Progress in Probability
74

Nathael Gozlan
Rafał Latała
Karim Lounici
Mokshay Madiman
Editors

# High Dimensional Probability VIII

## The Oaxaca Volume

Birkhäuser

Birkhäuser

# Progress in Probability
Volume 74



Progress in Probability is designed for the publication of workshops, seminars and conference proceedings on all aspects of probability theory and stochastic processes, as well as their connections with and applications to other areas such as mathematical statistics and statistical physics.

More information about this series at http://www.springer.com/series/4839

Nathael Gozlan • Rafał Latała • Karim Lounici • Mokshay Madiman
Editors

# High Dimensional Probability VIII

## The Oaxaca Volume

Birkhäuser

*Editors*
Nathael Gozlan
MAP 5
Université Paris Descartes
Paris, France

Rafał Latała
Institute of Mathematics
University of Warsaw
Warsaw, Poland

Karim Lounici
Centre de Mathématiques Appliquées
Ecole Polytechnique
Palaiseau, France

Mokshay Madiman
Department of Mathematical Sciences
University of Delaware
Newark, DE, USA

ISSN 1050-6977 ISSN 2297-0428 (electronic)
Progress in Probability
ISBN 978-3-030-26390-4 ISBN 978-3-030-26391-1 (eBook)
https://doi.org/10.1007/978-3-030-26391-1



This book is published under the imprint Birkhäuser, www.birkhauser-science.com, by the registered company Springer Nature Switzerland AG.
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

# Preface

The history of the High-Dimensional Probability (HDP) conferences dates back to the 1975 International Conference on Probability in Banach Spaces in Oberwolfach, Germany. After eight Probability in Banach Spaces meetings, in 1994 it was decided to give the series its current name: the International Conference on High-Dimensional Probability.

The present volume is an outgrowth of the Eighth High-Dimensional Probability Conference (HDP VIII), which was held at the Casa Matemática Oaxaca (Mexico) from May 28th to June 2nd, 2017. The scope and quality of the talks and contributed papers amply demonstrate that, now more than ever, high-dimensional probability is a very active area of mathematical research.

High-Dimensional Probability has its roots in the investigation of limit theorems for random vectors and regularity of stochastic processes. It was initially motivated by the study of necessary and sufficient conditions for the boundedness and continuity of trajectories of Gaussian processes and the extension of classical limit theorems, such as laws of large numbers, laws of the iterated logarithm and central limit theorems, to Hilbert and Banach space-valued random variables and empirical processes.

This resulted in the creation of powerful new tools: the methods of high-dimensional probability and especially its offshoots, the concentration of measure phenomenon and generic chaining techniques, were found to have a number of applications in various areas of mathematics, as well as statistics and computer science. These include random matrix theory, convex geometry, asymptotic geometric analysis, nonparametric statistics, empirical process theory, statistical learning theory, compressed sensing, strong and weak approximations, distribution function estimation in high dimensions, combinatorial optimization, random graph theory, stochastic analysis in infinite dimensions, and information and coding theory.

In recent years there has been substantial progress in the area. In particular, numerous important results have been obtained concerning the connections between various functional inequalities related to the concentration of measure phenomenon, application of generic chaining methods to study the suprema of stochastic processes and norms of random matrices, Malliavin–Stein theory of

Gaussian approximation, various stochastic inequalities and their applications in high-dimensional statistics and computer science. This breadth is duly reflected by the diverse contributions in the present volume.

The majority of the papers gathered here were presented at HDP VIII. The conference participants wish to express their gratitude for the support provided by the BIRS-affiliated mathematics research center Casa Matemática Oaxaca. In addition, the editors wish to thank Springer-Verlag for publishing the proceedings.

The book begins with a dedication to our departed and esteemed colleague, Jørgen Hoffmann-Jørgensen, whom we lost in 2017. This is followed by a collection of contributed papers that are divided into four general areas: inequalities and convexity, limit theorems, stochastic processes, and high-dimensional statistics. To give readers an idea of their scope, in the following we briefly describe them by subject area and in the order they appear in this volume.

**Dedication to Jørgen Hoffmann-Jørgensen (1942–2017)**

- *Jørgen Hoffmann-Jørgensen*, by M. B. Marcus, G. Peskir and J. Rosiński. This paper honors the memory, scientific career and achievements of Jørgen Hoffmann-Jørgensen.

**Inequalities and Convexity**

- *Moment estimation implied by the Bobkov-Ledoux inequality*, by W. Bednorz and G. Głowienko. The authors derive general bounds for exponential Orlicz norms of locally Lipschitz functions using the Bobkov-Ledoux entropic form of the Poincaré inequality.
- *Polar isoperimetry I—the case of the plane*, by S. G. Bobkov, N. Gozlan, C. Roberto and P.-M. Samson. This is the first part of a lecture notes series and offers preliminary remarks on the plane isoperimetric inequality and its applications to the Poincaré and Sobolev type inequalities in dimension one.
- *Iterated Jackknives and two-sided variance inequalities*, by O. Bousquet and C. Houdré. The authors revisit selected classical variance inequalities, such as the Efron–Stein inequality, and present refined versions.
- *A probabilistic characterization of negative definite functions*, by F. Gao. The author proves using Fourier transform tools that a continuous function $f$ on $\mathbb{R}^n$ is negative definite if and only if it is polynomially bounded and satisfies the inequality

$$\mathbb{E}f(X - Y) \leq \mathbb{E}f(X + Y)$$

  for all i.i.d. random vectors $X$ and $Y$ in $\mathbb{R}^n$.
- *Higher order concentration in presence of Poincaré type inequalities*, by F. Götze and H. Sambale. The authors obtain sharpened forms of the concentration of measure phenomenon that typically apply to differentiable functions with centered derivatives up to the order $d - 1$ and bounded derivatives of order $d$.

- *Rearrangement and Prékopa–Leindler type inequalities*, by J. Melbourne. The author obtains rearrangement sharpenings of several classical Prékopa–Leindler type functional inequalities.
- *Generalized semimodularity: order statistics*, by I. Pinelis. The author introduces a notion of generalized $n$-semimodularity, which extends that of (sub/super)modularity, and derives applications to correlation inequalities for order statistics.
- *Geometry of $\ell_p^n$-balls: Classical results and recent developments*, by J. Prochno, C. Thäle and N. Turchi. The paper presents a survey of asymptotic theorems for uniform measures on $\ell_p^n$-balls and cone measures on $\ell_p^n$-spheres.
- *Remarks on superconcentration and Gamma calculus. Applications to spin glasses*, by K. Tanguy. This paper explores applications of Bakry-Emery $\Gamma_2$ calculus to refined variant inequalities for several spin systems models.

**Limit Theorems**

- *Asymptotic behavior of Renyi entropy in the central limit theorem*, by S. G. Bobkov and A. Marsiglietti. The authors explore the asymptotic behavior and monotonicity of Renyi entropy along convolutions in the central limit theorem.
- *Uniform-in-bandwidth functional limit laws for multivariate empirical processes*, by P. Deheuvels. The author provides uniform-in-bandwidth functional limit laws for multivariate local empirical processes, with statistical applications to kernel density estimation.
- *Universality of limiting spectral distribution under projective criteria*, by F. Merlevède and M. Peligrad. The authors study the limiting empirical spectral distribution of an $n \times n$ symmetric matrix with dependent entries. For a class of generalized martingales, they show that the asymptotic behavior of the empirical spectral distribution depends only on the covariance structure.
- *Exchangeable pairs on Wiener chaos*, by I. Nourdin and G. Zheng. In this paper, the authors propose a new proof of a quantitative form of the fourth moment theorem in Gaussian approximation based on the construction of exchangeable pairs of Brownian motions.

**Stochastic Processes**

- *Permanental processes with kernels that are equivalent to a symmetric matrix*, by M. B. Marcus and J. Rosen. The authors consider $\alpha$-permanental processes whose kernel is of the form

  $$\widetilde{u}(x, y) = u(x, y) + f(y), \qquad x, y \in S,$$

  where $u$ is symmetric and $f$ has some good properties. In turn, they define conditions that determine whether the kernel $\widetilde{u}$ is symmetrizable or asymptotically symmetrizable.
- *Pointwise properties of martingales with values in Banach function spaces*, by M. Veraar and I. Yaroslavtsev. In this paper, the authors consider local

martingales with values in a UMD Banach function space and prove that such martingales have a version which is a martingale field. Moreover, a new Burkholder–Davis–Gundy type inequality is obtained.

**High-Dimensional Statistics**

- *Concentration inequalities for randomly permuted sums*, by M. Albert. The author proves a deviation inequality for random permutations and uses it to analyze the second kind error rate in a test of independence.
- *Uncertainty quantification for matrix compressed sensing and quantum tomography problems*, by A. Carpentier, J. Eisert, D. Gross and R. Nickl. The authors construct minimax optimal non-asymptotic confidence sets for low-rank matrix recovery algorithms such as the Matrix Lasso and Dantzig selector.
- *Uniform-in-bandwidth estimation of the gradient lines of a density*, by D. Mason and B. Pelletier. This paper exploits non parametric statistical techniques to estimate the gradient flow of a stochastic differential equation. The results can be of interest in clustering applications or the analysis of stochastic gradient schemes.

Paris, France — Nathael Gozlan
Warsaw, Poland — Rafał Latała
Palaiseau, France — Karim Lounici
Newark, DE, USA — Mokshay Madiman

# Contents

# Jørgen Hoffmann-Jørgensen (1942–2017)

**Michael B. Marcus, Goran Peskir, and Jan Rosiński**

Jørgen Hoffmann-Jørgensen, docent emeritus in the Department of Mathematics at Aarhus University, Denmark, died on the 8th of December 2017. He was 75 years old. He is survived by Karen, his wife of fifty years, his mother Ingeborg, his brother Bent and his niece Dorthe.

He was a devoted teacher and advisor, a wonderful, friendly person, and a very fine and prolific mathematician. His ties to Aarhus are legendary. Jørgen received his magister scientiarum degree from the Institute of Mathematics at Aarhus University in 1966. He began his research and teaching there in the previous year and continued through the academic ranks, becoming docent in 1988.

With a stroke of good luck he began his career as a probabilist under the most auspicious circumstances. Kiyoshi Itô was a professor at Aarhus from 1966 to 1969. Ron Getoor, who had been with Itô at Princeton, came to Aarhus as a visiting professor in the spring semester of 1969. Jørgen began his research career in the presence of these outstanding probabilists. He often commented that, more than any other mathematician, Itô had the greatest influence on his work.

There was widespread interest in sums of independent Banach space valued random variables at that time. The famous paper of Itô and Nisio, 'On the convergence of sums of independent Banach space valued random variables', appeared in 1968. Jean-Pierre Kahane's book, 'Some random series of functions'

M. B. Marcus
CUNY Graduate Center, New York, NY, USA
e-mail: mbmarcus@optonline.net

G. Peskir
Department of Mathematics, The University of Manchester, Manchester, UK
e-mail: goran@maths.man.ac.uk

J. Rosiński (✉)
Department of Mathematics, University of Tennessee, Knoxville, TN, USA
e-mail: jrosinsk@utk.edu



N. Gozlan et al. (eds.), *High Dimensional Probability VIII*,
Progress in Probability 74, https://doi.org/10.1007/978-3-030-26391-1_1

(first edition), mostly dealing with random Fourier series, also came out in 1968. Functional analysts in the circle of Laurent Schwartz were using properties of sums of independent Banach space valued random variables to classify Banach spaces.

Engaged in this work, Jørgen published his most cited papers, 'Sums of independent Banach space valued random variables', as a publication of the Institute of Mathematics in Aarhus in 1972, and a paper with the same title, in Studia Mathematica in 1974 (cf. [9]). The two papers overlap but each has material that is not in the other. They contain the important and very useful relationship, between the norm of the maximal term in a series and the norm of the series, that is now commonly referred to as 'Hoffmann-Jørgensen's inequality'.

Continuing in this study, Jørgen collaborated on two important papers; with Gilles Pisier on the law of large numbers and the central limit theorem in Banach spaces [12], and with Richard Dudley and Larry Shepp on the lower tails of Gaussian seminorms [13]. He returned repeatedly to the topics of these and his other early papers, examining them in more general and abstract spaces. In this vein Jørgen reexamined the concept of weak convergence from a new perspective that completely changed the paradigm of its applications in statistics. He formulated his new definition of weak convergence in the 1980s[1]. This is now referred to as 'weak convergence in Hoffmann-Jørgensen's sense'.

Jørgen remained an active researcher throughout his life. He was completing a paper with Andreas Basse-O'Connor and Jan Rosiński on the extension of the Itô-Nisio theorem to non-separable Banach spaces, when he died.

Jørgen was also a very fine teacher and advisor with great concern for his students. He wrote 10 sets of lecture notes for his courses, 2,620 pages in total, and a monumental 1,184 page, two volume, 'Probability with a view toward Statistics', published by Chapman and Hall in 1994. He was the principal advisor of seven Ph.D. students.

Reflecting the interest in sums of independent Banach space valued random variables, and the related field of Gaussian processes in Europe, Laurent Schwarz and Jacques Neveu organized an auspicious conference on Gaussian Processes in Strasbourg in 1973. This stimulated research and collaborations that continue to this day. The Strasbourg conference was followed, every two or three years, by nine conferences on Probability in Banach Spaces and eight conferences on High Dimensional Probability. The last one was in Oaxaca, Mexico in 2017. The change in the conference name reflected a broadening of the interests of the participants.

Jørgen was one of a core group, many of whom attended the 1973 conference, who took part in all or most of the eighteen conferences throughout their careers, and often were the conference organizers and editors of the conference proceedings. Most significantly, Jørgen was the principal organizer of three of these conferences in the beautiful, serene, conference center in Sandbjerg, Denmark in 1986, 1993 and 2002, and was an editor of the proceedings of these conferences. Moreover, his

[1] Some authors have claimed, as we did in [14], that this definition was introduced in Jørgen's paper *Probability in Banach space* [10] in 1977. However, after a careful reading of this paper, we do not think that this is correct.

influence on the study of probability in Europe extended beyond these activities. In total, Jørgen served on the conference committees of eighteen meetings in Croatia, Denmark, Italy, France and Germany. Jørgen also served as an editor of the Journal of Theoretical Probability.

Jørgen was one of the mathematicians at Aarhus University who made Aarhus a focal point for generations of probabilists. But it was not only the research that brought them to Aarhus. Just as important was Jørgen's warmth and wit and not least of all the wonderful hospitality he and his wife Karen extended to all of them. Who can forget the fabulous Danish meals at their house, and then, sitting around after dinner, exchanging mathematical gossip and arguing politics, with the mating calls of hump backed whales playing in the background[2].

We now present some of Jørgen's better known results. This is not an attempt to place him in the history of probability but merely to mention some of his work that has been important to us and to give the reader a glimpse of his achievements.

**Hoffmann-Jørgensen's Inequality** Let $(X_n)$ be a sequence of independent symmetric random variables with values in a Banach space $E$ with norm $\|\cdot\|$. We define

$$S_n = \sum_{j=1}^{n} X_j, \quad N = \sup_n \|X_n\|, \quad M = \sup_n \|S_n\|.$$

Hoffmann-Jørgensen's inequality states that

$$\mathsf{P}(M \geq 2t + s) \leq 2\mathsf{P}(N \geq s) + 8\mathsf{P}^2(M \geq t) \tag{1.1}$$

for all $t, s > 0$.

Note that since probabilities are less than 1 and the last term in this inequality is a square it suggests that if $M$ has sufficient regularity the distribution of $M$ is controlled by the distribution of $N$. This is a remarkable result.

Jørgen gives this inequality in his famous paper [9]. He does not highlight it. It simply appears in the proof of his Theorem 3.1 which is:

**Theorem 1** *Let $(X_n)$ be a sequence of independent $E$-valued random variables such that*

$$\mathsf{P}(M < \infty) = 1 \quad \textit{and} \quad \mathsf{E}(N^p) < \infty$$

*for some $0 < p < \infty$. Then $\mathsf{E}(M^p) < \infty$.*

[2]The material up to this point has appeared in [14].

This is how he uses the inequality to prove this theorem. Assume that the elements of $(X_n)$ are symmetric and let $R(t) = \mathsf{P}(M \geq t)$ and $Q(t) = \mathsf{P}(N \geq t)$ for $t \geq 0$. Using the relationship

$$\mathsf{E}(M^p) = \int_0^\infty px^{p-1}R(x)dx,$$

and similarly for $N$ and $Q$, it follows from (1.1) that for $A > 0$

$$\begin{aligned}\int_0^A px^{p-1}R(x)dx &= p\,3^p \int_0^{A/3} px^{p-1}R(3x)dx \qquad (1.2)\\ &\leq 2p\,3^p \int_0^{A/3} px^{p-1}Q(x)dx + 8p\,3^p \int_0^{A/3} px^{p-1}R^2(x)dx\\ &\leq 2p\,3^p\mathsf{E}(N^p) + 8p\,3^p \int_0^{A/3} px^{p-1}R^2(x)dx.\end{aligned}$$

Choose $t_0 > 0$ such that $R(t_0) < (16p3^p)^{-1}$. The condition that $\mathsf{P}(M < \infty) = 1$ implies that $t_0 < \infty$. Then choose $A > 3t_0$. Note that

$$\begin{aligned}\int_0^{A/3} px^{p-1}R^2(x)dx &= \int_0^{t_0} px^{p-1}R^2(x)dx + \int_{t_0}^{A/3} px^{p-1}R^2(x)dx\\ &\leq t_0^p + R(t_0)\int_{t_0}^{A/3} px^{p-1}R(x)dx. \qquad (1.3)\end{aligned}$$

Combining (1.2) and (1.3) we get

$$\int_0^A px^{p-1}R(x)dx \leq 2p\,3^p\mathsf{E}(N^p) + t_0^p + \frac{1}{2}\int_0^{A/3} px^{p-1}R(x)dx. \qquad (1.4)$$

It follows from (1.4) that when the elements of $(X_n)$ are symmetric and $\mathsf{E}(N^p) < \infty$, then $\mathsf{E}(M^p) < \infty$. Eliminating the condition that $(X_n)$ is symmetric is routine.

Inequalities for sums of independent random variables that relate the sum to the supremum of the individual terms are often referred to as Hoffmann-Jørgensen type inequalities. Jørgen's original inequality has been generalized and extended. Many of these results are surveyed in [5] which obtains Hoffmann-Jørgensen type inequalities for $U$ statistics. See [4] for a more recent treatment of Hoffmann-Jørgensen type inequalities in statistics.

**Weak Convergence in Hoffmann-Jørgensen's Sense** The classic concept of *convergence in distribution*, dating back to de Moivre's central limit theorem in 1737, admits the following well-known characterisation, traditionally referred to as *weak convergence* (cf. [3]).

Let $(\Omega, \mathcal{F}, \mathsf{P})$ be a probability space, let $S$ be a metric (topological) space, and let $\mathcal{B}(S)$ be the Borel $\sigma$-algebra on $S$. Let $X_1, X_2, \ldots$ and $X$ be measurable functions from $\Omega$ to $S$ with respect to $\mathcal{F}$ and $\mathcal{B}(S)$. If

$$\lim_{n\to\infty} \mathsf{E} f(X_n) = \mathsf{E} f(X) \tag{1.5}$$

for every bounded continuous function $f : S \to \mathbb{R}$, then we say that $X_n$ *converges weakly* to $X$, and following Jørgen's notation, write

$$X_n \overset{\sim}{\to} X \tag{1.6}$$

as $n \to \infty$. The expectation $\mathsf{E}$ in (1.5) is defined as the (Lebesgue-Stieltjes) integral with respect to the ($\sigma$-additive) probability measure $\mathsf{P}$.

The state space $S$ in classical examples is finite dimensional, e.g. $\mathbb{R}$ or $\mathbb{R}^n$ for $n \geq 2$. The main motivation for Jørgen's reconsideration of (1.5) and (1.6) comes from the empirical processes theory. Recall that the *empirical distribution function* is given by

$$F_n(t, \omega) := \frac{1}{n} \sum_{i=1}^{n} I(\xi_i(\omega) \leq t) \tag{1.7}$$

for $n \geq 1$, $t \in [0, 1]$ and $\omega \in \Omega$, where $\xi_1, \xi_2, \ldots$ are independent and identically distributed random variables on $\Omega$ taking values in $[0, 1]$ and having the common distribution function $F$. In this setting, motivated by the classical central limit theorem, one forms the *empirical process*

$$X_n(t, \omega) := \sqrt{n}\,\big(F_n(t, \omega) - F(t)\big) \tag{1.8}$$

and aims to establish that $X_n$ converges 'weakly' to a limiting process $X$ (of a Brownian bridge type) as $n \to \infty$. A substantial difficulty arises immediately because the mapping $X_n : \Omega \to S$ is *not measurable* when $S$ is taken to be the set of all right-continuous functions $x : [0, 1] \to \mathbb{R}$ with left-limits, equipped with the supremum norm $\|x\|_\infty = \sup_{t\in[0,1]} |x(t)|$ as a natural choice.

Skorokhod solved this measurability problem in 1956 by creating a different metric on $S$, for which the Borel $\sigma$-algebra coincides with the cylinder $\sigma$-algebra, so that each $X_n$ is measurable. For more general empirical processes

$$X_n(f, \omega) := \sqrt{n}\,\Big(\frac{1}{n} \sum_{i=1}^{n} f(X_i(\omega)) - \mathsf{E} f(X_1)\Big) \tag{1.9}$$

indexed by $f$ belonging to a family of functions, there is no obvious way to extend the Skorokhod approach. Jørgen solved this measurability problem in the most elegant way by simply replacing the first expectation $\mathsf{E}$ in (1.5) by the *outer* *expectation* $\mathsf{E}^*$, which is defined by

$$\mathsf{E}^* Y = \inf\{\, \mathsf{E} Z \mid Z \geq Y \text{ is measurable} \,\} \tag{1.10}$$

where $Y$ is any (not necessarily measurable) function from $\Omega$ to $\mathbf{R}$, and leaving the second expectation $\mathsf{E}$ in (1.5) unchanged (upon assuming that the limit $X$ is measurable).

This definition of *weak convergence in Hoffmann-Jørgensen's sense* is given for the first time in his monograph [11, page 149]. Although [11] was published in 1991, a draft of the monograph was available in Aarhus and elsewhere since 1984. Furthermore, the first paper [1] which uses Jørgen's new definition was published in 1985. Jørgen's definition of weak convergence became standard soon afterwards. It continues to be widely used.

It is now known that replacing the first $\mathsf{E}$ in (1.5) by $\mathsf{E}^*$ is equivalent to replacing it by $\mathsf{E}^Q$ where $Q$ is any *finitely additive* extension of $\mathsf{P}$ from $\mathcal{F}$ to $2^\Omega$ (see Theorem 4 in [2] for details). This revealing equivalence just adds to both simplicity and depth of Jørgen's thought when opting for $\mathsf{E}^*$ in his celebrated definition.

**Hoffmann-Jørgensen's Work on Measure Theory** As measure theory matured, difficult measurability problems arose in various areas of mathematics that could not be solved in general measure spaces. Consequently, new classes of measure spaces were introduced, such as *analytic spaces*, also called *Souslin spaces*, defined by Lusin and Souslin and further developed by Sierpiński, Kuratowski and others. For many years analytic spaces received little attention until important applications were found in potential theory by Choquet and group representation theory by Mackey. Analytic spaces were also found to be important in the theory of convex sets, and other branches of mathematics.

Stimulated by these developments, Jørgen undertook a deep study of analytic spaces early in his academic career, resulting in his monograph 'The Theory of Analytic Spaces' [7]. This monograph contains many original, and carefully presented results, that are hard to find elsewhere. For example, from Jørgen's Section Theorem, [7, Theorem 1, page 84], one can derive all of the most commonly used section and selection theorems in the literature.

The final chapter of the monograph is devoted to locally convex vector spaces, where it is shown that all of the locally convex spaces that are of interest to researchers are analytic spaces. As Jørgen wrote "The importance of analytic spaces lies in the fact that even though the category is sufficiently small to exclude all pathological examples ..., it is sufficiently large to include all (or almost all) interesting and important examples of topological measure spaces."

In one of his first papers [6] listed in Mathematical Reviews and Zentralblatt, Jørgen investigates extensions of regenerative events to continuous state spaces, a problem proposed to him by P.-A. Meyer. In his subsequent paper [8], he makes the surprising observation that the existence of a measurable modification of a stochastic process depends only on its 2-dimensional marginal distributions. He then gives necessary and sufficient conditions for the existence of such a modification for the process $(X_t)_{t \in T}$ with values in a complete separable metric space $K$, expressed in terms of the kernel

$$Q(s, t, A) = \mathsf{P}((X_s, X_t) \in A)$$

where $T$ is a separable metric space, $s, t \in T$, and $A \in \mathcal{B}(K^2)$. Jørgen's interest in measure theory aspects of probability continued throughout his career.

## References


1. N.T. Andersen, The central limit theorem for nonseparable valued functions. Z. Wahrsch. Verw. Gebiete **70**, 445–455 (1985)
2. P. Berti, P. Rigo, Convergence in distribution of nonmeasurable random elements. Ann. Probab. **32**, 365–379 (2004)
3. P. Billingsley, *Convergence of Probability Measures* (Willey, New York, 1968)
4. E. Giné, R. Nickl, *Mathematical Foundations of Infinite-Dimensional Statistical Models* (Cambridge University Press, New York, 2016)
5. R. Giné, E. Latała, J. Zinn, Exponential and moment inequalities for U-statistics, in *High Dimensional Probability II* (Seattle 1999). Programs and Probability, vol. 47 (Birkhäuser, Boston, 2000), pp. 13–38
6. J. Hoffmann-Jørgensen, Markov sets. Math. Scand. **24**, 145–166 (1969)
7. J. Hoffmann-Jørgensen, *The Theory of Analytic Spaces*, vol. 10 (Matematisk Institut, Aarhus Universitet, Aarhus, 1970)
8. J. Hoffmann-Jørgensen, Existence of measurable modifications of stochastic processes. Z. Wahrsch. Verw. Gebiete **25**, 205–207 (1973)
9. J. Hoffmann-Jørgensen, Sums of independent Banach space valued random variables. Stud. Math. **52**, 159–186 (1974)
10. J. Hoffmann-Jørgensen, *Probability in Banach Space*. Lecture Notes in Mathematics, vol. 598 (Springer, Berlin, 1977), pp. 1–186
11. J. Hoffmann-Jørgensen, *Stochastic Processes on Polish Spaces*, vol. 39 (Matematisk Institut, Aarhus Universitet, Aarhus, 1991)
12. J. Hoffmann-Jørgensen, G. Pisier, The law of large numbers and the central limit theorem in Banach spaces. Ann. Probab. **4**, 587–599 (1976)
13. J. Hoffmann-Jørgensen, L.A. Shepp, R.M. Dudley, On the lower tail of Gaussian seminorms. Ann. Probab. **7**, 319–342 (1979)
14. M.B. Marcus, G. Peskir, J. Rosiński, *Jørgen Hoffmann-Jørgensen (1942–2017)*, vol. 54 (Danish Mathematical Society, Matilde, 2018), pp. 14–15

# Chapter 2
# Moment Estimation Implied by the Bobkov-Ledoux Inequality

**Witold Bednorz and Grzegorz Głowienko**

**Abstract** In this paper we consider a probability measure on the high dimensional Euclidean space satisfying Bobkov-Ledoux inequality. Bobkov and Ledoux have shown in (Probab Theory Related Fields 107(3):383–400, 1997) that such entropy inequality captures concentration phenomenon of product exponential measure and implies Poincaré inequality. For this reason any measure satisfying one of those inequalities shares the same concentration result as the exponential measure. In this paper using B-L inequality we derive some bounds for exponential Orlicz norms for any locally Lipschitz function. The result is close to the question posted by Adamczak and Wolff in (Probab Theory Related Fields 162:531–586, 2015) regarding moments estimate for locally Lipschitz functions, which is expected to result from B-L inequality.

**Keywords** Concentration of measure · Poincaré inequality · Sobolev inequality

**Subject Classification** 60E15, 46N30

## 2.1 The Bobkov-Ledoux Inequality

Let $\mu$ be a probability measure on $\mathbb{R}^d$. We assume that $\mu$ satisfies Bobkov-Ledoux inequality i.e. with fixed $D > 0$, for any positive, locally Lipschitz function $f$ such that $|\nabla f|_\infty \leqslant f/2$ we have

$$\mathbf{Ent}_\mu f^2 \leqslant D\mathbf{E}_\mu |\nabla f|_2^2. \tag{2.1}$$

As noticed by Bobkov and Ledoux in [3] this modification of log-Sobolev inequality is satisfied by product exponential measure, but more importantly, it implies

W. Bednorz (✉) · G. Głowienko
Institute of Mathematics, University of Warsaw, Warszawa, Poland
e-mail: wbednorz@mimuw.edu.pl; g.glowienko@mimuw.edu.pl



N. Gozlan et al. (eds.), *High Dimensional Probability VIII*,
Progress in Probability 74, https://doi.org/10.1007/978-3-030-26391-1_2

subexponential concentration. It is also quite easy to show that it implies Poincaré inequality. For any smooth function $g$ we may take $f = 1 + \epsilon g$ and $\epsilon > 0$ such that $|\nabla f|_\infty \leqslant f/2$, which allows us to apply (2.1). In the next step divide both sides of inequality by $\epsilon^2$, consider standard Taylor expansion and take limit with $\epsilon$ tending to 0. As a result

$$\mathbf{Var}_\mu g \leqslant \frac{D}{2}\mathbf{E}_\mu |\nabla g|_2^2, \tag{2.2}$$

which is exactly the Poincaré inequality. Finally just notice that any locally Lipschitz function $f$ such that both $f$ and $|\nabla f|_2$ are square integrable w.r.t. $\mu$ may be approximated in (2.2) by smooth functions. The result means that B-L inequality (2.1) is stronger than Poincaré inequality (2.2), nevertheless both inequalities imply concentration phenomenon of product exponential measure, therefore any measure satisfying one of those inequalities shares the same concentration result. See [3] for more details regarding this subtle connection.

As we are dealing with big number of constants in the following section, it would be wise to adopt some useful convention. Therefore, let us denote by $D'$ numeric constant which may vary from line to line, but importantly, it is comparable to $D$ from log-Sobolev inequality (2.1). Similarly let $C$ be constant comparable to 1 and by $C(\alpha)$ denote one that depends on $\alpha$ only.

In [4] it was noticed by E. Milman that, Poincaré inequality (2.2) implies the following estimate for $p \geqslant 1$

$$\|f - \mathbf{E}_\mu f\|_p \leqslant \sqrt{D'}p\||\nabla f|_2\|_p, \tag{2.3}$$

with $f$ locally Lipschitz. It is easy to see that above results with the following bound

$$\|f - \mathbf{E}_\mu f\|_p \leqslant \sqrt{D'}p\sqrt{d}\||\nabla f|_\infty\|_p.$$

Adamczak and Wolff has conjectured in [1] that Bobkov-Ledoux inequality (2.1) imply

$$\|f - \mathbf{E}_\mu f\|_p \leqslant \sqrt{D'}\sqrt{p}\||\nabla f|_2\|_p + Cp\||\nabla f|_\infty\|_p.$$

They also proved following weaker form of the conjecture

$$\|f - \mathbf{E}_\mu f\|_p \leqslant \sqrt{D'}\sqrt{p}\||\nabla f|_2\|_p + Cp\||\nabla f|_\infty\|_\infty. \tag{2.4}$$

Their result is based on tricky modification of given function so that (2.1) could be used. In our paper we are trying to understand this phenomenon and apply its more advanced form.

## 2.2 Bounds for Moments

In this section we investigate possible estimates for $\|g\|_{p\alpha}$, with a given $\alpha > 0$, when we know that $g^\alpha$ is globally Lipschitz. This bounds will be useful when we start dealing with the exponential Orlicz norms.

**Theorem 2.1** *If measure $\mu$ satisfies (2.1), function $g$ is non-negative, locally Lipschitz and $p \geqslant 1$, then*
*for* $0 < \alpha \leqslant 1$

$$\|g\|_{p\alpha} \leqslant 2^{\frac{1}{\alpha}} \max\left\{p^{\frac{1}{\alpha}} \left\| |\nabla g^\alpha|_\infty \right\|_\infty^{\frac{1}{\alpha}}, \|g\|_{2\alpha}, \alpha p^{\frac{1}{2}} \sqrt{D'} \| |\nabla g|_2 \|_{p\alpha}\right\}$$

*and in case of* $\alpha > 1$

$$\|g\|_{p\alpha} \leqslant \max\left\{2^{\frac{1}{\alpha}} p^{\frac{1}{\alpha}} \left\| |\nabla g^\alpha|_\infty \right\|_\infty^{\frac{1}{\alpha}}, 2^{\frac{1}{\alpha}} \|g\|_{2\alpha}, \alpha p^{\frac{1}{2}} \sqrt{D'} \| |\nabla g|_2 \|_{p\alpha}\right\}.$$

*Proof* Consider $g^\alpha$ to be a non-negative Lipschitz function, otherwise estimate is trivial. Note that in case of $p \leqslant 2$ there is also nothing to prove, therefore we may take $p > 2$. For simplicity let us assume that $\| |\nabla g^\alpha|_\infty \|_\infty = 1$. If it happens to be

$$\|g\|_{p\alpha}^\alpha \leqslant 2p \left\| |\nabla g^\alpha|_\infty \right\|_\infty \tag{2.5}$$

then proof is once again trivial, therefore assume that

$$\|g\|_{p\alpha}^\alpha > 2p \left\| |\nabla g^\alpha|_\infty \right\|_\infty, \tag{2.6}$$

then following the idea of the proof of (2.4) from [1] we define function $h = \max\{g, c\}$, where $c = \|g\|_{p\alpha}/2^{\frac{1}{\alpha}}$. Obviously, for $2 \leqslant t \leqslant p$

$$\frac{|\nabla h^{\alpha t/2}|_\infty}{h^{\alpha t/2}} = \frac{t}{2} \frac{|\nabla h^\alpha|_\infty}{h^\alpha}.$$

Due to our definition $h \geqslant c$ and $|\nabla h^\alpha|_\infty \leqslant |\nabla g^\alpha|_\infty$, which gives us

$$\frac{|\nabla h^\alpha|_\infty}{h^\alpha} \leqslant \frac{2|\nabla g^\alpha|_\infty}{\|g\|_{p\alpha}^\alpha}.$$

Combining above with (2.6) we get

$$\left\| \frac{|\nabla h^{\alpha t/2}|_\infty}{h^{\alpha t/2}} \right\|_\infty \leqslant \frac{t}{2p} \leqslant \frac{1}{2}.$$

Therefore, we may apply (2.1) to the function $h^{\alpha t/2}$ and thus by the Aida Stroock [2] argument i.e.

$$\frac{d}{dt}\|h^\alpha\|_t^2 = \frac{2}{t^2}\big(\mathbf{E}h^{\alpha t}\big)^{2/t-1}\mathbf{Ent}(h^{\alpha t/2})^2 \leqslant \frac{D}{2}\big(\mathbf{E}h^{\alpha t}\big)^{2/t-1}\mathbf{E}|h^{\alpha t/2-\alpha}\nabla h^\alpha|_2^2,$$

combined with Hölder inequality with exponents $t/(t-2)$ and $t/2$ applied to the last term, gives us

$$\frac{d}{dt}\|h^\alpha\|_t^2 \leqslant \frac{D}{2}\big(\mathbf{E}h^{\alpha t}\big)^{2/t-1}\big(\mathbf{E}h^{\alpha t}\big)^{1-2/t}\big(\mathbf{E}|\nabla h^\alpha|_2^t\big)^{2/t} = \frac{D}{2}\big\||\nabla h^\alpha|_2\big\|_t^2.$$

The moment function (as function of $t$) is non-decreasing, therefore for $2 \leqslant t \leqslant p$ we get

$$\|h^\alpha\|_p^2 - \|h^\alpha\|_2^2 \leqslant \frac{D}{2}(p-2)\||\nabla h^\alpha|_2\|_p^2. \tag{2.7}$$

Now we have to consider two cases. First suppose that $\alpha \leqslant 1$ and then

$$\||\nabla h^\alpha|_2\|_p \leqslant \alpha\||\nabla g|_2 h^{\alpha-1}\|_p \leqslant \alpha c^{\alpha-1}\||\nabla g|_2\|_p$$

and combining this with (2.7), we infer

$$\|h^\alpha\|_p^2 \leqslant \|h^\alpha\|_2^2 + \frac{\alpha^2 D}{2}(p-2)c^{2\alpha-2}\||\nabla g|_2\|_p^2.$$

Now observe that $\|h^\alpha\|_p^2 \geqslant \|g^\alpha\|_p^2$ and furthermore

$$\|h^\alpha\|_2^2 \leqslant c^{2\alpha} + \|g^\alpha\|_2^2 \leqslant \frac{1}{4}\|g^\alpha\|_p^2 + \|g^\alpha\|_2^2,$$

which combined together gives us

$$\frac{3}{4}\|g^\alpha\|_p^2 \leqslant \|g^\alpha\|_2^2 + \frac{\alpha^2 D}{2}(p-2)c^{2\alpha-2}\||\nabla g|_2\|_p^2. \tag{2.8}$$

Noting that the case of

$$\|g\|_{p\alpha}^\alpha \leqslant 2\|g\|_{2\alpha}^\alpha, \tag{2.9}$$

is another trivial part, we assume conversely getting

$$\|g^\alpha\|_2^2 = \|g\|_{2\alpha}^{2\alpha} \leqslant \frac{1}{4}\|g\|_{p\alpha}^{2\alpha} = \frac{1}{4}\|g^\alpha\|_p^2$$

which together with (2.8) implies that

$$\|g\|_{p\alpha}^{2\alpha} \leqslant \alpha^2 D(p-2)c^{2\alpha-2}\||\nabla g|_2\|_p^2. \tag{2.10}$$

Reminding that $c^\alpha = 2^{-1}\|g\|_{p\alpha}^\alpha$ we infer

$$\|g\|_{p\alpha}^2 \leqslant 2^{\frac{2}{\alpha}-2}\alpha^2 D(p-2)\||\nabla g|_2\|_p^2$$

and rewriting it in simplified form

$$\|g\|_{p\alpha} \leqslant 2^{\frac{1}{\alpha}}\alpha\sqrt{D'}p^{\frac{1}{2}}\||\nabla g|_2\|_p. \tag{2.11}$$

Combining together (2.5), (2.9), and (2.11) implies the result in the case of $0 < \alpha \leqslant 1$.

