda(t,x) = \frac{1}{(4\pi t)^{d/2}} \exp\left(-\frac{|x|^2}{4t} \right), \, (t,x) \in [0,T] \times \mathbb{R}^d.$$

Hence, $\Lambda(t)(\mathbb{R}^d) = 1$ for any $t \in (0,T]$.

Let Γ be as in the previous example. Then, using the change of variable $\xi \mapsto \sqrt{s}\xi$, we have

$$\int_0^t \int_{\mathbb{R}^d} |\mathcal{F}\Lambda(s)(\xi)|^2 \mu(\mathrm{d}\xi) ds = \int_0^t \int_{\mathbb{R}^d} \exp(-4\pi^2 s |\xi|^2) |\xi|^{\beta - d} d\xi ds = C t^{\frac{2-\beta}{2}}. \quad (5.4)$$

Hence, the assumptions (A1), (A3) hold with $\gamma = \frac{2-\beta}{2}$ and $\delta = 1$.

Theorem 5.2 *Assume* (A2), (A4), (A5), (A6). *Then for all* $(t,x) \in (t_0, T] \times \mathbb{R}^d$,

$$\overline{\lim_{\varepsilon \downarrow 0}} \varepsilon^2 \log p^{\varepsilon}_{t,x}(y) \le -J(y),$$
$$\underline{\lim_{\varepsilon \downarrow 0}} \varepsilon^2 \log p^{\varepsilon}_{t,x}(y) \ge -I(y).$$

The upper bound holds for all $y \in \mathbb{R}$, *while the lower bound* (*with* I *defined in* (2.5)) *holds for any* y *in the interior of the support of* $u(t,x)$. *If in addition* σ *is bounded, then the lower bound holds for any* $y \in \mathbb{R}$.

Suppose that the function u_0 is bounded and Hölder continuous with exponent $\alpha \in (0,1]$. Then Lemma 4.2 in [7] implies (A2).

Finally, we give some remarks on the hypothesis (A6). In the literature, there are several results on large deviations for different types of stochastic heat equations with boundary conditions. For example, [2] deals with a heat equation with $d = 1$ on a bounded domain with either Neumann or Dirichlet boundary conditions, driven by a space-time white noise. In [19], the dimension d is arbitrary, the boundary conditions are of Dirichlet type, and the noise is spatially correlated. Additional relevant references are [33], where non-Gaussian noises are considered; [30] in the framework of evolution equations; [1] illustrates the variational method on reaction-diffusion equations. In [21], Varadhan estimates have been obtained for the stochastic heat equation in spatial dimension one with space-time white noise on bounded domains.

We are not aware of any reference on large deviations for Eq. (5.3) in the present setting. Nevertheless, we believe that using a similar approach as in [28], such a result could be proved and that the rate function coincides with I. If this intuition is correct, the assumption (A6) of Theorem 5.2 could be removed and we will have an equality like (5.2).

Appendix

This section is devoted to the proof of some auxiliary results used in the paper. In the first part, we state a theorem on existence and uniqueness of a random field solution to a class of SPDEs which applies to the different types of equations that appear in the paper. In the second part, we prove an estimate on the $\mathcal{H}_T$-norm of the deterministic Malliavin matrix.

A.1 A Result on Existence and Uniqueness of Solution

Let H_1 and H_2 be separable Hilbert spaces. Consider mappings

$$\tilde{A}, \tilde{B}, \tilde{G} : H_2 \times H_1 \to H_1$$

satisfying

- for all $y, y' \in H_1$,

$$\sup_{x \in H_2} \Big(\|\tilde{A}(x,y) - \tilde{A}(x,y')\|_{H_1} + \|\tilde{B}(x,y) - \tilde{B}(x,y')\|_{H_1} + \|\tilde{G}(x,y) - \tilde{G}(x,y')\|_{H_1} \Big) \le C \|y - y'\|_{H_1}.$$

- There exists $q \in [1, \infty)$, and for all $x \in H_2$,

$$\|\tilde{A}(x,0)\|_{H_1} + \|\tilde{B}(x,0)\|_{H_1} + \|\tilde{G}(x,0)\|_{H_1} \le C\big(1 + \|x\|_{H_2}^q\big).$$