Consider now case of $\alpha > 1$, following the same reasoning as in previous case, up to the (2.7) after that Hölder inequality is used, we get

$$\||\nabla h^\alpha|_2\|_p \leqslant \alpha\||\nabla g|_2 h^{\alpha-1}\|_p \leqslant \alpha\||\nabla g|_2\|_{p\alpha}\|h\|_{p\alpha}^{\alpha-1}.$$

Therefore, by (2.7)

$$\|h\|_{p\alpha}^2\Big(1-\frac{\|h\|_{2\alpha}^{2\alpha}}{\|h\|_{p\alpha}^{2\alpha}}\Big) \leqslant \alpha^2\frac{D}{2}(p-2)\||\nabla g|_2\|_{p\alpha}^2. \tag{2.12}$$

Again, either (2.9) holds or we have

$$\|h\|_{2\alpha}^{2\alpha} = \|h^\alpha\|_2^2 \leqslant c^{2\alpha} + \|g^\alpha\|_2^2 = \frac{1}{4}\|g^\alpha\|_p^2 + \frac{1}{4}\|g^\alpha\|_p^2 = \frac{1}{2}\|g\|_{p\alpha}^{2\alpha}.$$

Since obviously $\|h\|_{p\alpha}^{2\alpha} \geqslant \|g\|_{p\alpha}^{2\alpha}$, we get

$$\|h\|_{p\alpha}^2\Big(1-\frac{\|h\|_{2\alpha}^{2\alpha}}{\|h\|_{p\alpha}^{2\alpha}}\Big) \geqslant 2^{-1}\|g\|_{p\alpha}^2$$

and combining above with (2.12) gives us

$$\|g\|_{p\alpha} \leqslant \alpha\sqrt{D'}p^{\frac{1}{2}}\||\nabla g|_2\|_p. \tag{2.13}$$

Clearly (2.5), (2.9), and (2.13) cover the case of $\alpha > 1$, which ends whole proof. ■

Next step of the reasoning is to apply previous result to $g = |f - \mathbf{E}_\mu f|$ and combine it with Poincaré inequality. Let us gather everything together in form of

**Corollary 2.1** *If measure $\mu$ satisfies (2.1), function $f$ is locally Lipschitz and $p \geqslant 1$, then for $0 < \alpha \leqslant 1$*

$$\|f - \mathbf{E}_\mu f\|_{p\alpha} \leqslant 2^{\frac{1}{\alpha}} \max\Big\{p^{\frac{1}{\alpha}} \Big\| |\nabla| f - \mathbf{E}_\mu f|^\alpha|_\infty \Big\|_\infty^{\frac{1}{\alpha}},$$
$$\sqrt{D'}\big\| |\nabla f|_2 \big\|_2, \alpha p^{\frac{1}{2}} \sqrt{D'} \big\| |\nabla f|_2 \big\|_{p\alpha}\Big\}.$$

*and in case of $\alpha > 1$*

$$\|f - \mathbf{E}_\mu f\|_{p\alpha} \leqslant \max\Big\{2^{\frac{1}{\alpha}} p^{\frac{1}{\alpha}} \Big\| |\nabla| f - \mathbf{E}_\mu f|^\alpha|_\infty \Big\|_\infty^{\frac{1}{\alpha}},$$
$$2^{\frac{1}{\alpha}} \alpha \sqrt{D'} \big\| |\nabla f|_2 \big\|_{2\alpha}, \alpha p^{\frac{1}{2}} \sqrt{D'} \big\| |\nabla f|_2 \big\|_{p\alpha}\Big\}.$$

*Proof* If we fix $g = |f - \mathbf{E}_\mu f|$ then by the Poincaré inequality

$$\|f - \mathbf{E}_\mu f\|_{2\alpha} \leqslant (\alpha \vee 1)\sqrt{D'} \big\| |\nabla f|_2 \big\|_{2(\alpha \vee 1)}.$$

Note also that

$$\big\| |\nabla g|_2 \big\|_{p\alpha} = \big\| |\nabla f|_2 \big\|_{p\alpha},$$

then applying Theorem 2.1 statement easily follows. ■

## 2.3 Bounds for Exponential Orlicz Norms

First, let us recall the notion of exponential Orlicz norms. For any $\alpha > 0$

$$\|f\|_{\varphi(\alpha)} = \inf\{s > 0 : \ \mathbf{E}_\mu \exp(|f|^\alpha / s^\alpha) \leqslant 2\}.$$

Obviously, $\|f\|_{\varphi(\alpha)}$ is a norm in case of $\alpha \geqslant 1$ only, otherwise there is a problem with the triangle inequality. Moreover, we have $\|f\|_{\varphi(\alpha)} = \||f|^\alpha\|_{\varphi(1)}^{\frac{1}{\alpha}}$. Nevertheless, in case of $0 < \alpha < 1$ one can use

$$\|f + g\|_{\varphi(\alpha)} = \||f + g|^\alpha\|_{\varphi(1)}^{\frac{1}{\alpha}}$$
$$\leqslant \||f|^\alpha + |g|^\alpha\|_{\varphi(1)}^{\frac{1}{\alpha}} \leqslant (\||f|^\alpha\|_{\varphi(1)} + \||g|^\alpha\|_{\varphi(1)})^{\frac{1}{\alpha}}$$
$$= (\|f\|_{\varphi(\alpha)}^\alpha + \|g\|_{\varphi(\alpha)}^\alpha)^{\frac{1}{\alpha}} \leqslant 2^{\frac{1}{\alpha}-1}(\|f\|_{\varphi(\alpha)} + \|g\|_{\varphi(\alpha)}).$$

It is worth to know that $\|f\|_{\varphi(\alpha)}$ is always comparable with $\sup_{k\geqslant 1}\frac{\|f\|_{k\alpha}}{k^{1/\alpha}}$. More precisely, observe that for all $k \geqslant 1$ and a positive $g$

$$\frac{\|g\|_{k\alpha}^{k\alpha}}{k!} \leqslant \|g\|_{\varphi(\alpha)}^{k\alpha}.$$

Note that, just by the definition of $\|g\|_{\varphi(\alpha)}$, there exists $k \geqslant 1$ for which

$$\frac{\|g\|_{k\alpha}^{k\alpha}}{k!} \geqslant 2^{-k}\|g\|_{\varphi(\alpha)}^{k\alpha}.$$

Let us denote the set of such $k \geqslant 1$ by $J(g,\alpha)$ and note that for any $k \in J(g,\alpha)$

$$(k!)^{-\frac{1}{k\alpha}}\|g\|_{k\alpha} \leqslant \|g\|_{\varphi(\alpha)} \leqslant 2^{\frac{1}{\alpha}}(k!)^{-\frac{1}{k\alpha}}\|g\|_{k\alpha}. \tag{2.14}$$

Next let $M \geqslant e$ be such a constant that $(k!)^{\frac{1}{k}} \geqslant k/M$ for all $k \geqslant 1$. We have following crucial observation namely for all $k \in J(g,\alpha)$

$$\|g\|_{\varphi(\alpha)} \leqslant (2M)^{\frac{1}{\alpha}}\frac{\|g\|_{k\alpha}}{k^{\frac{1}{\alpha}}}. \tag{2.15}$$

Therefore, we may use Theorem 2.1 in order to obtain

**Corollary 2.2** *If $\mu$ satisfies (2.1) and $g$ is non-negative locally Lipschitz function, then for any $k \in J(g,\alpha)$ in case of* $0 < \alpha \leqslant 1$

$$\|g\|_{\varphi(\alpha)} \leqslant (4M)^{\frac{1}{\alpha}}\max\left\{\left\||\nabla g^{\alpha}|_{\infty}\right\|_{\infty}^{\frac{1}{\alpha}}, k^{-\frac{1}{\alpha}}\|g\|_{2\alpha}, \alpha k^{\frac{1}{2}-\frac{1}{\alpha}}\sqrt{D'}\left\||\nabla g|_2\right\|_{k\alpha}\right\}.$$

*and for* $1 < \alpha \leqslant 2$

$$\|g\|_{\varphi(\alpha)} \leqslant (4M)^{\frac{1}{\alpha}}\max\left\{\left\||\nabla g^{\alpha}|_{\infty}\right\|_{\infty}^{\frac{1}{\alpha}}, k^{-\frac{1}{\alpha}}\|g\|_{2\alpha}, 2^{-\frac{1}{\alpha}}\alpha k^{\frac{1}{2}-\frac{1}{\alpha}}\sqrt{D'}\left\||\nabla g|_2\right\|_{k\alpha}\right\}.$$

Note that set $J(g,\alpha)$ is stable with respect to $g \mapsto h$, where $h = \max\{g,c\}$ i.e. if $c$ is comparable to $\|g\|_{\varphi(\alpha)}$ there exists $C \geqslant 1$ such that for $k \in J(g,\alpha)$

$$\frac{\|h\|_{k\alpha}^{k\alpha}}{k!} \geqslant \frac{1}{C^k}\|h\|_{\varphi(\alpha)}^{k\alpha},$$

which means that we cannot easily improve the result using the trick.

In the same way as we have established Corollary 2.1 we can deduce the following result.

**Corollary 2.3** *If $\mu$ satisfies (2.1) and $g$ is locally Lipschitz function, then for any $k \in J(g,\alpha)$ in case of $0 < \alpha \leqslant 1$*

$$\|f - \mathbf{E}_\mu f\|_{\varphi(\alpha)} \leqslant (4M)^{\frac{1}{\alpha}} \max\Big\{\Big\| |\nabla|f - \mathbf{E}_\mu f|^\alpha|_\infty \Big\|_\infty^{\frac{1}{\alpha}},$$
$$k^{-\frac{1}{\alpha}}\sqrt{D'}\big\||\nabla f|_2\big\|_2, \alpha k^{\frac{1}{2}-\frac{1}{\alpha}}\sqrt{D'}\big\||\nabla f|_2\big\|_{k\alpha}\Big\}.$$

*and for $\alpha > 1$*

$$\|f - \mathbf{E}_\mu f\|_{\varphi(\alpha)} \leqslant (4M)^{\frac{1}{\alpha}} \max\Big\{\Big\| |\nabla|f - \mathbf{E}_\mu f|^\alpha|_\infty \Big\|_\infty^{\frac{1}{\alpha}},$$
$$k^{-\frac{1}{\alpha}}\sqrt{D'}\big\||\nabla f|_2\big\|_2, 2^{-\frac{1}{\alpha}}\alpha k^{\frac{1}{2}-\frac{1}{\alpha}}\sqrt{D'}\big\||\nabla f|_2\big\|_{k\alpha}\Big\}.$$

A simple consequence of the above is

**Corollary 2.4** *If $\mu$ satisfies (2.1) and $0 < \alpha \leqslant 2$, then for any locally Lipschitz function $f$*

$$\|f - \mathbf{E}_\mu f\|_{\varphi(\alpha)} \leqslant C(\alpha)\Big(\Big\| |\nabla|f - \mathbf{E}_\mu f|^\alpha|_\infty \Big\|_\infty^{\frac{1}{\alpha}} + \sqrt{D'}\big\||\nabla f|_2\big\|_{\varphi(\frac{2\alpha}{2-\alpha})}\Big).$$

The result shows that at least for globally Lipschitz function $|f|^\alpha$, $\alpha \leqslant 1$ the exponential moment $\|f - \mathbf{E}_\mu f\|_{\varphi(\alpha)}$ has to bounded, though it is still far from replacement of $\||\nabla|f - \mathbf{E}_\mu f|^\alpha|_\infty\|_\infty^{\frac{1}{\alpha}}$ by the expected $\||\nabla f|_\infty\|_{\varphi(\frac{\alpha}{1-\alpha})}$.

Note that it is not possible to simply replace the constant $C(\alpha) \sim (4M)^{\frac{1}{\alpha}}$ in Corollary 2.4 by 1 which would be a natural choice for the question. In the next section we will show another approach which allows to obtain such a result.

## 2.4 Another Approach

**Theorem 2.2** *If $\mu$ satisfies (2.1) and $0 < \alpha \leqslant 2$, then for any locally Lipschitz function $f$*

$$\|f - \mathbf{E}_\mu f\|_{\varphi(\alpha)} \leqslant \Big\| |\nabla|f - \mathbf{E}_\mu f|^\alpha|_\infty \Big\|_\infty^{\frac{1}{\alpha}} + C(\alpha)\sqrt{D'}\big\||\nabla f|_2\big\|_{\varphi(\frac{2\alpha}{2-\alpha})}$$

*where $C(\alpha) = \alpha\left(\frac{2}{\ln 2}\right)^{\frac{1}{\alpha}}$.*

*Proof* Let $g^\alpha$ be a non-negative Lipschitz function, we may assume that

$$\big\||\nabla g^\alpha|_\infty\|_\infty = \alpha\big\||\nabla g|_\infty g^{\alpha-1}\|_\infty \leqslant 1.$$

Then for any $t \leqslant 1$ and a function $h = \exp(g^\alpha t/2)$ we can apply (2.1), indeed

$$\left\| \frac{|\nabla h|_\infty}{h} \right\|_\infty = \frac{t}{2} \left\| |\nabla g^\alpha|_\infty \right\|_\infty \leqslant \frac{1}{2}$$

In fact there are three possibilities we should acknowledge.

The first case we should consider is $\mathbf{E}_\mu \exp(g^\alpha) \leqslant 2$, but then

$$\|g\|_{\varphi(\alpha)} \leqslant 1. \tag{2.16}$$

Otherwise there must exist $t_* \leqslant 1$ such that $\mathbf{E}\exp(g^\alpha t_*) = 2$. Clearly $1/t_*^{\frac{1}{\alpha}} = \|g\|_{\varphi(\alpha)}$. For simplicity let us denote $V(t) = \ln \mathbf{E}\exp(g^\alpha t)$, $t \geqslant 0$. It is well known that $V$ is convex, increasing and $V(0) = 0$. Now we use (2.1), in order to get for all $t \in [0, 1]$

$$\left(\frac{V(t)}{t}\right)' \leqslant \frac{D}{4} \mathbf{E}_\mu |\nabla g^\alpha|_2^2 \exp(g^\alpha t - V(t)). \tag{2.17}$$

Note that $V(0)' = \mathbf{E}_\mu g^\alpha$. Moreover, for $0 \leqslant t \leqslant t_*$ we have $\frac{1}{2} \leqslant \exp(-V(t)) \leqslant 1$, so we can rewrite (2.17) in the following form

$$\left(\frac{V(t)}{t}\right)' \leqslant \frac{D}{4} \mathbf{E}_\mu |\nabla g^\alpha|_2^2 e^{g^\alpha t}. \tag{2.18}$$

Since $V$ is convex $V(0) = 0$ we know that $V(t)/t$ is increasing and also $V'(0) = \mathbf{E}_\mu g^\alpha$. Consequently, integrating (2.18) on $[0, t_*]$

$$\frac{V(t_*)}{t_*} - \mathbf{E}_\mu g^\alpha \leqslant \frac{D}{4} \sum_{k=0}^\infty \mathbf{E}_\mu |\nabla g^\alpha|_2^2 \frac{g^{k\alpha} t_*^{k+1}}{(k+1)!}.$$

Note that $V(t_*) = \ln 2$, so

$$\ln 2 \leqslant t_* \mathbf{E}_\mu g^\alpha + \frac{D}{4} \sum_{k=0}^\infty \frac{t_*^{k+2}}{(k+1)!} \mathbf{E}_\mu |\nabla g^\alpha|_2^2 g^{k\alpha}.$$

The second case which should be considered is when $t_*$ is very close to $\mathbf{E}_\mu g^\alpha$. If $t_* \mathbf{E}_\mu g^\alpha > \frac{1}{2} \ln 2$, then

$$\|g\|_{\varphi(\alpha)} = \frac{1}{t_*^{\frac{1}{\alpha}}} \leqslant \left(\frac{2}{\ln 2}\right)^{\frac{1}{\alpha}} \|g\|_\alpha. \tag{2.19}$$

For the last part of the proof we assume that

$$t_*\mathbf{E}_\mu g^\alpha \leqslant \frac{1}{2}\ln 2.$$

Obviously, we have then

$$\frac{\ln 2}{2} \leqslant \frac{D}{4}\sum_{k=0}^{\infty}\frac{t_*^{k+2}}{(k+1)!}\mathbf{E}_\mu|\nabla g^\alpha|_2^2 g^{k\alpha}.$$

Using the Hölder inequality, we get

$$\begin{aligned}&\mathbf{E}_\mu|\nabla g^\alpha|_2^2 g^{k\alpha} = \alpha^2\mathbf{E}_\mu|\nabla g|_2^2 g^{(k+2)\alpha-2}\\ &\leqslant \alpha^2\||\nabla g|_2\|_{(k+2)\alpha}^2\|g\|_{(k+2)\alpha}^{(k+2)\alpha-2} = \frac{\alpha^2\||\nabla g|_2\|_{(k+2)\alpha}^2}{\|g\|_{(k+2)\alpha}^2}\mathbf{E}_\mu g^{(k+2)\alpha}.\end{aligned}$$

Therefore,

$$\frac{1}{2}\ln 2 \leqslant \frac{D\alpha^2}{4}\sum_{k=0}^{\infty}\frac{(k+2)\||\nabla g|_2\|_{(k+2)\alpha}^2}{\|g\|_{(k+2)\alpha}^2}\frac{t_*^{k+2}\mathbf{E}_\mu g^{(k+2)\alpha}}{(k+2)!}. \tag{2.20}$$

Now we split all the indices $k$ into two classes.

$$I = \{k \geqslant 0:\ \|g\|_{(k+2)\alpha} \leqslant \frac{(k+2)^{\frac{1}{\alpha}}}{M^{\frac{1}{\alpha}}t_*^{\frac{1}{\alpha}}}\},\quad J = \{k \geqslant 0:\ \|g\|_{(k+2)\alpha} > \frac{(k+2)^{\frac{1}{\alpha}}}{M^{\frac{1}{\alpha}}t_*^{\frac{1}{\alpha}}}\},$$

where the constant $M \geqslant 1$ will be chosen later. First, we bound summands over the set $I$, i.e.

$$\begin{aligned}&\sum_{k\in I}\frac{(k+2)\||\nabla g|_2\|_{(k+2)\alpha}^2}{\|g\|_{(k+2)\alpha}^2}\frac{t_*^{k+2}\mathbf{E}_\mu g^{(k+2)\alpha}}{(k+2)!}\\ &\leqslant \max_{k\geqslant 0}(k+2)\||\nabla g|_2\|_{(k+2)\alpha}^2 t_*^2 M^2(k+2)^{-\frac{2}{\alpha}}\sum_{k\in I}\frac{(k+2)^{k+2}}{M^{k+2}(k+2)!}t_*^{\frac{2}{\alpha}}M^{\frac{2}{\alpha}}.\end{aligned}$$

Obviously it is easy to choose $M$ close to $2e$ so that $\sum_{k\in I}\frac{(k+2)^{k+2}}{M^{k+2}(k+2)!} \leqslant 1$. Thus, we may state our bound over $I$ in the following form

$$\sum_{k\in I}\frac{(k+2)\||\nabla g|_2\|_{(k+2)\alpha}^2}{\|g\|_{(k+2)\alpha}^2}\frac{t_*^{k+2}\mathbf{E}_\mu g^{(k+2)\alpha}}{(k+2)!} \leqslant K^2 t_*^{\frac{2}{\alpha}}M^{\frac{2}{\alpha}}, \tag{2.21}$$

where $K = \max_{k\geqslant 1} \frac{\||\nabla g|_2\|_{k\alpha}}{k^{\frac{1}{\alpha}-\frac{1}{2}}}$. On the set $J$ we do as follows

$$\sum_{k\in J} \frac{(k+2)\||\nabla g|_2\|^2_{(k+2)\alpha}}{\|g\|^2_{(k+2)\alpha}} \frac{t_*^{k+2}\mathbf{E}_\mu g^{(k+2)\alpha}}{(k+2)!}$$

$$\leqslant t_*^{\frac{2}{\alpha}} M^{\frac{2}{\alpha}} \max_{k\geqslant 0} \frac{(k+2)\||\nabla g|_2\|^2_{(k+2)\alpha}}{(k+2)^{\frac{2}{\alpha}}} \sum_{k\in J} \frac{t_*^{k+2}\mathbf{E}_\mu g^{(k+2)\alpha}}{(k+2)!}.$$

But now

$$\sum_{k\in J} \frac{t_*^{k+2}\mathbf{E}_\mu g^{(k+2)\alpha}}{(k+2)!} \leqslant \sum_{k\geqslant 0} \frac{t_*^{k+2}\mathbf{E}_\mu g^{(k+2)\alpha}}{(k+2)!} \leqslant e^{V(t_*)} - 1 = 1.$$

Thus, our bound on $J$ is

$$\sum_{k\in J} \frac{(k+2)\||\nabla g|_2\|^2_{(k+2)\alpha}}{\|g\|^2_{(k+2)\alpha}} \frac{t_*^{k+2}\mathbf{E}_\mu g^{(k+2)\alpha}}{(k+2)!} \leqslant M^{\frac{2}{\alpha}} K^2 t_*^{\frac{2}{\alpha}}. \tag{2.22}$$

Combining bounds (2.21), (2.22), and (2.20) we get

$$\frac{2\ln 2}{D\alpha^2} \leqslant M^{\frac{2}{\alpha}} K^2 t_*^{\frac{2}{\alpha}}$$

but this implies

$$\frac{1}{t^*} \leqslant (D')^{\frac{\alpha}{2}} \alpha^\alpha K^\alpha.$$

Note that $K$ is comparable with $\||\nabla g|_2\|_{\varphi(\frac{2\alpha}{2-\alpha})}$. It leads to the formula

$$\|g\|_{\varphi(\alpha)} = \frac{1}{t_*^{\frac{1}{\alpha}}} \leqslant \alpha\sqrt{D'}\||\nabla g|_2\|_{\varphi(\frac{2\alpha}{2-\alpha})}. \tag{2.23}$$

Bound (2.16), (2.19), and (2.23) implies that for any positive $g$

$$\|g\|_{\varphi(\alpha)} \leqslant \max\left\{\left\||\nabla g^\alpha|_\infty^{\frac{1}{\alpha}}\right\|_\infty, \left(\frac{2}{\ln 2}\right)^{\frac{1}{\alpha}} \|g\|_\alpha, \alpha\sqrt{D'}\||\nabla g|_2\|_{\varphi(\frac{2\alpha}{2-\alpha})}\right\}. \tag{2.24}$$

If we now fix $g = |f - \mathbf{E}_\mu f|$ then by the Poincaré inequality

$$\|f - \mathbf{E}_\mu f\|_\alpha \leqslant \sqrt{D'}\||\nabla f|_2\|_2.$$

Note also that

$$\||\nabla g|_2\|_{\varphi(\frac{2\alpha}{2-\alpha})} = \||\nabla f|_2\|_{\varphi(\frac{2\alpha}{2-\alpha})}.$$

Thus, by (2.24) we obtain

$$\|f-\mathbf{E}_\mu f\|_{\varphi(\alpha)} \leqslant \max\left\{\left\|\left|\nabla|f-\mathbf{E}_\mu f|^\alpha\right|_\infty\right\|_\infty^{\frac{1}{\alpha}}, \left(\frac{2}{\ln 2}\right)^{\frac{1}{\alpha}}\|f-\mathbf{E}_\mu f\|_\alpha, \alpha\sqrt{D'}\||\nabla f|_2\|_{\varphi(\frac{2\alpha}{2-\alpha})}\right\}.$$

It ends the proof. ■

**Acknowledgement** This research was partially supported by NCN Grant UMO-2016/21/B/ST1/01489.

## References

1. R. Adamczak, P. Wolff, Concentration inequalities for non-Lipschitz functions with bounded derivatives of higher order. Probab. Theory Related Fields **162**, 531–586 (2015)
2. S. Aida, D. Stroock, Moment estimates derived from Poincare and logarithmic Sobolev inequality. Math. Res. Lett. **1**, 75–86 (1994)
3. S. Bobkov, M. Ledoux, Poincaré's inequalities and Talagrand's concentration phenomenon for the exponential distribution. Probab. Theory Related Fields **107**(3), 383–400 (1997)
4. E. Milman, On the role of convexity in isoperimetry, spectral gap and concentration. Invent. Math. **177**, 1–4 (2009)

# Chapter 3
# Polar Isoperimetry. I: The Case of the Plane

**Sergey G. Bobkov, Nathael Gozlan, Cyril Roberto, and Paul-Marie Samson**

**Abstract** This is the first part of the notes with preliminary remarks on the plane isoperimetric inequality and its applications to the Poincaré and Sobolev-type inequalities in dimension one. Links with informational quantities of Rényi and Fisher are briefly discussed.

**Keywords** Isoperimetry · Sobolev-type inequalities · Rényi divergence power · Relative Fisher information

## 3.1 Isoperimetry on the Plane and the Upper Half-Plane

The paper by Diaz et al. [4] contains the following interesting Sobolev-type inequality in dimension one.

**Proposition 3.1.1** *For any smooth real-valued function* $f$ *on* $[0, 1]$,

$$\int_0^1 \sqrt{f(x)^2 + \frac{1}{\pi^2} f'(x)^2}\, dx \geq \Big( \int_0^1 f(x)^2\, dx \Big)^{1/2}. \tag{3.1}$$

S. G. Bobkov (✉)
School of Mathematics, University of Minnesota, Minneapolis, MN, USA
e-mail: bobkov@math.umn.edu

N. Gozlan
Université Paris Descartes, MAP5, UMR 8145, Paris Cedex, France
e-mail: natael.gozlan@parisdescartes.fr

C. Roberto
Université Paris Nanterre, MODAL'X, EA 3454, Nanterre, France
e-mail: croberto@math.cnrs.fr

P.-M. Samson
LAMA, Univ Gustave Eiffel, UPEM, Univ Paris Est Creteil, CNRS, F-77447, Marne-la-Vallée, France
e-mail: paul-marie.samson@univ-mlv.fr

 
N. Gozlan et al. (eds.), *High Dimensional Probability VIII*,
Progress in Probability 74, https://doi.org/10.1007/978-3-030-26391-1_3

More precisely, this paper mentions without proof that (3.1) is a consequence of the isoperimetric inequality on the plane $\mathbb{R}^2$. Let us give an argument, which is actually based on the isoperimetric inequality

$$\mu^+(A) \geq \sqrt{2\pi}\,(\mu(A))^{1/2}, \qquad A \subset \mathbb{R}^2_+ \;\; (A \textit{ is Borel}), \tag{3.2}$$

in the upper half-plane $\mathbb{R}^2_+ = \{(x_1, x_2) \in \mathbb{R}^2 : x_2 \geq 0\}$. Here, $\mu$ denotes the Lebesgue measure restricted to this half-plane, which generates the corresponding notion of the perimeter

$$\mu^+(A) = \liminf_{\varepsilon \to 0} \frac{\mu(A + \varepsilon B_2) - \mu(A)}{\varepsilon}$$

(cf. e.g. [2]).

Inequality (3.2) follows from the Brunn-Minkowski inequality in $\mathbb{R}^2$

$$\mu(A + B)^{1/2} \geq \mu(A)^{1/2} + \mu(B)^{1/2}$$

along the same arguments as in the case of its application to the usual isoperimetric inequality. Indeed, applying it with a Borel set $A \subset \mathbb{R}^2_+$ and $B = \varepsilon B_2$ ($\varepsilon > 0$), we get

$$\begin{aligned} \mu(A + \varepsilon B_2) &\geq \big[\,\mu(A)^{1/2} + \mu(\varepsilon B_2)^{1/2}\big]^2 \\ &= \Big[\mu(A)^{1/2} + \Big(\frac{\pi}{2}\Big)^{1/2} \varepsilon\,\Big]^2 \\ &= \mu(A) + \sqrt{2\pi}\,(\mu(A))^{1/2}\varepsilon + O(\varepsilon^2), \end{aligned}$$

and therefore (3.2) from the definition of the perimeter.

The relation (3.2) is sharp and is attained for the upper semi-discs

$$A_\rho = \{(x_1, x_2) \in \mathbb{R}^2 : x_1^2 + x_2^2 \leq \rho^2,\; x_2 \geq 0\}, \qquad \rho > 0.$$

In this case, $\mu(A_\rho) = \frac{1}{2}\pi\rho^2$ is the area size between the upper part of the circle $x_1^2 + x_2^2 = \rho^2$ and the $x_1$-axis $x_2 = 0$, while the $\mu$-perimeter is just the length of the half-circle $\mu^+(A_\rho) = \pi\rho$.

To derive (3.1), one may assume that the function $f$ is non-negative and is not identically zero on [0, 1]. Then we associate with it the set in $\mathbb{R}^2_+$ described in polar coordinates as

$$A = \{(x_1, x_2) : 0 \leq r \leq f(t),\; 0 \leq t \leq 1\}$$

with $x_1 = r\cos(\pi t)$, $x_2 = r\sin(\pi t)$. Integration in polar coordinates indicates that, for any non-negative Borel function $u$ on $\mathbb{R}^2$,

$$\iint_{\mathbb{R}^2} u(x_1, x_2)\, dx_1\, dx_2 \ = \ \pi \int_{-1}^{1} \left[ \int_0^\infty u\big(r\cos(\pi t), r\sin(\pi t)\big)\, r dr \right] dt. \tag{3.3}$$

Applying it to the indicator function $u = 1_A$, we get

$$\mu(A) = \frac{\pi}{2} \int_0^1 f(t)^2\, dt.$$

On the other hand, $\mu^+(A)$ represents the length of the curve $C = \{(x_1(t), x_2(t)) : 0 \leq t \leq 1\}$ parameterized by

$$x_1(t) = f(t)\cos(\pi t), \qquad x_2(t) = f(t)\sin(\pi t).$$

Since

$$x_1'(t)^2 + x_2'(t)^2 \ = \ f'(t)^2 + \pi^2 f(t)^2,$$

we find that

$$\mu^+(A) = \int_0^1 \sqrt{x_1'(t)^2 + x_2'(t)^2}\, dt = \int_0^1 \sqrt{f'(t)^2 + \pi^2 f(t)^2}\, dt.$$

As a result, the isoperimetric inequality (3.2) takes the form

$$\int_0^1 \sqrt{f'(t)^2 + \pi^2 f(t)^2}\, dt \ \geq \ \sqrt{2\pi}\, \Big(\frac{\pi}{2} \int_0^1 f(t)^2\, dt\Big)^{1/2}.$$

which is the same as (3.1). Note that the condition $f \geq 0$ may easily be removed in the resulting inequality. □

One can reverse the argument and obtain the isoperimetric inequality (3.2) on the basis of (3.1) for the class of star-shaped sets in the upper half-plane.

The same argument may be used on the basis of the classical isoperimetric inequality

$$\mu^+(A) \geq \sqrt{4\pi}\, (\mu(A))^{1/2} \qquad (A \text{ is Borel}) \tag{3.4}$$

in the whole plane $\mathbb{R}^2$ with respect to the Lebesgue measure $\mu$. It is attained for the discs

$$A_\rho = \{(x_1, x_2) \in \mathbb{R}^2 : x_1^2 + x_2^2 \leq \rho^2\}, \qquad \rho > 0,$$

in which case $\mu(A_\rho) = \pi\rho^2$ and $\mu^+(A_\rho) = 2\pi\rho$.

Starting from a smooth non-negative function $f$ on $[-1, 1]$ such that $f(-1) = f(1)$, one may consider the star-shaped region

$$A = \{(x_1, x_2) : 0 \le r \le f(t),\ -1 \le t \le 1\}, \qquad x_1 = r\cos(\pi t),\ x_2 = r\sin(\pi t),$$

enclosed by the curve $C = \{(x_1(t), x_2(t)) : -1 \le t \le 1\}$ with the same functions $x_1(t) = f(t)\cos(\pi t)$, $x_2(t) = f(t)\sin(\pi t)$. Integration in polar coordinates (3.3) then yields a similar formula as before,

$$\mu(A) = \frac{\pi}{2} \int_{-1}^{1} f(t)^2\, dt,$$

and also the perimeter $\mu^+(A)$ represents the length of $C$, i.e.,

$$\mu^+(A) = \int_{-1}^{1} \sqrt{x_1'(t)^2 + x_2'(t)^2}\, dt = \int_{-1}^{1} \sqrt{f'(t)^2 + \pi^2 f(t)^2}\, dt.$$

As a result, the isoperimetric inequality (3.4) takes the form

$$\int_{-1}^{1} \sqrt{f'(t)^2 + \pi^2 f(t)^2}\, dt \ \ge\ \sqrt{4\pi}\, \Big(\frac{\pi}{2} \int_{-1}^{1} f(t)^2\, dt\Big)^{1/2},$$

or equivalently,

$$\frac{1}{2} \int_{-1}^{1} \sqrt{\frac{1}{\pi^2}\, f'(t)^2 + f(t)^2}\, dt \ \ge\ \Big(\frac{1}{2} \int_{-1}^{1} f(t)^2\, dt\Big)^{1/2}. \tag{3.5}$$

To compare with (3.1), let us restate (3.5) on the unit interval $[0, 1]$ by making the substitution $f(t) = u(\frac{1+t}{2})$. Then it becomes

$$\frac{1}{2} \int_{-1}^{1} \sqrt{\frac{1}{4\pi^2}\, u'\Big(\frac{1+t}{2}\Big)^2 + u\Big(\frac{1+t}{2}\Big)^2}\, dt \ \ge\ \Big(\frac{1}{2} \int_{-1}^{1} u\Big(\frac{1+t}{2}\Big)^2\, dt\Big)^{1/2}.$$

Changing $x = \frac{1+t}{2}$, replacing $u$ again with $f$, and removing the unnecessary condition $f \ge 0$, we arrive at:

**Proposition 3.1.2** *For any smooth real-valued function $f$ on $[0, 1]$ such that $f(0) = f(1)$,*

$$\int_0^1 \sqrt{f(x)^2 + \frac{1}{4\pi^2}\, f'(x)^2}\, dx \ \ge\ \Big(\int_0^1 f(x)^2\, dx\Big)^{1/2}. \tag{3.6}$$

As we can see, an additional condition $f(0) = f(1)$ allows one to improve the coefficient in front of the derivative, in comparison with (3.1). It should also be clear that (3.6) represents an equivalent form of the isoperimetric inequality (3.4) for the class of star-shaped regions.

## 3.2 Relationship with Poincaré-type Inequalities

It would be interesting to compare Propositions 3.1.1–3.1.2 with other popular Sobolev-type inequalities such as the Poincaré-type and logarithmic Sobolev inequalities. Starting from (3.1) and (3.6), a simple variational argument yields:

**Corollary 3.2.1** *For any smooth real-valued function $f$ on* $[0, 1]$,

$$\mathrm{Var}_\mu(f) \leq \frac{1}{\pi^2} \int_0^1 f'(x)^2 \, dx, \tag{3.7}$$

*where the variance is understood with respect to the uniform probability measure $d\mu(x) = dx$ on the unit segment. Moreover, if $f(0) = f(1)$, then*

$$\mathrm{Var}_\mu(f) \leq \frac{1}{4\pi^2} \int_0^1 f'(x)^2 \, dx. \tag{3.8}$$

The constants $\frac{1}{\pi^2}$ and $\frac{1}{4\pi^2}$ in (3.7)–(3.8) are optimal and are respectively attained for the functions $f(x) = \cos(\pi x)$ and $f(x) = \sin(2\pi x)$ (cf. also [1]).

For the proof, let us note that an analytic inequality of the form

$$\int_0^1 \sqrt{f(x)^2 + cf'(x)^2} \, dx \ \geq \ \Big(\int_0^1 f(x)^2 \, dx\Big)^{1/2} \tag{3.9}$$

with a constant $c > 0$ becomes equality for $f = 1$. So, one may apply it to $f_\varepsilon = 1 + \varepsilon f$, and letting $\varepsilon \to 0$, one may compare the coefficients in front of the powers of $\varepsilon$ on both sides. First,

$$\int_0^1 f_\varepsilon(x)^2 \, dx = 1 + 2\varepsilon \int_0^1 f(x) \, dx + \varepsilon^2 \int_0^1 f(x)^2 \, dx,$$

so, by Taylor's expansion, as $\varepsilon \to 0$,

$$\begin{aligned}
\Big(\int_0^1 f_\varepsilon(x)^2 \, dx\Big)^{1/2} &= 1 + \varepsilon \int_0^1 f(x) \, dx + \frac{\varepsilon^2}{2} \int_0^1 f(x)^2 \, dx \\
&\quad - \frac{1}{8} \Big(2\varepsilon \int_0^1 f(x) \, dx + \varepsilon^2 \int_0^1 f(x)^2 \, dx\Big)^2 + O(\varepsilon^3) \\
&= 1 + \varepsilon \int_0^1 f(x) \, dx + \frac{\varepsilon^2}{2} \int_0^1 f(x)^2 \, dx - \frac{\varepsilon^2}{2} \Big(\int_0^1 f(x) \, dx\Big)^2 + O(\varepsilon^3).
\end{aligned}$$

On the other hand, since

$$f_\varepsilon(x)^2 + cf_\varepsilon'(x)^2 = 1 + 2\varepsilon f(x) + \varepsilon^2 \left(f(x)^2 + cf'(x)^2\right),$$

we have

$$\begin{aligned}\left(f_\varepsilon(x)^2 + cf_\varepsilon'(x)^2\right)^{1/2} &= 1 + \varepsilon f(x) + \frac{\varepsilon^2}{2}\left(f(x)^2 + cf'(x)^2\right)\\ &\quad -\frac{1}{8}\Big(2\varepsilon f(x) + \varepsilon^2\left(f(x)^2 + cf'(x)^2\right)\Big)^2 + O(\varepsilon^3)\\ &= 1 + \varepsilon f(x) + \frac{c\varepsilon^2}{2}\, f'(x)^2 + O(\varepsilon^3).\end{aligned}$$

Hence

$$\int_0^1 \left(f_\varepsilon(x)^2 + cf_\varepsilon'(x)^2\right)^{1/2} dx \;=\; 1 + \varepsilon \int_0^1 f(x)\,dx + \frac{c\varepsilon^2}{2}\int f'(x)^2\,dx + O(\varepsilon^3).$$

Inserting both expansions in (3.9), we see that the linear coefficients coincide, while comparing the quadratic terms leads to the Poincaré-type inequality

$$c\int f'(x)^2\,dx \;\geq\; \int_0^1 f(x)^2\,dx - \Big(\int_0^1 f(x)\,dx\Big)^2.$$

□

Thus, the isoperimetric inequality on the upper half-plane implies the Poincaré-type inequality (3.7) on [0, 1], while the isoperimetric inequality on the whole plane implies the restricted Poincaré-type inequality (3.8), with optimal constants in both cases.

## 3.3 Sobolev Inequalities

If $f$ is non-negative, then $f(x) = 0 \Rightarrow f'(x) = 0$ and thus $f(x)^2 + cf'(x)^2 = 0$. Hence, applying Cauchy's inequality, from (3.9) we get

$$\begin{aligned}\int_0^1 f(x)^2\,dx &\leq \bigg(\int_0^1 \sqrt{f(x)}\sqrt{f(x) + c\frac{f'(x)^2}{f(x)}}\,1_{\{f(x)>0\}}\,dx\bigg)^2\\ &\leq \int_0^1 f(x)\,dx\,\Big(\int_0^1 f(x)\,dx + c\int_0^1 \frac{f'(x)^2}{f(x)}\,1_{\{f(x)>0\}}\,dx\Big).\end{aligned}$$

Therefore, Propositions 3.1.1–3.1.2 also yield:

**Proposition 3.3.1** *For any non-negative smooth function* $f$ *on* $[0,1]$ *with* $\int_0^1 f(x)\,dx = 1$,

$$\mathrm{Var}_\mu(f) \ \le \ \frac{1}{\pi^2} \int_0^1 \frac{f'(x)^2}{f(x)}\, 1_{\{f(x)>0\}}\, dx, \tag{3.10}$$

*where the variance is with respect to the uniform probability measure* $\mu$ *on the unit segment. Moreover, if* $f(0) = f(1)$, *then*

$$\mathrm{Var}_\mu(f) \ \le \ \frac{1}{4\pi^2} \int_0^1 \frac{f'(x)^2}{f(x)}\, 1_{\{f(x)>0\}}\, dx. \tag{3.11}$$

Recall that there is a general relation between the entropy functional

$$\mathrm{Ent}_\mu(f) = \int f \log f \, d\mu - \int f \, d\mu \, \log \int f \, d\mu \qquad (f \ge 0)$$

and the variance, namely

$$\mathrm{Ent}_\mu(f) \int f \, d\mu \ \le \ \mathrm{Var}_\mu(f). \tag{3.12}$$

It is rather elementary; assume by homogeneity that $\int f\,d\mu = 1$. Since $\log t \le t-1$ and therefore $t \log t \le t(t-1)$ for all $t \ge 0$, we have

$$f(x) \log f(x) \le f(x)^2 - f(x).$$

After integration it yields (3.12).

Using the latter in (3.10)–(3.11), we arrive at the logarithmic Sobolev inequalities.

**Corollary 3.3.2** *For any non-negative smooth function* $f$ *on* $[0,1]$, *with respect to the uniform probability measure* $\mu$ *on the unit segment we have*

$$\mathrm{Ent}_\mu(f) \ \le \ \frac{1}{\pi^2} \int_0^1 \frac{f'(x)^2}{f(x)}\, 1_{\{f(x)>0\}}\, dx. \tag{3.13}$$

*Moreover, if* $f(0) = f(1)$, *then*

$$\mathrm{Ent}_\mu(f) \ \le \ \frac{1}{4\pi^2} \int_0^1 \frac{f'(x)^2}{f(x)}\, 1_{\{f(x)>0\}}\, dx. \tag{3.14}$$

Replacing here $f$ by $(1+\varepsilon f)^2$ and letting $\varepsilon \to 0$, we return to the Poincaré-type inequalities (3.7) and (3.8) with an extra factor of 2. The best constant in (3.13) is however $\frac{1}{2\pi^2}$ and in (3.14) is $\frac{1}{8\pi^2}$ [1, Proposition 5.7.5]. On the other hand, the inequalities (3.10)–(3.11) are much stronger than (3.13)–(3.14).

## 3.4 Informational Quantities and Distances

The inequalities (3.13)–(3.14) may be stated equivalently in terms of informational distances to the uniform measure $\mu$ on the unit segment. Let us recall that, for random elements $X$ and $Z$ in an abstract measurable space $\Omega$ with distributions $\nu$ and $\mu$ respectively, the Rényi divergence power or the Tsallis distance from $\nu$ to $\mu$ of order $\alpha > 0$ is defined by

$$T_\alpha(X||Z) = T_\alpha(\nu||\mu) = \frac{1}{\alpha - 1}\bigg[\int \Big(\frac{p}{q}\Big)^\alpha p\, d\lambda - 1\bigg] = \frac{1}{\alpha - 1}\bigg[\int f^\alpha\, d\mu - 1\bigg],$$

where $p$ and $q$ are densities of $\nu$ and $\mu$ with respect to some (any) $\sigma$-finite dominating measure $\lambda$ on $\Omega$, with $f = p/q$ being the density of $\nu$ with respect to $\mu$ (the definition does not depend on the choice of $\lambda$). If $\alpha = 1$, we arrive at the Kullback–Leibler distance or an informational divergence

$$T_1(X||Z) = D(X||Z) = \int p \log \frac{p}{q}\, d\lambda = \int f \log f\, d\mu,$$

which is the same as $\mathrm{Ent}_\mu(f)$. For $\alpha = 2$ the Tsallis $T_2$-distance is the same as the $\chi^2$-distance. If $\alpha \geq 1$, necessarily $T_\alpha(X||Z) = \infty$ as long as $\nu$ is not absolutely continuous with respect to $\mu$. In any case, the function $\alpha \to T_\alpha$ is non-decreasing; we refer an interested reader to the survey [6] (cf. also [3]).

In the case of the real line $\Omega = \mathbb{R}$, and when the densities $p$ and $q$ are absolutely continuous, the relative Fisher information or the Fisher information distance from $\nu$ to $\mu$ is defined by

$$I(X||Z) = I(\nu||\mu) = \int_{-\infty}^{\infty} \Big(\frac{p'}{p} - \frac{q'}{q}\Big)^2 p\, d\lambda = \int_{-\infty}^{\infty} \frac{f'^2}{f}\, d\mu,$$

still assuming that the probability measure $\nu$ is absolutely continuous with respect to $\mu$ and has density $f = p/q$. This definition is commonly used when $q$ is supported and is positive on an interval $\Delta \subset \mathbb{R}$, finite or not, with the above integration restricted to $\Delta$. With these notations, Proposition 3.3.1 corresponds to the order $\alpha = 2$ and therefore takes the form

$$T_2(X||Z) \leq \frac{1}{\pi^2}\, I(X||Z), \qquad T_2(X||Z) \leq \frac{1}{4\pi^2}\, I(X||Z), \tag{3.15}$$

holding true for an arbitrary random variable $X$ with values in $[0, 1]$. Here the random variable $Z$ has a uniform distribution $\mu$ on $[0, 1]$, and we use an additional constraint $f(0) = f(1)$ in the second relation.

There is also another non-distance formulation of (3.15) in terms of classical informational quantities such as the Rényi entropy power and the Fisher information

$$N_\alpha(X) = \Big(\int_{-\infty}^{\infty} p(x)^\alpha \, dx\Big)^{-\frac{2}{\alpha-1}}, \qquad I(X) = \int_{-\infty}^{\infty} \frac{p'(x)^2}{p(x)} \, dx.$$

Here the case $\alpha = 2$ defines the quadratic Rényi entropy power $N_2(X)$. If $\mu$ is supported and has an absolutely continuous positive density $q$ on the interval $\Delta \subset \mathbb{R}$, one may also define the restricted Fisher information

$$I_0(X) = \int_\Delta \frac{p'(x)^2}{p(x)} \, dx.$$

For example, if $Z$ is uniformly distributed in the unit interval, so that $q(x) = 1$ for $0 < x < 1$, we have $I(Z) = \infty$, while $I_0(Z) = 0$. In this case, if $X$ has values in $[0, 1]$, we have

$$T_2(X||Z) = \int_0^1 p(x)^2 \, dx - 1 = N_2(X)^{-1/2} - 1, \qquad I(X||Z) = I_0(X).$$

Hence, the first inequality in (3.15) may be written as the following.

**Corollary 3.4.1** *For any random variable $X$ with values in* $[0, 1]$*, having there an absolutely continuous density, we have*

$$N_2(X)\Big(1 + \frac{1}{\pi^2}\, I_0(X)\Big)^2 \geq 1. \tag{3.16}$$

This relation is analogous to the well-known isoperimetric inequality for entropies,

$$N(X)\, I(X) \geq 2\pi e,$$

where $N(X) = N_1(X) = e^{2h(X)}$ is the entropy power, corresponding to the Shannon differential entropy

$$h(X) = -\int_{-\infty}^{\infty} p(x) \log p(x) \, dx.$$

The functional $I_0(X)$ may be replaced with $I(X)$ in (3.16) (since $I_0 \leq I$), and then one may remove the assumption on the values of $X$. Moreover, with the functional $I(X)$, this inequality may be considerably strengthened. Indeed, the relation $N_2(X)(1 + \frac{1}{\pi^2} I(X))^2 \geq 1$ is not 0-homogeneous with respect to $X$, and therefore it admits a self-refinement when applying it to the random variables $\lambda X$, $\lambda > 0$. Optimizing over this parameter, we will obtain an equivalent 0-homogeneous relation

$$N_2(X) I(X) \geq c, \tag{3.17}$$

with $c = \pi/4$. But, it is obviously true that with $c = 1$. To see this, first note that, by the Cauchy inequality, for all $x \in \mathbb{R}$,

$$p(x) = \int_{-\infty}^{x} p'(y)\,dy \leq \int_{p(y)>0} |p'(y)|\,dy = \int_{p(y)>0} \frac{|p'(y)|}{\sqrt{p(y)}} \sqrt{p(y)}\,dy$$
$$\leq \Big( \int_{p(y)>0} \frac{p'(y)^2}{p(y)}\,dy \Big)^{1/2} \Big( \int_{p(y)>0} p(y)\,dy \Big)^{1/2} = \sqrt{I(X)}.$$

Therefore,

$$\int_{-\infty}^{\infty} p(x)^2\,dx \leq \sqrt{I(X)},$$

that is, $N_2(X) I(X) \geq 1$.

Observe that another inequality involving the quadratic Rényi entropy power $N_2(X)$ and some generalisation of Fisher information can be extracted from [5], namely for all $1 \leq q < \infty$, $N_2(X)^q \int |p'|^q p \geq C_q$ for an optimal constant $C_q$. However it's unclear how to related this inequality to (3.17).

**Acknowledgements** Research was partially supported by the NSF grant DMS-1855575 and by the Bézout Labex, funded by ANR, reference ANR-10-LABX-58, the Labex MME-DII funded by ANR, reference ANR-11-LBX-0023-01, and the ANR Large Stochastic Dynamic, funded by ANR, reference ANR-15-CE40-0020-03-LSD.

## References

1. D. Bakry, I. Gentil, M. Ledoux, Analysis and geometry of Markov diffusion operators, in *Grundlehren der Mathematischen Wissenschaften (Fundamental Principles of Mathematical Sciences)*, vol. 348 (Springer, Cham, 2014), p. xx+552
2. Yu.D. Burago, V.A. Zalgaller, Geometric inequalities, in *Translated from the Russian by A. B. Sosinskii. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer Series in Soviet Mathematics, vol. 285 (Springer, Berlin, 1988), p. xiv+331
3. A. Dembo, T.M. Cover, J.A. Thomas, Information-theoretic inequalities. IEEE Trans. Inform. Theory **37**(6), 1501–1518 (1991)
4. A. Diaz, N. Harman, S. Howe, D. Thompson, Isoperimetric problems in sectors with density. Adv. Geom. **12**(4), 589–619 (2012)

5. E. Lutwak, D. Yang, G. Zhang, Cramer-Rao and moment-entropy inequalities for Renyi entropy and generalized Fisher information. IEEE Trans. Inform. Theory **51**(2), 473–478 (2005)
6. T. van Erven, P. Harremoës, Rényi divergence and Kullback-Leibler divergence. IEEE Trans. Inform. Theory **60**(7), 3797–3820 (2014)

# Chapter 4
# Iterated Jackknives and Two-Sided Variance Inequalities

**Olivier Bousquet and Christian Houdré**

**Abstract** We consider the variance of a function of $n$ independent random variables and provide inequalities that generalize previous results obtained for i.i.d. random variables. In particular we obtain upper and lower bounds on the variance based on iterated jackknife statistics that can be considered generalizations of the Efron–Stein inequality.

## 4.1 Introduction

The properties of functions of $n$ independent random variables, and in particular the estimation of their moments from the moments of their increments (i.e. when replacing a random variable by an independent copy) have been thoroughly studied (see, e.g., [2] for a comprehensive overview). We focus here on the variance and consider how to refine and generalize known extensions of the Efron–Stein inequality in the non-symmetric, non-iid case.

But first, let us review some of the existing results. Let $X_1, X_2, \ldots, X_n$ be iid random variables and let $S : \mathbb{R}^n \to \mathbb{R}$ be a statistic of interest which is symmetric, i.e., invariant under any permutation of its arguments, and square integrable. The (original) Efron–Stein inequality [3], states that the jackknife estimates of variance is biased upwards, i.e., denoting by $\tilde{X}$ an independent copy of $X_1, \ldots, X_n$, and setting $S_i = S(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n, \tilde{X})$, $i = 1, \ldots, n$, and $S_{n+1} = S$, then

$$\mathrm{Var}\, S \le \mathbb{E} J_1, \tag{4.1}$$

O. Bousquet
Google Research, Brain Team, Zürich, Switzerland

C. Houdré (✉)
Georgia Institute of Technology, School of Mathematics, Atlanta, GA, USA
e-mail: houdre@math.gatech.edu



N. Gozlan et al. (eds.), *High Dimensional Probability VIII*,
Progress in Probability 74, https://doi.org/10.1007/978-3-030-26391-1_4

where

$$J_1 = \sum_{i=1}^{n+1}(S_i - \bar{S})^2 = \frac{1}{(n+1)} \sum\sum_{1\le i<j\le n+1} (S_i - S_j)^2, \tag{4.2}$$

and $\bar{S} = \sum_{i=1}^{n+1} S_i/(n+1)$. Beyond the original framework, the inequality (4.1) has seen many extensions and generalizations with different proofs which are well described in [2], whose notation we essentially adopt and to which we refer for a more complete bibliography and many instances of applications. Let us just say that (4.1) can be seen as the "well known" tensorization property of the variance which asserts that if $X_1, X_2, \dots, X_n$ are independent random variables with $X_i \sim \mu_i$, then

$$\mathrm{Var}_{\mu^n} S \le \mathbb{E}_{\mu^n} \sum_{i=1}^{n} \mathrm{Var}_{\mu_i} S, \tag{4.3}$$

where $\mathbb{E}_{\mu^n}$ and $\mathrm{Var}_{\mu^n}$ are respectively the expectation and variance with respect to $\mu^n$, the joint law of $X_1, X_2, \dots, X_n$, while $\mathrm{Var}_{\mu_i} S$ is the variance of $S$ with respect to $\mu_i$, the law of $X_i$. In fact, if for each $i = 1, 2, \dots, n$, $\tilde{X}_i \sim \tilde{\mu}_i$ is an independent copy of $X_i$, then (4.3) can be rewritten as

$$\begin{aligned}\mathrm{Var}_{\mu^n} S &\le \frac{1}{2}\mathbb{E}_{\mu^n} \sum_{i=1}^{n} \mathbb{E}_{\mu_i\otimes\tilde{\mu}_i} (S - S_i)^2 \\ &= \frac{1}{2}\mathbb{E}_{\mu^n} \sum_{i=1}^{n} \mathbb{E}_{\tilde{\mu}_i} (S - S_i)^2, \end{aligned} \tag{4.4}$$

where $S_i = S(X_1, \dots, X_{i-1}, \tilde{X}_i, X_{i+1}, \dots, X_n)$.

Neither (4.1) nor (4.4), whose proof can be obtained, for example, by induction, require $S$ to be symmetric. In case $S$ is symmetric, and the random variables are identically distributed, the right-hand side of (4.4) becomes $n\mathbb{E}_{\mu^n\otimes\tilde{\mu}_1}(S - S_1)^2/2$ while, via (4.2), the right-hand side of (4.1) becomes $\binom{n-1}{2}\mathbb{E}(S_1 - S_2)^2/(n+1) = n\mathbb{E}(S_1 - S_2)^2/2$, and (4.4) and (4.1) are identical.

Since the jackknife estimate of variance is biased upwards, it is natural to try to estimate the bias $\mathbb{E}J_1 - \mathrm{Var}\, S$, and such an attempt is already presented in [5] via the "iterated jackknives". Let us recall what was meant there: Resampling the jackknife statistics, introduce for any $k = 2, \dots, n$, the iterated jackknives $J_2, J_3, \dots, J_n$, leading to both upper and lower bounds on $\mathrm{Var}\, S$, showing, in particular, that

$$\frac{1}{2}\mathbb{E}J_2 - \frac{1}{6}\mathbb{E}J_3 \le \mathbb{E}J_1 - \mathrm{Var}\, S \le \frac{1}{2}\mathbb{E}J_2. \tag{4.5}$$

In [5], the inequalities (4.1) and (4.5) were viewed as statistical versions of generalized (multivariate) Gaussian Poincaré inequalities previously obtained in [6]. Indeed, setting $\nabla S := (S - S_1, S - S_2, \dots, S - S_n)$, then $\mathbb{E}J_1 = \mathbb{E}\|\nabla S\|^2$. If instead of looking at the vector of first differences, one looks at second and third ones, then the corresponding norms will lead to (4.5). Throughout the years, it was asked whether or not an inequality such as (4.5) would have a general version and a positive answer had been informally given. The aim of the present note is to provide a synthetic proof of these, removing the iid and symmetry assumptions in (4.5) and its generalizations, leading to generic inequalities. This could be useful, as these dormant inequalities seem to have found, in recent times, some new life, e.g., see [1, 8, 9].

## 4.2 Iterated Jackknife Bounds

Throughout and unless otherwise noted, $X_1, \dots, X_n$ are independent random variables and $S : \mathbb{R}^n \to \mathbb{R}$ is a Borel function such that $\mathbb{E}S^2(X_1, \dots, X_n) < +\infty$. Next, and if $S$ is short for $S(X_1, \dots, X_n)$, let, for any $i = 1, \dots, n$, $\mathbb{E}^{(i)}$ denote the conditional expectation with respect to the $\sigma$-field generated by $X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n$. Hence,

$$\begin{aligned}\mathbb{E}^{(i)}S &:= \mathbb{E}(S \mid X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n)\\ &= \int_{-\infty}^{+\infty} S(X_1, \dots, X_{i-1}, x_i, X_{i+1}, \dots, X_n)\mu_i(dx_i), \end{aligned} \tag{4.6}$$

where $\mu_i$ is the law of $X_i$. By convention, $\mathbb{E}^{(0)}$ is the identity operator and so $\mathbb{E}^{(0)}S = S$. Iterating the above, it is clear that

$$\begin{aligned}\mathbb{E}^{(i)}\mathbb{E}^{(j)}S = \mathbb{E}^{(j)}\mathbb{E}^{(i)}S &= \mathbb{E}(S \mid X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_{j-1}, X_{j+1}, \dots, X_n) \\ &:= \mathbb{E}^{(i,j)}S = \mathbb{E}^{(j,i)}S,\end{aligned} \tag{4.7}$$

for any $i, j = 1, \dots, n$ and that for $i = 0, 1, \dots, n$,

$$\mathbb{E}^{(i)}\mathbb{E}^{(0)}S = \mathbb{E}^{(0)}\mathbb{E}^{(i)}S := \mathbb{E}^{(i,0)}S = \mathbb{E}^{(0,i)}S = \mathbb{E}^{(i)}S.$$

Next, let

$$\mathrm{Var}^{(i)}S := \mathbb{E}^{(i)}(S - \mathbb{E}^{(i)}S)^2 = \mathbb{E}^{(i)}S^2 - (\mathbb{E}^{(i)}S)^2,$$

$i = 0, 1, \dots, n$, and for any $i, j = 0, 1, \dots, n$, set

$$\mathrm{Var}^{(i,j)}S := \mathbb{E}^{(i)}\mathrm{Var}^{(j)}S - \mathrm{Var}^{(j)}\mathbb{E}^{(i)}S = \mathrm{Var}^{(j,i)}S \geq 0. \tag{4.8}$$

where, above, the rightmost equality follows from the commutativity property of the conditional expectations, as given in (4.7), while the inequality follows from convexity, and more precisely from the conditional Hölder's inequality.

Continuing with our notation, for any $i = 1, \ldots, n$, let throughout $\mathbb{E}_i$ denote the conditional expectation with respect to the $\sigma$-field generated by $X_1, \ldots, X_i$, i.e., $\mathbb{E}_i S := \mathbb{E}(S \mid X_1, \ldots, X_i)$, while this time $\mathbb{E}_0 S = \mathbb{E}S$.

At this point we also note that although $\mathrm{Var}^{(i)}$ is the conditional variance with respect to the $\sigma$-field generated by $X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n$, $\mathrm{Var}^{(i,j)}$ is not the conditional variance with respect to the $\sigma$-field generated by $X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_{j-1}, X_{j+1}, \ldots, X_n$. Indeed,

$$\mathrm{Var}^{(i,j)} S = \mathbb{E}^{(i,j)}(S - \mathbb{E}^{(i,j)} S)^2 - \mathrm{Var}^{(i)}\mathbb{E}^{(j)} S - \mathrm{Var}^{(j)}\mathbb{E}^{(i)} S. \tag{4.9}$$

Further iterating, for $i_1, i_2, \ldots, i_k \in \{0, 1, 2, \ldots, n\}$, then $\mathbb{E}^{(i_1)} \cdots \mathbb{E}^{(i_k)} := \mathbb{E}^{(i_1, i_2 \ldots, i_k)}$ is uniquely defined, i.e., the order in which the indices are taken is irrelevant, in particular $\mathbb{E}^{(1,2,\ldots,n)} S = \mathbb{E}S$. Still, iterating, set

$$\mathrm{Var}^{(i_1,i_2,\ldots,i_k)} S := \mathbb{E}^{(i_1)}\mathrm{Var}^{(i_2,\ldots,i_k)} S - \mathrm{Var}^{(i_2,\ldots,i_k)}\mathbb{E}^{(i_1)} S, \tag{4.10}$$

where again, above, the order in which the indices $i_1, i_2, \ldots, i_k \in \{0, 1, 2, \ldots, n\}$ are taken is irrelevant, and further, by convexity, (4.10) is non-negative, i.e.,

$$\mathrm{Var}^{(i_1,i_2,\ldots,i_k)} S \geq 0.$$

With the help of the above definitions, and in view of [5], let us now introduce the iterated jackknives,

$$J_k := \sum_{1 \leq i_1 \neq i_2 \cdots \neq i_k \leq n} \mathrm{Var}^{(i_1,\ldots,i_k)} S = k! \sum_{1 \leq i_1 < i_2 < \cdots < i_k \leq n} \mathrm{Var}^{(i_1,\ldots,i_k)} S.$$

Clearly, $J_1 = \sum_{i=1}^n \mathrm{Var}^{(i)} S$ and in view of (4.6), (4.3) can just be rewritten as:

$$\mathrm{Var} S \leq \mathbb{E} \sum_{i=1}^n \mathrm{Var}^{(i)} S = \mathbb{E} J_1. \tag{4.11}$$

Still in view of the results of [5], we now intend to prove:

**Theorem 4.2.1** *For any* $p = 1, 2, \ldots, [n/2]$,

$$\sum_{k=1}^{2p} \frac{(-1)^{k+1}}{k!} \mathbb{E} J_k \leq \mathrm{Var}\ S \leq \sum_{k=1}^{2p-1} \frac{(-1)^{k+1}}{k!} \mathbb{E} J_k, \tag{4.12}$$

*and*

$$\mathrm{Var}\ S = \sum_{k=1}^{n} \frac{(-1)^{k+1}}{k!} \mathbb{E} J_k. \tag{4.13}$$

*Proof* The proof of (4.13) is a simple decomposition/induction, while that of (4.12) further uses convexity. For $k = 1, 2, \ldots, n$, let

$$R_k = \sum_{1 \le i_1 < \cdots < i_k \le n} \mathrm{Var}^{(i_1,\ldots,i_k)}(\mathbb{E}^{(1,\ldots,i_1-1)} S),$$

with the understanding that for $i = 1$, $\mathbb{E}^{(1,i-1)} S = \mathbb{E}^{(0)} S = S$. Then, first note that,

$$\begin{aligned}\mathbb{E} R_1 &= \mathbb{E} \sum_{i_1=1}^{n} \left( (\mathbb{E}^{(1,\ldots,i_1-1)} S)^2 - (\mathbb{E}^{(1,\ldots,i_1)} S)^2 \right) \\ &= \mathbb{E}(S^2 - (\mathbb{E} S)^2) = \mathrm{Var} S. \end{aligned} \tag{4.14}$$

Notice further that for $2 \le k \le n-1$,

$$\begin{aligned}\mathbb{E} R_k &= \mathbb{E} \sum_{1 \le i_1 < \cdots < i_k \le n} \mathrm{Var}^{(i_1,\ldots,i_k)}(\mathbb{E}^{(1,\ldots,i_1-1)} S) \\ &= \mathbb{E} \sum_{1 \le i_1 < \cdots < i_k \le n} \left( \mathrm{Var}^{(i_2,\ldots,i_k)}(\mathbb{E}^{(1,\ldots,i_1-1)} S) - \mathrm{Var}^{(i_2,\ldots,i_k)}(\mathbb{E}^{(1,\ldots,i_1)} S) \right) \\ &= \mathbb{E} \sum_{1 < i_2 < \cdots < i_k \le n} \sum_{i_1=1}^{i_2-1} \left( \mathrm{Var}^{(i_2,\ldots,i_k)}(\mathbb{E}^{(1,\ldots,i_1-1)} S) - \mathrm{Var}^{(i_2,\ldots,i_k)}(\mathbb{E}^{(1,\ldots,i_1)} S) \right) \\ &= \mathbb{E} \sum_{1 \le i_2 < \cdots < i_k \le n} \left( \mathrm{Var}^{(i_2,\ldots,i_k)} S - \mathrm{Var}^{(i_2,\ldots,i_k)}(\mathbb{E}^{(1,\ldots,i_2-1)} S) \right) \\ &= \frac{\mathbb{E} J_{k-1}}{(k-1)!} - \mathbb{E} R_{k-1}. \end{aligned} \tag{4.15}$$

Finally, it is clear that, $R_n = \mathrm{Var}^{(1,\ldots,n)} S$, and so $n! \mathbb{E} R_n = \mathbb{E} J_n$. Combining the last three identities, gives (4.13). To obtain (4.12), note first that by convexity and for any $1 \le i_1 < i_2 < \cdots < i_k \le n$,

$$\mathbb{E}^{(1,\ldots,i_1-1)} \mathrm{Var}^{(i_1,\ldots,i_k)} S \ge \mathrm{Var}^{(i_1,\ldots,i_k)}(\mathbb{E}^{(1,\ldots,i_1-1)} S). \tag{4.16}$$

Hence, taking expectation and summing gives $\mathbb{E} J_k \ge k! \mathbb{E} R_k$, which when combined with (4.15) finishes the proof. □

*Remark 4.2.2*

(i) In case $S$ is symmetric, i.e., invariant under any permutation of its arguments, $J_k = n(n-1)\dots(n-k+1)\mathrm{Var}^{(1,\dots,k)}S$, then $\mathbb{E}J_k = n(n-1)\dots(n-k+1)\mathbb{E}\mathrm{Var}^{(1,\dots,k)}S$, and (4.13) and (4.12) precisely recover corresponding results in [5].

(ii) The inequalities (4.12) can be viewed as martingale inequalities.

(iii) As in [2] or [1], one could also rewrite (4.12) using only the positive or negative parts of the involved quantities.

(iv) It is natural to wonder whether or not the above inequalities have $\mathbf{\Phi}$-entropic versions; this will be explored and presented elsewhere.

Let us now further refine (4.12) providing, in particular, a non-trivial lower bound on the bias $\mathbb{E}J_1 - \mathrm{Var}\, S$ improving upon (4.5). To do so, denote by $\overline{(i_1,\dots,i_k)}$ the complement of the indices $(i_1,\dots,i_k)$ (i.e., the ordered sequence of elements of the set $\{1,\dots,n\}\backslash\{i_1,\dots,i_k\}$, and introduce the following quantities:

$$K_k := k! \sum_{1\le i_1<i_2<\cdots<i_k\le n} \mathrm{Var}^{(i_1,\dots,i_k)}\mathbb{E}^{\overline{(i_1,\dots,i_k)}}S.$$

It is clear that by Jensen's inequality and the convexity of $\mathrm{Var}^{(i_1,\dots,i_k)}$ we have

$$\mathbb{E}K_k \le \mathbb{E}J_k\,.$$

**Theorem 4.2.3** *For any* $p = 1, 2, \dots, [n/2]$,

$$\sum_{k=1}^{2p} \frac{(-1)^{k+1}}{k!}\mathbb{E}J_k + \frac{1}{(2p+1)!}\mathbb{E}K_{2p+1} \le \mathrm{Var}\, S \le \sum_{k=1}^{2p-1} \frac{(-1)^{k+1}}{k!}\mathbb{E}J_k - \frac{1}{(2p)!}\mathbb{E}K_{2p} \tag{4.17}$$

*Proof* The only modification compared to the proof of Theorem 4.2.1 is that instead of using the bound $\mathbb{E}J_k \ge k!\mathbb{E}R_k$ we use the fact that

$$\mathbb{E}K_k \le k!\mathbb{E}R_k\,,$$

which follows from the convexity of $\mathrm{Var}^{(i_1,\dots,i_k)}$. □

In particular, from Theorems 4.2.1 and 4.2.3, the following inequalities hold true:

$$\frac{1}{2}\mathbb{E}K_2 \le \mathbb{E}J_1 - \mathrm{Var}\, S \le \frac{1}{2}\mathbb{E}J_2. \tag{4.18}$$

## 4.3 Relationship with the Hoeffding Decomposition

Let us recall the notion of a Hoeffding decomposition [4] (see [7] Section 2 for the general non-symmetric non-iid case). Given a function $f(X) \in L_1(\mathbb{P})$, it is the unique decomposition

$$\begin{aligned} f(X_1, \ldots, X_n) &= \mathbb{E} f(X) + \sum_{1 \le i \le n} h_i(X_i) + \sum_{1 \le i < j \le n} h_{ij}(X_i, X_j) + \ldots \\ &= f_0 + f_1 + \ldots + f_n, \end{aligned}$$

such that $\mathbb{E}^{(i_s)} h_{i_1,\ldots,i_k}(X_{i_1}, \ldots, X_{i_k}) = 0$ whenever $1 \le i_1 < \ldots < i_k \le n$, $s = 1, \ldots, k$. The term $f_d$ is called the Hoeffding term of degree $d$ and these terms form an orthogonal decomposition of $f$ in $L_2(\mathbb{P})$ (provided, of course, $f \in L_2(\mathbb{P})$), so that $\mathrm{Var} f = \sum_{k=1}^n \mathrm{Var}\, f_k = \sum_{I \subset \{1,\ldots,n\}} \mathbb{E} h_I^2$

The following lemma provides a relationship between the previously introduced iterated jackknives and the variance of the Hoeffding terms.

**Lemma 4.3.1** *For any k such that* $1 \le k \le n$*,*

$$\frac{1}{k!} \mathbb{E} J_k(f) = \sum_{j \ge k} \binom{j}{k} Var\, f_j,$$

*and*

$$\frac{1}{k!} \mathbb{E} K_k(f) = Var\, f_k\,.$$

*Proof* Let us rewrite the Hoeffding decomposition of $f$ as $f = \sum_{I \subset \{1,2,\ldots,n\}} h_I$. We have $\mathbb{E}^{(i)} h_I = 0$ whenever $i \in I$, and $\mathbb{E}^{(i)} h_I = h_I$ otherwise. Hence, $\mathrm{Var}^{(i)} h_I = \mathbb{E}^{(i)} h_I^2$ if $i \in I$ and 0 otherwise. Therefore, $\mathbb{E}\mathrm{Var}^{(i)} S = \sum_{i \in I} \mathbb{E} h_I^2$.

Continuing with the same reasoning, we can see that $\mathrm{Var}^{(i)} \mathbb{E}^{(j)} h_I = \mathbb{E}^{(i)} h_I^2$, if $i \in I$ and $j \notin I$ and 0 otherwise, thus $\mathbb{E}\mathrm{Var}^{(i)} \mathbb{E}^{(j)} S = \sum_{i \in I, j \notin I} \mathbb{E} h_I^2$ so that $\mathbb{E}\mathrm{Var}^{(i,j)} S = \sum_{\{i,j\} \subset I} \mathbb{E} h_I^2$ and by induction, we get that

$$\mathbb{E}\mathrm{Var}^{(i_1,\ldots,i_k)} S = \sum_{\{i_1,\ldots,i_k\} \subset I} \mathbb{E} h_I^2\,.$$

If we now sum over the possible sets of indices, since each term $\mathbb{E} h_I^2$ appears as many times as there are subsets of size $k$ of $I$, this implies that $\mathbb{E} J_k = k! \sum_{|I| \ge k} \binom{|I|}{k} \mathbb{E} h_I^2 = k! \sum_{j \ge k} \binom{j}{k} \mathrm{Var}\, f_j$ and gives the first statement.

To prove the second statement of the lemma, observe that $\mathbb{E}^{\overline{(i_1,\ldots,i_k)}}S = \sum_{I\subset\{i_1,\ldots,i_k\}} h_I$ so that $\mathbb{E}\mathrm{Var}^{(i_1,\ldots,i_k)}\mathbb{E}^{\overline{(i_1,\ldots,i_k)}}S = \mathbb{E}h^2_{i_1,\ldots,i_k}$, and therefore $\mathbb{E}K_k = k!\sum_{|I|=k}\mathbb{E}h_I^2 = k!\mathrm{Var}\, f_k$. □

It is easily verified that (4.13) can be recovered as a consequence of Lemma 4.3.1. Also, from Lemma 4.3.1 it is easy to get the following corollary obtained in [1] (as part of their Theorem 1.8).

**Corollary 4.3.2** *Let $S$ have Hoeffding decomposition of type $S = \mathbb{E}S + \sum_{k=d}^n S_k$, i.e., such that $f_k = 0$, for $1 \le k < d$, then*

$$\mathrm{Var}\ S \le \frac{1}{d!}\mathbb{E}J_d. \tag{4.19}$$

*Proof* Using the fact that $f_k = 0$, for $1 \le k < d$, we have

$$\mathrm{Var}\, S = \sum_{j=d}^n \mathrm{Var}\, f_j \le \sum_{j=d}^n \binom{j}{d}\mathrm{Var}\, f_j = \frac{1}{d!}\mathbb{E}J_d,$$

where the last equality follows from Lemma 4.3.1. □

**Acknowledgements** C. Houdré was partially supported by the grant # 524678 from the Simons Foundation and by a Simons Fellowship grant # 267336. This material is also based in part upon material supposed by NSF grant No. 1440140 while this author was in residence at MSRI in Berkeley, CA, during Fall, 2017.

## References

1. S.G. Bobkov, F. Götze, H. Sambale, Higher order concentration of measure. Commun. Contemp. Math. **21**, 1850043 (2019). arXiv:1709.06838
2. S. Boucheron, G. Lugosi, P. Massart, *Concentration Inequalities. A nonasymptotic Theory of Independence* (Oxford University Press, Oxford, 2013)
3. B. Efron, C. Stein, The jackknife estimate of variance. Ann. Stat. **9**, 586–596 (1981)
4. W. Hoeffding, A class of statistics with asymptotically normal distribution. Ann. Math. Statist. **19**, 293–325 (1948)
5. C. Houdré, The iterated jackknife estimate of variance. Statist. Probab. Lett. **35**, 197–201 (1997)
6. C. Houdré, A. Kagan, Variance inequalities for functions of Gaussian variables. J. Theoret. Probab. **8**, 23–30 (1995)
7. S. Karlin, Y. Rinott, Applications of ANOVA type decompositions for comparisons of conditional variance statistics including jackknife estimates. Ann. Stat. **10**, 485–501 (1982)
8. A. Maurer, A Bernstein-type inequality for functions of bounded interaction. Bernoulli **25**, 1451–1471 (2019). arXiv: 1701.06191
9. A. Maurer, M. Pontil, Empirical bounds for functions with weak interactions, in *COLT 2018*. arXiv:1803.03934

# Chapter 5
# A Probabilistic Characterization of Negative Definite Functions

**Fuchang Gao**

**Abstract** It is proved that a continuous function $f$ on $\mathbb{R}^n$ is negative definite if and only if it is polynomially bounded and satisfies the inequality $\mathbb{E}f(X-Y) \le \mathbb{E}f(X+Y)$ for all i.i.d. random vectors $X$ and $Y$ in $\mathbb{R}^n$. The proof uses Fourier transforms of tempered distributions. The "only if" part has been proved earlier by Lifshits et al. (A probabilistic inequality related to negative definite functions. Progress in probability, vol. 66 (Springer, Basel, 2013), pp. 73–80).

**Keywords** Negative definite function · Lévy–Khintchine representation · Fourier inversion theorem · Polynomially bounded

**2010 Mathematics Subject Classification** Primary: 60E15, 42A82; Secondary: 42B10, 60E10

## 5.1 Introduction

A real-valued function $f$ on $\mathbb{R}^n$ is said to be negative definite if for every sequence of vectors $x_1, x_2, \ldots, x_m$ in $\mathbb{R}^n$, the matrix

$$(f(x_i) + f(x_j) - f(x_i - x_j))_{1 \le i,j \le m} \tag{5.1}$$

is positive definite [3]; or equivalently [8], for every sequence of vectors $x_1, x_2, \ldots, x_m$ in $\mathbb{R}^n$, and every sequence of real numbers $\rho_1, \rho_2, \ldots, \rho_m$ satisfying $\sum_{i=1}^m \rho_i = 0$, the following inequality holds

$$\sum_{i=1}^m \sum_{j=1}^m f(x_i - x_j)\rho_i \rho_j \le 0. \tag{5.2}$$

F. Gao (✉)
Department of Mathematics, University of Idaho, Moscow, ID, USA
e-mail: fuchang@uidaho.edu



N. Gozlan et al. (eds.), *High Dimensional Probability VIII*,
Progress in Probability 74, https://doi.org/10.1007/978-3-030-26391-1_5

The conditions (5.1) and (5.2) are difficult to check in general. If in addition $f$ is continuous, then by the well known Lévy–Khintchine representation (cf. [7]), $f$ can be uniquely expressed as

$$f(\xi) = f(0) + \langle Q\xi, \xi\rangle + \int_{\mathbb{R}^n\setminus\{0\}} (1 - \cos\langle u, \xi\rangle)\, d\nu(u), \tag{5.3}$$

where $Q$ is a positive semidefinite matrix, and $d\nu$ is a Borel measure on $\mathbb{R}^n \setminus \{0\}$ satisfying

$$\int_{\mathbb{R}^n\setminus\{0\}} \min\{\|u\|^2, 1\} d\nu(u) < \infty,$$

where and in the rest of the paper, $\|\cdot\|$ means the Euclidean norm.

Negative definite functions have many applications in potential theory, statistics, and the theory of probability. For example, they are closely related to Lévy processes. The purpose of this paper is to provide a probabilistic characterization of negative definite functions. This study is motivated by a recent work of Lifshits et al. [5], in which it was proved that if $f$ is a continuous real-valued negative definite function on $\mathbb{R}^n$, then the inequality

$$\mathbb{E} f(X - Y) \le \mathbb{E} f(X + Y) \tag{5.4}$$

holds for all i.i.d. random vectors $X$ and $Y$ in $\mathbb{R}^n$. (Here and throughout the paper, $A \le B$ means either $A \le B < \infty$, or $B = +\infty$.) The main idea of the proof of [5] is as follows: If $f$ is a continuous negative definite function on $\mathbb{R}^n$, then by the Lévy–Khintchine representation (5.3) we can write

$$f(X + Y) - f(X - Y) = 4\langle QX, Y\rangle + 2\int_{\mathbb{R}^n\setminus\{0\}} (\sin\langle u, X\rangle \sin\langle u, Y\rangle) d\nu(u).$$

Taking expectation and using Fubini's Theorem, one obtains the desired inequality (5.4). What seems to be a bit surprising is that under some growth rate assumptions on $f$, the validity of the inequality (5.4) for all i.i.d. random vectors $X$ and $Y$ in $\mathbb{R}^n$ also implies that $f$ is negative definite. This is the main contribution of the current paper. The proof uses Fourier transforms of tempered distributions.

Inequalities relating $X + Y$ and $X - Y$ have attracted interest from different communities. We refer interested readers to a recent article of Li and Madiman [4] in which inequalities on small ball probabilities of $X+Y$ and $X-Y$ were established, where $X$ and $Y$ are i.i.d. random variables taking values in an abelian topological group.

## 5.2 Statement of Results

A function $f$ on $\mathbb{R}^n$ is said to be polynomially bounded if there exist a positive constant $C$ and a positive integer $k$ such that $|f(x)| \le C\|x\|^k$ for all $\|x\| \ge 1$. The main result of this paper is the following characterization of negative definite functions on $\mathbb{R}^n$.

**Theorem 5.1** *If $f$ is a negative definite function on $\mathbb{R}^n$, then the inequality $\mathbb{E}f(X-Y) \le \mathbb{E}f(X+Y)$ holds for all i.i.d. random vectors $X$ and $Y$ that take finitely many values in $\mathbb{R}^n$. If $f$ is a continuous function on $\mathbb{R}^n$, then $f$ is negative definite if and only if $f$ is polynomially bounded and satisfies $\mathbb{E}f(X-Y) \le \mathbb{E}f(X+Y)$ for all i.i.d. random vectors $X$ and $Y$ in $\mathbb{R}^n$.*

*Remark 5.2* In Theorem 5.1 and throughout the rest of the paper, $\mathbb{E}f(X-Y) \le \mathbb{E}f(X+Y)$ means either $\mathbb{E}f(X-Y) \le \mathbb{E}f(X+Y) < \infty$ or $\mathbb{E}f(X+Y) = \infty$.

In applications, one may need an inequality $\mathbb{E}f(X-Y) \le \mathbb{E}f(X+Y)$ for functions $f$ which have a singularity at 0. Note that if $f$ has singularity at 0, the matrix $(f(x_i)+f(x_j)-f(x_i-x_j))$ makes no sense, while the expectation $\mathbb{E}f(X-Y) - \mathbb{E}f(X+Y)$ may still make sense for continuous random variables. From the proof of Theorem 5.1 the following result can be easily observed:

**Proposition 5.3** *Let $f$ be continuous everywhere except at its unique singular point at 0. Suppose $f$ is polynomially bounded. If the Fourier transform of $f$ is negative (i.e., as a linear functional on the space of Schwartz test functions, the Fourier transform $\widehat{f}$ maps every positive Schwartz test function into a non-positive number), then, $\mathbb{E}f(X-Y) \le \mathbb{E}f(X+Y)$ for all i.i.d. random vectors $X$ and $Y$.*

Note that the negativity of $\widehat{f}$ is related to the negative definiteness of the generalized function $f$, cf. [2]. As an example, we consider the function $f(x) = -\|x\|^{-\beta}$ on $\mathbb{R}^n \setminus \{0\}$, where $0 < \beta < n$. A direct computation shows that its Fourier transform is negative:

$$\widehat{f}(\xi) = -\frac{\pi^{n/2} 2^{n-\beta} \Gamma(\frac{n-\beta}{2})}{\Gamma(\frac{\beta}{2})} \|\xi\|^{\beta-n} < 0.$$

Thus, by Proposition 5.3, for all i.i.d. random vectors $X$ and $Y$ in $\mathbb{R}^n$,

$$\mathbb{E}\|X+Y\|^{-\beta} \le \mathbb{E}\|X-Y\|^{-\beta}. \tag{5.5}$$

Note that the function $\|x\|^{-\beta}$ does not have a Lévy–Khintchine representation, but it has the following Fourier transform representation:

$$\|x\|^{-\beta} = \frac{\Gamma\left(\frac{n-\beta}{2}\right)}{2^{\beta}\pi^{n/2}\Gamma(\frac{\beta}{2})} \int_{\mathbb{R}^n} \cos\langle x, \xi\rangle \, \|\xi\|^{\beta-n} d\xi,$$

from which (5.5) can be easily derived. (Note that the integral converges when $0 < \beta < n$.)

Theorem 5.1 should be compared with the following characterization of functions satisfying inequality (5.4) for all i.i.d. random vectors:

**Theorem 5.4** *The inequality $\mathbb{E}f(X-Y) \le \mathbb{E}f(X+Y)$ holds for all i.i.d. random vectors $X$ and $Y$ that take finitely many values in $\mathbb{R}^n$ if and only if for every sequence of vectors $x_1, x_2, \dots, x_m$ in $\mathbb{R}^n$, the matrix*

$$(f(x_i + x_j) - f(x_i - x_j))_{1 \le i,j \le m} \tag{5.6}$$

*is positive definite. If $f$ is continuous, then the inequality $\mathbb{E}f(X-Y) \le \mathbb{E}f(X+Y)$ holds for all i.i.d. random vectors $X$ and $Y$ in $\mathbb{R}^n$ if and only if the matrices in (5.6) are positive definite.*

As we will see in the proof that Theorem 5.4 remains valid if the continuity of $f$ is replaced by local integrability of $f(X+Y)$ and $f(X-Y)$. Also note that the matrix $(f(x_i + x_j) - f(x_i - x_j))$ in Theorem 5.4 has some similarity with the matrix $(f(x_i) + f(x_j) - f(x_i - x_j))$ in (5.1) by which negative definite functions are defined. Because by Lévy–Khintchine representation every continuous function satisfying (5.1) is polynomially bounded, one might wonder if the requirement of polynomial boundedness in Theorem 5.1 is redundant. The following theorem says that it is not the case.

**Theorem 5.5** *For all i.i.d. random variables $X$ and $Y$ in $\mathbb{R}$,*

$$\mathbb{E}e^{|X-Y|} \le \mathbb{E}e^{|X+Y|};$$

*while the function $e^{|x|}$ is not a negative definite function on $\mathbb{R}$.*

## 5.3 Proofs

We first prove Theorem 5.4, one of the properties proved in the proof of Theorem 5.4 will be used in the proof of Theorem 5.1.

*Proof of Theorem 5.4* Suppose $\mathbb{E}f(X+Y) \ge \mathbb{E}f(X-Y)$ for all i.i.d. random vectors $X$ and $Y$ that take finitely many values in $\mathbb{R}^n$. We first show that $f$ is an even function. Indeed, for any fixed $y \in \mathbb{R}^n$, let $X$ and $Y$ be i.i.d. random vectors such that $\mathbb{P}(X = y) = p$ and $\mathbb{P}(X = 0) = 1 - p$. Then

$$\begin{aligned}\mathbb{E}f(X+Y) - \mathbb{E}f(X-Y) &= [p^2 f(2y) + 2p(1-p)f(y) + (1-p)^2 f(0)] \\ &\quad - [p^2 f(0) + (1-p)^2 f(0) + p(1-p)f(y) + p(1-p)f(-y)] \\ &= p[p(f(2y) - f(0)) + (1-p)(f(y) - f(-y))].\end{aligned}$$

Because $\mathbb{E}f(X+Y) \geq \mathbb{E}f(X-Y)$, we obtain

$$p(f(2y) - f(0)) + (1-p)(f(y) - f(-y)) \geq 0.$$

Letting $p \to 0^+$, we obtain $f(y) \geq f(-y)$. Since $y$ is arbitrary, we also have $f(-y) \geq f(y)$. Hence $f(-y) = f(y)$ for all $y \in \mathbb{R}^n$, and therefore $f$ is even.

Next, we show that the matrix $(f(x_i + x_j) - f(x_i - x_j))$ is positive definite for every sequence of vectors $x_1, x_2, \ldots, x_m$ in $\mathbb{R}^n$. Take any sequence of real numbers $\rho_1, \rho_2, \ldots, \rho_m$ that are not all 0. Without loss of generality, we assume $\sum_{i=1}^m |\rho_i| = 1$. Let $X$ and $Y$ be i.i.d. random vectors such that

$$\mathbb{P}(X = \mathrm{sign}(\rho_i)x_i) = |\rho_i|.$$

Then, by using the fact that $f$ is an even function in the second equality below, we have

$$\begin{aligned}
0 &\leq \mathbb{E}f(X+Y) - \mathbb{E}f(X-Y) \\
&= \sum_{i=1}^m \sum_{j=1}^m [f(\mathrm{sign}(\rho_i)x_i + \mathrm{sign}(\rho_j)x_j) - f(\mathrm{sign}(\rho_i)x_i - \mathrm{sign}(\rho_j)x_j)]|\rho_i||\rho_j| \\
&= \sum_{\rho_i\rho_j \geq 0} [f(x_i + x_j) - f(x_i - x_j)]\rho_i\rho_j - \sum_{\rho_i\rho_j < 0} [f(x_i - x_j) - f(x_i + x_j)]\rho_i\rho_j \\
&= \sum_{i=1}^m \sum_{j=1}^m [f(x_i + x_j) - f(x_i - x_j)]\rho_i\rho_j.
\end{aligned}$$

Hence, the matrix $(f(x_i + x_j) - f(x_i - x_j))$ is positive definite. (This method has been used in the proof of Theorem 2.3 of [1].)

On the other hand, suppose the matrix $(f(x_i + x_j) - f(x_i - x_j))$ is positive definite for every sequence of vectors $x_1, x_2, \ldots, x_m$. For every pair of i.i.d. random vectors $X$ and $Y$ that take finitely many values in $\mathbb{R}^n$, suppose $\sum_{i=1}^k \mathbb{P}(X = z_i) = 1$. By letting $m = k$, $x_i = z_i$, and $\rho_i = \mathbb{P}(X = x_i)$ for $1 \leq i \leq m$, we have

$$\mathbb{E}f(X+Y) - \mathbb{E}f(X-Y) = \sum_{i=1}^m \sum_{j=1}^m [f(x_i + x_j) - f(x_i - x_j)]\rho_i\rho_j \geq 0.$$

This finishes the proof of the first statement in Theorem 5.4.

For the second statement in Theorem 5.4, we only need to prove the "if" part. Since the matrix $(f(x_i + x_j) - f(x_i - x_j))_{1 \leq i,j \leq m}$ is positive definite for every sequence of vectors $x_1, x_2, \ldots, x_m$, by choosing $m = 1$, we see that $f(x) \geq f(0)$ for all $x \in \mathbb{R}^n$. By otherwise replacing $f(x)$ by $f(x) - f(0)$, we can now assume $f \geq 0$.

Let $X, Y$ be i.i.d. continuous random vectors in $\mathbb{R}^n$. Fix a large number $M$. We partition $[-M, M]^n$ into $M^{2n}$ small cubes of side length $2/M$. Let $x_1, x_2, \dots x_m$ be the center of these small cubes. Thus, $[-M, M]^n = \cup_{i=1}^m (x_i + [-1/M, 1/M]^n)$. Let $\rho_i = \mathbb{P}(X \in x_i + [-1/M, 1/M]^n)$. If $\mathbb{E} f(X+Y) < \infty$, then by the "if" part of the first statement in Theorem 5.4, we have

$$\limsup_{M\to\infty} \sum_{i=1}^m \sum_{j=1}^m f(x_i - x_j)\rho_i\rho_j \le \limsup_{M\to\infty} \sum_{i=1}^m \sum_{j=1}^m f(x_i + x_j)\rho_i\rho_j = \mathbb{E} f(X+Y),$$

which implies that $\mathbb{E} f(X-Y) \le \mathbb{E} f(X+Y)$.