Combining these two estimates yields

$$\|\tilde{A}(x,y)\|_{H_1} + \|\tilde{B}(x,y)\|_{H_1} + \|\tilde{G}(x,y)\|_{H_1} \le C\big(1 + \|y\|_{H_1} + \|x\|_{H_2}^q\big). \quad \text{(A.1)}$$

Let $V = (V(t,x), (t,x) \in [0,T] \times \mathbb{R}^d)$ be a predictable H_2-valued stochastic process such that for all $p \in [1, \infty)$,

$$\sup_{(t,x)\in[0,T]\times\mathbb{R}^d} \mathbb{E}\left[\|V(t,x)\|_{H_2}^p\right] < \infty. \tag{A.2}$$

Let $U_0 = (U_0(t,x), (t,x) \in [0,T] \times \mathbb{R}^d)$ be a predictable H_1-valued stochastic process such that for all $p \in [1,\infty)$,

$$\sup_{(t,x)\in[0,T]\times\mathbb{R}^d} \mathbb{E}\left[\|U_0(t,x)\|_{H_1}^p\right] < \infty.$$

Let h be an $\mathcal{H}_T$-valued random variable such that

$$\sup_{\omega\in\Omega} \|h(\omega)\|_{\mathcal{H}_T} < \infty. \tag{A.3}$$

Consider the equation on H_1,

$$\begin{aligned} U(t,x) = U_0(t,x) &+ \int_0^t \int_{\mathbb{R}^d} \Lambda(t-s, x-y)\tilde{A}(V(s,y), U(s,y))M(\mathrm{d}s, \mathrm{d}y) \\ &+ \left\langle \Lambda(t-\cdot, x-*)\tilde{G}(V(\cdot,*), U(\cdot,*)), h\right\rangle_{\mathcal{H}_T} \\ &+ \int_0^t \int_{\mathbb{R}^d} \Lambda(t-s, x-y)\tilde{B}(V(s,y), U(s,y))\mathrm{d}y\mathrm{d}s. \end{aligned} \tag{A.4}$$

The following statement is a generalization of [7, Theorem 4.3] and [31, Theorem 6.2].

Theorem A.1 *Assume* (A1) *and let $\tilde{A}$, $\tilde{B}$, $\tilde{G}$, V, U_0, h, be given as above. There exists a unique predictable stochastic process* $(U(t,x), (t,x) \in [0,T] \times \mathbb{R}^d)$ *with values in H_1 satisfying* (A.4) *for all* $(t,x) \in [0,T] \times \mathbb{R}^d$*, a.s. This solution satisfies*

$$\sup_{(t,x)\in[0,T]\times\mathbb{R}^d} \mathbb{E}\left[\|U(t,x)\|_{H_1}^p\right] < \infty, \tag{A.5}$$

for all $p \in [1,\infty)$ *and is continuous in* $L^2(\Omega)$.

Proof We use the classical approach based on Picard's iterations, as in [31, Theorem 6.2], and similar ideas as in [7, Theorem 4.3], extended to a Hilbert space setting. In comparison with [31, Theorem 6.2], Equation (A.4) has the extra term $\left\langle \Lambda(t-\cdot, x-*)\tilde{G}(V(\cdot,*), U(\cdot,*)), h\right\rangle_{\mathcal{H}_T}$. We illustrate with an example how to deal with it.

Fix $p \in [1,\infty)$. The Cauchy-Schwarz inequality and (A.1) yield

$$
\begin{aligned}
&\mathbb{E}\Big[\|\langle\Lambda(t-\cdot,x-*)\tilde{G}(V(\cdot,*),U(\cdot,*)),h\rangle_{\mathcal{H}_T}\|_{H_1}^p\Big]\\
&\le \sup_{\omega\in\Omega}\|h(\omega)\|_{\mathcal{H}_T}^p\,\mathbb{E}\big[\|\Lambda(t-\cdot,x-*)\tilde{G}(V(\cdot,*),U(\cdot,*))\|_{H_1\otimes\mathcal{H}_T}^p\big]\\
&\le C\int_0^t\Bigg(1+\sup_{(r,y)\in[0,s]\times\mathbb{R}^d}\mathbb{E}\big[\|U(r,y)\|_{H_1}^p\big]+\sup_{(r,y)\in[0,s]\times\mathbb{R}^d}\mathbb{E}\big[\|V(r,y)\|_{H_1}^{pq}\big]\Bigg)\\
&\qquad\times J_1(t-s)\mathrm{d}s\\
&\le C\int_0^t\Bigg(1+\sup_{(r,y)\in[0,s]\times\mathbb{R}^d}\mathbb{E}\big[\|U(r,y)\|_{H_1}^p\big]\Bigg)J_1(t-s)\mathrm{d}s,
\end{aligned}
$$