*Proof of Theorem 5.1* Suppose $f$ is negative definite on $\mathbb{R}^n$. Then, by definition, the matrix $(f(x_i) + f(x_j) - f(x_i - x_j))$ is positive definite for every sequence of vectors $x_1, x_2, \dots, x_m$. In particular, it implies that the matrix is symmetric. So, we have $f(x_i - x_j) = f(x_j - x_i)$, and hence $f$ is an even function.

We prove that for any i.i.d. random vectors $X$ and $Y$ that take finitely many values in $\mathbb{R}^n$, the inequality $\mathbb{E} f(X-Y) - \mathbb{E} f(X+Y) \le 0$ holds.

Suppose $\mathbb{P}(X = x_i) > 0$, $1 \le i \le m$ and $\sum_{i=1}^m \mathbb{P}(X = x_i) = 1$. Denote

$$T = \{x_1, x_2, \dots, x_m\} \cup \{-x_1, -x_2, \dots, -x_m\},$$

and relabel it as $\{z_1, z_2, \dots, z_k\}$. Because $f$ is an even function, we have

$$\begin{aligned}
&\mathbb{E} f(X-Y) - \mathbb{E} f(X+Y) \\
&= \frac{1}{2} \sum_{x\in T} \sum_{y\in T} f(x-y)[\mathbb{P}(X=x) - \mathbb{P}(X=-x)][\mathbb{P}(Y=y) - \mathbb{P}(Y=-y)] \\
&= \frac{1}{2} \sum_{i=1}^k \sum_{j=1}^k f(z_i - z_j)[\mathbb{P}(X=z_i) - \mathbb{P}(X=-z_i)][\mathbb{P}(Y=z_j) - \mathbb{P}(Y=-z_j)].
\end{aligned} \tag{5.7}$$

Let $\rho_i = \mathbb{P}(X = z_i) - \mathbb{P}(X = -z_i)$. Because

$$\sum_{i=1}^m \rho_i = \sum_{i=1}^k [\mathbb{P}(X = z_i) - \mathbb{P}(X = -z_i)] = 1 - 1 = 0,$$

the right-hand side of (5.7) is non-positive by (5.2). This proves the first statement of Theorem 5.1.

Now, we assume that $f$ is a continuous function on $\mathbb{R}^n$. Because a continuous negative definite function is necessarily polynomial bounded, we only need to prove the validity of the inequality. Just as in the proof of second statement of Theorem 5.4, the "only if" part of the second statement of Theorem 5.1 follows by approximating continuous random vectors using random vectors that take finitely many values in $\mathbb{R}^n$, and taking limit. It is also a simple application of Lévy–Khintchine representation Theorem as we discussed in the introduction. Thus, we only need to prove the "if" part.

Note that for every continuous function $f$ that is polynomially bounded, and every $x \in \mathbb{R}^n$, we have

$$f(x) = \lim_{\sigma \to 0} \int_{\mathbb{R}^n} f(x+u) G_\sigma(u) du, \tag{5.8}$$

where

$$G_\sigma(u) = \frac{1}{(\sqrt{2\pi}\sigma)^n} e^{-\frac{1}{2\sigma^2}\|u\|^2}.$$

This is an elementary fact in approximation of unit (cf. Theorem 6.32 in [6]), and can be easily proved. Indeed, because $f$ is continuous at $x$, for any $\varepsilon > 0$, there exists $0 < \delta < 1$ such that for every $\|u\| < \delta$, we have $|f(x+u) - f(x)| \le \varepsilon$. Let $M(x) = \sup_{\delta < \|u\| \le 1 + \|x\|} |f(x+u)|$. Then,

$$\begin{aligned}
\left| \int_{\mathbb{R}^n} f(x+u) G_\sigma(u) du - f(x) \right| &= \left| \int_{\mathbb{R}^n} [f(x+u) - f(x)] G_\sigma(u) du \right| \\
\le \int_{\|u\| \le \delta} \varepsilon G_\sigma(u) du &+ \int_{\|u\| > \delta} |f(x)| G_\sigma(u) du \\
+ \int_{\delta < |u\| \le 1 + \|x\|} |f(x+u)| G_\sigma(u) du &+ \int_{\|u\| > 1 + \|x\|} |f(x+u)| G_\sigma(u) du \\
\le \varepsilon + (|f(x)| + M(x)) \int_{\|u\| > \delta} G_\sigma(u) du &+ \int_{\|u\| > 1 + \|x\|} C(2\|u\|)^k G_\sigma(u) du.
\end{aligned} \tag{5.9}$$

It is easy to check that the last two terms on the right-hand side go to 0 as $\sigma \to 0$.

Now, we use expression (5.8) to show that if $f$ is polynomially bounded and satisfies the inequality (5.4) for all i.i.d. random vectors $X$ and $Y$, then $f$ is negatively definite; or equivalently, for every sequence of vectors $x_1, x_2, \ldots, x_m$ in $\mathbb{R}^n$, and every sequence of real numbers $\rho_1, \rho_2, \ldots, \rho_m$, with $\sum_{i=1}^m \rho_i = 0$, we have

$$\Sigma := \sum_{i=1}^m \sum_{j=1}^m f(x_i - x_j) \rho_i \rho_j \le 0. \tag{5.10}$$

We claim that we only need to consider the case when $m \ge n$ and

$$\operatorname{span}\{x_1, x_2, \ldots, x_m\} = \mathbb{R}^n. \tag{5.11}$$

Indeed, if (5.10) has been proved under the assumption (5.11). For any sequence $x_1, x_2, \ldots, x_m$ in $\mathbb{R}^n$, if $\text{span}\{x_1, x_2, \ldots, x_m\} \neq \mathbb{R}^n$, we can find some extra vectors $x_{m+1}, x_{m+2} \ldots, x_{m+s}$, such that

$$\text{span}\{x_1, x_2, \ldots, x_{m+s}\} = \mathbb{R}^n.$$

By choosing $\rho_{m+1} = \cdots = \rho_{m+s} = 0$, we have

$$\sum_{i=1}^{m}\sum_{j=1}^{m} f(x_i - x_j)\rho_i\rho_j = \sum_{i=1}^{m+s}\sum_{j=1}^{m+s} f(x_i - x_j)\rho_i\rho_j \leq 0.$$

This means that the inequality (5.10) continues to hold without the assumption (5.11).

To prove (5.10) under the assumption (5.11), without loss of generality, we can assume $\sum_{i=1}^{m} |\rho_i| = 1$. By using (5.8), we can write

$$\Sigma = \lim_{\sigma\to 0}\int_{\mathbb{R}^n}\left[\sum_{i=1}^{m}\sum_{j=1}^{m} f(x_i - x_j + u)\rho_i\rho_j\right] G_\sigma(u)du.$$

Since $f$ is a tempered distribution, and $G_\sigma$ is a Schwartz test function, by Parseval identity [9] and the fact that $f$ is even (which was proved in the first paragraph of the proof of Theorem 5.4), we have

$$\int_{\mathbb{R}^n} f(x_i - x_j + u)G_\sigma(u)du = \frac{1}{(2\pi)^n}\int_{\mathbb{R}^n}\widehat{f}(\xi)\cos\left\langle x_i - x_j, \xi\right\rangle e^{-\frac{\sigma^2}{2}\|\xi\|^2}d\xi.$$

Thus,

$$\begin{aligned}
\Sigma &= \lim_{\sigma\to 0}\frac{1}{(2\pi)^n}\int_{\mathbb{R}^n}\widehat{f}(\xi)\left[\sum_{i=1}^{m}\sum_{j=1}^{m}\cos\left\langle x_i - x_j, \xi\right\rangle\rho_i\rho_j\right] e^{-\frac{\sigma^2}{2}\|\xi\|^2}d\xi \\
&= \lim_{\sigma\to 0}\frac{1}{(2\pi)^n}\int_{\mathbb{R}^n}\widehat{f}(\xi)\left[\sum_{i=1}^{m}\sum_{j=1}^{m}[\cos\langle x_i, \xi\rangle\cos\left\langle x_j, \xi\right\rangle + \sin\langle x_i, \xi\rangle\sin\left\langle x_j, \xi\right\rangle]\rho_i\rho_j\right] e^{-\frac{\sigma^2}{2}\|\xi\|^2}d\xi \\
&= \lim_{\sigma\to 0}\frac{1}{(2\pi)^n}\int_{\mathbb{R}^n}\widehat{f}(\xi)\left[\left(\sum_{i=1}^{m}\cos\langle x_i, \xi\rangle\,\rho_i\right)^2 + \left(\sum_{i=1}^{m}\sin\langle x_i, \xi\rangle\,\rho_i\right)^2\right] e^{-\frac{\sigma^2}{2}\|\xi\|^2}d\xi \\
&= \lim_{\sigma\to 0}\frac{1}{(2\pi)^n}\int_{\mathbb{R}^n}\widehat{f}(\xi)\left[\left(\sum_{i=1}^{m}\rho_i(1 - \cos\langle x_i, \xi\rangle)\right)^2 + \left(\sum_{i=1}^{m}\rho_i\sin\langle x_i, \xi\rangle\right)^2\right] e^{-\frac{\sigma^2}{2}\|\xi\|^2}d\xi,
\end{aligned}$$

where in the last equality we used the fact that $\sum_{i=1}^{m}\rho_i = 0$.

On the other hand, if $X$ and $Y$ are i.i.d. random vectors with a density $g$ which is a Schwartz test function, then Parseval identity together with the fact that $f$ is an even function gives

$$\mathbb{E}f(X-Y)-\mathbb{E}f(X+Y)=\frac{1}{2(2\pi)^n}\int_{\mathbb{R}^n}\widehat{f}(\xi)|\widehat{g}(\xi)-\widehat{g}(-\xi)|^2 d\xi. \tag{5.12}$$

Indeed,

$$\begin{aligned}\mathbb{E}f(X-Y)-\mathbb{E}f(X+Y)&=\int_{\mathbb{R}^n}\left[\int_{\mathbb{R}^n}[f(x-y)-f(x+y)]g(x)dx\right]g(y)dy\\&=\int_{\mathbb{R}^n}\left[\frac{1}{(2\pi)^n}\int_{\mathbb{R}^n}\widehat{f}(\xi)(e^{-\langle y,\xi\rangle i}-e^{\langle y,\xi\rangle i})\overline{\widehat{g}(\xi)}d\xi\right]g(y)dy\\&=\frac{1}{(2\pi)^n}\int_{\mathbb{R}^n}\widehat{f}(\xi)\left(\widehat{g}(\xi)-\widehat{g}(-\xi)\right)\overline{\widehat{g}(\xi)}d\xi. \end{aligned}\tag{5.13}$$

Since $f$ is even, so is $\widehat{f}$. By changing variable $\xi$ to $-\xi$, we can rewrite (5.13) as

$$\mathbb{E}f(X-Y)-\mathbb{E}f(X+Y)=-\frac{1}{(2\pi)^n}\int_{\mathbb{R}^n}\widehat{f}(\xi)[\widehat{g}(\xi)-\widehat{g}(-\xi)]\overline{\widehat{g}(-\xi)}d\xi. \tag{5.14}$$

Averaging (5.13) and (5.14), we obtain (5.12).

In particular, if we define a random vector $W$ such that

$$\mathbb{P}(W=\operatorname{sign}(\rho_i)x_i)=|\rho_i|,\ i=1,2,\ldots,m,$$

and let $X$ and $Y$ be i.i.d. random vectors with the same distribution as $W+\frac{\sigma}{\sqrt{2}}Z$, where $Z$ is a standard centered Gaussian random vector independent of $W$, then it is straightforward to check that the density function $g$ of $X$ and $Y$ is a Schwartz test function:

$$g(x)=\sum_{i=1}^m|\rho_i|\frac{1}{(\sqrt{\pi}\sigma)^n}e^{-\frac{1}{\sigma^2}\|x-\operatorname{sign}(\rho_i)x_i\|^2}.$$

Thus,

$$\widehat{g}(\xi)=\sum_{i=1}^m|\rho_i|e^{-\frac{\sigma^2}{4}\|\xi\|^2}e^{-i\langle\operatorname{sign}(\rho_i)x_i,\xi\rangle},$$

which implies that

$$\widehat{g}(\xi) - \widehat{g}(-\xi) = -2i \sum_{i=1}^{m} e^{-\frac{\sigma^2}{4}\|\xi\|^2} \rho_i \sin\langle x_i, \xi\rangle .v$$

Thus,

$$|\widehat{g}(\xi) - \widehat{g}(-\xi)|^2 = 4\left(\sum_{i=1}^{m} \rho_i \sin\langle x_i, \xi\rangle\right)^2 e^{-\frac{\sigma^2}{2}\|\xi\|^2}.$$

Thus, by using the expression (5.12), we have

$$\frac{1}{(2\pi)^n}\int_{\mathbb{R}^n} \widehat{f}(\xi)\left(\sum_{i=1}^{m} \rho_i \sin\langle x_i, \xi\rangle\right)^2 e^{-\frac{\sigma^2}{2}\|\xi\|^2} d\xi = \mathbb{E}f(X-Y) - \mathbb{E}f(X+Y) \le 0. \tag{5.15}$$

Next, we will use a similar method to show that

$$\frac{1}{(2\pi)^n}\int_{\mathbb{R}^n} \widehat{f}(\xi)\left(\sum_{i=1}^{m} \rho_i (1-\cos\langle x_i, \xi\rangle)\right)^2 e^{-\frac{\sigma^2}{2}\|\xi\|^2} d\xi \le 0,$$

from which Theorem 5.1 follows.

For $c > 0$, define

$$T(c;t) = \tanh(c(t_1 + t_2 + \cdots + t_n)) \sum_{i=1}^{m} \rho_i (1-\cos\langle x_i, t\rangle) e^{-\frac{\sigma^2}{8}\|t\|^2}.$$

Because $\lim_{c\to\infty} |\tanh(c(t_1+t_2+\cdots+t_n))| = 1$ except on $\{t_1+t_2+\cdots+t_n = 0\}$ which has measure 0, it suffices to show that for each $c > 0$,

$$\int_{\mathbb{R}^n} \widehat{f}(\xi) T^2(c;\xi) e^{-\frac{\sigma^2}{4}\|\xi\|^2} d\xi \le 0. \tag{5.16}$$

For notational simplicity, for fixed $c$, we denote $T(c;t)$ simply by $T(t)$. Let $\widehat{T}$ be the Fourier transform of $T(t)$. Since $T(t)$ is odd and continuously differentiable, the cosine term does not appear. Integrating by parts and using the fact that $T(t)$ is a Schwartz test function, we obtain

$$\begin{aligned}\widehat{T}(y) &= -i\int_{\mathbb{R}^n} T(t)\sin\langle y, t\rangle\, dt_1 dt_2 \cdots dt_n \\ &= \frac{(-1)^{n+1} i}{y_1^2 y_2^2 \cdots y_n^2}\int_{\mathbb{R}^n} \frac{\partial^{2n} T(t)}{\partial t_1^2 \partial t_2^2 \cdots \partial t_n^2} \sin\langle y, t\rangle\, dt_1 dt_2 \cdots dt_n.\end{aligned} \tag{5.17}$$

In particular, this implies that $i\widehat{T}$ is a real-valued absolutely integrable function on $\mathbb{R}^n$, and there exists a positive constant $\lambda$ such that $\lambda|\widehat{T}(\xi)|$ is a probability density.

Let $W$ be a random vector with density function $\lambda|\widehat{T}|$. Let $X$ and $Y$ be i.i.d. random variables having the same distribution as $\text{sign}(i\widehat{T}(W))W + \frac{\sigma}{2}Z$, where $Z$ is a standard centered Gaussian random vector. Then, the density function $g$ of $X$ and $Y$ is a Schwartz test function:

$$g(x) = \int_{\mathbb{R}^n} \frac{(\sqrt{2})^n}{(\sqrt{\pi}\sigma)^n} e^{-\frac{2}{\sigma^2}\|x - \text{sign}(i\widehat{T}(t))t\|^2} \lambda|\widehat{T}(t)|dt.$$

Furthermore,

$$\widehat{g}(\xi) = \int_{\mathbb{R}^n} e^{-\frac{\sigma^2}{8}\|\xi\|^2} e^{-i\langle \text{sign}(i\widehat{T}(t))t, \xi\rangle} \lambda|\widehat{T}(t)|dt,$$

which implies that

$$\widehat{g}(\xi) - \widehat{g}(-\xi) = 2\lambda e^{-\frac{\sigma^2}{8}\|\xi\|^2} \int_{\mathbb{R}^n} \widehat{T}(t) \sin\langle t, \xi\rangle\, dt. \tag{5.18}$$

Because both $T$ and $i\widehat{T}$ are absolutely integrable in $\mathbb{R}^n$, by using Fourier inversion theorem, we have

$$\int_{\mathbb{R}^n} \widehat{T}(t) \sin\langle t, \xi\rangle\, dt = (2\pi)^n T(-\xi) = -(2\pi)^n T(\xi).$$

Plugging into (5.18), we obtain

$$|\widehat{g}(\xi) - \widehat{g}(-\xi)|^2 = 4(2\pi)^{2n}\lambda^2 T^2(\xi) e^{-\frac{\sigma^2}{4}\|\xi\|^2}.$$

Thus, by using (5.12) in the second equality below, we have

$$\begin{aligned}\int_{\mathbb{R}^n} \widehat{f}(\xi) T^2(\xi) e^{-\frac{\sigma^2}{4}\|\xi\|^2} d\xi =& \frac{1}{4(2\pi)^{2n}\lambda^2} \int_{\mathbb{R}^n} \widehat{f}(\xi)|\widehat{g}(\xi) - \widehat{g}(-\xi)|^2 d\xi \\ =& \frac{1}{2\lambda^2(2\pi)^n}\left[\mathbb{E}f(X - Y) - \mathbb{E}f(X + Y)\right] \\ \leq& 0.\end{aligned}$$

This finishes the proof of (5.16), and therefore the proof Theorem 5.1 as well.

*Remark 5.6* If $f$ is not polynomially bounded, (5.17) fails, and consequently, $\lambda|\widehat{T}(\xi)|$ may no longer be a probability density for any $\lambda$.

*Proof of Theorem 5.5* By applying Theorem 5.4, we only need to show that for every sequence of real numbers $x_1, x_2, \ldots, x_m$, the matrix $(e^{|x_i + x_j|} - e^{|x_i - x_j|})$

is positive definite. By Sylvester's criterion, we only need to show that all its leading principal minors are non-negative, or equivalently, for *every* sequence of real numbers $z_1, z_2, \ldots, z_k$, the determinant $\det(e^{|z_i+z_j|} - e^{|z_i-z_j|})$ is non-negative, which can be directly verified as

$$\det((e^{|z_i+z_j|} - e^{|z_i-z_j|})_{1\le i,j\le m}) = \prod_{i=1}^{m} (\mathrm{sign}(z_i))^2 \left| e^{2|z_i|} - e^{2|z_{i-1}|} \right|,$$

where we set $z_0 = 0$. Alternatively, we notice that $(e^{|x_i+x_j|} - e^{|x_i-x_j|})_{1\le i,j\le m}$ is the Gram matrix of the system of functions $g_i(\cdot)$, $1 \le i \le m$, on $L^2(\mathbb{R})$, where

$$g_i(t) = 2\,\mathrm{sgn}(x_i) e^{|x_i|-t} 1_{[0,|x_i|]}(t).$$

Hence, the matrix $(e^{|x_i+x_j|} - e^{|x_i-x_j|})$ is positive definite.

The function $f(x) = e^{|x|}$ is not negative definite because any continuous negative definite function is necessarily polynomially bounded. We can also see that from the definition of negative definiteness. Indeed, the matrix $(f(x_i) + f(x_j) - f(x_i - x_j))_{1\le i,j\le n}$ in the definition (5.1) is not positive definite. For example, by choosing $x_1 = -\ln 4$ and $x_2 = \ln 4$, the corresponding matrix

$$\begin{pmatrix} f(x_1) + f(x_1) - f(x_1 - x_1) & f(x_1) + f(x_2) - f(x_1 - x_2) \\ f(x_2) + f(x_1) - f(x_2 - x_1) & f(x_2) + f(x_2) - f(x_2 - x_2) \end{pmatrix} = \begin{pmatrix} 7 & -8 \\ -8 & 7 \end{pmatrix}$$

is clearly not positive definite.

**Acknowledgements** This research was partially supported by a grant from the Simons Foundation, #246211, and an NIH grant P20GM104420.

## References

1. A. Buja, B.F. Logan, J.A. Reeds, L.A. Shepp, Inequalities and positive-definite functions arising from a problem in multidimensional scaling. Ann. Stat. **22**(1), 406–438 (1994)
2. I.M. Gel'fand, N.Ya. Vilenkin, *Generalized Functions*. Applications of Harmonic Analysis, vol. 4 Translated by Amiel Feinstein (Academic, New York, 1964)
3. N. Jacob, *Pseudo Differential Operators and Markov Processes*. Fourier Analysis and Semigroups, vol. I (Imperial College Press, London, 2001)
4. J. Li, M. Madiman, A combinatorial approach to small ball inequalities for sums and differences. Comb. Probab. Comput. **28**(1), 100–129 (2019). https://arxiv.org/abs/1601.03927
5. M. Lifshits, R. Schilling, I. Tyurin, *A Probabilistic Inequality Related to Negative Definite Functions*. Progress in Probability, vol. 66 (Springer, Basel, 2013), pp. 73–80
6. W. Rudin, *Functional Analysis*. International Series in Pure and Applied Mathematics, 2nd edn. (McGraw-Hill, New York, 1991)
7. K. Sato, *Lévy Processes and Infinitely Divisible Distributions*. Cambridge Studies in Advanced Mathematics, vol. 68 (Cambridge University Press, Cambridge, 1999)

8. I.J. Schoenberg, Metric spaces and positive definite functions. Trans. Am. Math. Soc. **44**(3), 522–536 (1938)
9. R.S. Strichartz, *A Guide to Distribution Theory and Fourier Transforms*. Reprint of the 1994 original (World Scientific, River Edge, 2003)

# Chapter 6
# Higher Order Concentration in Presence of Poincaré-Type Inequalities

**Friedrich Götze and Holger Sambale**

**Abstract** We show sharpened forms of the concentration of measure phenomenon typically centered at stochastic expansions of order $d-1$ for any $d \in \mathbb{N}$. Here we focus on differentiable functions on the Euclidean space in presence of a Poincaré-type inequality. The bounds are based on $d$-th order derivatives.

**Keywords** Concentration of measure phenomenon · Poincaré inequalities

**1991 Mathematics Subject Classification** Primary 60E15, 60F10; Secondary 60B20

## 6.1 Introduction

In this note, we study higher order versions of the concentration of measure phenomenon. Instead of the classical problem of deviations of $f$ around the mean $\mathbb{E}f$, we study potentially smaller fluctuations of $\tilde{f}_d := f - \mathbb{E}f - f_1 - \ldots - f_d$, where $f_1, \ldots, f_d$ are "lower order terms" of $f$ with respect to a suitable decomposition, such as a Taylor-type decomposition of $f$. In order to study the concentration of $\tilde{f}_d$ around 0, which we call higher order concentration of measure, we use derivatives up to order $d$.

Previous work includes Adamczak and Wolff [2], who exploited certain Sobolev-type inequalities or subGaussian tail conditions to derive exponential tail inequalities for functions with bounded higher-order derivatives (evaluated in terms of some tensor-product matrix norms). This approach was continued by Adamczak, Bednorz and Wolff for measures satisfying modified logarithmic Sobolev inequalities in [3]. While in [2], concentration around the mean is studied, the idea of sharpening concentration inequalities for Gaussian and related measures by requiring orthogonality

F. Götze (✉) · H. Sambale
Faculty of Mathematics, Bielefeld University, Bielefeld, Germany
e-mail: goetze@math.uni-bielefeld.de; hsambale@math.uni-bielefeld.de



N. Gozlan et al. (eds.), *High Dimensional Probability VIII*,
Progress in Probability 74, https://doi.org/10.1007/978-3-030-26391-1_6

to linear functions also appears in Wolff [16] as well as in Cordero-Erausquin et al. [9]. For a detailed overview of the concentration of measure phenomenon, see [8, 14].

Our research started with second order results for functions on the $n$-sphere orthogonal to linear functions [6], with an approach which has been extended in [10] for measures satisfying logarithmic Sobolev inequalities. This includes discrete models as well as differentiable functions on open subsets of $\mathbb{R}^n$. These results were extended to arbitrary higher orders in [7].

While in [7], measures satisfying a logarithmic Sobolev inequality were considered, the aim of this note is to prove similar results for measures satisfying a Poincaré-type inequality, i.e. a weaker assumption. To this end, let us recall that a Borel probability measure $\mu$ on an open set $G \subset \mathbb{R}^n$ is said to satisfy a *Poincaré-type inequality* with constant $\sigma^2 > 0$ if for any bounded smooth function $f$ on $G$ with gradient $\nabla f$,

$$\mathrm{Var}_\mu(f) \le \sigma^2 \int |\nabla f|^2 \, d\mu. \tag{6.1}$$

Here, $\mathrm{Var}_\mu(f) = \int f^2 \, d\mu - (\int f \, d\mu)^2$ denotes the variance. When considering $\sigma$ instead of $\sigma^2$ itself, we will always assume it to be positive.

Given a function $f \in \mathcal{C}^d(G)$, we define $f^{(d)}$ to be the (hyper-) matrix whose entries

$$f^{(d)}_{i_1 \dots i_d}(x) = \partial_{i_1 \dots i_d} f(x), \qquad d = 1, 2, \dots \tag{6.2}$$

represent the $d$-fold (continuous) partial derivatives of $f$ at $x \in G$. By considering $f^{(d)}(x)$ as a symmetric multilinear $d$-form, we define operator-type norms by

$$|f^{(d)}(x)|_{\mathrm{Op}} = \sup \left\{ f^{(d)}(x)[v_1, \dots, v_d] \colon |v_1| = \dots |v_d| = 1 \right\}. \tag{6.3}$$

For instance, $|f^{(1)}(x)|_{\mathrm{Op}}$ is the Euclidean norm of the gradient $\nabla f(x)$, and $|f^{(2)}(x)|_{\mathrm{Op}}$ is the operator norm of the Hessian $f''(x)$. Furthermore, we will use the short-hand notation

$$\|f^{(d)}\|_{\mathrm{Op},p} = \left( \int_G |f^{(d)}|_{\mathrm{Op}}^p \, d\mu \right)^{1/p}, \qquad p \in (0, \infty]. \tag{6.4}$$

For $p = \infty$, the right-hand side has to be read as the $L^\infty$-norm of $|f^{(d)}|_{\mathrm{Op}}$.

We now have the following:

**Theorem 6.1.1** *Let $\mu$ be a probability measure on an open set $G \subset \mathbb{R}^n$ satisfying a Poincaré-type inequality with constant $\sigma^2 > 0$, and let $f \colon G \to \mathbb{R}$ be a $\mathcal{C}^d$-smooth* *function with* $\int_G f\, d\mu = 0$*. Assuming the conditions*

$$\|f^{(k)}\|_{\mathrm{Op},2} \le \sigma^{d-k} \qquad \forall k = 1, \ldots, d-1, \tag{6.5}$$

$$\|f^{(d)}\|_{\mathrm{Op},\infty} \le 1, \tag{6.6}$$

*there exists some universal constant* $c > 0$ *such that*

$$\int_G \exp\Big(\frac{c}{\sigma}\, |f|^{1/d}\Big)\, d\mu \;\le\; 2.$$

Here, a possible choice is $c = 1/(12\mathrm{e})$. Comparing Theorem 6.1.1 to its analogue in presence of a logarithmic Sobolev inequality, i.e. Theorem 1.6 in [7], we see that under the same assumptions (6.5) and (6.6), logarithmic Sobolev inequalities yield exponential moment bounds for $|f|^{2/d}$, whereas Poincaré-type inequalities provide exponential moments for $|f|^{1/d}$ only. This corresponds to the well-known behaviour in case of $d = 1$.

If $f$ has centered partial derivatives of order up to $d-1$, it is possible to replace (6.5) by a somewhat simpler condition. To this end, we need to involve Hilbert–Schmidt-type norms $|f^{(d)}(x)|_{\mathrm{HS}}$ defined as the Euclidean norm of $f^{(d)}(x) \in \mathbb{R}^{n^d}$. Similarly to (6.4), $\|f^{(d)}\|_{\mathrm{HS},2}$ then denotes the $L^2$-norm of $|f^{(d)}|_{\mathrm{HS}}$. In detail:

**Theorem 6.1.2** *Let* $\mu$ *be a probability measure on an open set* $G \subset \mathbb{R}^n$ *satisfying a Poincaré-type inequality with constant* $\sigma^2$*, and let* $f : G \to \mathbb{R}$ *be a* $\mathcal{C}^d$*-smooth function such that*

$$\int_G f\, d\mu = 0 \qquad \text{and} \qquad \int_G \partial_{i_1 \ldots i_k} f\, d\mu = 0$$

*for all* $k = 1, \ldots, d-1$ *and* $1 \le i_1, \ldots, i_k \le n$*. Assuming that*

$$\|f^{(d)}\|_{\mathrm{HS},2} \le 1 \qquad \text{and} \qquad \|f^{(d)}\|_{\mathrm{Op},\infty} \le 1,$$

*there exists some universal constant* $c > 0$ *such that*

$$\int_G \exp\Big(\frac{c}{\sigma}\, |f|^{1/d}\Big)\, d\mu \le 2.$$

Here again, a possible choice is $c = 1/(12\mathrm{e})$.

By Chebyshev's inequality, Theorem 6.1.1 immediately yields

$$\mu(|f| \ge t) \le 2\mathrm{e}^{-ct^{1/d}/\sigma}$$

for any $t \ge 0$. For small values of $t$, it is possible to obtain refined tail estimates in the spirit of Adamczak [1], Theorem 7, or Adamczak and Wolff [2], Theorem 3.3 (with $\gamma = 1$ using their notation), by analyzing the proof of Theorem 6.1.1:

**Corollary 6.1.3** *Let $\mu$ be a probability measure on an open set $G \subset \mathbb{R}^n$ satisfying a Poincaré-type inequality with constant $\sigma^2 > 0$, and let $f: G \to \mathbb{R}$ be a $\mathcal{C}^d$-smooth function with $\int_G f\, d\mu = 0$. For any $t \geq 0$, set*

$$\eta_f(t) := \min\Big(\frac{\sqrt{2}t^{1/d}}{\sigma\|f^{(d)}\|_{\mathrm{Op},\infty}^{1/d}}, \min_{k=1,\ldots,d-1} \frac{\sqrt{2}t^{1/k}}{\sigma\|f^{(k)}\|_{\mathrm{Op},2}^{1/k}}\Big).$$

*Then,*

$$\mu(|f| \geq t) \leq \mathrm{e}^2 \exp(-\eta_f(t)/(d\mathrm{e})).$$

As a generalization of these bounds, we may consider measures satisfying weighted Poincaré-type inequalities. Recall that a Borel probability measure $\mu$ on an open set $G \subset \mathbb{R}^n$ is said to satisfy a *weighted Poincaré-type inequality* if for any bounded smooth function $f$ on $G$ with gradient $\nabla f$,

$$\mathrm{Var}_\mu(f) \leq \int |\nabla f|^2 w^2\, d\mu, \tag{6.7}$$

where $w: G \to [0, \infty)$ is some measurable function. Examples include Cauchy measures and Beta distributions. For a detailed discussion see Bobkov and Ledoux [5].

In these cases we cannot expect exponential integrability as in Theorem 6.1.1 any more, since distributions satisfying (6.7) may have a slow, say, polynomial, decay at infinity. Nevertheless, it is still possible to obtain higher order concentration results by controlling the $L^p$-norms of $f$ and its derivatives. In detail:

**Proposition 6.1.4** *Let $\mu$ be a probability measure on an open set $G \subset \mathbb{R}^n$ satisfying a weighted Poincaré-type inequality (6.7), and let $f: G \to \mathbb{R}$ be a $\mathcal{C}^d$-smooth function with $\int_G f\, d\mu = 0$. Then, for any $p \geq 2$,*

$$\begin{aligned} \|f\|_p &\leq \sum_{k=1}^{d-1} (2^{\frac{k-2}{2}} p \|w\|_{2^k p})^k \, \|f^{(k)}\|_{\mathrm{Op},2} + (2^{\frac{d-2}{2}} p)^d \, \|w\|_{2^{d-1}p}^{d-1} \, \|w|f^{(d)}|_{\mathrm{Op}}\|_{2^{d-1}p} \\ &\leq \sum_{k=1}^{d-1} (2^{\frac{k-2}{2}} p \|w\|_{2^k p})^k \, \|f^{(k)}\|_{\mathrm{Op},2} + (2^{\frac{d-2}{2}} p \|w\|_{2^d p})^d \, \|f^{(d)}\|_{\mathrm{Op},2^d p}. \end{aligned}$$

Proposition 6.1.4 should be compared to (6.15) from the proof of Theorem 6.1.1 in Sect. 6.2. In particular, if the weight function $w$ is bounded by some real number $\sigma > 0$, $\mu$ clearly satisfies a Poincaré-type inequality (6.1) with constant $\sigma^2$. In this case, Proposition 6.1.4 implies a slightly weaker version of (6.15), and it is possible to derive Theorem 6.1.1 again though with a somewhat weaker constant $c = c_d > 0$.

Suitable conditions on the weight function $w$ may still yield exponential-type tails at least in certain intervals. For instance, the following higher order analogue of Corollary 4.2 in [5] holds:

**Corollary 6.1.5** *Let $\mu$ be a probability measure on an open set $G \subset \mathbb{R}^n$ satisfying a weighted Poincaré-type inequality (6.7), and let $f\colon G \to \mathbb{R}$ be a $\mathcal{C}^d$-smooth function with $\int_G f d\mu = 0$ and such that (6.5) (with $\sigma^2 = 1$) and (6.6) from Theorem 6.1.1 hold. Assume $\|w\|_{2^d p} \le C$ for some $p \ge 2$ and some $C \ge 2^{-(d-1)/2}$. Then, for any $0 \le t \le (2^{\frac{d+5}{2}} C\mathrm{e}\, p)^d$,*

$$\mu(|f| \ge t) \le \mathrm{e}^{d/\mathrm{e}} \exp(-dt^{1/d}/(2^{\frac{d+5}{2}} C\mathrm{e})).$$

Hence, we obtain exponential-type tail bounds on an interval of length proportional to $p^d$. Note that if $t > (2^{\frac{d+5}{2}} C\mathrm{e}\, p)^d$, we may still give bounds on $\mu(|f| \ge t)$ by taking (6.23) for $q = p$ from the proof of Corollary 6.1.5. We omit details at this point. The assumption $C \ge 2^{-(d-1)/2}$ is needed for technical reasons. In fact, it guarantees that the quantities $(2^{\frac{k-1}{2}} C)^k$, $k \le d-1$, are bounded by $(2^{\frac{d-1}{2}} C)^d$. For $d = 1$ it can be removed. It is possible to adapt the proof for $0 < C < 2^{-(d-1)/2}$ and obtain similar bounds.

For $d = 1$, Corollary 6.1.5 gives back a version of Corollary 4.2 from [5] up to constants, though with a boundedness condition on $\|w\|_{2p}$ rather than $\|w\|_p$. This may be adjusted by working with the first inequality from Proposition 6.1.4, in which case we directly get back the [5] result. In the same way, it is possible to derive a result similar to Corollary 6.1.5 which requires bounds on $\|w\|_{2^{d-1}p}$. We have chosen to work with the second inequality from Proposition 6.1.4 instead (and thus need bounds on $\|w\|_{2^d p}$) since this is technically slightly more convenient.

Under stronger moment conditions on the weight function $w$, e. g. $\int \mathrm{e}^{w^2/\alpha} d\mu \le 2$ for some $\alpha > 0$, it is possible to obtain exponential-type tail bounds even on the whole positive half-line, cf. Corollary 4.3 in [5].

**Outline** In Sect. 6.2, we give the proofs of the results stated above. In Sect. 6.3, we provide some applications, including homogeneous multilinear polynomials of order $d$ and linear eigenvalue statistics in random matrix theory.

## 6.2 Proofs

Given a continuous function on an open subset $G \subset \mathbb{R}^n$, the equality

$$|\nabla f(x)| = \limsup_{x \to y} \frac{|f(x) - f(y)|}{|x - y|}, \qquad x \in G, \tag{6.8}$$

may be used as definition of the generalized modulus of the gradient of $f$. The function $|\nabla f|$ is Borel measurable, and if $f$ is differentiable at $x$, the generalized modulus of the gradient agrees with the Euclidean norm of the usual gradient. This operator preserves many identities from calculus in form of inequalities, such as a "chain rule inequality"

$$|\nabla T(f)| \le |T'(f)||\nabla f|, \tag{6.9}$$

where $|T'|$ is understood according to (6.8) again.

As shown in [7], Lemma 4.1, using the generalized modulus of the gradient, the operator norms of the derivatives of consecutive orders are related as follows:

**Lemma 6.2.1** *Given a $\mathcal{C}^d$-smooth function $f\colon G \to \mathbb{R}$, $d \in \mathbb{N}$, at all points $x \in G$,*

$$|\nabla |f^{(d-1)}(x)|_{\mathrm{Op}}| \le |f^{(d)}(x)|_{\mathrm{Op}}.$$

*Proof* Indeed, for any $h \in \mathbb{R}^n$, by the triangle inequality,

$$\big|\, |f^{(d-1)}(x+h)|_{\mathrm{Op}} - |f^{(d-1)}(x)|_{\mathrm{Op}} \big| \le |f^{(d-1)}(x+h) - f^{(d-1)}(x)|_{\mathrm{Op}}$$
$$= \sup\{(f^{(d-1)}(x+h) - f^{(d-1)}(x))[v_1, \ldots, v_{d-1}]\colon v_1, \ldots, v_{d-1} \in S^{n-1}\},$$

while, by the Taylor expansion,

$$(f^{(d-1)}(x+h) - f^{(d-1)}(x))[v_1, \ldots, v_{d-1}] = f^{(d)}(x)[v_1, \ldots, v_{d-1}, h] + o(|h|)$$

as $h \to 0$. Here, the $o$-term can be bounded by a quantity which is independent of $v_1, \ldots, v_{d-1} \in S^{n-1}$. As a consequence,

$$\limsup_{h\to 0} \frac{|\,|f^{(d-1)}(x+h)|_{\mathrm{Op}} - |f^{(d-1)}(x)|_{\mathrm{Op}}|}{|h|}$$
$$\le \sup\{f^{(d)}(x)[v_1, \ldots, v_{d-1}, v_d]\colon v_1, \ldots, v_d \in S^{n-1}\} = |f^{(d)}(x)|_{\mathrm{Op}}.$$

□

Following the scheme of proof developed in [7], we moreover need to establish a recursion for the $L^p$-norms of the derivatives of $f$ of consecutive orders. To this end, we recall a classical result on the moments of Lipschitz functions in the presence of Poincaré-type inequalities. Here, similarly to (6.4), we write

$$\|\nabla g\|_{\mathrm{Op},p} = \left(\int_G |\nabla g|^p \, d\mu\right)^{1/p}, \qquad p \in (0, \infty],$$

for any locally Lipschitz function $g$ on $G$ with generalized modulus of gradient $|\nabla g|$. In detail:

**Lemma 6.2.2** *Let $\mu$ be a probability measure on an open set $G \subset \mathbb{R}^n$ satisfying a Poincaré-type inequality with constant $\sigma^2 > 0$, and let $g\colon G \to \mathbb{R}$ be locally Lipschitz with $\int_G g\,d\mu = 0$. Then, for any $p \geq 2$,*

$$\int_G |g|^p d\mu \leq \Big(\frac{\sigma p}{\sqrt{2}}\Big)^p \int_G |\nabla g|^p d\mu. \tag{6.10}$$

*In particular, for any $g\colon G \to \mathbb{R}$ locally Lipschitz,*

$$\|g\|_p \;\leq\; \|g\|_2 + \frac{\sigma p}{\sqrt{2}}\,\|\nabla g\|_p. \tag{6.11}$$

Note that in (6.11), $g$ is not required to have mean 0. For the reader's convenience, let us briefly recall the proof.

*Proof* By standard arguments, we may assume $g$ to be $\mathcal{C}^1$-smooth and bounded. Moreover, by the subadditivity property of the variance functional, the Poincaré-type inequality for the probability measure $\mu$ on $G$ is extended to the same relation on $G \times G$, i.e.

$$\mathrm{Var}_{\mu^2}(u) \leq \sigma^2 \iint |\nabla u(x,y)|^2 d\mu(x) d\mu(y) \tag{6.12}$$

for the product measure $\mu^2 = \mu \otimes \mu$. Here, for any $\mathcal{C}^1$-smooth function $u = u(x,y)$, the modulus of the gradient is given by

$$|\nabla u(x,y)|^2 = |\nabla_x u(x,y)|^2 + |\nabla_y u(x,y)|^2.$$

Now consider the function

$$u(x,y) = |g(x) - g(y)|^{\frac{p}{2}} \,\mathrm{sign}(g(x) - g(y)),$$

which is $\mathcal{C}^1$-smooth for $p > 2$ with modulus of gradient

$$|\nabla u(x,y)| = \frac{p}{2} |g(x) - g(y)|^{\frac{p}{2}-1} \sqrt{|\nabla g(x)|^2 + |\nabla g(y)|^2}.$$

Since $u$ has a symmetric distribution under $\mu^2$, applying (6.12) together with Hölder's inequality yields

$$\begin{aligned}
&\frac{1}{\sigma^2} \iint |g(x) - g(y)|^p d\mu^2(x,y) \\
&\leq \frac{p^2}{4} \iint |g(x) - g(y)|^{p-2} \big(|\nabla g(x)|^2 + |\nabla g(y)|^2\big) d\mu^2(x,y) \\
&\leq \frac{p^2}{4} \Big(\iint |g(x) - g(y)|^p d\mu^2(x,y)\Big)^{\frac{p-2}{p}} \Big(\iint \big(|\nabla g(x)|^2 + |\nabla g(y)|^2\big)^{\frac{p}{2}} d\mu^2(x,y)\Big)^{\frac{2}{p}}.
\end{aligned}$$

By Jensen's inequality, the last integral may be bounded by

$$2^{\frac{p}{2}-1}\iint(|\nabla g(x)|^p+|\nabla g(y)|^p)d\mu^2(x,y)=2^{\frac{p}{2}}\int|\nabla g|^p d\mu.$$

Consequently,

$$\Big(\iint|g(x)-g(y)|^p d\mu^2(x,y)\Big)^{\frac{2}{p}}\le\frac{\sigma^2p^2}{2}\Big(\int|\nabla g|^p d\mu\Big)^{\frac{2}{p}},$$

or, equivalently,

$$\iint|g(x)-g(y)|^p d\mu^2(x,y)\le\Big(\frac{\sigma p}{\sqrt{2}}\Big)^p\int|\nabla g|^p d\mu.$$

In particular, the latter inequality shows that any locally Lipschitz function $g$ such that the right-hand side is finite is integrable (if $g$ is unbounded, we may perform a simple truncation argument). If $\int g d\mu=0$, it follows from Jensen's inequality that the left integral can be bounded below by $\int|g|^p d\mu$, which proves (6.10). To see (6.11), it remains to note that by the triangle inequality,

$$\Big\|g-\int g d\mu\Big\|_p\ge\|g\|_p-\Big|\int g d\mu\Big|\ge\|g\|_p-\|g\|_2.$$

□

Combining Lemma 6.2.1 and (6.11), we are able to prove Theorem 6.1.1. Recall that if a relation of the form

$$\|f\|_k\le\gamma k\qquad(k\in\mathbb{N})\tag{6.13}$$

holds true with some constant $\gamma>0$, then $f$ has sub-exponential tails, i.e. $\int e^{c|f|}d\mu\le 2$ for some constant $c=c(\gamma)>0$, e. g. $c=\frac{1}{2\gamma e}$. Indeed, using $k!\ge(\frac{k}{e})^k$, we have

$$\int\exp(c|f|)d\mu=1+\sum_{k=1}^{\infty}c^k\frac{\int|f|^k d\mu}{k!}\le 1+\sum_{k=1}^{\infty}(c\gamma)^k\frac{k^k}{k!}\le 1+\sum_{k=1}^{\infty}(c\gamma e)^k=2.$$

*Proof of Theorem 6.1.1* Using (6.11) with $f$ replaced by $|f^{(k-1)}|_{\mathrm{Op}}$, $2\le k\le d$, we get

$$\begin{aligned}\|f^{(k-1)}\|_{\mathrm{Op},p}&\le\|f^{(k-1)}\|_{\mathrm{Op},2}+\frac{\sigma p}{\sqrt{2}}\,\|\nabla|f^{(k-1)}|_{\mathrm{Op}}\|_p\\&\le\|f^{(k-1)}\|_{\mathrm{Op},2}+\frac{\sigma p}{\sqrt{2}}\,\|f^{(k)}\|_{\mathrm{Op},p},\end{aligned}\tag{6.14}$$

where Lemma 6.2.1 was applied on the last step. Consequently, using (6.10) and then (6.14) iteratively,

$$\|f\|_p \;\le\; \sum_{k=1}^{d-1} \Big(\frac{\sigma p}{\sqrt{2}}\Big)^k \,\|f^{(k)}\|_{\mathrm{Op},2} + \Big(\frac{\sigma p}{\sqrt{2}}\Big)^d \,\|f^{(d)}\|_{\mathrm{Op},p}. \tag{6.15}$$

Since $\|f^{(k)}\|_{\mathrm{Op},2} \le \sigma^{d-k}$ for all $k = 1,\dots,d-1$ and $\|f^{(d)}\|_{\mathrm{Op},\infty} \le 1$ by assumption, we obtain

$$\|f\|_p \;\le\; \sigma^d \sum_{k=1}^{d} (p/\sqrt{2})^k \;\le\; \frac{1}{1-(p/\sqrt{2})^{-1}}\,(\sigma p/\sqrt{2})^d \;\le\; 4\,(\sigma p/\sqrt{2})^d \tag{6.16}$$

and therefore $\|f\|_p \le (3\sigma p)^d$ for all $p \ge 2$. Moreover, $\|f\|_p \le \|f\|_2 \le (6\sigma)^d$ for $p < 2$. It follows that

$$\||f|^{1/d}\|_k = \|f\|_{k/d}^{1/d} \le \gamma k$$

for all $k \in \mathbb{N}$, i.e. (6.13) holds with $\gamma = 6\sigma$ (and $|f|^{1/d}$ in place of $f$). This yields the assertion of the theorem. □

*Proof of Theorem 6.1.2* Starting as in the proof of Theorem 6.1.1, we arrive at

$$\|f\|_p \le \sum_{k=1}^{d-1} (\sigma p/\sqrt{2})^k \,\|f^{(k)}\|_{\mathrm{HS},2} + (\sigma p/\sqrt{2})^d \,\|f^{(d)}\|_{\mathrm{Op},p}, \tag{6.17}$$

where we used that operator norms are dominated by Hilbert–Schmidt norms. Moreover, since $\int_G \partial_{i_1\dots i_k} f\, d\mu = 0$, by the Poincaré-type inequality,

$$\int_G (\partial_{i_1\dots i_k} f)^2\, d\mu \le \sigma^2 \sum_{j=1}^{n} \int_G (\partial_{i_1\dots i_k j} f)^2\, d\mu$$

whenever $1 \le i_1,\dots,i_k \le n$, $k \le d-1$. Summing over all $1 \le i_1,\dots,i_k \le n$, we get

$$\|f^{(k)}\|_{\mathrm{HS},2}^2 = \int_G |f^{(k)}|_{\mathrm{HS}}^2\, d\mu \le \sigma^2 \int_G |f^{(k+1)}|_{\mathrm{HS}}^2\, d\mu = \sigma^2\, \|f^{(k+1)}\|_{\mathrm{HS},2}^2. \tag{6.18}$$

Using (6.18) in (6.17) and iterating, we thus obtain

$$\|f\|_p \leq \sum_{k=1}^{d-1} \sigma^d (p/\sqrt{2})^k \,\|f^{(d)}\|_{\mathrm{HS},2} + (\sigma p/\sqrt{2})^d \,\|f^{(d)}\|_{\mathrm{Op},p}.$$

Noting that $\|f^{(d)}\|_{\mathrm{HS},2} \leq 1$ and $\|f^{(d)}\|_{\mathrm{Op},\infty} \leq 1$, we arrive at (6.16), from where we may proceed as in the proof of Theorem 6.1.1. □

*Proof of Corollary 6.1.3* First note that by Chebyshev's inequality, for any $p \geq 1$

$$\mu(|f| \geq \mathrm{e}\,\|f\|_p) \leq \mathrm{e}^{-p}. \tag{6.19}$$

Moreover, if $p \geq 2$, it follows from (6.15) that

$$\mathrm{e}\,\|f\|_p \leq \mathrm{e}\Big(\sum_{k=1}^{d-1} (\sigma p/\sqrt{2})^k \,\|f^{(k)}\|_{\mathrm{Op},2} + (\sigma p/\sqrt{2})^d \,\|f^{(d)}\|_{\mathrm{Op},\infty}\Big).$$

Assuming $\eta_f(t) \geq 2$, we therefore arrive at

$$\mathrm{e}\,\|f\|_{\eta_f(t)} \leq \mathrm{e}\big(\sum_{k=1}^{d-1} t + t\big) = (d\mathrm{e})t.$$

Hence, applying (6.19) to $p = \eta_f(t)$ (if $p \geq 2$) yields

$$\mu(|f| \geq (d\mathrm{e})t) \leq \mu(|f| \geq \mathrm{e}\,\|f\|_{\eta_f(t)}) \leq \exp(-\eta_f(t)).$$

Using a trivial estimate provided that $p = \eta_f(t) < 2$, we obtain

$$\mu(|f| \geq (d\mathrm{e})t) \leq \mathrm{e}^2 \exp(-\eta_f(t))$$

for all $t \geq 0$. The proof now easily follows by rescaling $f$ by $d\mathrm{e}$ and using that $\eta_{d\mathrm{e}\,f}(t) \geq \eta_f(t)/(d\mathrm{e})$. □

In order to prove Proposition 6.1.4, we have to adapt the first steps of the proof of Theorem 6.1.1. First, we have the following generalization of Lemma 6.2.2 (in fact, this is a version of Theorem 4.1 in [5]):

**Lemma 6.2.3** *Let $\mu$ be a probability measure on an open set $G \subset \mathbb{R}^n$ satisfying a weighted Poincaré-type inequality (6.7), and let $g\colon G \to \mathbb{R}$ be locally Lipschitz with $\int_G g\,d\mu = 0$. Then, for any $p \geq 2$,*

$$\int_G |g|^p \,d\mu \leq \Big(\frac{p}{\sqrt{2}}\Big)^p \int_G |\nabla g|^p w^p \,d\mu. \tag{6.20}$$

*In particular, for any* $g\colon G \to \mathbb{R}$ *locally Lipschitz,*

$$\|g\|_p \;\le\; \|g\|_2 + \frac{p}{\sqrt{2}}\,\|w|\nabla g|\|_p. \tag{6.21}$$

The proof of Lemma 6.2.3 uses similar arguments as the proof of Lemma 6.2.2, and we therefore omit it. In particular, by Hölder's inequality, (6.21) implies

$$\|g\|_p \;\le\; \|g\|_2 + \frac{p}{\sqrt{2}}\,\|w\|_{2p}\|\nabla g\|_{2p}. \tag{6.22}$$

Starting with (6.20)–(6.22) and iterating as in (6.14) and (6.15), we obtain

$$\|f\|_p \;\le\; \sum_{k=1}^{d-1} 2^{\binom{k}{2}}\Big(\frac{p\|w\|_{2^k p}}{\sqrt{2}}\Big)^k\,\|f^{(k)}\|_{\mathrm{Op},2} + 2^{\binom{d}{2}}\Big(\frac{p\|w\|_{2^{d-1}p}}{\sqrt{2}}\Big)^d\,\|w|f^{(d)}|_{\mathrm{Op}}\|_{2^{d-1}p},$$

hence we easily arrive at the conclusions of Proposition 6.1.4. Again, we omit the details.

Finally, the proof of Corollary 6.1.5 is similar to the proof of Corollary 4.2 in [5].

*Proof of Corollary 6.1.5* First let $2 \le q \le p$. Using the assumptions and Proposition 6.1.4, we arrive at

$$\|f\|_q \;\le\; \sum_{k=1}^{d-1}(2^{\frac{k-2}{2}}qC)^k + (2^{\frac{d-2}{2}}qC)^d$$

and hence

$$\|f\|_q \le 4\,(2^{\frac{d-1}{2}}Cq)^d \le (2^{\frac{d+3}{2}}Cq)^d$$

(this follows as in (6.16), substituting $\sigma$ by $2^{\frac{d-1}{2}}C \ge 1$). Moreover, if $0 < q \le 2$, we have

$$\|f\|_q \le \|f\|_2 \le (2^{\frac{d+5}{2}}C)^d.$$

Since the function $q \mapsto \mathrm{e}^{d/\mathrm{e}}q^{dq}$, $q > 0$, is minimized at $q = 1/\mathrm{e}$ with minimum value 1, it follows that $\mathbb{E}|f|^q \le \mathrm{e}^{d/\mathrm{e}}\,(2^{\frac{d+5}{2}}Cq)^{dq}$ for all $0 < q \le p$. Therefore, for any $t > 0$ and any $0 < q \le p$,

$$\mu(|f| \ge t) \le \frac{\mathbb{E}|f|^q}{t^q} \le \mathrm{e}^{d/\mathrm{e}}\left(\frac{(2^{\frac{d+5}{2}}Cq)^d}{t}\right)^q. \tag{6.23}$$

Now set $s = t^{1/d}/(2^{\frac{d+5}{2}}C)$ and write $\mu(|f| \ge t) \le \mathrm{e}^{d/\mathrm{e}}\mathrm{e}^{-\varphi(q)}$ with $\varphi(q) = dq(\log(s) - \log(q))$. It is easy to check that $\varphi$ is a concave function on $(0,\infty)$

which attains its maximum at $q_0 = s/\mathrm{e}$ with $\varphi(q_0) = ds/\mathrm{e} = dt^{1/d}/(2^{\frac{d+5}{2}} C\mathrm{e})$. Noting that $q_0 \le p$ is equivalent with $t \le (2^{\frac{d+5}{2}} C\mathrm{e}\, p)^d$ completes the proof. □

## 6.3 Applications

Let $X_1, \ldots, X_n$ be independent random variables with distributions satisfying a Poincaré-type inequality (6.1) with common constant $\sigma^2 > 0$. For real numbers $a_{i_1 \ldots i_d}$, $i_1 < \ldots < i_d$, consider the function

$$f(X_1, \ldots, X_n) := \sum_{i_1 < \ldots < i_d} a_{i_1 \ldots i_d} X_{i_1} \cdots X_{i_d}, \tag{6.24}$$

which is a homogeneous multilinear polynomial of order $d$. For any $i_1 < \ldots < i_d$ and any permutation $\sigma \in S^d$, set $a_{\sigma(i_1) \ldots \sigma(i_d)} \equiv a_{i_1 \ldots i_d}$. Moreover, set $a_{i_1 \ldots i_d} = 0$ whenever the indexes $i_1, \ldots, i_d$ are not pairwise different. This gives rise to a hypermatrix $A = (a_{i_1 \ldots i_d}) \in \mathbb{R}^{n^d}$, whose Euclidean norm we denote by $\|A\|_{\mathrm{HS}}$. Moreover, set $\|A\|_\infty := \max_{i_1 < \ldots < i_d} |a_{i_1 \ldots i_d}|$.

As a first example, we may apply our results to functions of type (6.24). Here it is convenient to assume for the random variables $X_i$ to have mean zero:

**Proposition 6.3.1** *Let $X_1, \ldots, X_n$ be independent random variables with distributions satisfying a Poincaré-type inequality (6.1) with common constant $\sigma^2 > 0$. Assume $\mathbb{E}X_i = 0$ for all $i = 1, \ldots, n$. Let $d \in \mathbb{N}$, and consider a function $f$ of type (6.24). Then,*

$$\mathbb{E} \exp\Big( \frac{c}{\sigma \|A\|_{\mathrm{HS}}^{1/d}} |f|^{1/d} \Big) \le 2.$$

*Here, $\mathbb{E}$ denotes the expectation with respect to the random variables $X_1, \ldots, X_n$, and $c$ is the absolute constant appearing in Theorem 6.1.2. In particular,*

$$\mathbb{E} \exp\Big( \frac{c}{\sigma n^{1/2} \|A\|_\infty^{1/d}} |f|^{1/d} \Big) \le 2.$$

*Moreover, if $\mathbb{E}X_i^2 = 1$ for all $i = 1, \ldots, n$,*

$$\begin{aligned} \mathbb{P}(|f - \mathbb{E}f| \ge t) &\le \mathrm{e}^2 \exp\Big( -\frac{\sqrt{2}}{\sigma d \mathrm{e}} \min\Big( \frac{t}{\|A\|_{\mathrm{HS}}}, \frac{t^{1/d}}{\|A\|_{\mathrm{HS}}^{1/d}} \Big)\Big) \\ &\le \mathrm{e}^2 \exp\Big( -\frac{\sqrt{2}}{\sigma d \mathrm{e}} \min\Big( \frac{t}{n^{d/2}\|A\|_\infty}, \frac{t^{1/d}}{n^{1/2}\|A\|_\infty^{1/d}} \Big)\Big). \end{aligned}$$

Proposition 6.3.1 follows immediately from Theorem 6.1.2 and Corollary 6.1.3. Note that for non-centered random variables $X_1, \ldots, X_n$, applying Proposition 6.3.1 to the random variables $X_i - \mathbb{E}X_i$ means removing certain "lower order" terms in (6.24), which is in accordance with the ideas sketched in the introduction.

We may furthermore apply our results in the context of random matrix theory. Here we extend an example on second order concentration bounds for linear eigenvalue statistics in presence of a logarithmic Sobolev inequality [10], Proposition 1.10, to the situation where only a Poincaré-type inequality is available.

Indeed, let $\{\xi_{jk}, 1 \le j \le k \le N\}$ be a family of independent random variables on some probability space. Assume that the distributions of the $\xi_{jk}$'s all satisfy a (one-dimensional) Poincaré-type inequality (6.1) with common constant $\sigma^2$. Put $\xi_{jk} = \xi_{kj}$ for $1 \le k < j \le N$ and consider a symmetric $N \times N$ random matrix $\Xi = (\xi_{jk}/\sqrt{N})_{1 \le j,k \le N}$ and denote by $\mu^{(N)}$ the joint distribution of its ordered eigenvalues $\lambda_1 \le \ldots \le \lambda_N$ on $\mathbb{R}^N$ (in fact, $\lambda_1 < \ldots < \lambda_N$ a.s.). Recall that by a simple argument using the Hoffman–Wielandt theorem, $\mu^{(N)}$ satisfies a Poincaré-type inequality with constant

$$\sigma_N^2 = \frac{2\sigma^2}{N} \tag{6.25}$$

(see for instance Bobkov and Götze [4]). Note that similar observations also hold for Hermitian random matrices.

Considering the probability space $(\mathbb{R}^N, \mathbb{B}^N, \mu^{(N)})$, if $f: \mathbb{R} \to \mathbb{R}$ is a $\mathcal{C}^1$-smooth function, it is well-known that asymptotic normality

$$S_N = \sum_{j=1}^{N} (f(\lambda_j) - \mathbb{E}f(\lambda_j)) \Rightarrow \mathcal{N}(0, \sigma_f^2) \tag{6.26}$$

holds for the self-normalized linear eigenvalue statistics $S_N$. Here, "$\Rightarrow$" denotes weak convergence, $\mathbb{E}$ means taking the expectation with respect to $\mu^{(N)}$ and $\mathcal{N}(0, \sigma_f^2)$ denotes a normal distribution with mean zero and variance $\sigma_f^2$ depending on $f$. This result was established by Johansson [12] for the case of $\beta$-ensembles and, for general Wigner matrices, by Khorunzhy et al. [13] as well as Sinai and Soshnikov [15]. Concentration of measure results have been studied by Guionnet and Zeitouni [11], in particular proving fluctuations of order $\mathcal{O}_{\mathbb{P}}(1)$. Our results yield a second order concentration bound:

**Proposition 6.3.2** *Let $\mu^{(N)}$ be the joint distribution of the ordered eigenvalues of $\Xi$. Let $f: \mathbb{R} \to \mathbb{R}$ be a $\mathcal{C}^2$-smooth function with $f'(\lambda_j) \in L^1(\mu^{(N)})$ and bounded second derivatives, and let*

$$\tilde{S}_N := S_N - \sum_{j=1}^{N} (\lambda_j - \mathbb{E}(\lambda_j))\mathbb{E}f'(\lambda_j)$$

*with $S_N$ as in (6.26). Then, we have*

$$\mathbb{E}\exp\Big(\frac{cN^{1/4}}{\sqrt{2}\sigma\|f''\|_\infty^{1/2}}|\tilde{S}_N|^{1/2}\Big) \le 2,$$

*where $c > 0$ is the absolute constant from Theorem 6.1.2.*

Since $\tilde{S}_N$ is "centered" in the sense of Theorem 6.1.2, Proposition 6.3.2 immediately follows from elementary calculus, using (6.25). Note that in view of the self-normalizing property of $S_N$, the fluctuation result for $\tilde{S}_N$ is of the next order, although the scaling is of order $\sqrt{N}$ only. Comparing Proposition 6.3.2 to [10], Proposition 1.10, we see that we essentially arrive at the same result though for $|\tilde{S}_N|^{1/2}$ instead of $|\tilde{S}_N|$ due to the assumption of a Poincaré-type inequality.

Using Corollary 6.1.3, we can in fact slightly sharpen the results on the tail behavior of $S_N$. Indeed, an easy calculation yields

$$\mu_N(|S_N| \ge t) \le \mathrm{e}^2\exp\Big(-\frac{1}{\sigma d\mathrm{e}}\min\Big(\frac{tN^{1/2}}{(\int\sum_i(f'(\lambda_i))^2 d\mu_N)^{1/2}}, \frac{t^{1/2}N^{1/4}}{\|f''\|_\infty^{1/2}}\Big)\Big)$$

for any $t \ge 0$. Similar results may be obtained for higher orders $d \ge 3$.

**Acknowledgement** This research was supported by CRC 1283.

## References

1. R. Adamczak, Moment inequalities for $U$-statistics. Ann. Probab. **34**(6), 2288–2314 (2006)
2. R. Adamczak, P. Wolff, Concentration inequalities for non-Lipschitz functions with bounded derivatives of higher order. Probab. Theory Relat. Fields **162**(3), 531–586 (2015)
3. R. Adamczak, W. Bednorz, P. Wolff, Moment estimates implied by modified log-Sobolev inequalities. ESAIM Probab. Stat. **21**, 467–494 (2017)
4. S.G. Bobkov, F. Götze, Concentration of empirical distribution functions with applications to non-i.i.d. models. Bernoulli **16**(4), 1385–1414 (2010)
5. S.G. Bobkov, M. Ledoux, Weighted Poincaré-type inequalities for Cauchy and other convex measures. Ann. Probab. **37**(2), 403–427 (2009)
6. S.G. Bobkov, G.P. Chistyakov, F. Götze, Second order concentration on the sphere. Commun. Contemp. Math. **19**(5), 1650058, 20 pp. (2017)
7. S.G. Bobkov, F. Götze, H. Sambale, Higher order concentration of measure. Commun. Contemp. Math. **21**(3), 1850043, 36 (2019)
8. S. Boucheron, G. Lugosi, P. Massart, *Concentration Inequalities. A Nonasymptotic Theory of Independence* (Oxford University Press, Oxford, 2013)
9. D. Cordero-Erausquin, M. Fradelizi, B. Maurey, The (B) conjecture for the Gaussian measure of dilates of symmetric convex sets and related problems. J. Funct. Anal. **214**(2), 410–427 (2004)
10. F. Götze, H. Sambale, Second order concentration via logarithmic Sobolev inequalities. Bernoulli (to appear). arXiv:1605.08635 (preprint)
11. A. Guionnet, O. Zeitouni, Concentration of the spectral measure for large matrices. Electron. Commun. Prob. **5**, 119–136 (2000)

12. K. Johansson, On fluctuations of eigenvalues of random Hermitian matrices. Duke Math. J. **91**(1), 151–204 (1998)
13. A.M. Khorunzhy, B.A. Khoruzhenko, L.A. Pastur, Asymptotic properties of large random matrices with independent entries. J. Math. Phys. **37**, 5033–5060 (1996)
14. M. Ledoux, *The Concentration of Measure Phenomenon* (American Mathematical Society, Providence, 2001)
15. Ya. Sinai, A. Soshnikov, A central limit theorem for traces of large random matrices with independent matrix elements. Bol. Soc. Brasil. Mat. **29**, 1–24 (1998)
16. P. Wolff, On some Gaussian concentration inequality for non-Lipschitz functions. High Dimens. Probab. VI **66**, 103–110 (2013)

# Chapter 7
# Rearrangement and Prékopa–Leindler Type Inequalities

**James Melbourne**

**Abstract** We investigate the interactions of functional rearrangements with Prékopa–Leindler type inequalities. It is shown that certain set theoretic rearrangement inequalities can be lifted to functional analogs, thus demonstrating that several important integral inequalities tighten on functional rearrangement about "isoperimetric" sets with respect to a relevant measure. Applications to the Borell–Brascamp–Lieb, Borell–Ehrhard, and the recent polar Prékopa–Leindler inequalities are demonstrated. It is also proven that an integrated form of the Gaussian log-Sobolev inequality sharpens on rearrangement.

## 7.1 Introduction

The Prékopa–Leindler inequality (PLI) stated below has become a useful tool in the study of log-concave distributions in probability and statistics, particularly in high dimension, and a point of interest and unification between probabilists and convex geometers.

**Theorem 7.1.1 (Prékopa–Leindler)** *For* $f, g : \mathbb{R}^d \to [0, \infty)$ *Borel measurable and* $t \in (0, 1)$, *define*

$$f \Box g(z) := \sup_{(1-t)x+ty=z} f^{1-t}(x) g^t(y)$$

A portion of this work relevant to information theory was announced at 56th Annual Allerton Conference on Communication, Control, and Computing [43].

J. Melbourne (✉)
Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN, USA
e-mail: melbo013@umn.edu



N. Gozlan et al. (eds.), *High Dimensional Probability VIII*,
Progress in Probability 74, https://doi.org/10.1007/978-3-030-26391-1_7

*then*

$$\int_{\mathbb{R}^d} f\Box g(z)dz \geq \left(\int_{\mathbb{R}^d} g(z)dz\right)^{1-t}\left(\int_{\mathbb{R}^d} h(z)dz\right)^{t}.$$

The inequality can be motivated from a convex geometric perspective as a functional generalization of the dimension free statement of the Brunn–Minkowski inequality (BMI), which we recall as the fact that for $A, B$ compact in $\mathbb{R}^d$ and $|\cdot|_d$ the $d$-dimensional Lebesgue volume,

$$|(1-t)A + tB|_d \geq |A|_d^{1-t}|B|_d^t.$$

Indeed, by taking $f = \mathbb{1}_A$ and $g = \mathbb{1}_B$, we have $f\Box g = \mathbb{1}_{(1-t)A+tB}$. PLI implies that integration preserves the inequality and the result follows.

The BMI has an elegant qualitative formulation; the volume of sum-sets decreases on spherical symmetrization. More explicitly, if $A$ and $B$ are compact sets, with $A^*$ and $B^*$ Euclidean balls satisfying $|A^*|_d = |A|_d$, $|B^*|_d = |B|_d$, then

$$|A + B|_d \geq |A^* + B^*|_d. \tag{7.1}$$

Our first main result (Theorem 7.3.1) contains a functional generalization of (7.1). We will show PLI "sharpens" on rearrangement in the sense that

$$\int f\Box g(z)dz \geq \int f^*\Box g^*(z)dz, \tag{7.2}$$

where $*$ denotes a functional rearrangement to be defined below. In fact we will prove that for $\psi$ increasing,

$$\int \psi(f\Box g(z))dz \geq \int \psi(f^*\Box g^*(z))dz. \tag{7.3}$$

Our methods are reasonably general and Theorem 7.4.6 will give a class of set theoretic inequalities that admit functional generalization in the sense of (7.3). As a consequence, we will show that analogs of (7.3) can be given to sharpen not only the PLI, but the Borell–Brascamp–Lieb inequalities [15, 18], the Borell–Ehrhard inequality in the Gaussian setting [16, 24], and a recent Polar Prékopa–Leindler [1].

These results can also be motivated from an information theoretic perspective, where the BMI can be considered a Rényi entropy power inequality. There has been considerable recent work (see [6, 7, 10, 29, 31, 33, 45]) developing Rényi entropy [46] generalizations of the classical entropy power inequality (EPI) of Shannon–Stam [47, 48]. One should compare the sharpening of PLI here to [50], where Madiman and Wang show that while spherically symmetric decreasing rearrangements of random variables preserve their Rényi entropy, they decrease the Rényi entropy of independent sums of random variables. One information theoretic application of the rearrangement result is the reduction of Rényi generalizations of the EPI to the spherically symmetric case, see for example [39] where the Madiman–Wang result is used to sharpen the Rényi EPI put forth in [40]. See [36] to find an extension and application of [50] for the $\infty$-Rényi entropy. It should be mentioned that the connections between BMI and entropy power inequalities are not new. The analogy between the two inequalities was first observed in [21], and a unified proof was given in [23] drawing on the work of [4, 17, 34]. The reader is directed to [35] where a further development of Rényi entropy power inequalities and their connections to convex geometry are given.

In the Gaussian case, the strict convexity of the potential gives a result stronger than PLI, and we are able to adapt the rearrangement ideas to approach the Gaussian log-Sobolev inequality. We show in Theorem 7.6.5 that for the Gaussian measure, the "integrated" log-Sobolev inequality derived in [8] by Bobkov and Ledoux, and understood as reverse hypercontractivity of the Hamilton–Jacobi equations in [12], sharpens on half space rearrangement.

An alternative motivation for this investigation is the Brascamp–Lieb–Barthe inequality's relationship to the Brascamp–Lieb–Luttinger rearrangement inequalities [19]. The Brascamp–Lieb inequality [18] enjoys the Brascamp–Lieb–Luttinger inequality as a rearrangement analog. In [2] Barthe used an optimal transport argument to prove Brascamp–Lieb and simultaneously demonstrated a dual inequality that includes PLI as a special case. It is natural to ask for a rearrangement inequality analog of Barthe's result, to provide a dual to the Brascamp–Lieb–Luttinger rearrangement inequality. This work represents a confirmation of such an inequality in the special case corresponding to PLI.

The paper is organized in the following manner; in Sect. 7.2, we will give definitions and background on a notion of rearrangement. In Sect. 7.3, we give a rearrangement inequality for PLI, before giving a general version in Sect. 7.4. In Sect. 7.5, we give applications of the theorem derived in Sect. 7.4 to special cases. In Sect. 7.6, we give a sharpening of an integrated Gaussian log-Sobolev inequality via half-space rearrangement. Finally, in Sect. 7.7, we discuss connections with the work of Barthe and Brascamp-Lieb-Luttinger closing with an open problem.

## 7.2 Preliminaries

For a set $A$, we will use the notation $\mathbb{1}_A$ to denote the indicator function of $A$, taking the value 1 on $A$, and 0 elsewhere. For $x \in \mathbb{R}^d$, $|x|$ will denote the usual Euclidean norm. We use $\mathbb{Q}_+$ to denote the non-negative rational numbers. We use $\gamma_d$ to denote both the standard Gaussian measure on $\mathbb{R}^d$ and its density function

$$\gamma_d(x) = \frac{e^{-|x|^2/2}}{(2\pi)^{\frac{d}{2}}}.$$

When $d = 1$, and there is no risk of confusion, we will omit the subscript and write $\gamma$. We denote the Gaussian distribution function

$$\Phi(x) = \int_{-\infty}^{x} \gamma(y)dy$$

and its inverse $\Phi^{-1}$.

### 7.2.1 Spherically Symmetric Decreasing Rearrangements

Given a nonempty measurable set $A \subseteq \mathbb{R}^d$, we define its spherically symmetric rearrangement $A^*$ to be the origin centered ball of equal volume,

$$A^* := \left\{ x : |x| < (|A|_d/\omega_d)^{\frac{1}{d}} \right\},$$

where $\omega_d$ is the volume of the $d$-dimensional unit ball, with the understanding that $A^* = \varnothing$ in the case that $|A|_d = 0$ and $A^* = \mathbb{R}^d$ when $|A|_d = \infty$.

We can extend this notion of symmetrization to functions via the layer-cake decomposition of a non-negative function $f$,

$$f(x) = \int_0^{f(x)} 1dt = \int_0^{\infty} \mathbb{1}_{\{y:f(y)>t\}}(x)dt.$$

**Definition 7.2.1** For a measurable non-negative function $f$, define its decreasing symmetric rearrangement $f^*$ by

$$f^*(x) := \int_0^{\infty} \mathbb{1}_{\{y:f(y)>t\}^*}(x)dt. \tag{7.4}$$

Note that *decreasing* is used here in the non-strict sense, synonymous with non-increasing.

**Proposition 7.2.2** *$f^*$ is characterized by the equality*

$$\{f^* > \lambda\} = \{f > \lambda\}^*. \tag{7.5}$$

The proof will be given in greater generality in the following section.

**Corollary 7.2.3** *$f^*$ is lower semi-continuous, spherically symmetric and non-increasing in the sense that $|x| \leq |y|$ implies $f^*(x) \geq f^*(y)$.*

*Proof* $f^*$ has open super level sets by Eq. (7.5) and is thus lower semi-continuous. To prove non-increasingness observe that using the characterization above $f^*(y) > \lambda$ iff $y \in \{f > \lambda\}^*$ which implies by $|x| \leq |y|$ that $x \in \{f > \lambda\}^*$, and thus

$f^*(x) > \lambda$. Applying this to $\lambda_n$ increasing to $f^*(y)$ yields our result. Observe that this implies spherical symmetry by applying the preceding argument in the opposite direction $f(x) = f(y)$ when $|x| = |y|$. □

### *7.2.2 More General Rearrangements*

**Definition 7.2.4** For Polish measure spaces $(M, \mu)$ and $(N, \alpha)$, with Borel $\sigma$-algebra, we will call a set map from the Borel $\sigma$-algebra of $M$ to the Borel $\sigma$-algebra of $N$ a *rearrangement* when it satisfies the following,

1. $*(A)$ is an open set satisfying $\alpha(*(A)) = \mu(A)$
2. $\mu(A) \leq \mu(B)$ implies $*(A) \subseteq *(B)$
3. For a sequence $A_i \subseteq A_{i+1}$, $*(\cup_{i=1}^{\infty} A_i) = \cup_{i=1}^{\infty} *(A_i)$.

Notice that in 3, $\cup_j *(A_j) \subseteq *(\cup_j A_j)$ holds from 2, so the assumption is only $\cup_j *(A_j) \supseteq *(\cup_j A_j)$. For brevity of notation, we write $A^* = *(A)$ and note the following extension to functions.

**Definition 7.2.5** For a rearrangement $*$ and Borel measurable $f : M \to [0, \infty)$ define $f^* : N \to [0, \infty)$,

$$f^*(x) := \int_0^{\infty} \mathbb{1}_{\{f>t\}^*}(x)dt.$$

Rearrangement is in general non-linear, however, we do have linear behavior in the following special case.

**Lemma 7.2.6** *For a simple function $s$, expressed as $s = \sum_{i=1}^{n} a_i \mathbb{1}_{A_i}$ with $a_i > 0$ and $A_i \subsetneq A_{i-1}$,*

$$s^* = \sum_{i=1}^{n} a_i \mathbb{1}_{A_i^*}.$$

*Proof* Let us give more explicit formulas for both quantities.

$$\sum_{i=1}^{n} a_i \mathbb{1}_{A_i^*}(z) = \sum_{i=1}^{m_z} a_i$$

where $m_z = \max\{i : z \in A_i^*\}$, and the formula

$$s^*(z) = \sup\{t : z \in \{s > t\}^*\},$$

which holds not just for simple functions but general $f$. If $z \in A^*_{m_z}$ with $m_z$ maximal, then for $t < \sum_{i=1}^{m_z} a_i$, $A_{m_z} \subseteq \{s > t\}$, which in turn gives $A^*_{m_z} \subseteq \{s > t\}^*$. Thus $z \in \{s > t\}^*$ for all $t < \sum_{i=1}^{m_z} a_i$ and we have

$$s^*(z) = \sup_t \{z \in \{s > t\}^*\} \geq \sum_{i=1}^{m_z} a_i = \sum_{i=1}^{n} a_i \mathbb{1}_{A_i^*}(z).$$

For the reverse inequality, assume $s^*(z) > 0$ (else there is nothing to prove) and take $t$ such that $z \in \{s > t\}^*$. Since $\{s > t\} = A_{k_t}$ where $k_t = \min\{j : \sum_{i=1}^{j} a_i > t\}$, we have $\{s > t\}^* = A^*_{k_t}$. This implies that $\sum_{i=1} a_i \mathbb{1}_{A_i^*}(z) \geq \sum_{i=1}^{k_t} a_i > t$. Taking the supremum in $t$,

$$\sum_{i=1}^{n} a_i \mathbb{1}_{A_i^*}(z) \geq s^*(z).$$

□

**Proposition 7.2.7** *$f^*$ is characterized by the equality*

$$\{f^* > \lambda\} = \{f > \lambda\}^*. \tag{7.6}$$

*In particular $f^*$ is lower semi-continuous, and equi-measureable with $f$ in that $\mu\{f > \lambda\} = \alpha\{f^* > \lambda\}$.*

*Proof* First we prove the equality (7.6). Since $f^*(x) > \lambda$ implies $\int_0^\infty \mathbb{1}_{\{f>t\}^*}(x)\, dt > \lambda$, which in turn, by the monotonicity of $\mathbb{1}_{\{f>t\}^*}$ implies the existence of $t > \lambda$ such that $x \in \{f > t\}^*$. From this it follows that

$$\{f^* > \lambda\} \subseteq \{f > \lambda\}^*.$$

For the converse, first assume that $f = s$ is a simple function, expressed as

$$s = \sum_{i=1}^{n} a_i \mathbb{1}_{A_i}$$

with $a_i > 0$ and $A_i \subsetneq A_{i-1}$. By Lemma 7.2.6

$$s^* = \sum_{i=1}^{n} a_i \mathbb{1}_{A_i^*}.$$

Since $\{s > \lambda\} = A_k$ where $k = \min\{j : \sum_{i=1}^{j} a_i > \lambda\}$, $z \in \{s > \lambda\}^* = A_k^*$ implies $s^*(z) = \sum_{i=1}^{n} a_i \mathbb{1}_{A_i^*}(z) \geq \sum_{i=1}^{k} a_i > \lambda$. Thus $\{s > \lambda\}^* \subseteq \{s^* > \lambda\}$ holds for simple functions. Now take $s_n$ to be a sequence of increasing simple functions approximating $f$ pointwise and uniformly on sets where $f$ is bounded. Then

$$\{f > \lambda\}^* = \left(\bigcup_{n=1}^{\infty}\{s_n > \lambda\}\right)^* = \bigcup_{n=1}^{\infty}\{s_n > \lambda\}^* = \bigcup_{n=1}^{\infty}\{s_n^* > \lambda\}.$$

where the first equality is from the assumption of increasingness of the simple functions, the second is from Definition 7.2.4 item 3, and the third follows from the characterization just proven for simple functions. Since $f_1 \leq f_2$, implies $f_1^* \leq f_2^*$ it follows that $\cup\{s_n^* > \lambda\} \subseteq \{f^* > \lambda\}$, so that $\{f > \lambda\}^* \subseteq \{f^* > \lambda\}$.

If $g$ is another function satisfying $\{g > \lambda\} = \{f > \lambda\}^*$ for all $\lambda$, then

$$g(z) = \int_0^\infty \mathbb{1}_{\{g>\lambda\}} d\lambda = \int_0^\infty \mathbb{1}_{\{f>\lambda\}^*} d\lambda = \int_0^\infty \mathbb{1}_{\{f^*>\lambda\}} d\lambda = f^*(z).$$

The fact that $f$ is lower semi-continuous follows from item (1) of our definition, that $A^*$ is open. Equimeasurability is given by $\alpha\{f^* > \lambda\} = \alpha\{f > \lambda\}^* = \mu\{f > \lambda\}$. □

**Proposition 7.2.8** *For an open convex set* $K \subseteq \mathbb{R}^d$ *with closure containing the origin. The set map* $*_K$ *defined by*

$$A^{*_K} := \left(\frac{|A|_d}{|K|_d}\right)^{\frac{1}{d}} K,$$

*is a rearrangement with* $(M, \mu) = (N, \alpha) = (\mathbb{R}^d, |\cdot|_d)$.

*Proof* It is immediate that $A^{*_K}$ is open and the homogeneity of the Lebesgue measure ensures that $|A^{*_K}|_d = |A|_d$, hence (1) follows. To prove (2), note that for $0 < |A| \leq |B|$, by the definition of $*_K$, $A^{*_K} = tK$ and $B^{*_K} = sK$ for some $0 < t \leq s$. Let $x = tk$ for $k \in K$ and $k_n$ a sequence in $K$ converging to 0. Then

$$x = s\left(\frac{t}{s}\left(k - \left(\frac{s}{t} - 1\right)k_n\right) + \left(1 - \frac{t}{s}\right)k_n\right).$$

By $K$ open, $k - (\frac{s}{t} - 1)k_n$ belongs to $K$ for large $n$, and when this holds, by convexity $(\frac{t}{s}(k - (\frac{s}{t} - 1)k_n) + (1 - \frac{t}{s})k_n) \in K$. It follows that $x \in sK$, as such, $A^{*_K} \subseteq B^{*_K}$. The continuity condition in (3) holds, since both sets are origin symmetric balls of the same volume. □

Observe that the qualitative statement of Brunn–Minkowski (7.1), for Borel $A$, $B$

$$|A + B|_d \geq |A^{*_K} + B^{*_K}|_d, \tag{7.7}$$

is preserved. In the following section, we will extend this qualitative result to the functional setting.

**Proposition 7.2.9** *For a fixed coordinate $i$, the set function $*$ defined on a Polish space $M$ with probability measure $\mu$ and $(N, \alpha) = (\mathbb{R}^d, \gamma_d)$ by*

$$A^* = \{x : x_i < \Phi^{-1}(\mu(A))\}$$

*is a rearrangement.*

*Proof* $A^*$ is open by definition, and $\gamma_d(A^*) = \Phi(\Phi^{-1}(\mu(A))) = \mu(A)$. Conditions (2) and (3) follow from the monotonicity and continuity of $\Phi$. □

We will not pursue examples in discrete spaces here. We direct the interested reader to [37, 38] for recent information theoretic work regarding rearrangement on discrete spaces and [25, 26, 44] for discrete PLI investigations.

## 7.3 Rearrangement and Prékopa–Leindler

We begin with a special case of a more general result to build some intuition for the abstractions to follow. For $f, g : \mathbb{R}^d \to [0, \infty)$ and $t \in [0, 1]$ recall

$$f \Box g(z) = \sup_{(1-t)x+ty=z} f^{1-t}(x) g^t(y). \tag{7.8}$$

**Theorem 7.3.1** *For $f, g : \mathbb{R}^d \to [0, \infty)$ Borel, $t \in (0, 1)$, and $*$ denoting a rearrangement to a fixed open convex set with closure containing the origin,*

$$\int_{\mathbb{R}^d} f \Box g(z) dz \geq \int_{\mathbb{R}^d} f^* \Box g^*(z) dz \geq \left( \int f dz \right)^{1-t} \left( \int g dz \right)^t. \tag{7.9}$$

*What is more, when $\psi$ is a non-negative and non-decreasing function*

$$\int_{\mathbb{R}^d} \psi(f \Box g)(z) dz \geq \int_{\mathbb{R}^d} \psi(f^* \Box g^*)(z) dz. \tag{7.10}$$

The universal measurability of $f \Box g$ will follow from the proof, which gives the universal measurability of $\psi(f \Box g)$ as a consequence.

*Proof* For $\lambda \in (0, \infty)$, define

$$S_0 = S_0(\lambda) = \{s \in \mathbb{Q}_+^2 : s_1^{1-t} s_2^t > \lambda\}. \tag{7.11}$$

Observe,

$$\{f \Box g > \lambda\} = \bigcup_{s \in S_0(\lambda)} (1-t)\{f > s_1\} + t\{g > s_2\}. \tag{7.12}$$

Indeed, it is routine to check that $z \in \cup_{s \in S_0}(1-t)\{f > s_1\} + t\{g > s_2\}$ implies $f\Box g(z) > \lambda$. Conversely, if $f\Box g(z) > \lambda$, then there exists a pair of $x$ and $y$ such that $(1-t)x + ty = z$ and $f^{1-t}(x)g^t(y) > \lambda$. By the continuity of the map $(u, v) \mapsto u^{1-t}v^t$, there exists $(s_1, s_2)$ rational satisfying $s_1 < f(x)$, $s_2 < g(y)$, and $s_1^{1-t}s_2^t > \lambda$, which proves the claim.