where in the second inequality we have applied (A.1), and in the last one (A.2). We leave it to the reader to complete all the details of the proof. □

In the preceding sections, the following particular cases of Eq. (A.4) have been considered.

(1) $H_1 = H_2 = \mathbb{R}$, $h = 0$, $\tilde{A} = \varepsilon\sigma$ and $\tilde{B} = b$ do not depend on the first coordinate, $U_0 = w$. Then $U = u^\varepsilon$ (see (2.1)).
(2) $H_1 = H_2 = \mathbb{R}$, $\tilde{A} = 0$, $\tilde{G} = \sigma$ and $\tilde{B} = b$ do not depend on the first coordinate, $U_0 = w$. Then $U := \Phi^h$ (see (2.6)).
(3) $H_1 = \mathcal{H}_T$, $H_2 = \mathbb{R}$, $h = 0$, $\tilde{A}(x,y) = \varepsilon\sigma'(x)y$, $\tilde{B}(x,y) = b'(x)y$, $U_0 = x\varepsilon\Lambda(t-\cdot,x-*)\sigma(u^\varepsilon(\cdot,*))$, $V = u^\varepsilon$. Then $U := Du^\varepsilon$ (see (3.2)).
(4) $H_1 = \mathcal{H}_T$, $H_2 = \mathbb{R}$, $\tilde{A} = 0$, $\tilde{G}(x,y) = \sigma'(x)y$, $\tilde{B}(x,y) = b'(x)y$, $U_0 = \Lambda(t-\cdot,x-*)\sigma(\Phi^h_{\cdot,*})$, $V = \Phi^h$. Then $U := \Xi^h$ (see (3.13)).
(5) $H_1 = H_2 = \mathbb{R}$, $\tilde{A} = \sigma$ does not depend on the second coordinate, $\tilde{G}(x,y) = \sigma'(x)y$, $\tilde{B}(x,y) = b'(x)y$, $U_0 = 0$, $V = \Phi^h$. Then $U := \mathcal{N}(h)$ (see (3.15)).

There is yet another particular case of Eq. (A.4) that we met in the proof of Lemma 3.7. It is obtained by shifting the sample paths ω by $\varepsilon^{-1}h$, where $h \in \mathcal{H}_T$, as it is shown in the next Lemma.

Lemma A.2 *The hypotheses are* (A1), (A2)*, that σ, b are Lipschitz continuous. Let $h \in H$ and set $u^{\varepsilon,h}(t,x;\omega) := u^\varepsilon(t,x;\omega+\varepsilon^{-1}h)$, $(t,x) \in [0,T]\times\mathbb{R}^d$, where u^ε is the solution to* (2.1)*. Then $(u^{\varepsilon,h}(t,x), (t,x) \in [0,T]\times\mathbb{R}^d)$ satisfies the equation*

$$
\begin{aligned}
u^{\varepsilon,h}(t,x) = w(t,x) &+ \varepsilon\int_0^t\int_{\mathbb{R}^d}\Lambda(t-s,x-z)\sigma(u^{\varepsilon,h}(s,z))M(\mathrm{d}s,\mathrm{d}z)\\
&+\int_0^t\int_{\mathbb{R}^d}\Lambda(t-s,x-z)b(u^{\varepsilon,h}(s,z))\mathrm{d}z\mathrm{d}s\\
&+\langle\Lambda(t-\cdot,x-*)\sigma(u^{\varepsilon,h}(\cdot,*)),h\rangle_{\mathcal{H}_T}. \qquad \text{(A.6)}
\end{aligned}
$$