Let us remark that the sum of Borel sets is universally measurable,[1] and hence $\{f\Box g > \lambda\}$ is as well. This shows we are well justified in our notation $\int_{\mathbb{R}^d} f\Box g(z)dz$. By Brunn–Minkowski and the characterizing property of rearrangements on super level sets

$$|(1-t)\{f > s_1\} + t\{g > s_2\}| \geq |(1-t)\{f > s_1\}^* + t\{g > s_2\}^*| \tag{7.13}$$

$$= |(1-t)\{f^* > s_1\} + t\{g^* > s_2\}|. \tag{7.14}$$

Now applying (7.12) to $f^*\Box g^*$ and observing that

$$(1-t)\{f^* > s_1\} + t\{g^* > s_2\}$$

is an origin centered ball in $\mathbb{R}^d$ for every $s \in S_0(\lambda)$, we see that

$$|\{f^*\Box g^* > \lambda\}| = \left|\bigcup_{s \in S_0(\lambda)} (1-t)\{f^* > s_1\} + t\{g^* > s_2\}\right|$$

$$= \sup_{s \in S_0} \left|(1-t)\{f^* > s_1\} + t\{g^* > s_2\}\right|.$$

Using (7.13),

$$\left|(1-t)\{f^* > s_1\} + t\{g^* > s_2\}\right| \leq \left|\bigcup_{s \in S_0(\lambda)} (1-t)\{f > s_1\} + t\{g > s_2\}\right|.$$

Thus it follows that

$$|\{f\Box g > \lambda\}| \geq |\{f^*\Box g^* > \lambda\}|. \tag{7.15}$$

Using the layer-cake decomposition of the integral,

$$\int_{\mathbb{R}^d} \psi(f\Box g)(z)dz = \int_0^\infty |\{\psi(f\Box g) > t\}|dt.$$

Notice that by the non-decreasingness, $\psi^{-1}(\lambda, \infty)$ is an interval of the form $[x, \infty)$ or $(x, \infty)$ for a non-negative $x$. From this, we can use (7.15) (and continuity of

[1]This follows from the fact that Borel sets are analytic, see [28], and analytic sets are closed under summation and universally measurable.

measure if the interval is closed) to obtain (7.10). To recover (7.9), note that the first inequality follows from setting $\psi(x) = x$, while the second is the application of PLI to $f^*$ and $g^*$ combined with the equimeasurability of the rearrangements ensuring $\int f^* = \int f$ and $\int g^* = \int g$. □

## 7.4 Functional Lifting of Rearrangements

In this section we show that in a general setting, certain set theoretic rearrangement inequalities can be extended to functional analogs, extending the rearrangement inequality proven for PLI in the previous section to more general operations than □ in (7.8). Let us make precise the set theoretic rearrangement inequality we will generalize.

**Definition 7.4.1** Let $m : M^n \to M$ and $\eta : N^n \to N$ be such that $m(A_1, \ldots, A_n) = \{x = m(a_1, \ldots, a_n) : a_i \in A_i\}$ and $\eta(B_1, \ldots, B_n) = \{y = \eta(b_1, \ldots, b_n) : b_i \in B_i\}$ are universally measurable for $A_i$ and $B_j$ Borel. Suppose further that $\{\eta(A_1^*, \ldots, A_n^*)\}_A$ indexed on $n$-tuples of Borel sets is totally ordered in the sense that for any Borel $A_1, \ldots, A_n$ and $A'_1, \ldots, A'_n$ we have either $\eta(A_1^*, \ldots, A_n^*) \subseteq \eta(A'^*_1, \ldots, A'^*_n)$ or $\eta(A_1^*, \ldots, A_n^*) \supseteq \eta(A'^*_1, \ldots, A'^*_n)$ we say that $*$ satisfies a set theoretic rearrangement inequality when the following holds

$$\mu(m(A_1, \ldots, A_n)) \geq \alpha(\eta(A_1^*, \ldots, A_n^*)).$$

We will focus on two main examples, the rearrangement to convex sets in Euclidean space and rearrangement to half-spaces in Gaussian space.

**Proposition 7.4.2** *When* $(M, m, \mu) = (N, \eta, \alpha) = (\mathbb{R}^d, m_t, dx)$, *and* $t = (t_1, \ldots, t_n) \in \mathbb{R}^n$, *defines a map* $m_t$ *by vector space operations,*

$$x = (x_1, \ldots, x_n) \mapsto \sum_{i=1}^{n} t_i x_i, \tag{7.16}$$

*then the* $*K$ *rearrangement, as in Sect. 7.2 for* $K$ *open, convex, and symmetric, satisfies a set theoretic rearrangement inequality. If the* $t_i$ *are assumed positive,* $*K$ *satisfies a set theoretic rearrangement without symmetry if* 0 *belongs to the closure of* $K$.

*Proof* Take $B_i = \operatorname{sgn}(t_i) A_i$ so that $t_1 A_1 + \cdots + t_n A_n = |t_1| B_1 + \cdots + |t_n| B_n$. Using the symmetry and convexity of $K$ and the definition of our rearrangement as a scaling of $K$, it follows that

$$t_1 A_1^* + \cdots + t_n A_n^* = \left( \sum_{i=1}^{n} |t_i| |A_i|^{\frac{1}{d}} \right) K$$

and hence, the images of $m_t$ are totally ordered. Brunn–Minkowski implies that

$$||t_1|B_1 + \cdots + |t_n|B_n| \geq ||t_1|B_1^* + \cdots + |t_n|B_n^*|.$$

It follows that

$$|t_1 A_1 + \cdots + t_n A_n| \geq |t_1 A_1^* + \cdots + A_n^*|.$$

When $t_i$ are positive, the proof is similar and simpler. □

**Proposition 7.4.3** *When $(M, m, \mu)$ is a centered Gaussian measure on a Banach space $M$ and $m$ defined as $x = (x_1, \ldots, x_n) \mapsto \sum_i t_i x_i$ for $t_i > 0$, $\sum_i t_i = 1$, and $(N, \eta, \alpha)$ with $N = \mathbb{R}^d$, $\eta$ defined by $y \mapsto \sum_i t_i y_i$ and $\alpha = \gamma_d$ the half-space rearrangement from Proposition 7.2.9 yields a set theoretic rearrangement inequality.*

This is the content of the Borell–Ehrhard theorem, which we will discuss in more detail in Sect. 7.5.2. Now let us generalize the geometric mean used in PLI.

**Definition 7.4.4** For $0 < T \leq \infty$, a function $\mathcal{M} : [0, T)^n \to [0, \infty]$ is *continuous coordinate increasing* when

1. $x, y \in \mathbb{R}^n$ satisfying $x_i > y_i$ for all $i$, necessarily satisfy $\mathcal{M}(x) > \mathcal{M}(y)$
2. $\mathcal{M}(x) = 0$ when $\prod_i x_i = 0$
3. $\mathcal{M}(x) = \sup_{y<x} \mathcal{M}(y)$ with the convention that $\sup_{y<x} \mathcal{M}(y) = 0$ when $\{y < x\}$ is empty.

By convention, in the case that $T$ is finite, we extend $\mathcal{M}$ to $[0, T]^n$ by $\mathcal{M}(x) = \sup_{y<x} \mathcal{M}(y)$. It should also be assumed, all $\mathcal{M}$ that follow are defined to be zero on $\{x : \prod_i x_i = 0\}$.

### 7.4.1 Examples

1. For $t = (t_1, \ldots, t_n)$ with $t_i > 0$ and $p \in [-\infty, 0) \cup (0, \infty]$ take for $u \in [0, \infty)^n$

$$\mathcal{M}_p^t(u) = \left(t_1 u_1^p + \cdots + t_n u_n^p\right)^{\frac{1}{p}}. \tag{7.17}$$

   with $M_{-\infty}^t(u) = \min_i u_i$ and $M_\infty^t(u) = \max_i u_i$
2. For $t = (t_1, \cdots, t_n)$ with $t_i > 0$ and $u \in [0, \infty)^n$,

$$\mathcal{M}_0^t(u) = \prod u_i^{t_i}. \tag{7.18}$$

   Note that in the case that $\sum_i t_i = 1$, $\mathcal{M}_0^t$ is the limiting case of the previous example.

3. Define for $t_i > 0$ and $u \in (0,1)^n$,

$$\mathcal{M}_{\Phi}^{t}(u) = \Phi(t_1 \Phi^{-1}(u_1) + \cdots + t_n \Phi^{-1}(u_n))$$

Now let us define the functional operation our set theoretic rearrangement inequalities may be generalized to.

**Definition 7.4.5** For $\mathcal{M}$, a continuous coordinate increasing function, $f = \{f_i\}_{i=1}^n$ with $f_i : M \to [0, T)$ and $m : M^n \to M$ define

$$\Box_{\mathcal{M},m} f(z) \coloneqq \sup_{m(x)=z} \mathcal{M}(f_1(x_1), \ldots, f_n(x_n)).$$

Let us further denote for a rearrangement $*$ satisfying a set theoretic rearrangement inequality, $f_* = \{f_i^*\}_{i=1}^n$, so that

$$\Box_{\mathcal{M},\eta} f_*(w) = \sup_{\eta(y)=w} \mathcal{M}(f_1^*(y_1), \ldots, f_n^*(y_n)).$$

When there is no risk of ambiguity we will suppress the notation for the mapping $m$ and write $\Box_{\mathcal{M}} f$ in place of $\Box_{\mathcal{M},m} f$.

Notice that Theorem 7.3.1 was the case that $m(x,y) = \eta(x,y) = (1-t)x + ty$ and $\mathcal{M}$ is the geometric mean as in (7.18).

**Theorem 7.4.6** *A set theoretic rearrangement inequality,*

$$\mu(m(A_1, \ldots, A_n)) \geq \alpha(\eta(A_1^*, \ldots, A_n^*))$$

*can be extended to functions in the sense that for* $f = \{f_i\}_{i=1}^n$*, with* $f_i$ *Borel measurable from* $M$ *to* $[0,\infty)$*,* $\mathcal{M}$ *a continuous coordinate increasing function, and a non-negative non-decreasing* $\psi$*,*

$$\int \psi(\Box_{\mathcal{M},m} f) d\mu \geq \int \psi(\Box_{\mathcal{M},\eta} f_*) d\alpha.$$

*Proof* For $\lambda > 0$, write

$$S_{\mathcal{M}}(\lambda) = \{q \in \mathbb{Q}_+^n : \mathcal{M}(q) > \lambda\}.$$

We will prove $\mu(\Box_{\mathcal{M}} f > \lambda) \geq \alpha(\Box_{\mathcal{M}} f_* > \lambda)$. First observe that by arguments similar to the proof of Theorem 7.3.1

$$\{\Box_{\mathcal{M}} f > \lambda\} = \bigcup_{q \in S_{\mathcal{M}}(\lambda)} m(\{f_1 > q_1\}, \ldots, \{f_n > q_n\}). \tag{7.19}$$

Indeed, suppose $\Box_{\mathcal{M}} f(z) > \lambda$. This implies the existence of some $x$ such that $m(x) = z$ and $\mathcal{M}(f_1(x_1), \ldots, f_n(x_n)) > \lambda$. By the continuity of $\mathcal{M}$, there exists $q \in S_{\mathcal{M}}(\lambda)$ such that $\mathcal{M}(q_1, \ldots, q_n) > \lambda$ and $f(x_i) > q_i$. The opposite direction is immediate. Observe that by our measurability assumptions on $m$ and (7.19), the superlevel sets of $\Box_{\mathcal{M},m} f$ are universally measurable. Since $\psi$ is necessarily Borel measurable by its monotonicity, its composition with $\Box_{\mathcal{M},m} f$ is indeed universally measurable. Analogously, (note that $f_i^*$ are Borel measurable, by lower semi-continuity),

$$\{\Box_{\mathcal{M}} f_* > \lambda\} = \bigcup_{q \in S_{\mathcal{M}}(\lambda)} \eta(\{f_1^* > q_1\}, \ldots, \{f_n^* > q_n\}). \tag{7.20}$$

This gives

$$\begin{aligned}
\mu\{\Box_{\mathcal{M}} f > \lambda\} &= \mu\left(\bigcup_{q \in S_{\mathcal{M}}(\lambda)} m(\{f_1 > q_1\}, \ldots, \{f_n > q_n\})\right). \\
&\geq \sup_{q \in S_{\mathcal{M}}(\lambda)} \mu(m(\{f_1 > q_1\}, \ldots, \{f_n > q_n\})) \\
&\geq \sup_{q \in S_{\mathcal{M}}(\lambda)} \alpha(\eta(\{f_1 > q_1\}^*, \ldots, \{f_n > q_n\}^*)) \\
&= \alpha\left(\bigcup_{q \in S_{\mathcal{M}}(\lambda)} \eta(\{f_1^* > q_1\}, \ldots, \{f_n^* > q_n\})\right) \\
&= \alpha\{\Box_{\mathcal{M}} f_* > \lambda\}.
\end{aligned}$$

The first inequality is obvious, the second is by the assumed set theoretic rearrangement inequality, and the following equality is by the assumption of total orderedness. The last equality is the from (7.20). □

## 7.5 Applications

### 7.5.1 *Borell–Brascamp–Lieb Type Inequalities*

In the case that $\lambda \in (0, 1)$ and $-\infty \leq p \leq \infty$, we recall from example (1) the following continuous coordinate increasing function,

$$\mathcal{M}(u, v) = \mathcal{M}_p^\lambda(u, v) = \begin{cases} ((1-\lambda)u^p + \lambda v^p)^{\frac{1}{p}} & \text{if } uv \neq 0 \\ 0 & \text{if } uv = 0. \end{cases} \tag{7.21}$$

The Borell–Brascamp–Lieb inequality generalizes the PLI with the understanding that $\mathcal{M}_0^\lambda(u, v) = u^{1-\lambda}v^\lambda$. Note that $\mathcal{M}_\infty^\lambda(u, v) = \max\{u, v\}$ and $\mathcal{M}_{-\infty}^\lambda(u, v) = \min\{u, v\}$ as defined in Eq. (7.17). If we define $f\square_{\mathcal{M}_p^\lambda} g$ using $m(x, y) = (1-\lambda)x + \lambda y$ as in Definition 7.4.5, we can state the inequality as the following.

**Theorem 7.5.1 (Borell–Brascamp–Lieb [15, 18])** *For $\lambda \in (0, 1)$ and Borel functions $f, g : \mathbb{R}^n \to [0, \infty)$,*

$$\int f\square_{\mathcal{M}_p^\lambda} g(x)\, dx \geq \mathcal{M}_{p/(np+1)}^\lambda \left( \int f(x)dx, \int g(x)dx \right)$$

*when $p \geq -1/n$.*

We present the following sharpening.

**Theorem 7.5.2** *For Borel functions $f, g : \mathbb{R}^n \to [0, \infty)$ and $*$ a rearrangement to a convex set,*

$$\begin{aligned}\int f\square_{\mathcal{M}_p^\lambda} g(x)\, dx &\geq \int f^*\square_{\mathcal{M}_p^\lambda} g^*(x)\, dx \\ &\geq \mathcal{M}_{p/(np+1)}^\lambda \left( \int f(x)dx, \int g(x)dx \right)\end{aligned}$$

*when $p \geq -1/n$.*

*Proof* As described in Proposition 7.4.2, the Brunn–Minkowski inequality shows that the usual Lebesgue measure with the map $(x, y) \mapsto (1-\lambda)x + ty)$ satisfy a set theoretic rearrangement inequality. The result then follows from Theorem 7.4.6. □

## 7.5.2 The Gaussian Case

For simplicity we restrict ourselves to the $\mathbb{R}^d$ case and employ the rearrangement $*$ from the Gaussian measure space $(\mathbb{R}^d, \gamma_d)$ to $(\mathbb{R}, \gamma_1)$, by

$$A^* = \{x \in \mathbb{R} : x < t\}$$

where $t = \Phi^{-1}(\gamma_d(A))$ is chosen to satisfy $\gamma_d(A) = \gamma(A^*)$. A functional half-space rearrangement by

$$f^*(x) = \int_0^\infty \mathbb{1}_{\{f>t\}^*}(x)dt.$$

The Borell–Ehrhard's inequality [16, 24] is usually stated as the assertion that $t \in (0,1)$, $A, B$ Borel in $\mathbb{R}^d$ imply

$$\gamma_d((1-t)A + tB) \geq \Phi((1-t)\Phi^{-1}(\mu(A)) + t\Phi^{-1}(\mu(B))).$$

It can be equivalently formulated in our terminology and notation .

**Theorem 7.5.3 (Borell–Ehrhard [16, 24])** *For $t \in (0,1)$, $m(x,y) = (1-t)x + ty$, $\eta(u,v) = (1-t)u + tv$, and $*$ our halfspace rearrangement from $(\mathbb{R}^d, \gamma_d)$ to $(\mathbb{R}, \gamma)$, satisfy a the set theoretic rearrangement inequality, explicitly for Borel $A$ and $B$*

$$\gamma_d((1-t)A + tB) \geq \gamma((1-t)A^* + tB^*).$$

We will extend Theorem 7.5.3 to a functional inequality by Theorem 7.4.6. However, it should be mentioned that the semigroup proof of Borell actually gave a functional inequality already. The argument was streamlined by Barthe and Huet, and it is their generalization below that we will sharpen.

**Theorem 7.5.4 (Barthe–Huet [3])** *Fix a set $I \subseteq \{1, 2, \ldots, n\}$ and positive numbers $\lambda_1, \ldots, \lambda_n$ satisfying $\sum \lambda_i \geq 1$ and $\lambda_j - \sum_{i \neq j} \lambda_i \leq 1$ for $j \notin I$. Then for Borel $f_1, \ldots, f_n$ from $\mathbb{R}^d$ to $[0,1]$ such that $\Phi^{-1} \circ f_i$ is concave for $i \in I$, and a Borel $h$ satisfying $h(\sum_i \lambda_i x_i) \geq \Phi(\sum_i \lambda_i \Phi^{-1}(f_i(x_i)))$, then*

$$\int h d\gamma_d \geq \Phi\left(\lambda_1 \Phi^{-1}\left(\int f_1 d\gamma_d\right) + \cdots + \lambda_n \Phi^{-1}\left(\int f_n d\gamma_d\right)\right).$$

A consequence of Theorem 7.5.4 (and actually proven equivalent to Theorem 7.5.4 in the same paper) is the following.

**Corollary 7.5.5** *Fix a set $I \subseteq \{1, 2, \ldots, n\}$ and set of positive numbers $\lambda_1, \ldots, \lambda_n$ satisfying $\sum \lambda_i \geq 1$ and $\lambda_j - \sum_{i \neq j} \lambda_i \leq 1$ for $j \notin I$. Then for Borel $A_j$,*

$$\begin{aligned}\gamma_d(\lambda_1 A_1 + \cdots + \lambda_n A_n) &\geq \Phi(\lambda_1 \Phi^{-1}(\gamma_d(A_1)) + \cdots + \lambda_n \Phi^{-1}(\gamma_d(A_n))) \\ &= \gamma(\lambda_1 A_1^* + \cdots + \lambda_n A_n^*)\end{aligned}$$

*holds, provided $A_i$ are convex when $i \in I$.*

Strictly speaking, unless $I$ is empty, the half-line rearrangement does not yield a set theoretic rearrangement inequality with the maps $m_\lambda(x) = \lambda_1 x_1 + \cdots + \lambda_n x_n$ and $\eta_\lambda(y) = \lambda_1 y_1 + \cdots + \lambda_n y_n$. However the proof of Theorem 7.4.6 can be adapted to achieve the following refinement of Barthe-Huet.

**Theorem 7.5.6** *For Borel $f_1, \dots, f_n$ from $\mathbb{R}^d$ to $[0, 1]$ such that $\Phi^{-1} \circ f_i$ is concave for $i \in I$ and*

$$\int \square_{\mathcal{M}_\Phi^\lambda} f d\gamma_d \geq \int \square_{\mathcal{M}_\Phi^\lambda} f_* d\gamma$$
$$\geq \mathcal{M}_\Phi^\lambda \left( \int f_1^* d\gamma, \dots, \int f_n^* d\gamma \right)$$
$$= \mathcal{M}_\Phi^\lambda \left( \int f_1 d\gamma, \dots, \int f_n d\gamma \right).$$

In analyzing the proof of Theorem 7.5.6, a loosening of the hypothesis can be achieved, requiring only that for $i \in I$, $f_i$ is quasi-concave and $\Phi^{-1} \circ f_i^*$ concave.

*Proof* Once it is observed that $\Phi^{-1} \circ f_i$ concave ensures $\{f_i > q_i\}$ is a convex set, so that one can apply Corollary 7.5.5, the first inequality can be derived following the proof of Theorem 7.4.6. The equality is immediate as well, following from our definition of rearrangement. Thus, to prove the result, we need only justify the second inequality, which follows from Theorem 7.5.4 once we know that the concavity of $\Phi^{-1} \circ f_i$ implies the concavity of $\Phi^{-1} \circ f_i^*$ as well. For this, we prove a general result below. □

**Definition 7.5.7** For a fixed $t \in (0, 1)$ and a convex set $K$ we will call $f : K \to \mathbb{R}$, $\Psi_t$*-concave* when there exists a continuous coordinate increasing function $\Psi_t$ such that

$$f((1-t)x_1 + tx_2) \geq \Psi_t(f(x_1), f(x_2)).$$

Notice that the concavity of $\Phi^{-1} \circ f$ is equivalent to the statement that $f$ is $\Psi_t$-concave with $\Psi_t(u_1, u_2) = \mathcal{M}_\Phi^t(u_1, u_2) = \Phi((1-t)\Phi^{-1}(u_1) + t\Phi^{-1}(u_2))$ for $t \in (0, 1)$.

**Proposition 7.5.8** *Suppose that $f, g, h$ are Borel functions on a space $(M, \mu)$ satisfying*

$$h((1-t)x + ty) \geq \Psi_t(f(x), g(y)) \tag{7.22}$$

*for $x, y \in M$, and that $*$ is a rearrangement from $(M, \mu)$ to a space $(N, \alpha)$ satisfying*

$$\mu((1-t)A + tB) \geq \alpha((1-t)A^* + tB^*). \tag{7.23}$$

*Additionally assume that the space of rearranged sets has a total ordering that respects Minkowski summation in the sense that $(1-t)A^* + tB^*$ and $C^*$ satisfy either*

$$(1-t)A^* + tB^* \subseteq C^* \text{ or } (1-t)A^* + tB^* \supseteq C^* \tag{7.24}$$

*then*

$$h^*((1-t)x+ty) \geq \Psi_t(f^*(x), g^*(y)) \tag{7.25}$$

*holds for* $x, y \in N$.

Note that Theorem 7.5.6 follows from the proposition by taking $f = g = h$ and $\Psi_t = \mathcal{M}_\Phi^t$. Indeed, since the half-line rearrangement satisfies (7.24), as half-lines are stable under convex combination, it follows that $f^*$ to be $\mathcal{M}_\Phi^t$-concave if $f$ is.

*Proof* Observe that inequality (7.22) can be equivalently stated as $\lambda_i \in \mathbb{R}$ implies

$$(1-t)\{f > \lambda_1\} + t\{g > \lambda_2\} \subseteq \{h > \Psi_t(\lambda_1, \lambda_2)\}. \tag{7.26}$$

which can be easily verified using our assumptions of continuity and monotonicity. Indeed, if (7.22) holds, then for $z = (1-t)x+ty$ for $x \in \{f > \lambda_1\}$ and $y \in \{g > \lambda_2\}$ we have $h(z) \geq \Psi_t(f(x), g(y)) > \Psi_t(\lambda_1, \lambda_2)$. For the converse, given $x, y$ take $\lambda_1 < f(x)$ and $\lambda_2 < g(y)$, then $z = (1-t)x + ty \in (1-t)\{f > \lambda_1\} + t\{g > \lambda_2\}$. By (7.26), $h(z) > \Psi_t(f(x), g(y))$, and by the continuity assumption on $\Psi_t$, $\Psi_t(f(x), g(y)) = \sup_\lambda \Psi_t(\lambda_1, \lambda_2) \leq h(z)$. Thus we will prove $(1-t)\{f^* > \lambda_1\} + t\{g^* > \lambda_2\} \subseteq \{h^* > \Psi_t(\lambda_1, \lambda_2)\}$, or equivalently

$$(1-t)\{f > \lambda_1\}^* + t\{g > \lambda_2\}^* \subseteq \{h > \Psi_t(\lambda_1, \lambda_2)\}^*.$$

By (7.24), it is enough to show

$$\alpha((1-t)\{f > \lambda_1\}^* + t\{g > \lambda_2\}^*) \leq \alpha(\{h > \Psi_t(\lambda_1, \lambda_2)\}^*).$$

By our assumptions (7.23) and (7.26),

$$\begin{aligned}\alpha((1-t)\{f > \lambda_1\}^* + t\{g > \lambda_2\}^*) &\leq \mu((1-t)\{f > \lambda_1\} + t\{g > \lambda_2\}) \\ &\leq \mu(\{h > \Psi_t(\lambda_1, \lambda_2)\}).\end{aligned}$$

Our result follows since

$$\mu(\{h > \Psi_t(\lambda_1, \lambda_2)\}) = \alpha(\{h > \Psi_t(\lambda_1, \lambda_2)\}^*).$$

□

Observe that Proposition 7.5.8 gives another proof of Theorem 7.3.1. Indeed, since $f\Box g((1-t)x+ty) \geq f^{1-t}(x)g^t(y)$ holds for all $x, y$, $(f\Box g)^*((1-t)x + ty)) \geq (f^*)^{1-t}(x)(g^*)^t(y)$ holds as well. This implies $(f\Box g)^* \geq f^*\Box g^*$ and hence

$$|\{f\Box g > \lambda\}| = \left|\left\{(f\Box g)^* > \lambda\right\}\right| \geq \left|\left\{f^*\Box g^* > \lambda\right\}\right|.$$

Let us also point out the corollary obtained by taking $f = g = h$, as it is of interest independent of the application to Theorem 7.5.6.

**Corollary 7.5.9** *If $f : \mathbb{R}^d \to [0, \infty)$ is $\Psi_t$-concave, and $*$ implies $f^*$ is as well.*

It follows immediately that the class of $d$-dimensional $s$-concave measures is stable under (convex set) rearrangement. See [11, 13] for background and [30, 32] for recent connections between $s$-concave measures and information theory.

### *7.5.3 Polar Prékopa–Leindler*

For fixed $t, \lambda \in (0, 1)$, define $\mathcal{M} : [0, \infty)^2 \to [0, \infty)$ by

$$\mathcal{M}(u, v) = \min\left\{u^{\frac{1-t}{1-\lambda}}, v^{\frac{t}{\lambda}}\right\},$$

and for $x, y \in \mathbb{R}^d$ define $m(x, y) = (1-t)x + ty$ so that

$$f \square_{\mathcal{M}} g(z) = \sup_{m(x,y)=z} \min\left\{f(x)^{\frac{1-t}{1-\lambda}}, g(y)^{\frac{t}{\lambda}}\right\}.$$

We can state the recent polar analog of Prékopa–Leindler due to Artstein-Avidan, Florentin, and Segal.

**Theorem 7.5.10 (Artstein-Avidan et al. [1])** *For $f, g : \mathbb{R}^d \to [0, \infty)$ Borel, and $\mu$ log-concave*

$$\int f \square_{\mathcal{M}} g(x) d\mu(x) \geq \mathcal{M}_{-1}^{\lambda}\left(\int f(x) d\mu(x), \int g(x) d\mu(x)\right).$$

In the case that $\mu$ is Lebesgue (with $*$ rearrangement to a convex set) or Gaussian (with $*$ rearrangement to a half-space), and $\eta(x, y) = (1-t)x + ty$, this can be sharpened to the following.

**Theorem 7.5.11** *For $f, g : \mathbb{R}^d \to [0, \infty)$ Borel, and $\mu$ either Gaussian, with $*$ the half space rearrangement, or Lebesgue with $*$ a convex set rearrangement, then*

$$\begin{aligned}\int f \square_{\mathcal{M}} g d\mu &\geq \int f^* \square_{\mathcal{M}} g^* d\mu \\ &\geq \mathcal{M}_{-1}^{\lambda}\left(\int f d\mu, \int g d\mu\right).\end{aligned}$$

*Proof* As we have seen, the map $(x, y) \mapsto (1-t)x + ty$ satisfies a set theoretic rearrangement inequality by Brunn–Minkowski with respect to Lebesgue measure and rearrangement to a convex set, and by Borell–Ehrhard with respect to Gaussian measure and rearrangement to a halfspace. The map $\mathcal{M}(u, v) = \min\{u^{\frac{1-t}{1-\lambda}}, v^{\frac{t}{\lambda}}\}$ is clearly continuous and coordinate increasing for $\lambda, t \in (0, 1)$. Thus in both cases, Gaussian and Lebesgue, we can invoke Theorem 7.4.6 to obtain the first inequality. The second inequality is obtained from the application of Theorem 7.5.10 to $f^*$ and $g^*$, and the equimeasurability of rearrangements. □

## 7.6 Gaussian Log-Sobolev Inequality

For a probability measure $\mu$ define the entropy functional[2] for a non-negative $f$ by

$$H_\mu(f) = \int f \log f d\mu - \int f d\mu \log \int f d\mu.$$

One formulation of the Gaussian log-Sobolev inequality is the following.

**Theorem 7.6.1 (Gaussian Log-Sobolev)** *For positive smooth $f$,*

$$H_{\gamma_d}(f) \leq \frac{1}{2} \int \frac{|\nabla f|^2}{f} d\gamma_d.$$

In this form the inequality is due to Gross [27]. Carlen [20] showed it to be equivalent to the earlier information theoretic Blachman–Stam inequality [5, 48]. The Gaussian log-Sobolev inequality was shown to be a consequence of a strengthened PLI for strongly log-concave measures by Bobkov–Ledoux [8], and it is this perspective that we now develop to motivate the main result of this section, a rearrangement sharpening of an integrated Gaussian log-Sobolev inequality. In this direction, let us recall that the PLI can be easily extended to the log-concave case.

**Theorem 7.6.2 (Log-Concave PLI)** *For measure $\mu$ with density $\varphi$ satisfying*

$$\varphi((1-t)x + ty) \geq \varphi^{1-t}(x)\varphi^t(y),$$

*the inequality for non-negative functions $u, v, w$*

$$u((1-t)x + ty) \geq v^{1-t}(x)w^t(y)$$

*implies*

$$\int u d\mu \geq \left(\int v d\mu\right)^{1-t} \left(\int w d\mu\right)^t. \tag{7.27}$$

[2]Note that when $f = \frac{d\nu}{d\mu}$ is the density function of a probability measure $\nu$ with respect to $\mu$, $H_\mu(f)$ is the Kullback–Liebler divergence $D(\nu||\mu)$ or relative entropy [22].

*Proof* Observing that the functions $\tilde{u}(z) = u(z)\varphi(z)$, $\tilde{v}(z) = v(z)\varphi(z)$, and $\tilde{w}(z) = w(z)\varphi(z)$ satisfy

$$\tilde{u}((1-t)x + ty) \geq \tilde{v}^{1-t}(x)\tilde{w}^t(y)$$

so that applying the ordinary PLI, we have

$$\int \tilde{u}(z)dz \geq \left( \int \tilde{v}(z)dz \right)^{1-t} \left( \int \tilde{w}(z)dz \right)^t,$$

which is exactly (7.27). □

The log-concave case corresponds to the case when the measure is given by a density corresponding to a convex potential, that is, $\varphi(x) = e^{-V(x)}$ with $V$ is convex. For the Gaussian measure something stronger is true. In this case, $V$ satisfies

$$V((1-t)x + ty) \leq (1-t)V(x) + tV(y) - t(1-t)|x-y|^2/2. \tag{7.28}$$

Note that in the case that $V$ is smooth, log-concavity is exactly $V'' \geq 0_d$ in the sense of positive semi-definite matrices, while (7.28) is $V'' \geq I_d$. Under these assumptions, Theorem 7.6.2 admits the following strengthening.

**Theorem 7.6.3 (Curved Prékopa–Leindler)** *For $t \in (0, 1)$, $\mu$ strongly log-concave in the sense of* (7.28)*, and $u, v, w : \mathbb{R}^d \to [0, \infty)$ satisfying*

$$u((1-t)x + ty) \geq e^{-t(1-t)|x-y|^2/2} v^{1-t}(x) w^t(y),$$

*for all $x, y \in \mathbb{R}^d$, then*

$$\int u d\mu \geq \left( \int v\, d\mu \right)^{1-t} \left( \int w d\mu \right)^t.$$

*Proof* The proof follows again from applying the Euclidean PLI to $\tilde{u}(z) = u(z)\varphi(z)$, $\tilde{v}(z) = v(z)\varphi(z)$. □

Following arguments of Bobkov–Ledoux [8] we pursue a specialization of Theorem 7.6.3 to a single function, revealing a log-Sobolev inequality as a consequence of a strengthened PLI. For a fixed $t \in (0, 1)$, and a strongly log-concave probability measure $\mu$, and $f$, take $w = f^{\frac{1}{t}}$, $v = 1$, then for any $u$, satisfying

$$u((1-t)x + ty) \geq e^{-t(1-t)|x-y|^2/2} f(y)$$

we have from Theorem 7.6.3

$$\int u\, d\mu \geq \left( \int f^{\frac{1}{t}} d\mu \right)^t.$$

With the interest of determining the optimal such $u$ achievable through the methods of PLI, it is natural to consider

$$u(z) = \sup_{\{(x,y):(1-t)x+ty=z\}} e^{-t(1-t)|x-y|^2/2} f(y).$$

Writing $\lambda = \frac{1-t}{t}$, note that the constraint on $x, y$ is equivalent to $y = z + \lambda(z - x)$, so that the $u(z)$ above can be expressed as $Q_\lambda f(z)$ in the following definition.

**Definition 7.6.4** For $\lambda \in (0, \infty)$ and $f$ non-negative and Borel measurable, define

$$\begin{aligned} Q_\lambda f(z) &= \sup_w f(z + \lambda w) e^{-\lambda |w|^2/2} \\ &= \sup_w f(z + w) e^{-|w|^2/2\lambda}. \end{aligned}$$

Writing $\|f\|_p = \left(\int |f|^p d\mu\right)^{\frac{1}{p}}$ we can collect the above as the following.

**Theorem 7.6.5 (Integrated Log-Sobolev)** *For $\mu$ a strongly log-concave probability measure, $\lambda \in (0, \infty)$ and $f$ non-negative and Borel measurable,*

$$\|Q_\lambda f\|_1 \geq \|f\|_{1+\lambda}.$$

The log-Sobolev inequality for strongly log-concave probability measures can be recovered as a corollary.

**Corollary 7.6.6 (Log-Sobolev Inequality)** *For $\mu$ strongly log-concave probability measure, and $f$ a positive smooth function*

$$H_\mu(f) \leq \frac{1}{2} \int \frac{|\nabla f|^2}{f} d\mu$$

A proof is given in [8] where the expressions are given in terms of $f^2$ rather than $f$. It follows as a limiting case of Theorem 7.6.5 with $\lambda \to 0$.

*Sketch of Proof* For smooth positive functions constant outside of a compact set, one observes that equality holds when $\lambda = 0$. Then the Taylor series expansion,

$$\|f\|_{1+\lambda} = \|f\|_1 + \lambda H_\mu(f) + o(\lambda)$$

and a derived inequality

$$\|Q_\lambda f\|_1 \leq \|f\|_1 + \frac{\lambda}{2} \int \frac{|\nabla f|^2}{f} d\mu + o(\lambda)$$

deliver the conclusion. A limiting argument gives the result for general functions. □

To state our main result of the section, let $\mu = \gamma_d$ a standard Gaussian and $*$ be the half-space rearrangement of a set under $\gamma_d$, as in Proposition 7.4.3.

**Theorem 7.6.7** *For non-negative Borel $f$ and $\lambda, s > 0$,*

$$\gamma_d(\{Q_\lambda f > s\}) \geq \gamma(\{Q_\lambda f^* > s\})$$

*where $f^*$ is the Gaussian half-line rearrangement of $f$.*

It will be a consequence of the proof that $Q_\lambda f$ is universally measurable.

*Proof* We first express $\{Q_\lambda f > s\}$ as the union of simpler sets. Denoting

$$S = S(s, q_1, q_2) = \{q = (q_1, q_2) \in \mathbb{Q}_+^2 : q_1 q_2 > s\},$$

it is straight forward to verify

$$\{Q_\lambda f > s\} = \bigcup_{q \in S} \left( \{x \in \mathbb{R}^d : f(x) > q_1\} + \left\{ y \in \mathbb{R}^d : |y| < \sqrt{2\lambda \ln \frac{1}{q_2}} \right\} \right). \tag{7.29}$$

Indeed, for $z$ belonging to the union, there exists rational $q_i$, and $x, y$ satisfying $f(x) > q_1$, $|y| < \sqrt{2\lambda \ln \frac{1}{q_2}}$, and $x + y = z$. Taking $w = -x = y - z$,

$$f(w) e^{-|w|^2/2\lambda} > q_1 q_2 > s,$$

so that $z \in \{Q_\lambda f > s\}$. Conversely if there exists a $w$ such that $f(z+w)e^{-|w|^2/2\lambda} > s$ then by continuity there exist rational $q_i$ satisfying $f(z+w) > q_1$, $e^{-|w|^2/2\lambda} > q_2$, and $q_1 q_2 > s$. Taking $x = z + w$ and $y = -w$ we see that $(q_1, q_2) \in S$ and

$$z \in \{f > q_1\} + \left\{ |y| < \sqrt{2\lambda \ln \frac{1}{q_2}} \right\}.$$

Notice that this gives $\{Q_\lambda f > s\}$ as a countable union of Minkowski sums of analytic sets. Since analytic sets are closed under such operations, $\{Q_\lambda f > s\}$ is an analytic set as well, and the universal measurability of $Q_t f$ follows.

Applying the Gaussian isoperimetric inequality [14, 49], which in our preferred formulation states that $\gamma_d(A + B_d) \geq \gamma(A^* + B_1)$ where $B_d$ and $B_1$ are origin symmetric Euclidean balls of equal radius (in $\mathbb{R}^d$ and $\mathbb{R}$ respectively), we have

$$\begin{aligned}\gamma_d(\{Q_\lambda f > s\}) &= \gamma_d\left(\bigcup_{q\in S}\{f > q_1\} + \left\{w \in \mathbb{R}^d : |w| < \sqrt{2\lambda \ln \frac{1}{q_2}}\right\}\right)\\ &\geq \sup_{q\in S}\gamma_d\left(\{f > q_1\} + \left\{w \in \mathbb{R}^d : |w| < \sqrt{2\lambda \ln \frac{1}{q_2}}\right\}\right)\\ &\geq \sup_{q\in S}\gamma\left(\{f > q_1\}^* + \left\{w \in \mathbb{R} : |w| < \sqrt{2\lambda \ln \frac{1}{q_2}}\right\}\right).\end{aligned}$$

But $\{f > q_1\}^* = \{f^* > q_1\}$ is a half-line and hence the sets $\{f^* > q_1\} + \left\{|w| < \sqrt{2\lambda \ln \frac{1}{q_2}}\right\}$, indexed by $S(\lambda, q_1, q_2)$, are totally ordered. Thus,

$$\sup_{q\in S}\gamma\left(\{f > q_1\}^* + \left\{|w| < \sqrt{2\lambda \ln \frac{1}{q_2}}\right\}\right) = \gamma\left(\bigcup_{q\in S}\{f^* > q_1\} + \left\{|w| < \sqrt{2\lambda \ln \frac{1}{q_2}}\right\}\right).$$

Applying (7.29),

$$\gamma\left(\bigcup_{q\in S}\{f^* > q_1\} + \left\{|w| < \sqrt{2\lambda \ln \frac{1}{q_2}}\right\}\right) = \gamma(\{Q_\lambda f^* > \lambda\}),$$

and our theorem follows. □

As an immediate consequence, we have a sharpening of Theorem 7.6.5.

**Corollary 7.6.8** *For $f$ non-negative and Borel, and norms taken with respect to $\gamma$,*

$$\int Q_\lambda f d\gamma \geq \int Q_\lambda f^* d\gamma \geq \|f^*\|_{1+\lambda} = \|f\|_{1+\lambda}.$$

*Proof* The first inequality is a consequence of Theorem 7.6.7, while the second is from Theorem 7.6.5. □

We also direct the reader to the articles [41, 42] of Martín and Milman, whose work on symmetrization, isoperimetry, and log-Sobolev inequalities the author learned of during the revision of this paper.

## 7.7 Barthe, Brascamp, Lieb and Rearrangement

The Brascamp–Lieb inequality is the following.

**Theorem 7.7.1 (Brascamp–Lieb [17])** *For natural numbers $n \leq m$, and $\{n_i\}_{i=1}^m$ with $n_i \leq n$ and $\{c_i\}_{i=1}^m$ a sequence of positive numbers such that $\sum_{i=1}^m c_i n_i = n$ then for surjective linear maps $B_i : \mathbb{R}^n \to \mathbb{R}^{n_i}$, with $\cap_i \ker(B_i) = 0$ and transposes denoted $B_i'$ satisfy the following,*

$$\int_{\mathbb{R}^n} \prod_{i=1}^m f_i^{c_i}(B_i x)dx \leq C^{-1/2} \prod \left( \int_{\mathbb{R}^{n_i}} f_i \right)^{c_i}$$

*for $f_i : \mathbb{R}^{n_i} \to [0, \infty)$ integrable, and*

$$C = \inf \left\{ \frac{\det(\sum_{i=1} c_i B_i' A_i B_i)}{\prod \det^{c_i} A_i} : A_i \text{ positive definite} \right\}.$$

The theorem enjoys a qualitative analog in the case that $n_i = d$, so that $n = md$ and $x \in \mathbb{R}^n$ can be expressed as $x = (x_1, \ldots, x_m)$ for $x_j \in \mathbb{R}^d$ and $B_i$ are of the form

$$B_i x = \sum_{j=1}^m B_{ij} x_j. \tag{7.30}$$

**Theorem 7.7.2 (Brascamp et al. [19])** *For $B_i$ satisfying (7.30),*

$$\int_{\mathbb{R}^n} \prod_{i=1}^m f_i(B_i x)dx \leq \int_{\mathbb{R}^n} \prod_{i=1}^m f_i^*(B_i x)dx,$$

*where $*$ represents the spherically symmetric decreasing rearrangement.*

Notice that when Theorem 7.7.2 applies, it gives an intermediary inequality to Theorem 7.7.1. Indeed, since $(f^{c_i})^* = (f^*)^{c_i}$, applying first Theorem 7.7.2, and then 7.7.1, gives

$$\begin{aligned} \int_{\mathbb{R}^n} \prod_{i=1}^m f^{c_i}(B_i x)dx &\leq \int_{\mathbb{R}^n} \prod_{i=1}^m (f^*)^{c_i}(B_i x)dx \\ &\leq C^{-1/2} \prod_{i=1}^m \left( \int_{\mathbb{R}^{n_i}} f \right)^{c_i}. \end{aligned}$$

Barthe gave the following reversal of Brascamp–Lieb, which serves as a dual inequality.