The Lemma relies on the formula

$$
\begin{aligned}
&\left(\int_0^t\int_{\mathbb{R}^d}\Lambda(s,y)Y(s,y)M(\mathrm{d}s,\mathrm{d}y)\right)(\omega+h)\\
&\quad=\langle\Lambda(\cdot,*)Y(s,y)(\omega+h),h\rangle_{\mathcal{H}_T}\\
&\qquad+\left(\int_0^t\int_{\mathbb{R}^d}\Lambda(s,y)Y(s,y)(\omega+h)M(\mathrm{d}s,\mathrm{d}y)\right)(\omega),
\end{aligned}
$$

where $h\in\mathcal{H}_T$ and $(Y(s,y),(s,y)\in[0,T]\times\mathbb{R}^d)$ is a predictable stochastic processes such that $(\Lambda(s,y)Y(s,y),(s,y)\in[0,T]\times\mathbb{R}^d)$ is integrable with respect to the martingale measure M (see [5, 7] for details). This is proved by considering first step processes g, given by $g(t,x,\omega)=1_{(a,b]}(t)1_A(x)X(\omega)$ for $0\le a<b\le T$, $A\in\mathcal{B}_b(\mathbb{R}^d)$ and X a bounded, $\mathcal{F}_a$-measurable random variable, and then passing to the limit.

A.2 Analysis of the Deterministic Malliavin Matrix

In this section we derive an assertion similar to [31, Lemma 8.2] for the Fréchet derivative of the function Φ, defined in (3.12), (3.13).

Proposition A.3 *The assumptions are* (A1)*, $\sigma,b\in\mathcal{C}^1$ with bounded Lipschitz continuous derivatives. Then, for all $\rho\in[0,t]$ and $t\in[0,T]$,*

$$
\sup_{(r,z)\in[0,t]\times\mathbb{R}^d}\|\bar{D}\Phi^h(r,z)\|^{2p}_{\mathcal{H}_{t-\rho,t}}\le C\big(g_1(\rho)\big)^p.
$$

Proof Fix $p\in[1,\infty)$, $t\in[0,T]$, $\rho\in[0,t]$ and $(r,y)\in[0,t]\times\mathbb{R}^d$. Recall that $\bar{D}\Phi^h(r,y)$ is an $\mathcal{H}_T$-valued random variable given by

$$
\begin{aligned}
\bar{D}_{\circ,\bullet}\Phi^h(r,y)=&\,\Lambda(r-\circ,y-\bullet)\sigma(\Phi^h_{\circ,\bullet})\\
&+\big\langle\Lambda(r-\cdot,y-*)\sigma'(\Phi^h_{\cdot,*})\bar{D}_{\circ,\bullet}\Phi^h(\cdot,*),h\big\rangle_{\mathcal{H}_T}\\
&+\int_0^r\int_{\mathbb{R}^d}\Lambda(r-s,y-z)b'(\Phi^h_{s,z})\bar{D}_{\circ,\bullet}\Phi^h(s,z)\mathrm{d}z\mathrm{d}s
\end{aligned}
$$

(see (3.12) and (3.13)).

We analyze each one of the three terms on the right-hand side of this equation separately.

For the first term, we have

$$
\begin{aligned}
&\left\| \Lambda(r-\circ, y-\bullet)\sigma(\Phi^h_{\circ,\bullet})\right\|^{2p}_{\mathcal{H}_{t-\rho,t}} \\
&\quad\le \left(\int_{t-\rho}^{r} J_1(r-s)\mathrm{d}s\right)^{p-1} \int_{t-\rho}^{r} \sup_{(v,z)\in[0,s]\times\mathbb{R}^d} \mathbb{E}\big[|\sigma(\Phi^h_{v,z})|^{2p}\big] J_1(r-s)\mathrm{d}s \\
&\quad\le C(g_1(\rho))^p,
\end{aligned}
$$

where in the last inequality we have used that σ is Lipschitz continuous and also that for each $h \in \mathcal{H}_T$, the function $(t,x) \in [0,T]\times\mathbb{R}^d \mapsto \Phi^h(t,x)$ is uniformly bounded. Indeed, this is a consequence of (A.5), since $\Phi^h(t,x)$ is deterministic.

For the second term, we apply first Schwarz' inequality and then Hölder's inequality. Using that σ' is bounded, we obtain

$$
\begin{aligned}
&\left\| \langle \Lambda(r-\cdot, y-*)\sigma'(\Phi^h(\cdot,*))\bar{D}_{\circ,\bullet}\Phi^h(\cdot,*), h\rangle_{\mathcal{H}_T}\right\|^{2p}_{\mathcal{H}_{t-\rho,t}} \\
&\quad\le C\|h\|^{2p}_{\mathcal{H}_T} \int_0^r \sup_{(v,z)\in[0,s]\times\mathbb{R}^d} \|\bar{D}_{\circ,\bullet}\Phi^h(v,z)\|^{2p}_{\mathcal{H}_{t-\rho,t}} J_1(r-s)\mathrm{d}s.
\end{aligned}
$$

Finally, for the last term we apply Hölder's inequality with respect to the finite measure $\Lambda(r-s,x-z)\mathrm{d}z\mathrm{d}s$ along with the boundedness of b'. We obtain,

$$
\begin{aligned}
&\left\| \int_0^r \int_{\mathbb{R}^d} \Lambda(r-s, y-z) b'(\Phi^h_{s,y})\bar{D}_{\circ,\bullet}\Phi^h(s,z)\mathrm{d}z\mathrm{d}s\right\|^{2p}_{\mathcal{H}_{t-\rho,t}} \\
&\qquad\le C\int_0^r \sup_{(v,z)\in[0,s]\times\mathbb{R}^d} \|\bar{D}_{\circ,\bullet}\Phi^h(v,z)\|^{2p}_{\mathcal{H}_{t-\rho,t}} J_2(r-s)\mathrm{d}s.
\end{aligned}
$$

By applying Gronwall's lemma to the real function

$$
s \mapsto \sup_{(r,z)\in[0,s]\times\mathbb{R}^d} \|\bar{D}\Phi^h_{\circ,\bullet}(r,z)\|^{2p}_{\mathcal{H}_{t-\rho,t}},
$$

we have

$$
\sup_{(r,z)\in[0,t]\times\mathbb{R}^d} \|\bar{D}_{\circ,\bullet}\Phi^h(r,z)\|^{2p}_{\mathcal{H}_{t-\rho,t}} \le C\left(\int_0^\rho J_1(s)\mathrm{d}s\right)^p = C\,(g_1(\rho))^p\,,
$$