**Theorem 7.7.3 (Barthe [2])** *For $n$, $m$, $\{n_i\}_{i=1}^m$, $\{c_i\}_{i=1}^m$, $B_i$, and $C$ as in Theorem 7.7.1 then the inequality*

$$C^{1/2}\prod_{i=1}^m \left(\int_{\mathbb{R}^{n_i}} f_i\right)^{c_i} \leq \int_{\mathbb{R}^n} \sup\left\{\prod_{i=1}^m f_i^{c_i}(y_i) : \sum_i c_i B_i' y_i = x\right\} dx,$$

*holds for $f_i : \mathbb{R}^{n_i} \to [0,\infty)$ integrable.*

Taking $m = 2$, $c_1 = (1-t)$, $c_2 = t$ and $n_i = n$ and $B_i$ to be the identity map yields $C = 1$ and we recover the Prekopa–Liendler inequality. We ask if further extensions of our work here exist.

*Question 7.7.4* Suppose that $B_i$ are of the form (7.30), and $f_i : \mathbb{R}^d \to [0,\infty)$, when is it true that

$$\int_{\mathbb{R}^n} \sup\left\{\prod_{i=1}^m f_i(y_i) : \sum_i B_i' y_i = x\right\} dx \geq \int_{\mathbb{R}^n} \sup\left\{\prod_{i=1}^m f_i^*(y_i) : \sum_i B_i' y_i = x\right\} dx \tag{7.31}$$

holds?

The results presented here verify the inequality for general Borel $f_i$ in the case that $B_i$ are scalar multiples of the identity. Note that the case $f_i = \mathbb{1}_{A_i}$ asks if the following generalization of BMI holds

$$\left|\sum_i B_i' A_i\right|_n \geq \left|\sum_i B_i' A_i^*\right|_n, \tag{7.32}$$

where

$$\sum_i B_i' A_i = \left\{z = \sum_i B_i' x_i : x_i \in A_i\right\}.$$

In the case that $B_i' : \mathbb{R} \to \mathbb{R}^d$, inequality (7.32) was proven by Zamir and Feder [51].

## References


1. S. Artstein-Avidan, D. Florentin, A. Segal, Polar Prékopa–Leindler inequalities (2017). arXiv preprint. arXiv:1707.08732
2. F. Barthe, On a reverse form of the Brascamp-Lieb inequality. Invent. Math. **134**(2), 335–361 (1998)

3. F. Barthe, N. Huet, On Gaussian Brunn–Minkowski inequalities. Stud. Math. **191**(3), 283–304 (2009)
4. W. Beckner, Inequalities in Fourier analysis. Ann. Math. **102**(1), 159–182 (1975)
5. N.M. Blachman, The convolution inequality for entropy powers. IEEE Trans. Inf. Theory **11**, 267–271 (1965)
6. S.G. Bobkov, G.P. Chistyakov, Bounds for the maximum of the density of the sum of independent random variables. Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. **408**, 62–73, 324 (2012)
7. S.G. Bobkov, G.P. Chistyakov, Entropy power inequality for the Rényi entropy. IEEE Trans. Inf. Theory **61**(2), 708–714 (2015)
8. S.G. Bobkov, M. Ledoux, From Brunn-Minkowski to Brascamp-Lieb and to logarithmic Sobolev inequalities. Geom. Funct. Anal. **10**(5), 1028–1052 (2000)
9. S.G. Bobkov, M. Ledoux, Weighted Poincaré-type inequalities for Cauchy and other convex measures. Ann. Probab. **37**(2), 403–427 (2009)
10. S. Bobkov, A. Marsiglietti, Variants of the entropy power inequality. IEEE Trans. Inf. Theory **63**(12), 7747–7752 (2017)
11. S. Bobkov, J. Melbourne, Hyperbolic measures on infinite dimensional spaces. Probab. Surv. **13**, 57–88 (2016)
12. S.G. Bobkov, I. Gentil, M. Ledoux, Hypercontractivity of Hamilton–Jacobi equations. J. Math. Pures Appl. **80**(7), 669–696 (2001)
13. C. Borell, Convex measures on locally convex spaces. Ark. Mat. **12**, 239–252 (1974)
14. C. Borell, The Brunn-Minkowski inequality in Gauss space. Invent. Math. **30**(2), 207–216 (1975)
15. C. Borell, Convex set functions in $d$-space. Period. Math. Hungar. **6**(2), 111–136 (1975)
16. C. Borell, Inequalities of the Brunn-Minkowski type for Gaussian measures. Probab. Theory Relat. Fields **140**(1–2), 195–205 (2008)
17. H.J. Brascamp, E.H. Lieb, Best constants in Young's inequality, its converse, and its generalization to more than three functions. Adv. Math. **20**(2), 151–173 (1976)
18. H.J. Brascamp, E.H. Lieb, On extensions of the Brunn-Minkowski and Prékopa-Leindler theorems, including inequalities for log concave functions, and with an application to the diffusion equation. J. Funct. Anal. **22**(4), 366–389 (1976)
19. H.J. Brascamp, E.H. Lieb, J.M. Luttinger, A general rearrangement inequality for multiple integrals. J. Funct. Anal. **17**, 227–237 (1974)
20. E.A. Carlen, Superadditivity of Fisher's information and logarithmic Sobolev inequalities. J. Funct. Anal. **101**(1), 194–211 (1991)
21. M.H.M. Costa, T.M. Cover, On the similarity of the entropy power inequality and the Brunn-Minkowski inequality. IEEE Trans. Inf. Theory **30**(6), 837–839 (1984)
22. T.M. Cover, J.A. Thomas, *Elements of Information Theory* (Wiley, New York, 1991)
23. A. Dembo, T.M. Cover, J.A. Thomas, Information-theoretic inequalities. IEEE Trans. Inf. Theory **37**(6), 1501–1518 (1991)
24. A. Ehrhard, Symétrisation dans l'espace de Gauss. Math. Scand. **53**(2), 281–301 (1983)
25. N. Gozlan, C. Roberto, P.M. Samson, P. Tetali, Displacement convexity of entropy and related inequalities on graphs. Probab. Theory Relat. Fields **160**(1–2), 47–94 (2014)
26. N. Gozlan, C. Roberto, P.M. Samson, P. Tetali, Transport proofs of some discrete variants of the Prékopa-Leindler inequality (2019). arXiv preprint. arXiv:1905.04038
27. L. Gross, Logarithmic Sobolev inequalities. Am. J. Math. **97**(4), 1061–1083 (1975)
28. A. Kechris, *Classical Descriptive Set Theory*, vol. 156 (Springer Science & Business Media, New York, 2012)
29. J. Li, Rényi entropy power inequality and a reverse (2017). Preprint. arXiv:1704.02634
30. J. Li, J. Melbourne, Further investigations of the maximum entropy of the sum of two dependent random variables, in *2018 IEEE International Symposium on Information Theory (ISIT)* (IEEE, Piscataway, 2018), pp. 1969–1972
31. J. Li, A. Marsiglietti, J. Melbourne, Entropic central limit theorem for rényi entropy, in *2018 IEEE International Symposium on Information Theory (ISIT)* (IEEE, Piscataway, 2019)

32. J. Li, A. Marsiglietti, J. Melbourne, Further investigations of Rényi entropy power inequalities and an entropic characterization of s-concave densities, in *Geometric Aspects of Functional Analysis – Israel Seminar (GAFA) 2017–2019*. Lecture Notes in Mathematics, vol. 2256 (to appear)
33. J. Li, A. Marsiglietti, J. Melbourne, Rényi entropy power inequalities for $s$-concave densities, in *2019 IEEE International Symposium on Information Theory (ISIT)* (IEEE, Piscataway, 2019)
34. E.H. Lieb, Proof of an entropy conjecture of Wehrl. Commun. Math. Phys. **62**(1), 35–41 (1978)
35. M. Madiman, J. Melbourne, P. Xu, Forward and reverse entropy power inequalities in convex geometry, in *Convexity Concentration* (Springer, New York, 2017), pp. 427–485
36. M. Madiman, J. Melbourne, P. Xu, Rogozin's convolution inequality for locally compact groups (2017). Preprint, arXiv:1705.00642
37. M. Madiman, L. Wang, J.O. Woo, Entropy inequalities for sums in prime cyclic groups (2017). arXiv preprint. arXiv:1710.00812
38. M. Madiman, L. Wang, J.O. Woo, Majorization and Rényi entropy inequalities via Sperner theory. Discrete Math. **342**, 2911–2923 (2019)
39. A. Marsiglietti, J. Melbourne, A Rényi entropy power inequality for log-concave vectors and parameters in [0, 1], in *2018 IEEE International Symposium on Information Theory (ISIT)* (IEEE, Piscataway, 2018), pp. 1964–1968
40. A. Marsiglietti, J. Melbourne, On the entropy power inequality for the Rényi entropy of order [0, 1]. IEEE Trans. Inf. Theory **65**, 1387–1396 (2019)
41. J. Martín, M. Milman, Isoperimetry and symmetrization for logarithmic Sobolev inequalities. J. Funct. Anal. **256**(1), 149–178 (2009)
42. J. Martín, M. Milman, Pointwise symmetrization inequalities for Sobolev functions and applications. Adv. Math. **225**(1), 121–199 (2010)
43. J. Melbourne, Rearrangements and information theoretic inequalities, in *2018 56th Annual Allerton Conference on Communication, Control, and Computing (Allerton)* (IEEE, Piscataway, 2018)
44. Y. Ollivier, C. Villani, A curved Brunn–Minkowski inequality on the discrete hypercube, or: what is the Ricci curvature of the discrete hypercube? SIAM J. Discrete Math. **26**(3), 983–996 (2012)
45. E. Ram, I. Sason, On rényi entropy power inequalities. IEEE Trans. Inf. Theory **62**(12), 6800–6815 (2016)
46. A. Rényi, On measures of entropy and information, in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1* (University of California Press, Berkeley, 1961), pp. 547–561
47. C.E. Shannon, A mathematical theory of communication. Bell Syst. Tech. J. **27**, 379–423, 623–656 (1948)
48. A.J. Stam, Some inequalities satisfied by the quantities of information of Fisher and Shannon. Inf. Control **2**, 101–112 (1959)
49. V.N. Sudakov, B.S. Tsirel'son, Extremal properties of half-spaces for spherically invariant measures. Zap. Nauch. Sem. LOMI **41**, 14–24 (1974). Translated in J. Soviet Math. 9, 9–18 (1978)
50. L. Wang, M. Madiman, Beyond the entropy power inequality, via rearrangements. IEEE Trans. Inf. Theory **60**(9), 5116–5137 (2014)
51. R. Zamir, M. Feder, A generalization of the entropy power inequality with applications. IEEE Trans. Inf. Theory **39**(5), 1723–1728 (1993)

# Chapter 8
# Generalized Semimodularity: Order Statistics

**Iosif Pinelis**

**Abstract** A notion of generalized $n$-semimodularity is introduced, which extends that of (sub/super)modularity in four ways at once. The main result of this paper, stating that every generalized $(n : 2)$-semimodular function on the $n$th Cartesian power of a distributive lattice is generalized $n$-semimodular, may be considered a multi/infinite-dimensional analogue of the well-known Muirhead lemma in the theory of Schur majorization. This result is also similar to a discretized version of the well-known theorem due to Lorentz, which latter was given only for additive-type functions. Illustrations of our main result are presented for counts of combinations of faces of a polytope; one-sided potentials; multiadditive forms, including multilinear ones—in particular, permanents of rectangular matrices and elementary symmetric functions; and association inequalities for order statistics. Based on an extension of the FKG inequality due to Rinott & Saks and Aharoni & Keich, applications to correlation inequalities for order statistics are given as well.

**Keywords** Semimodularity · Submodularity · Supermodularity · FKG-type inequalities · Association inequalities · Correlation inequalities

**2010 Mathematics Subject Classification** Primary 06D99, 26D15, 26D20, 60E15; Secondary 05A20, 05B35, 06A07, 60C05, 62H05, 62H10, 82D99, 90C27

## 8.1 Summary and Discussion

As pointed out e.g. in [3, 4], the notion of submodularity has become useful in various areas: combinatorial optimization, with many applications in operations research; machine learning; computer vision; electrical networks; signal processing;

I. Pinelis (✉)
Department of Mathematical Sciences, Michigan Technological University, Houghton, MI, USA
e-mail: ipinelis@mtu.edu



N. Gozlan et al. (eds.), *High Dimensional Probability VIII*,
Progress in Probability 74, https://doi.org/10.1007/978-3-030-26391-1_8

several areas of theoretical computer science, such as matroid theory; economics. One may also note the use of this notion in potential theory [6], as a capacity is a submodular function.

Let $L$ be any distributive lattice; for definitions and facts pertaining to lattices, see e.g. [10].

A function $\lambda\colon L \to \mathbb{R}$ is called submodular if

$$\lambda(f) + \lambda(g) \geq \lambda(f \vee g) + \lambda(f \wedge g) \tag{8.1}$$

for all $f$ and $g$ in $L$. A function $\lambda$ is called supermodular if the function $-\lambda$ is submodular, and $\lambda$ is called modular if it is both submodular and supermodular. See e.g. [4, 9, 18, 19, 24, 25]. Let us say that a function $\mu$ is log-submodular if $\ln\mu$ is submodular. The log-submodularity condition and the corresponding log-supermodularity condition were referred to in Karlin and Rinott [13, 14] as the multivariate total positivity of order 2 ($\mathrm{MTP}_2$) and the multivariate reverse rule of order 2 ($\mathrm{MRR}_2$), respectively. As noted by Choquet [6, §14.3], a nondecreasing function $\lambda$ is alternating of order 2 iff it satisfies inequality (8.1), that is, $\lambda$ is submodular; it was also shown in [6] that the classical Newtonian capacity is such a function.

The log-supermodularity condition is the condition under which the famous Fortuin–Kasteleyn–Ginibre (FKG) correlation inequality [8] holds. Therefore, using inequality (8.17) together with the FKG inequality and its generalizations, we will be able to obtain the corresponding applications, in Corollaries 8.2.11 and 8.2.12.

More generally, let $\mathcal{R}$ be any set, endowed with a transitive relation $\bowtie$, so that for any $a, b, c$ in $\mathcal{R}$ one has the implication $a \bowtie b \;\&\; b \bowtie c \implies a \bowtie c$. For any natural $n$, let us say that a function $\Lambda\colon L^n \to \mathcal{R}$ is *generalized n-semimodular* if

$$\Lambda(f_1, \dots, f_n) \bowtie \Lambda(f_{n:1}, \dots, f_{n:n})$$

for all $f = (f_1, \dots, f_n) \in L^n$, where $f_{n:1}, \dots, f_{n:n}$ are the "order statistics" for $f$ defined by the formula

$$f_{n:j} = \bigwedge\Big\{\bigvee_{i\in J} f_i\colon J \in \binom{[n]}{j}\Big\} \tag{8.2}$$

for $j \in [n] := \overline{1, n}$, with $\binom{[n]}{j}$ denoting the set of all subsets $J$ of the set $[n]$ such that the cardinality of $J$ is $j$. Here and in the sequel we use the notation $\overline{\alpha, \beta} := \{j \in \mathbb{Z}\colon \alpha \leq j \leq \beta\}$. In particular, $f_{n:1} = f_1 \wedge \cdots \wedge f_n$ and $f_{n:n} = f_1 \vee \cdots \vee f_n$.

For any function $\lambda\colon L \to \mathbb{R}$, let the function $\Lambda_\lambda\colon L^2 \to \mathbb{R}$ be given by the formula $\Lambda_\lambda(f, g) := \lambda(f)+\lambda(g)$ for $f$ and $g$ in $L$. Then, obviously, $\lambda$ is submodular or supermodular or modular if and only if $\Lambda_\lambda$ is generalized 2-semimodular with the relation "$\bowtie$" being "$\geq$" or "$\leq$" or "$=$", respectively.

Thus, the notion of generalized $n$-semimodularity extends that of (sub/super)modularity in four ways at once: (1) the function $\Lambda$ may be a function of any natural number $n$ of arguments, whereas $\lambda$ is a function of only one argument; (2) in contrast with a general form of dependence of $\Lambda(f_1, \dots, f_n)$ on $f_1, \dots, f_n$, the function $\Lambda_\lambda$ of two arguments is of the special form, linear in $\lambda(f)$ and $\lambda(g)$; (3) whereas the values of $\lambda$ are real numbers, those of $\Lambda$ may be in any set $\mathcal{R}$; and (iv) we now have an arbitrary transitive relation $\bowtie$ over $\mathcal{R}$ instead of one of the three particular relations "$\ge$" or "$\le$" or "$=$" over $\mathbb{R}$.

For any $k \in [n]$, let us say that a function $\Lambda\colon L^n \to \mathbb{R}$ is *generalized* $(n\colon k)$*-semimodular* if for each $j \in \overline{0, n-k}$ and each $(n-k)$-tuple $(f_i\colon i \in [n] \setminus \overline{j+1, j+k}) \in L^{n-k}$ the function $L^k \ni (f_{j+1}, \dots, f_{j+k}) \mapsto \Lambda(f_1, \dots, f_n)$ is generalized $k$-semimodular. In particular, $\Lambda$ is generalized $(n\colon n)$-semimodular if and only if it is generalized $n$-semimodular.

Whenever the relation "$\bowtie$" is denoted as "$\ge$" or "$\le$" or "$=$", let us replace "semi" in the above definitions by "sub", "super", and "", respectively. For instance, "generalized $n$-modular" will stand for "generalized $n$-semimodular" with the relation "$\bowtie$" being "$=$".

The main result of this note is

**Theorem 8.1.1** *Again, let $L$ be any distributive lattice. If a function $\Lambda\colon L^n \to \mathcal{R}$ is generalized $(n\colon 2)$-semimodular, then it is generalized $n$-semimodular.*

The necessary proofs will be given in Sect. 8.3.

As will be seen from the proof of Theorem 8.1.1, the condition that the function $\Lambda$ be generalized $(n\colon 2)$-semimodular can be relaxed to the following: for each $j \in \overline{1, n-1}$ and each $f = (f_1, \dots, f_n) \in L^n$ *such that* $f_1 \le \dots \le f_j$, one has $L(f_1, \dots, f_n) \bowtie L(f_1, \dots, f_{j-1}, f_j \wedge f_{j+1}, f_j \vee f_{j+1}, f_{j+2}, \dots, f_n)$.

*Remark 8.1.2* Theorem 8.1.1 will not hold in general if the lattice $L$ is not assumed to be distributive. For instance, let $L$ be defined by the set $[5] = \{1, 2, 3, 4, 5\}$ with the partial order being the subset of the natural order $\le$ on the set [5] with elements 2, 3, 4 now considered non-comparable with one another, so that the resulting order relation is the set $\{(f, f)\colon f \in [5]\} \cup \{(1, 2), (1, 3), (1, 4), (2, 5), (3, 5), (4, 5), (1, 5)\}$; then, in particular, $2 \wedge 3 = 1$ and $2 \vee 3 = 5$. This lattice is one of the simplest examples of non-distributive lattices. It is isomorphic to the diamond lattice $\mathsf{M}_3$—see e.g. [10, p. 110]. Let $n = 3$, $\mathcal{R} = \mathbb{R}$, and define the function $\Lambda\colon L^3 \to \mathbb{R}$ by the formula $\Lambda(f_1, f_2, f_3) := 12 f_1 f_2 + 3 f_2 f_3 + 5 f_1 f_3$ for all $f = (f_1, f_2, f_3) \in L^3$. Then one can verify directly—by a straightforward but tedious calculation consisting in checking $2 \times 5^3 = 250$ inequalities, two inequalities for each $f = (f_1, f_2, f_3) \in [5]^3$— that this function $\Lambda$ is generalized $(3\colon 2)$-submodular. However, $\Lambda$ is not generalized 3-submodular, because for $f = (2, 3, 4)$ one has $(f_{3:1}, f_{3:2}, f_{3:3}) = (1, 5, 5)$ and $\Lambda(f_1, f_2, f_3) = \Lambda(2, 3, 4) = 148 \not\ge 160 = \Lambda(1, 5, 5) = \Lambda(f_{3:1}, f_{3:2}, f_{3:3})$. □

*Remark 8.1.3* A well-known fact, which will be crucial in the proof of Theorem 8.1.1, is the representation theorem due to Birkhoff and Stone stating that any distributive lattice $L$ is isomorphic to a lattice of subsets of (and hence to a lattice of nonnegative real-valued functions on) a certain set $S$, depending on $L$ (see e.g. [10, Theorem 119]). For such a lattice of functions, the "order statistics" $f_{n:1}, \dots, f_{n:n}$

are uniquely determined by the condition that

$$f_{n:1}(s) \le \cdots \le f_{n:n}(s) \quad \text{and} \quad \{\{f_{n:1}(s), \dots, f_{n:n}(s)\}\} = \{\{f_1(s), \dots, f_n(s)\}\} \tag{8.3}$$

for each $s \in S$, where the double braces are used to denote multisets, with appropriate multiplicities. To quickly see why this is true, one may reason as follows: Let us now use condition (8.3) to *define* $f_{n:1}, \dots, f_{n:n}$. Note that the value of the right-hand side (rhs) of (8.2) at any point $s \in S$ is invariant with respect to all permutations of the values $f_1(s), \dots, f_n(s)$. So, the value of the rhs of (8.2) at $s$ will not change if one replaces there $f_1, \dots, f_n$ by $f_{n:1}, \dots, f_{n:n}$, and this value will equal $f_{n:j}(s)$. Thus, the definition of $f_{n:1}, \dots, f_{n:n}$ by means of formula (8.3) is equivalent to the one given by (8.2), if the lattice $L$ is already a lattice of real-valued functions on $S$. Moreover, it is clear now that, if the lattice $L$ is distributive, then definition (8.2) can be rewritten in the dual form, as

$$f_{n:j} = \bigvee \Big\{ \bigwedge_{i \in J} f_i \colon J \in \binom{[n]}{n+1-j} \Big\} \tag{8.4}$$

for all $j \in [n]$.

On the other hand, it can be seen that, if $L$ is not distributive, then this duality can be lost and each of the definitions (8.2) and (8.4) of $f_{n:j}$ can be rather unnaturally skewed up or down. For instance, in the counterexample given in Remark 8.1.2, for $f = (2, 3, 4)$ we had $(f_{3:1}, f_{3:2}, f_{3:3}) = (1, 5, 5)$ according to definition (8.2), but we would have $(f_{3:1}, f_{3:2}, f_{3:3}) = (1, 1, 5)$ according to (8.4).

However, one may note that the right-hand side of (8.4) is always $\le$ than that of (8.2); this follows because for any $J \in \binom{[n]}{n+1-j}$ and any $K \in \binom{[n]}{j}$ there is some $k \in J \cap K$, and then $\bigwedge_{i \in J} f_i \le f_k \le \bigvee_{i \in K} f_i$. □

In view of the lattice representation theorem cited in Remark 8.1.3, Theorem 8.1.1 may be considered a multi/infinite-dimensional analogue of the well-known Muirhead lemma in the theory of Schur majorization (cf. e.g. [17, Lemma 2.B.1, p. 32]), which may be stated as follows: for vectors $x$ and $y$ in $\mathbb{R}^n$ such that $x \prec y$ (that is, $x$ is majorized by $y$), there exist finitely many vectors $x_0, \dots, x_m$ in $\mathbb{R}^n$ such that $x = x_0 \prec \cdots \prec x_m = y$ and for each $j \in \overline{0, m-1}$ the vectors $x_j$ and $x_{j+1}$ differ only in two coordinates. However, no direct multi-dimensional extension of the Muirhead lemma seems to exist, even in two dimensions (see e.g. [20, p. 11]).

For functions that are "infinite-dimensional" counterparts of the "$m$-dimensional" function $\Lambda \colon L^m \to \mathbb{R}$ given by the formula of the additive form

$$\Lambda(g_1, \dots, g_m) = \sum_{j=1}^{m} \lambda_j(g_j), \tag{8.5}$$

Lorentz [16] obtained a result similar to Theorem 8.1.1; for readers' convenience, let us reproduce it here: For each $j \in [n]$, let $f_j^*$ denote the equimeasurable decreasing rearrangement [11] of a function $f_j\colon (0,1) \to \mathbb{R}$. Let a real-valued expression $\Phi(x, u_1, \dots, u_n)$ be continuous in $(x, u_1, \dots, u_n) \in (0,1) \times [0,\infty) \times \cdots \times [0,\infty)$. Then the inequality

$$\int_0^1 \Phi(x, f_1(x), \dots, f_n(x))\,dx \le \int_0^1 \Phi(x, f_1^*(x), \dots, f_n^*(x))\,dx \tag{8.6}$$

holds for all bounded positive measurable functions $f_1, \dots, f_n$ from $(0,1)$ to $\mathbb{R}$ if and only if the following two conditions hold:

$$\Phi(u_i + h, u_j + h) - \Phi(u_i + h, u_j) - \Phi(u_i, u_j + h) + \Phi(u_i, u_j) \ge 0 \tag{8.7}$$

and

$$\int_0^\delta \big[\Phi(x-t, u_i+h) - \Phi(x-t, u_i) - \Phi(x+t, u_i+h) + \Phi(x+t, u_i)\big]\,dt \ge 0 \tag{8.8}$$

for all $h > 0$, $x \in (0,1)$, $\delta \in (0, x \wedge (1-x))$, $(u_1, \dots, u_n) \in [0,\infty)^n$, and $i, j$ in $[n]$ such that $i < j$; here, in each of inequalities (8.7) and (8.8), the arguments of $\Phi$ that are the same for all the four instances of $\Phi$ are omitted, for brevity.

To establish the connection between Lorentz's result and our Theorem 8.1.1, suppose e.g. that each of the functions $f_1, \dots, f_n$ in [16] is a step function, constant on each of the intervals $(\frac{j-1}{m}, \frac{j}{m}]$ for $j \in [m]$, and then let $g_j(s) := f_s(\frac{j}{m})$ for $j \in [m]$ and $s \in S := [n]$. In fact, in the proof in [16] the result is first established for such step functions $f_1, \dots, f_n$. It is also shown in [16] that, for such "infinite-dimensional" counterparts of the functions given by the "additive" formula (8.5), the sufficient condition is also necessary. In turn, as pointed out in [16], the result there generalizes an inequality in [23]. Another proof of a special case of the result in [16] was given in [5].

## 8.2 Illustrations and Applications

### *8.2.1 A General Construction of Generalized n-Submodular Functions from Submodular Ones*

Recall here some basics of majorization theory [17]. For $x = (x_1, \dots, x_n)$ and $y = (y_1, \dots, y_n)$ in $\mathbb{R}^n$, write $x \prec y$ if $x_1 + \cdots + x_n = y_1 + \cdots + y_n$ and $x_{n:1} + \cdots + x_{n:k} \ge y_{n:1} + \cdots + y_{n:k}$ for all $k \in [n]$. For any $D \subseteq \mathbb{R}^n$, a function $F\colon D \to \mathbb{R}$ is called Schur-concave if for any $x$ and $y$ in $D$ such that $x \prec y$ one has $F(x) \ge F(y)$. If $D = I^n$ for some open interval $I \subseteq \mathbb{R}$ and the function

$F$ is continuously differentiable then, by Schur's theorem [17, Theorem A.4], $F$ is Schur-concave iff $(\frac{\partial F}{\partial x_i} - \frac{\partial F}{\partial x_j})(x_i - x_j) \le 0$ for all $x = (x_1, \ldots, x_n) \in D$.

**Proposition 8.2.1** *Suppose that a real-valued function $\lambda$ defined on a distributive lattice $L$ is submodular and nondecreasing, and a function $\mathbb{R}^n \ni x = (x_1, \ldots, x_n) \to F(x_1, \ldots, x_n)$ is nondecreasing in each of its $n$ arguments and Schur-concave. Then the function $\Lambda = \Lambda_{\lambda,F} : L^n \to \mathbb{R}$ defined by the formula*

$$\Lambda(f_1, \ldots, f_n) := \Lambda_{\lambda,F}(f_1, \ldots, f_n) := F(\lambda(f_1), \ldots, \lambda(f_n)) \tag{8.9}$$

*for $(f_1, \ldots, f_n) \in L^n$ is generalized $(n\colon 2)$-submodular and hence generalized $n$-submodular.*

A rather general construction of submodular functions on rings of sets is provided by [6, §23.2], which implies that $\cup$-homomorpisms preserve the property of being alternating of a given order, and the proposition at the end of [6, §23.1], which describes general $\cup$-homomorpisms as maps of the form

$$S \supseteq A \mapsto G(A) := \{t \in T : (s, t) \in G \text{ for some } s \in A\},$$

where $S$ and $T$ are sets and $G \subseteq S \times T$; in the case when $G$ is (the graph of) a map, the above notation $G(A)$ is of course consistent with that for the image of a set $A$ under the map $G$; according to the definition in the beginning of [6, §23], a $\cup$-homomorpism is a map $\varphi$ of set rings defined by the condition $\varphi(A \cup B) = \varphi(A) \cup \varphi(B)$ for all relevant sets $A$ and $B$.

Therefore and because an additive function on a ring of sets is modular and hence submodular, we conclude that functions of the form

$$A \mapsto \mu(G(A)) \tag{8.10}$$

are submodular, where $\mu$ is a measure or, more generally, an additive function (say on a discrete set, to avoid matters of measurability). From this observation, one can immediately obtain any number of corollaries of Proposition 8.2.1 such as the following:

**Corollary 8.2.2** *Let $P$ be a polytope of dimension $d$. For each $\alpha \in \overline{0, d}$, let $\mathcal{F}_\alpha$ denote the set of all $\alpha$-faces (that is, faces of dimension $\alpha$) of $P$. For any distinct $\alpha, \beta, \gamma$ in $\overline{0, d}$, let $G = G_{\alpha,\beta,\gamma}$ be the set of all pairs $\big(f_\alpha, (f_\beta, f_\gamma)\big) \in \mathcal{F}_\alpha \times (\mathcal{F}_\beta \times \mathcal{F}_\gamma)$ such that $f_\alpha \cap f_\beta \ne \emptyset$, $f_\alpha \cap f_\gamma \ne \emptyset$, and $f_\beta \cap f_\gamma \ne \emptyset$. Let $L$ be a lattice of subsets of $\mathcal{F}_\alpha$. Let a function $\mathbb{R}^n \ni x = (x_1, \ldots, x_n) \to F(x_1, \ldots, x_n)$ be nondecreasing in each of its $n$ arguments and Schur-concave. Then the function $\Lambda = \Lambda_{\alpha,\beta,\gamma} : L^n \to \mathbb{R}$ defined by the formula*

$$\Lambda(A_1, \ldots, A_n) := F(\operatorname{card} G(A_1), \ldots, \operatorname{card} G(A_n))$$

*for $(A_1, \ldots, A_n) \in L^n$ is generalized $(n\colon 2)$-submodular and hence generalized $n$-submodular.*

For readers' convenience, here is a direct verification of the fact that maps of the form (8.10) are submodular: noting that $G(A \cup B) = G(A) \cup G(B)$ and $G(A \cap B) \subseteq G(A) \cap G(B)$ and using the additivity of $\mu$, we have

$$\mu(G(A \cup B)) + \mu(G(A \cap B)) \le \mu(G(A) \cup G(B)) + \mu(G(A) \cap G(B)) = \mu(G(A)) + \mu(G(B))$$

for all relevant sets $A$ and $B$.

### 8.2.2 Generalized One-Sided Potential

Let here $L$ be the lattice of all measurable real-valued functions on a measure space $(S, \Sigma, \mu)$, with the pointwise lattice operations $\vee$ and $\wedge$. Consider the function $\Lambda \colon L^n \to \mathcal{R}$ given by the formula

$$\Lambda(f_1, \ldots, f_n) := \Lambda_{\varphi,\psi}(f_1, \ldots, f_n) := \sum_{j,k=1}^{n} \Psi(f_j - f_k) \tag{8.11}$$

for all $f = (f_1, \ldots, f_n) \in L^n$, where

$$\Psi(g) := \psi\Big(\int_S (\varphi \circ g)\, \mathrm{d}\mu\Big) \tag{8.12}$$

for all $g \in L$, $\varphi \colon \mathbb{R} \to [0, \infty]$ is a nondecreasing or nonincreasing function, and $\psi \colon [0, \infty] \to (-\infty, \infty]$ is a concave function. Thus, the function $\Lambda = \Lambda_{\varphi,\psi}$ may be referred to as a generalized one-sided potential, since the function $\varphi$ is assumed to be monotonic.

**Proposition 8.2.3** *The function $\Lambda = \Lambda_{\varphi,\psi}$ defined by formula (8.11) is generalized $(n\colon 2)$-submodular and hence generalized $n$-submodular.*

### 8.2.3 Symmetric Sums of Nonnegative Multiadditive Functions

Let $k$ be a natural number. Let $L$ be a sublattice of the lattice $\mathbb{R}^S$ of all real-valued functions on a set $S$. Let us say that the lattice $L$ is *complementable* if $f \setminus g := f - f \wedge g \in L$ for any $f$ and $g$ in $L$, so that $f = f \wedge g + f \setminus g$. Assuming that $L$ is complementable, let us say that a function $m \colon L \to \mathbb{R}$ is *additive* if

$$m(f) = m(f \wedge g) + m(f \setminus g)$$

for all $f$ and $g$ in $L$; further, let us say that a function $m \colon L^k \to \mathbb{R}$ is *multiadditive* or, more specifically, *$k$-additive* if $m$ is additive in each of its $k$ arguments, that is, if for each $j \in [k]$ and each $(k-1)$-tuple $(f_i : i \in [k] \setminus \{j\})$ the function $L \ni f_j \mapsto m(f_1, \dots, f_k)$ is additive.

To state the main result of this subsection, we shall need the following notation: for any set $J$, let $\Pi_k^J$ denote the set of all $k$-permutations of $J$, that is, the set of all injective maps of the set $[k]$ to $J$.

**Proposition 8.2.4** *Suppose that $k$ and $n$ are natural numbers such that $k \le n$, $L$ is a complementable sublattice of $\mathbb{R}^S$, and $m : L^k \to \mathbb{R}$ is a nonnegative multiadditive function. Then the function $\Lambda_m : L^n \to \mathbb{R}$ defined by the formula*

$$\Lambda_m(f_1, \dots, f_n) := \sum_{\pi \in \Pi_k^{[n]}} m(f_{\pi(1)}, \dots, f_{\pi(k)}) \tag{8.13}$$

*for $(f_1, \dots, f_n) \in L^n$ is generalized $(n:2)$-submodular and hence generalized $n$-submodular.*

Formula (8.13) can be rewritten in the following symmetrized form:

$$\Lambda_m(f_1, \dots, f_n) = k! \sum_{I \in \binom{[n]}{k}} \overline{m}(f_I), \tag{8.14}$$

where, for $I = \{i_1, \dots, i_k\}$ with $1 \le i_1 < \dots < i_k \le n$,

$$\overline{m}(f_I) := \overline{m}(f_{i_1}, \dots, f_{i_k}) := \frac{1}{k!} \sum_{\pi \in \Pi_k^I} m(f_{\pi(1)}, \dots, f_{\pi(k)}); \tag{8.15}$$

note that the so-defined function $\overline{m} : L^k \to \mathbb{R}$ is multiadditive and nonnegative, given that $m$ is so. Also, $\overline{m}$ is permutation-symmetric in the sense that $\overline{m}(f_{\pi(1)}, \dots, f_{\pi(k)}) = \overline{m}(f_1, \dots, f_k)$ for all $(f_1, \dots, f_k) \in L^k$ and all permutations $\pi \in \Pi_k^{[k]}$.

*Example 8.2.5* If $V$ is a vector sublattice of the lattice $\mathbb{R}^S$ and $L$ is the lattice of all nonnegative functions in $V$ then, clearly, $L$ is complementable and the restriction to $L^k$ of any multilinear function from $V^k$ to $\mathbb{R}$ is multiadditive.

In particular, if $\mu$ is a measure on a $\sigma$-algebra $\Sigma$ over $S$, $V$ is a vector sublattice of $L^k(S, \Sigma, \mu)$, and $L$ is the lattice of all nonnegative functions in $V$, then the function $m : L^k \to \mathbb{R}$ given by the formula

$$m(f_1, \dots, f_k) := \int_S f_1 \cdots f_k \, d\mu$$

for $(f_1, \dots, f_k) \in L^k$ is multiadditive.

So, by Proposition 8.2.4, the functions $\Lambda_m$ corresponding to the functions $m$ presented above in this example are generalized $(n:2)$-submodular and hence generalized $n$-submodular.

Let now $B = (b_{i,j})$ be a $d \times p$ matrix with $d \le p$ and nonnegative entries $b_{i,j}$. The permanent of $B$ is defined by the formula

$$\operatorname{perm} B := \sum_{J \in \binom{[p]}{d}} \operatorname{perm} B_{\cdot J},$$

where $B_{\cdot J}$ the square submatrix of $B$ consisting of the columns of $B$ with column indices in the set $J \in \binom{[p]}{d}$; and for a square $d \times d$ matrix $B = (b_{i,j})$,

$$\operatorname{perm} B := \sum_{\pi \in \Pi_d^{[d]}} b_{1,\pi(1)} \cdots b_{d,\pi(d)}.$$

So, perm $B$ is a multilinear function of the $d$-tuple $(b_{1,\cdot}, \dots, b_{d,\cdot})$ of the rows of $B$. Also, if $d = p$, then perm $B$ is a multilinear function of the $d$-tuple $(b_{\cdot,1}, \dots, b_{\cdot,d})$ of the columns of $B$. If $d \ge p$, then perm $B$ may be defined by the requirement that the permanent be invariant with respect to transposition.

Thus, from Proposition 8.2.4 we immediately obtain

**Corollary 8.2.6** *Assuming that the entries $b_{i,j}$ of the $d \times p$ matrix $B$ are nonnegative,* perm $B$ *is a generalized $d$-submodular function of the $d$-tuple $(b_{1,\cdot}, \dots, b_{d,\cdot})$ of its rows and a generalized $p$-submodular function of the $p$-tuple $(b_{\cdot,1}, \dots, b_{\cdot,p})$ of its columns (with respect to the standard lattice structures on $\mathbb{R}^{1\times p}$ and $\mathbb{R}^{d\times 1}$, respectively):*

$$\operatorname{perm}\begin{pmatrix} b_{d:1,\cdot} \\ \vdots \\ b_{d:d,\cdot} \end{pmatrix} \le \operatorname{perm}\begin{pmatrix} b_{1,\cdot} \\ \vdots \\ b_{d,\cdot} \end{pmatrix} [= \operatorname{perm} B],$$

$$\operatorname{perm}(b_{\cdot,p:1}, \dots, b_{\cdot,p:}) \le \operatorname{perm}(b_{\cdot,1}, \dots, b_{\cdot,p})[= \operatorname{perm} B].$$

Note that the condition $d \le p$ is not needed or assumed in Corollary 8.2.6.

Yet another way in which multilinear and hence multiadditive functions may arise is via the elementary symmetric polynomials. Let $n$ be any natural number, and let $k \in [n]$. The elementary symmetric polynomials are defined by the formula

$$e_k(x_1, \dots, x_n) := \sum_{J \in \binom{[n]}{k}} \prod_{j \in J} x_j.$$

In particular, $e_1(x_1, \dots, x_n) := \sum_{j\in[n]} x_j$ and $e_n(x_1, \dots, x_n) := \prod_{j\in[n]} x_j$.

Let $f = (f_1, \dots, f_n)$ be the vector of measurable functions $f_1, \dots, f_n$ defined on a measure space $(S, \Sigma, \mu)$ with values in the interval $[0, \infty)$. Then it is not hard to see that the "order statistics" are nonnegative measurable functions as well. As usual, let $\mu(h) := \int_S h \, \mathrm{d}\mu$.

If the measure $\mu$ is a probability measure, then the functions $f_1, \dots, f_n$ are called random variables (r.v.'s) and, in this case, $f_{n:1}, \dots, f_{n:n}$ will indeed be what is commonly referred to as the order statistics based on the "random sample" $f = (f_1, \dots, f_n)$; cf. e.g. [7]. In contrast with settings common in statistics, in general we do not impose any conditions on the joint or individual distributions of the r.v.'s $f_1, \dots, f_n$—except that these r.v.'s be nonnegative.

Then we have the following.

**Corollary 8.2.7**

$$e_k\big(\mu(f_1), \dots, \mu(f_n)\big) \geq e_k\big(\mu(f_{n:1}), \dots, \mu(f_{n:n})\big). \tag{8.16}$$

*In particular,*

$$\mu(f_1) \cdots \mu(f_n) \geq \mu(f_{n:1}) \cdots \mu(f_{n:n}). \tag{8.17}$$

This follows immediately from Proposition 8.2.4 and formula (8.14), since the product $\mu(f_1) \cdots \mu(f_k)$ is clearly multilinear and hence multiadditive in $(f_1, \dots, f_k)$.

To deal with cases when some of the $\mu(f_j)$'s (or the $\mu(f_{n:j})$'s) equal 0 and other ones equal $\infty$, let us assume here the convention $0 \cdot \infty := 0$. One may note that, if the nonnegative functions $f_1, \dots, f_n$ are scalar multiples of one another or, more generally, if $f_{\pi(1)} \leq \cdots \leq f_{\pi(n)}$ for some permutation $\pi$ of the set $[n]$, then inequality (8.16) turns into the equality.

As mentioned above, in Corollary 8.2.7 it is not assumed that $f_1, \dots, f_n$ are independent r.v.'s. However, if $\mu$ is a probability measure and the r.v.'s $f_1, \dots, f_n$ are independent (but not necessarily identically distributed), then $\mu(f_1) \cdots \mu(f_n) = \mu(f_1 \cdots f_n) = \mu(f_{n:1} \cdots f_{n:n})$ by the second part of (8.3), and so, (8.17) can then be rewritten as the following positive-association-type inequality for the order statistics:

$$\mu(f_{n:1} \cdots f_{n:n}) \geq \mu(f_{n:1}) \cdots \mu(f_{n:n}). \tag{8.18}$$

Let now $\psi$ be any monotone (that is, either nondecreasing or nonincreasing) function from $[0, \infty]$ to $[0, \infty]$. For $f = (f_1, \dots, f_n)$ as before, let

$$\psi \bullet f := (\psi \circ f_1, \dots, \psi \circ f_n).$$

Then for $j \in [n]$ one has $(\psi \bullet f)_{n:j} = \psi \circ f_{n:j}$ if $\psi$ is nondecreasing and $(\psi \bullet f)_{n:j} = \psi \circ f_{n:n+1-j}$ if $\psi$ is nonincreasing. Thus, we have the following ostensibly more general forms of (8.17) and (8.18):

**Corollary 8.2.8**

$$\mu(\psi \circ f_1) \cdots \mu(\psi \circ f_n) \geq \mu\big((\psi \bullet f)_{n:1}\big) \cdots \mu\big((\psi \bullet f)_{n:n}\big). \tag{8.19}$$

*If* $\mu$ *is a probability measure and the r.v.'s* $f_1, \dots, f_n$ *are independent, then*

$$\mu\big((\psi \bullet f)_{n:1} \cdots (\psi \bullet f)_{n:n}\big) \geq \mu\big((\psi \bullet f)_{n:1}\big) \cdots \mu\big((\psi \bullet f)_{n:n}\big). \tag{8.20}$$

The property of the order statistics $f_{n:1}, \cdots, f_{n:n}$ given by inequality (8.20) may be called the diagonal positive orthant dependence—cf. e.g. Definition 2.3 in [12] of the negative orthant dependence.