for all $t\in[0,T]$. This yields the assertion. $\square$

References

1. A. Budhiraja, P. Dupuis, Large deviations for infinite dimensional stochastic dynamical systems. Ann. Probab. **36**(4), 1390–1420 (2008)
2. F. Chenal, A. Millet, Uniform large deviations for parabolic spdes and applications. Stochast. Process Appl. **72**(2), 161–186 (1997)
3. F. Chenal, Principes des grandes dévations pour des équations aux dérivées partielles stochastiques et applications. PhD thesis, Université Paris 6 (2000)
4. R.C. Dalang, N. Frangos, The stochastic wave equation in two spatial dimensions. Ann. Probab. **26**(1), 187–212 (1998)
5. R.C. Dalang, Extending martingale measure stochastic integral with applications to spatially homogeneous SPDEs. Electron. J. Probab. **4**, 1–29 (1999)
6. R. Dalang, M. Sanz-Solé, Hölder-Sobolev regularity of the solution to the stochastic wave equation in dimension three. Mem. Am. Math. Soc. **199**(931), 1–70 (2009)
7. R. Dalang, L. Quer-Sardanyons, Stochastic integrals for SPDEs: a comparison. Expos. Math. **29**, 67–109 (2011)
8. R. Dalang, M. Sanz-Solé, Hitting probabilities for non-linear systems of stochastic waves. Mem. Am. Math. Soc. (2012) (to appear). arXiv:1205.3041
9. G. Da Prato, J. Zabczyk, Stochastic equations in infinite dimensions, in *Encyclopedia of Mathematics and its Applications*, vol. 45 (Cambridge University Press, Cambridge, 2008)
10. F. Delgado-Vences, M. Sanz-Solé, Approximation of a stochastic wave equation in dimension three, with application to a support theorem in Hölder norm. Bernoulli **16**(1), 165–186 (2010). arXiv:1203.1188
11. A. Dembo, O. Zeitouni, *Large Deviations Techniques and Applications* (John and Bartlett Publishers, Boston, 1993)
12. P. Dupuis, R.S. Ellis, *A Weak Convergence Approach to the Theory of Large Deviations* (Wiley, New York, 1997)
13. G.B. Folland, *Introduction to Partial Differential Equations* (Princeton University Press, Princeton, 1976)
14. A. Kohatsu-Higa, D. Márquez-Carreras, M. Sanz-Solé, Logarithmic estimates for the density of hypoelliptic two-parameter diffusions. J. Funct. Anal. **190**, 481–506 (2002)
15. R. Léandre, Majoration en temps petit de la densité d'une diffusion dégénérée. Probab. Theory Relat. Fields **74**, 289–294 (1987)
16. R. Léandre, Minoration en temps petit de la densité d'une diffusion dégénérée. J. Funct. Anal. **74**, 399–414 (1987)
17. R. Léandre, F. Russo, Estimations de Varadhan pour des diffusions à deux paramètres. Probab. Theory Relat. Fields **84**, 429–451 (1990)
18. D. Márquez-Carreras, M. Sarrà, Behavior of the density in perturbed SPDE's with spatially correlated noise. Bull. Sci. Math. **127**, 348–367 (2003)
19. D. Márquez-Carreras, M. Sarrà, Large deviation principle for a stochastic heat equation with spatially correlated noise. Electron. J. Probab. **8**(12), (2003)
20. M. Métivier, *Semimartingales* (de Gruyter, Berlin, 1982)
21. A. Millet, M. Sanz-Solé, *Varadhan Estimates for the Density of the Solution to a Parabolic Stochastic Partial Differential Equation* (World Scientific Publishing, Singapore, 1996)
22. A. Millet, M. Sanz-Solé, Approximation and support theorem for a wave equation in two space dimensions. Bernoulli **6**(5), 887–915 (2000)
23. J. Munkres, *Topology*, 2nd edn. (Prentice Hall, Upper Saddle River, 2000)
24. D. Nualart, Analysis on the Wiener space and anticipating calculus, in *École d'Été de Probabilités de Saint-Flour XXV, 1995*. Lecture Notes in Mathematics, vol. 1690 (Springer, Heidelberg, 1998)
25. D. Nualart, *The Malliavin Calculus and Related Topics*, 2nd edn. (Springer, New York, 2006)
26. D. Nualart, L. Quer-Sardanyons, Existence and smoothness of the density for spatially homogeneous SPDEs. Potential Anal. **27**, 281–299 (2007)

27. E. Nualart, On the density of systems of non-linear spatially homogeneous spdes. Stoch. Stoch. Rep. **85**, 48–70 (2013)
28. V. Ortiz-López, M. Sanz-Solé, A Laplace principle for a stochastic wave equation in spatial dimension three, in *Stochastic Analysis 2010*, ed. by D. Crisan (Springer, Berlin, 2010)
29. V. Ortiz-López, Principi de Grans Desviacions per a l'equació en derivades parcials estocàstica d'ones en dimensió espacial 3, PhD Thesis, Universitat de Barcelona (2011)
30. S. Peszat, Large deviation principle for stochastic evolution equations. Probab. Theory Relat. Fields **98**(1), 113–136 (1994)
31. M. Sanz-Solé, *Malliavin Calculus with Applications to Stochastic Partial Differential Equations* (EPFL Press, Lausanne, 2005)
32. L. Schwartz, *Théorie des Distributions*, 2nd edn. (Hermann, Paris, 2010)
33. R.B. Sowers, Large deviations for a reaction-diffusion equation with non-gaussian perturbations. Ann. Probab. **20**(1), 504–537 (1992)
34. S.R.S. Varadhan, On the behavior of the fundamental solution of the heat equations with variable coefficients. Commun. Pure Appl. Math. **20**, 431–455 (1967)
35. S.R.S. Varadhan, Diffusion processes in a small time interval. Commun. Pure Appl. Math. **20**, 659–685 (1967)

Author Index

D. Crisan et al. (eds.), *Stochastic Analysis and Applications 2014*,
Springer Proceedings in Mathematics & Statistics 100,
DOI 10.1007/978-3-319-11292-3