Immediately from Theorem 8.1.1 or from inequality (8.19) in Corollary 8.2.8, one obtains

**Corollary 8.2.9** *Take any* $p \in \mathbb{R} \setminus \{0\}$. *Then*

$$\mu(f_1^p)^r \cdots \mu(f_n^p)^r \geq \mu(f_{n:1}^p)^r \cdots \mu(f_{n:n}^p)^r \tag{8.21}$$

*for any* $r \in (0, \infty)$, *and*

$$\mu(f_1^p)^r \cdots \mu(f_n^p)^r \leq \mu(f_{n:1}^p)^r \cdots \mu(f_{n:n}^p)^r \tag{8.22}$$

*for any* $r \in (-\infty, 0)$. *Here we use the conventions* $0^t := \infty$ *and* $\infty^t := 0$ *for* $t \in (-\infty, 0)$. *We also the following conventions:* $0 \cdot \infty := 0$ *concerning* (8.21) *and* $0 \cdot \infty := \infty$ *concerning* (8.22).

Consider now the special case of Corollary 8.2.9 with $r = 1/p$. Letting then $p \to \infty$, we see that (8.21) will hold with the $\mu(f_j^p)^r$'s and $\mu(f_{n:j}^p)^r$'s replaced there by $\mu$-ess sup $f_j$ and $\mu$-ess sup $f_{n:j}$, respectively, where $\mu$-ess sup denotes the essential supremum with respect to measure $\mu$. This follows because $\mu(h^p)^{1/p} \underset{p\to\infty}{\longrightarrow} \mu\text{-ess}\sup h$. Similarly, letting $p \to -\infty$, we see that (8.22) will hold with the $\mu(f_j^p)^r$'s and $\mu(f_{n:j}^p)^r$'s replaced there by $\mu$-ess inf $f_j$ and $\mu$-ess inf $f_{n:j}$, respectively, where $\mu$-ess inf denotes the essential infimum with respect to $\mu$. Moreover, considering (say) the counting measures $\mu$ on finite subsets of the set $S$ and noting that $\sup h = \sup_S h$ coincides with the limit of the net $(\max_J h)$ over the filter of all finite subsets $J$ of $S$, we conclude that (8.21) will hold with the $\mu(f_j^p)^r$'s and $\mu(f_{n:j}^p)^r$'s replaced there by $\sup f_j$ and $\sup f_{n:j}$, respectively. (The statement about the limit can be spelled out as follows: $\sup_S h \geq \max_J h$ for all finite $J \subseteq S$, and for each real $c$ such that $c < \sup h$ there is some finite set $J_c \subseteq S$ such that for all finite sets $J$ such that $J_c \subseteq J \subseteq S$ one has $\max_J h > c$.) Similarly, (8.22) will hold with the $\mu(f_j^p)^r$'s and $\mu(f_{n:j}^p)^r$'s replaced there by $\inf f_j$ and $\inf f_{n:j}$, respectively. Thus, we have

**Corollary 8.2.10**

$$(\sup f_1) \cdots (\sup f_n) \geq (\sup f_{n:1}) \cdots (\sup f_{n:n}) \tag{8.23}$$

*and*

$$(\inf f_1) \cdots (\inf f_n) \leq (\inf f_{n:1}) \cdots (\inf f_{n:n}). \tag{8.24}$$

*Here we use the following conventions:* $0\cdot\infty := 0$ *concerning* (8.23) *and* $0\cdot\infty := \infty$ *concerning* (8.24).

Alternatively, one can obtain (8.23) and (8.24) directly from Theorem 8.1.1.

Also, of course there is no need to assume in Corollary 8.2.10 that the functions $f_1, \dots, f_n$ are measurable.

The special cases of inequalities (8.22) and (8.24) for $n = 2$ mean that the functions $h \mapsto \mu(h^p)^r$ and $h \mapsto \inf h$ are log-supermodular functions on the distributive lattice (say $\mathcal{L}_\Sigma$) of all nonnegative $\Sigma$-measurable functions on $S$ and on the distributive lattice (say $\mathcal{L}$) of all nonnegative functions on $S$, respectively.

At this point, let us recall the famous Fortuin–Kasteleyn–Ginibre (FKG) correlation inequality [8], which states that for any log-supermodular function $\nu$ on a finite distributive lattice $L$ and any nondecreasing functions $F$ and $G$ on $L$ we have

$$\nu(FG)\nu(1) \geq \nu(F)\nu(G),$$

where $\nu(F) := \sum_{f\in L} \nu(f)$.

Then we immediately obtain

**Corollary 8.2.11** *Let* $\mathcal{L}_\Sigma^\circ$ *be any finite sub-lattice of the lattice* $\mathcal{L}_\Sigma$*, and let* $F$ *and* $G$ *be nondecreasing functions from* $\mathcal{L}_\Sigma^\circ$ *to* $\mathbb{R}$*. Then*

$$\Big(\sum_{h\in\mathcal{L}_\Sigma^\circ} F(h)G(h)\mu(h)^r\Big)\Big(\sum_{h\in\mathcal{L}_\Sigma^\circ} \mu(h)^r\Big) \geq \Big(\sum_{h\in\mathcal{L}_\Sigma^\circ} F(h)\mu(h)^r\Big)\Big(\sum_{h\in\mathcal{L}_\Sigma^\circ} G(h)\mu(h)^r\Big)$$

*for any* $r \in (-\infty, 0)$*. Similarly, let* $\mathcal{L}^\circ$ *be any finite sub-lattice of the lattice* $\mathcal{L}$*, and let* $F$ *and* $G$ *be nondecreasing functions from* $\mathcal{L}^\circ$ *to* $\mathbb{R}$*. Then*

$$\Big(\sum_{h\in\mathcal{L}_\Sigma^\circ} F(h)G(h)\inf h\Big)\Big(\sum_{h\in\mathcal{L}_\Sigma^\circ} \inf h\Big) \geq \Big(\sum_{h\in\mathcal{L}_\Sigma^\circ} F(h)\inf h\Big)\Big(\sum_{h\in\mathcal{L}_\Sigma^\circ} G(h)\inf h\Big).$$

As shown by Ahlswede and Daykin [2, pp. 288–289], their inequality [2, Theorem 1] almost immediately implies, and is in a sense sharper than, the FKG inequality. Furthermore, Rinott and Saks [21, 22] and Aharoni and Keich [1] independently obtained a more general inequality "for $n$-tuples of nonnegative functions on a distributive lattice, of which the Ahlswede–Daykin inequality is the case $n = 2$." More specifically, in notation closer to that used in the present paper, [1, Theorem 1.1] states the following:

Let $\alpha_1, \dots, \alpha_n, \beta_1, \dots, \beta_n$ be nonnegative functions defined on a distributive lattice $L$ such that

$$\prod_{j=1}^n \alpha_j(f_j) \leq \prod_{j=1}^n \beta_j(f_{n:j})$$

for all $f_1, \dots, f_n$ in $L$. Then for any finite subsets $F_1, \dots, F_n$ of $L$

$$\prod_{j=1}^{n} \sum_{f_j \in F_j} \alpha_j(f_j) \le \prod_{j=1}^{n} \sum_{g_j \in F_{n:j}} \beta_j(g_j),$$

where

$$F_{n:j} := \{f_{n:j} : f = (f_1, \dots, f_n) \in F_1 \times \dots \times F_n\}.$$

Note that the definition of the "order statistics" used in [1] is different from (8.2) in that their "order statistics" go in the descending, rather than ascending, order; also, the term "order statistics" is not used in [1].

In view of this result of [1] and our Corollaries 8.2.9 and 8.2.10, one immediately obtains the following statement, which generalizes and strengthens Corollary 8.2.11:

**Corollary 8.2.12** *Let $\mathcal{F}_1, \dots, \mathcal{F}_n$ be any finite subsets of the lattice $\mathcal{L}_\Sigma$. For each $j \in [n]$, let*

$$\mathcal{F}_{n:j} := \{f_{n:j} : f = (f_1, \dots, f_n) \in \mathcal{F}_1 \times \dots \times \mathcal{F}_n\}.$$

*Then*

$$\prod_{j=1}^{n} \sum_{f_j \in \mathcal{F}_j} \mu(f_j)^r \le \prod_{j=1}^{n} \sum_{h_j \in \mathcal{F}_{n:j}} \mu(h_j)^r \tag{8.25}$$

*for any $r \in (-\infty, 0)$.*

*Similarly, let now $\mathcal{F}_1, \dots, \mathcal{F}_n$ be any finite subsets of the lattice $\mathcal{L}$. Then*

$$\prod_{j=1}^{n} \sum_{f_j \in \mathcal{F}_j} \inf f_j \le \prod_{j=1}^{n} \sum_{h_j \in \mathcal{F}_{n:j}} \inf h_j.$$

Comparing inequalities (8.21) and (8.22) in Corollary 8.2.9 or inequalities (8.23) and (8.24) in Corollary 8.2.10, one may wonder whether the FKG-type inequalities stated in Corollaries 8.2.11 and 8.2.12 for the functions $h \mapsto \mu(h)^r$ with $r < 0$ and $h \mapsto \inf h$ admit of the corresponding reverse analogues for the functions $h \mapsto \mu(h)^r$ with $r > 0$ and $h \mapsto \sup h$. However, it is not hard to see that such FKG-type inequalities are not reversible in this sense, a reason being that the sets $\mathcal{F}_{n:j}$ may be much larger than the sets $\mathcal{F}_j$.

E.g., suppose that $n = 2$, $S = \mathbb{R}$, $\mu$ is a Borel probability measure on $\mathbb{R}$, $0 < \varepsilon < \delta < 1$, $N$ is a natural number, $\mathcal{F}_1$ is the set of $N$ pairwise distinct constant functions $f_1, \dots, f_N$ on $\mathbb{R}$ such that $1 - \varepsilon < f_j < 1 + \varepsilon$ for all $j \in [n]$, and $\mathcal{F}_2 = \{g_1, \dots, g_N\}$, where $g_j := (1-\delta)\mathbf{1}_{(-\infty, j]} + (1+\delta)\mathbf{1}_{(j,\infty)}$ and $\mathbf{1}_A$ denotes the indicator of a set $A$. Then it is easy to see that each of the sets $\mathcal{F}_{2:1}$ and $\mathcal{F}_{2:2}$ is of cardinality $N^2$. So, letting $\delta \downarrow 0$ (so that $\varepsilon \downarrow 0$ as well), we see that, for any real $r$, the right-hand side of (8.25) goes to $N^4$ whereas its left-hand side goes to $N^2$, which is much less than $N^4$ if $N$ is large.

*Example 8.2.13* Closely related to Example 8.2.5 is as follows. Suppose that $(S, \Sigma)$ is a measurable space, $\mu$ is a measure on the product $\sigma$-algebra $\Sigma^{\otimes k}$, and $L$ is a subring of $\Sigma$. Then $L$ is complementable and the function $m\colon L^k \to \mathbb{R}$ given by the formula

$$m(A_1, \dots, A_k) := \mu(A_1 \times \dots \times A_k) \tag{8.26}$$

for $(A_1, \dots, A_k) \in L^k$ is multiadditive.

A particular case of formula (8.26) is

$$m(A_1, \dots, A_k) := \operatorname{card}\big(G \cap (A_1 \times \dots \times A_k)\big), \tag{8.27}$$

where card stands for the cardinality and $G$ is an arbitrary subset of $S^k$. If $G$ is symmetric in the sense that $(s_1, \dots, s_k) \in G$ iff $(s_{\pi(1)}, \dots, s_{\pi(k)}) \in G$ for all permutations $\pi$ of the set $[k]$, then $G$ represents the set (say $E$) of all hyperedges of a $k$-uniform hypergraph over $S$, in the sense that $(s_1, \dots, s_k) \in G$ iff $\{s_1, \dots, s_k\} \in E$.

We now have another immediate corollary of Proposition 8.2.4:

**Corollary 8.2.14** *Suppose that $k$ and $n$ are natural numbers such that $k \le n$, $(S, \Sigma)$ is a measurable space, $\mu$ is a measure on the product $\sigma$-algebra $\Sigma^{\otimes k}$, and $L$ is a subring of $\Sigma$. Then*

$$\sum_{\pi \in \Pi_k^{[n]}} \mu(A_{n:\pi(1)} \times \dots \times A_{n:\pi(k)}) \le \sum_{\pi \in \Pi_k^{[n]}} \mu(A_{\pi(1)} \times \dots \times A_{\pi(k)}) \tag{8.28}$$

*for all $(A_1, \dots, A_n) \in L^n$.*

## 8.3 Proofs

One may note that formula (8.31) in the proof of Theorem 8.1.1 below defines a step similar to a step in the process of the so-called insertion search (cf. e.g. [15, Section 5.2.1] (also called the sifting or sinking technique)—except that here we do the pointwise comparison of functions (rather than numbers) and therefore we do not stop when the right place of the value $f_{n+1}(s)$ of the "new" function $f_{n+1}$ among the already ordered values $f_{n:1}(s), \dots, f_{n:n}(s)$ at a particular point $s \in S$ has been found, because this place will in general depend on $s$. So, the proof that (8.31) implies (8.34) may be considered as (something a bit more than) a rigorous proof of the validity of the insertion search algorithm, avoiding such informal, undefined terms as swap, moving, and interleaving.

*Proof of Theorem 8.1.1* Let us prove the theorem by induction in $n$. For $n = 1$, the result is trivial. To make the induction step, it suffices to prove the following: For any natural $n \geq 2$, if the function $\Lambda\colon L^n \to \mathcal{R}$ is generalized $(n\colon 2)$-semimodular and the function $L^{n-1} \ni (f_1, \dots, f_{n-1}) \mapsto \Lambda(f_1, \dots, f_n)$ is generalized $(n-1)$-semimodular for each $f_n \in L$, then $\Lambda$ is generalized $n$-semimodular. Thus, we are assuming that the function $\Lambda\colon L^n \to \mathcal{R}$ is generalized $(n\colon 2)$-semimodular and

$$\Lambda(f_1, \dots, f_n) \bowtie \Lambda(f_{n-1:1}, \dots, f_{n-1:n-1}, f_n) \tag{8.29}$$

for all $(f_1, \dots, f_n) \in L^n$, where $f_{n-1:1}, \dots, f_{n-1:n-1}$ are the "order statistics" based on $(f_1, \dots, f_{n-1})$.

Take indeed any $(f_1, \dots, f_n) \in L^n$. Define the rectangular array of functions $(g_{k,j}\colon k \in \overline{0, n-1},\ j \in [n])$ recursively, as follows:

$$(g_{0,1}, \dots, g_{0,n-1}, g_{0,n}) := (f_{n-1:1}, \dots, f_{n-1:n-1}, f_n) \tag{8.30}$$

and, for $k \in \overline{1, n-1}$ and $j \in [n]$,

$$g_{k,j} := \begin{cases} g_{k-1,j} & \text{if } \ j \in \overline{1, n-k-1} \cup \overline{n-k+2, n}, \\ g_{k-1,n-k} \wedge g_{k-1,n-k+1} & \text{if } \ j = n-k, \\ g_{k-1,n-k} \vee g_{k-1,n-k+1} & \text{if } \ j = n-k+1. \end{cases} \tag{8.31}$$

By (8.29) and (8.30),

$$\Lambda(f_1, \dots, f_n) \bowtie \Lambda(g_{0,1}, \dots, g_{0,n-1}, g_{0,n}). \tag{8.32}$$

Moreover, for each $k \in \overline{1, n-1}$,

$$\Lambda(g_{k-1,1}, \dots, g_{k-1,n}) \bowtie \Lambda(g_{k,1}, \dots, g_{k,n}), \tag{8.33}$$

since $\Lambda$ is generalized $(n\colon 2)$-semimodular.

It follows from (8.32) and (8.33) that

$$\Lambda(f_1, \dots, f_n) \bowtie \Lambda(g_{n-1,1}, \dots, g_{n-1,n}).$$

It remains to verify the identity

$$(g_{n-1,1}, \dots, g_{n-1,n}) \overset{(?)}{=} (f_{n:1}, \dots, f_{n:n}). \tag{8.34}$$

In accordance with Remark 8.1.3, we may and shall assume that the distributive lattice $L$ is a lattice of nonnegative real-valued functions on a set $S$, so that (8.3) holds for each $s \in S$.

In the remainder of the proof, fix any $s \in S$. Then

$$\{\{g_{0,1}(s), \dots, g_{0,n}(s)\}\} = \{\{f_1(s), \dots, f_n(s)\}\},$$

by (8.30) and the second part of (8.3) used with $n-1$ in place of $n$; also, for each $k \in \overline{1, n-1}$,

$$\{\{g_{k,1}(s), \dots, g_{k,n}(s)\} = \{\{g_{k-1,1}(s), \dots, g_{k-1,n}(s)\}\},$$

by (8.31). So,

$$\{\{g_{n-1,1}(s), \dots, g_{n-1,n}(s)\}\} = \{\{f_1(s), \dots, f_n(s)\}\}.$$

Therefore, to complete the proof of (8.34) and thus that of Theorem 8.1.1, it remains to show that

$$g_{n-1,1}(s) \overset{(?)}{\le} \cdots \overset{(?)}{\le} g_{n-1,n}(s), \tag{8.35}$$

which will follow immediately from

**Lemma 8.3.1** *For each $k \in \overline{1, n-1}$, the following assertion is true for all $s \in S$:*

$$\begin{aligned} &g_{k,j}(s) \le g_{k,j+1}(s) \textit{ for all } j \in \overline{1, n-k-2} \cup \overline{n-k, n-1}; \\ &\textit{also, } g_{k,n-k-1}(s) \le g_{k,n-k+1}(s) \textit{ if } k \le n-2. \end{aligned} \tag{$A_k$}$$

Indeed, (8.35) is the first clause in assertion $(A_k)$ with $k = n-1$. Thus, what finally remains to prove Theorem 8.1.1 is to present the following.

*Proof of Lemma 8.3.1* For simplicity, let us be dropping $(s)$—thus writing $g_{k,j}, f_n, \dots$ in place of $g_{k,j}(s), f_n(s), \dots$. We shall prove Lemma 8.3.1 by induction in $k \in \overline{1, n-1}$. Assertion $(A_1)$ means that $g_{1,1} \le \cdots \le g_{1,n-2}$, $g_{1,n-1} \le g_{1,n}$, and $g_{1,n-2} \le g_{1,n}$ if $1 \le n-2$. So, in view of (8.31) and (8.30), $(A_1)$ can be rewritten as follows: $f_{n-1:1} \le \cdots \le f_{n-1:n-2}$, $f_{n-1:n-1} \wedge f_n \le f_{n-1:n-1} \vee f_n$, and $f_{n-1:n-2} \le f_{n-1:n-1} \vee f_n$; all these inequalities are obvious. So, $(A_1)$ holds.

Take now any $k \in \overline{2, n-1}$ and suppose that $(A_{k-1})$ holds. We need to show that then $(A_k)$ holds.

For all $j \in \overline{1, n-k-2} \cup \overline{n-k+2, n-1}$, we have $j+1 \in \overline{1, n-k-1} \cup \overline{n-k+2, n}$, whence, by (8.31) and the first clause of $(A_{k-1})$, $g_{k,j} = g_{k-1,j} \le g_{k-1,j+1} = g_{k,j+1}$. So,

$$g_{k,j} \le g_{k,j+1} \quad \text{for } j \in \overline{1, n-k-2} \cup \overline{n-k+2, n-1}. \tag{8.36}$$

If $j = n-k$ then, by (8.31), $g_{k,j} = g_{k-1,n-k} \wedge g_{k-1,n-k+1} \le g_{k-1,n-k} \vee g_{k-1,n-k+1} = g_{k,j+1}$.

If $j = n - k + 1$ then the condition $k \in \overline{2, n-1}$ implies $j \le n - 1$, and so, by (8.31) and the second and first clauses of $(A_{k-1})$, $g_{k,j} = g_{k-1,n-k} \vee g_{k-1,n-k+1} \le g_{k-1,n-k+2} = g_{k-1,j+1} = g_{k,j+1}$.

Thus, in view of (8.36), the first clause of $(A_k)$ holds. Also, if $k \le n - 2$ then, by (8.31) and the first clause of $(A_{k-1})$, $g_{k,n-k-1} = g_{k-1,n-k-1} \le g_{k-1,n-k} \le g_{k-1,n-k} \vee g_{k-1,n-k+1} = g_{k,n-k+1}$, so that the second clause of $(A_k)$ holds as well. This completes the proof of Lemma 8.3.1. □

Thus, Theorem 8.1.1 is proved. □

*Proof of Proposition 8.2.1* Take any $(f_1, \dots, f_n) \in L^n$. Corollary B.3 in [17] states that $x \prec y$ iff $x$ is in the convex hull of the set of all points obtained by permuting the coordinates of the vector $y$. Also, since the function $\lambda$ is nondecreasing, we have $\lambda(f_1 \vee f_2) \ge \lambda(f_1) \vee \lambda(f_2)$. For any real $a, b, c$ such that $c \ge a \vee b$, we have $(a, b) = (1-t)(a+b-c, c) + t(c, a+b-c)$ for $t = \frac{c-b}{2c-a-b} \in [0, 1]$ if $c > (a+b)/2$ and for any $t \in [0, 1]$ otherwise (that is, if $a = b = c$). So, the point $(a, b)$ is a convex combination of points $(a+b-c, c)$ and $(c, a+b-c)$. Using this fact for $a = \lambda(f_1)$, $b = \lambda(f_2)$, $c = \lambda(f_1 \vee f_2)$, we see that

$$(\lambda(f_1), \dots, \lambda(f_n)) \prec (\lambda(f_1) + \lambda(f_2) - \lambda(f_1 \vee f_2), \lambda(f_1 \vee f_2), \lambda(f_3), \dots, \lambda(f_n)).$$

Also, $\lambda(f_1 \wedge f_2) \le \lambda(f_1) + \lambda(f_2) - \lambda(f_1 \vee f_2)$, by the submodularity of $\lambda$. Therefore and because $F$ is nondecreasing (in each of its $n$ arguments) and Schur-concave, we conclude that

$$\begin{aligned} &F(\lambda(f_1 \wedge f_2), \lambda(f_1 \vee f_2), \lambda(f_3), \dots, \lambda(f_n)) \\ &\qquad \le F(\lambda(f_1) + \lambda(f_2) - \lambda(f_1 \vee f_2), \lambda(f_1 \vee f_2), \lambda(f_3), \dots, \lambda(f_n)) \\ &\qquad\qquad \le F(\lambda(f_1), \dots, \lambda(f_n)). \end{aligned}$$

Quite similarly,

$$\begin{aligned} &F(\lambda(f_1), \dots, \lambda(f_{i-1}), \lambda(f_i \wedge f_{i+1}), \lambda(f_i \vee f_{i+1}), \lambda(f_{i+2}), \dots, \lambda(f_n)) \\ &\qquad\qquad \le F(\lambda(f_1), \dots, \lambda(f_n)) \end{aligned}$$

for all $i \in \overline{1, n-1}$, so that the function $F$ is indeed generalized $(n:2)$-submodular and hence, by Theorem 8.1.1, generalized $n$-submodular. □

*Proof of Proposition 8.2.3* In view of Theorem 8.1.1, it is enough to show that the function $\Lambda = \Lambda_{\varphi,\psi}$ is generalized $(n:2)$-submodular. Without loss of generality (w.l.o.g.), we may and shall assume that the function $\varphi$ is nondecreasing, since $\Lambda_{\varphi^-,\psi} = \Lambda_{\varphi,\psi}$, where $\varphi^-(u) := \varphi(-u)$ for all real $u$. Also, w.l.o.g. $\psi(0) = 0$ and hence $\Psi(0) = 0$.

Take any $f = (f_1, \dots, f_n) \in L^n$. Then, letting

$$\tilde{\Psi}(g) := \Psi(g) + \Psi(-g) \tag{8.37}$$

for $g \in L$, one has

$$\Lambda(f_1, f_2, f_3, \dots, f_n) = \tilde{\Psi}(f_1 - f_2) + \sum_{j=3}^{n} \big(\tilde{\Psi}(g_j) + \tilde{\Psi}(h_j)\big) + R, \tag{8.38}$$

where $g_j := f_1 - f_j$, $h_j := f_2 - f_j$, and $R := \sum_{3\le j<k\le n}^{n} \tilde{\Psi}(f_j - f_k)$. Since $f_1 \wedge f_2 - f_1 \vee f_2 = -|f_1 - f_2|$, one similarly has

$$\Lambda(f_1 \wedge f_2, f_1 \vee f_2, f_3, \dots, f_n) = \tilde{\Psi}(|f_1 - f_2|) + \sum_{j=3}^{n} \big(\tilde{\Psi}(g_j \wedge h_j) + \tilde{\Psi}(g_j \vee h_j)\big) + R. \tag{8.39}$$

Next,

$$\tilde{\Psi}(f_1 - f_2) = \psi\Big(\int_S \varphi \circ (f_1 - f_2)\,\mathrm{d}\mu\Big) + \psi\Big(\int_S \varphi \circ (f_2 - f_1)\,\mathrm{d}\mu\Big), \tag{8.40}$$

$$\tilde{\Psi}(|f_1 - f_2|) = \psi\Big(\int_S \varphi \circ |f_1 - f_2|\,\mathrm{d}\mu\Big) + \psi\Big(\int_S \varphi \circ (-|f_2 - f_1|)\,\mathrm{d}\mu\Big), \tag{8.41}$$

$\varphi \circ (f_1 - f_2) + \varphi \circ (f_2 - f_1) = \varphi \circ |f_1 - f_2| + \varphi \circ (-|f_1 - f_2|)$ and hence

$$\int_S \varphi \circ (f_1 - f_2)\,\mathrm{d}\mu + \int_S \varphi \circ (f_2 - f_1)\,\mathrm{d}\mu = \int_S \varphi \circ |f_1 - f_2|\,\mathrm{d}\mu + \int_S \varphi \circ (-|f_2 - f_1|)\,\mathrm{d}\mu. \tag{8.42}$$

Also, since $\varphi$ is nondecreasing, $\varphi \circ (f_1 - f_2) \vee \varphi \circ (f_2 - f_1) \le \varphi \circ |f_1 - f_2|$ and hence

$$\int_S \varphi \circ (f_1 - f_2)\,\mathrm{d}\mu \;\vee\; \int_S \varphi \circ (f_2 - f_1)\,\mathrm{d}\mu \le \int_S \varphi \circ |f_1 - f_2|\,\mathrm{d}\mu. \tag{8.43}$$

Since the function $\psi$ is convex, it follows from (8.40)–(8.43) that

$$\tilde{\Psi}(f_1 - f_2) \le \tilde{\Psi}(|f_1 - f_2|). \tag{8.44}$$

Further, take any $j \in \overline{3, n}$. Then $\varphi \circ g_j + \varphi \circ h_j = \varphi \circ (g_j \wedge h_j) + \varphi \circ (g_j \vee h_j)$. So,

$$\int_S (\varphi \circ g_j)\,\mathrm{d}\mu + \int_S (\varphi \circ h_j)\,\mathrm{d}\mu = \int_S \varphi \circ (g_j \wedge h_j)\,\mathrm{d}\mu + \int_S \varphi \circ (g_j \vee h_j)\,\mathrm{d}\mu.$$

Moreover, since $\varphi$ is nondecreasing, $\int_S \varphi \circ (g_j \vee h_j)\,\mathrm{d}\mu$ is no less than each of the integrals $\int_S (\varphi \circ g_j)\,\mathrm{d}\mu$ and $\int_S (\varphi \circ h_j)\,\mathrm{d}\mu$. So, in view of (8.12) and the convexity of the function $\psi$, one has $\Psi(g_j) + \Psi(h_j) \le \Psi(g_j \wedge h_j) + \Psi(g_j \vee h_j)$. Similarly, because $\int_S \varphi \circ (-(g_j \wedge h_j))\,\mathrm{d}\,\mu$ is no less than each of the integrals $\int_S \varphi \circ (-g_j)\,\mathrm{d}\,\mu$ and $\int_S \varphi \circ (-h_j)\,\mathrm{d}\,\mu$, one has $\Psi(-g_j)+\Psi(-h_j) \le \Psi(-(g_j \wedge h_j))+\Psi(-(g_j \vee h_j))$. So, by (8.37), $\tilde{\Psi}(g_j) + \tilde{\Psi}(h_j) \le \tilde{\Psi}(g_j \wedge h_j) + \tilde{\Psi}(g_j \vee h_j)$.

Therefore, by (8.38), (8.39), and (8.44), $\Lambda(f_1, f_2, f_3, \dots, f_n) \le \Lambda(f_1 \wedge f_2, f_1 \vee f_2, f_3, \dots, f_n)$. Similarly, $\Lambda(f_1, \dots, f_{j-1}, f_j, f_{j+1}, f_{j+2}, \dots, f_n) \le \Lambda(f_1, \dots, f_{j-1}, f_j \wedge f_{j+1}, f_j \vee f_{j+1}, f_{j+2}, \dots, f_n)$ for all $j \in \overline{1, n-1}$.

Thus, the function $\Lambda$ is generalized $(n\colon 2)$-supermodular, and so, by Theorem 8.1.1, it is generalized $n$-supermodular. □

*Proof of Proposition 8.2.4* Fix any $(f_1, \dots, f_n) \in L^n$. Then, in view of the permutation symmetry of $\overline{m}$ defined by (8.15),

$$\frac{1}{k!}\,\Lambda_m(f_1, \dots, f_n) = \lambda_2(f_{n-1}, f_n) + \lambda_1(f_{n-1}) + \lambda_1(f_n) + \lambda_0, \tag{8.45}$$

where

$$\lambda_2(f, g) := \sum_{1 \le i_1 < \cdots < i_{k-2} \le n-2} \overline{m}(f_{i_1}, \dots, f_{i_{k-2}}, f, g),$$

$$\lambda_1(f) := \sum_{1 \le i_1 < \cdots < i_{k-1} \le n-2} \overline{m}(f_{i_1}, \dots, f_{i_{k-1}}, f),$$

$$\lambda_0 := \sum_{1 \le i_1 < \cdots < i_k \le n-2} \overline{m}(f_{i_1}, \dots, f_{i_k}),$$

Similarly,

$$\frac{1}{k!}\,\Lambda_m(f_1, \dots, f_{n-2}, f_{n-1} \wedge f_n, f_{n-1} \vee f_n) = \lambda_2(f_{n-1} \wedge f_n, f_{n-1} \vee f_n) + \lambda_1(f_{n-1} \wedge f_n) + \lambda_1(f_{n-1} \vee f_n) + \lambda_0. \tag{8.46}$$

Note that the function $\lambda_2\colon L^2 \to \mathbb{R}$ is 2-additive and permutation-symmetric, and the function $\lambda_1\colon L^2 \to \mathbb{R}$ is additive. Take any $f$ and $g$ in $L$. Then $(f \vee g) \wedge f = f$ and $(f \vee g) \setminus f = g \setminus f$. So, by the additivity of $\lambda_1$ we have $\lambda_1(f \vee g) = \lambda_1(f) + \lambda_1(g \setminus f)$, whereas $\lambda_1(f \wedge g) + \lambda_1(g \setminus f) = \lambda_1(g)$. So,

$$\lambda_1(f \wedge g) + \lambda_1(f \vee g) = \lambda_1(f \wedge g) + \lambda_1(f) + \lambda_1(g \setminus f) = \lambda_1(f) + \lambda_1(g). \tag{8.47}$$

By the 2-additivity and permutation symmetry of $\lambda_2$ and because the function $\lambda_2$ is 2-additive, permutation-symmetric, and nonnegative, we have

$$\begin{aligned}\lambda_2(f \wedge g, f \vee g) &= \lambda_2(f \wedge g, f \setminus g) + \lambda_2(f \wedge g, g)\\ &= \lambda_2(f \wedge g, f \setminus g) + \lambda_2(f, g) - \lambda_2(f \setminus g, g)\\ &= \lambda_2(f \wedge g, f \setminus g) + \lambda_2(f, g) - \lambda_2(f \setminus g, g \wedge f) - \lambda_2(f \setminus g, g \setminus f)\\ &= \lambda_2(f, g) - \lambda_2(f \setminus g, g \setminus f)\\ &\le \lambda_2(f, g).\end{aligned} \tag{8.48}$$

It follows from (8.45), (8.46), (8.47), and (8.48) (with $f = f_{n-1}$ and $g = f_n$) that

$$\Lambda_m(f_1, \ldots, f_{n-2}, f_{n-1} \wedge f_n, f_{n-1} \vee f_n) \le \Lambda_m(f_1, \ldots, f_n).$$

Therefore, being permutation-symmetric, the function $\Lambda_m$ is indeed generalized $(n:2)$-submodular. Hence, by Theorem 8.1.1, $\Lambda_m$ is generalized $n$-submodular. □

## References


1. R. Aharoni, U. Keich, A generalization of the Ahlswede-Daykin inequality. Discrete Math. **152**(1–3), 1–12 (1996)
2. R. Ahlswede, D.E. Daykin, Inequalities for a pair of maps $S \times S \to S$ with $S$ a finite set. Math. Z. **165**(3), 267–289 (1979)
3. F. Bach, Learning with submodular functions: a convex optimization perspective. Found. Trends® Mach. Learn. **6**(2–3), 145–373 (2013)
4. F. Bach, Submodular functions: from discrete to continuous domains. Math. Program. **175**(1–2), 419–459 (2019)
5. C. Borell, A note on an inequality for rearrangements. Pac. J. Math. **47**, 39–41 (1973)
6. G. Choquet, Theory of capacities. Ann. Inst. Fourier **5**, 131–295 (1954)
7. H.A. David, H.N. Nagaraja, *Order Statistics*. Wiley Series in Probability and Statistics, 3rd edn. (Wiley, Hoboken, 2003)
8. C.M. Fortuin, P.W. Kasteleyn, J. Ginibre, Correlation inequalities on some partially ordered sets. Commun. Math. Phys. **22**, 89–103 (1971)
9. S. Fujishige, *Submodular Functions and Optimization*. Annals of Discrete Mathematics, vol. 47 (North-Holland, Amsterdam, 1991)
10. G. Grätzer, *Lattice Theory: Foundation* (Birkhäuser, Basel, 2011)
11. G.H. Hardy, J.E. Littlewood, G. Pólya, *Inequalities*, 2nd edn. (Cambridge University Press, Cambridge, 1952)
12. K. Joag-Dev, F. Proschan, Negative association of random variables, with applications. Ann. Stat. **11**(1), 286–295 (1983)
13. S. Karlin, Y. Rinott, Classes of orderings of measures and related correlation inequalities. I. Multivariate totally positive distributions. J. Multivariate Anal. **10**(4), 467–498 (1980)
14. S. Karlin, Y. Rinott, Classes of orderings of measures and related correlation inequalities. II. Multivariate reverse rule distributions. J. Multivariate Anal. **10**(4), 499–516 (1980)

15. D.E. Knuth, *The Art of Computer Programming*, vol. 3, 2nd edn. (Addison-Wesley, Reading, 1998). Sorting and searching [MR0445948]
16. G.G. Lorentz, An inequality for rearrangements. Am. Math. Mon. **60**, 176–179 (1953)
17. A.W. Marshall, I. Olkin, B.C. Arnold, *Inequalities: Theory of Majorization and Its Applications*. Springer Series in Statistics, 2nd edn. (Springer, New York, 2011)
18. P. Milgrom, J. Roberts, Rationalizability, learning, and equilibrium in games with strategic complementarities. Econometrica **58**(6), 1255–1277 (1990)
19. H. Narayanan, *Submodular Functions and Electrical Networks*. Annals of Discrete Mathematics, vol. 54 (North-Holland, Amsterdam, 1997)
20. I. Pinelis, Optimal binomial, Poisson, and normal left-tail domination for sums of nonnegative random variables. Electron. J. Probab. **21**(20), 19 (2016)
21. Y. Rinott, M. Saks, On FKG-type and permanental inequalities, in *Stochastic Inequalities (Seattle, WA, 1991)*. IMS Lecture Notes Monograph Series, vol. 22 (Institute of Mathematical Statistics, Hayward, 1992), pp. 332–342
22. Y. Rinott, M. Saks, Correlation inequalities and a conjecture for permanents. Combinatorica **13**(3), 269–277 (1993)
23. H.D. Ruderman, Two new inequalities. Am. Math. Mon. **59**, 29–32 (1952)
24. D.M. Topkis, Equilibrium points in nonzero-sum $n$-person submodular games. SIAM J. Control Optim. **17**(6), 773–787 (1979)
25. D.M. Topkis, *Supermodularity and Complementarity*. Frontiers of Economic Research (Princeton University Press, Princeton, 1998).

# Chapter 9
# Geometry of $\ell_p^n$-Balls: Classical Results and Recent Developments

**Joscha Prochno, Christoph Thäle, and Nicola Turchi**

**Abstract** In this article we first review some by-now classical results about the geometry of $\ell_p$-balls $\mathbb{B}_p^n$ in $\mathbb{R}^n$ and provide modern probabilistic arguments for them. We also present some more recent developments including a central limit theorem and a large deviations principle for the $q$-norm of a random point in $\mathbb{B}_p^n$. We discuss their relation to the classical results and give hints to various extensions that are available in the existing literature.

**Keywords** Asymptotic geometric analysis · $\ell_p^n$-Balls · Central limit theorem · Law of large numbers · Large deviations · Polar integration formula

**2010 Mathematics Subject Classification** 46B06, 47B10, 60B20, 60F10

## 9.1 Introduction

The geometry of the classical $\ell_p$ sequence spaces and their finite-dimensional versions is nowadays quite well understood. It has turned out that it is often a probabilistic point of view that shed (new) light on various geometric aspects and characteristics of these spaces and, in particular, their unit balls. In this survey we want to take a fresh look at some of the classical results and also on some more recent developments. The probabilistic approach to study the geometry of $\ell_p^n$-balls will be an asymptotic one. In particular, our aim is to demonstrate the usage of various limit theorems from probability theory, such as laws of large numbers, central limit theorems or large deviation principles. While the law of large numbers

J. Prochno (✉)
Institut für Mathematik und Wissenschaftliches Rechnen, Karl-Franzens-Universität Graz, Graz, Austria
e-mail: joscha.prochno@uni-graz.at

C. Thäle · N. Turchi
Fakultät für Mathematik, Ruhr-Universität Bochum, Bochum, Germany
e-mail: christoph.thaele@rub.de; nicola.turchi@rub.de

 
N. Gozlan et al. (eds.), *High Dimensional Probability VIII*,
Progress in Probability 74, https://doi.org/10.1007/978-3-030-26391-1_9

and the central limit theorem are already part of the—by now—classical theory (see, e.g., [21, 23, 24]), the latter approach via large deviation principles was introduced only recently in the theory of asymptotic geometric analysis by Gantert et al. in [9]. Most of the results we present below are not new and we shall always give precise references to the original papers. On the other hand, we provide detailed arguments at those places where we present generalizations of existing results that cannot be found somewhere else. For some of the other results the arguments are occasionally sketched as well.

Our text is structured as follows. In Sect. 9.2 we collect some preliminary material. In particular, we introduce our notation (Sect. 9.2.1), the class of $\ell_p^n$-balls (Sect. 9.2.2), and also rephrase some background material on Grassmannian manifolds (Sect. 9.2.3) and large deviation theory (Sect. 9.2.4). In Sect. 9.3 we introduce a number of probability measures that can be considered in connection with a convex body. We do this for the case of $\ell_p^n$-balls (Sect. 9.3.1), but also more generally for symmetric convex bodies (Sect. 9.3.2). The usage of the central limit theorem and the law of large numbers in the context of $\ell_p^n$-balls is demonstrated in Sect. 9.4. We rephrase there some more classical results of Schechtman and Schmuckenschläger (Sect. 9.4.1) and also consider some more recent developments (Sect. 9.4.2) including applications of the multivariate central limit theorem. We also take there an outlook to the matrix-valued set-up. The final Sect. 9.5 is concerned with various aspects of large deviations. We start with the classical concentration inequalities of Schechtman and Zinn (Sect. 9.5.1) and then describe large deviation principles for random projections of $\ell_p^n$-balls (Sect. 9.5.2).

## 9.2 Preliminaries

In this section we shall provide the basics from both asymptotic geometric analysis and probability theory that are used throughout this survey article. The reader may also consult [3, 5–7, 14] for detailed expositions and additional explanations when necessary.

### *9.2.1 Notation*

We shall denote with $\mathbb{N} = \{1, 2, \dots\}$, $\mathbb{R}$ and $\mathbb{R}^+$ the set of natural, real and real non-negative numbers, respectively. Given $n \in \mathbb{N}$, let $\mathbb{R}^n$ be the $n$-dimensional vector space on the real numbers, equipped with the standard inner product denoted by $\langle \cdot, \cdot \rangle$. We write $\mathcal{B}(\mathbb{R}^n)$ for the $\sigma$-field of all Borel subsets of $\mathbb{R}^n$. Analogously, for a subset $S \subseteq \mathbb{R}^n$, we denote by $\mathcal{B}(S) := \{A \cap S : A \in \mathcal{B}(\mathbb{R}^n)\}$ the corresponding trace $\sigma$-field of $\mathcal{B}(\mathbb{R}^n)$. Given a set $A$, we write $\#A$ for its cardinality. For a set $A \subseteq \mathbb{R}^n$, we shall write $\mathbf{1}_A : \mathbb{R}^n \to \{0, 1\}$ for the indicator function of $A$. Given $A \in \mathcal{B}(\mathbb{R}^n)$, we write $|A|$ for its $n$-dimensional Lebesgue measure and frequently refer to this as the volume of $A$.

Given sets $I \subseteq \mathbb{R}^+$ and $A \subseteq \mathbb{R}^n$, we define the set $IA$ as follows,

$$IA := \{rx \in \mathbb{R}^n : r \in I, x \in A\}.$$

If $I = \{r\}$, we also write $rA$ instead of $\{r\}A$. Note that $\mathbb{R}^+A$ is usually called the cone spanned by $A$.

We say that $K \subseteq \mathbb{R}^n$ is a convex body if it is a convex, compact set with non-empty interior. We indicate with $\partial K$ its boundary.

Fix now a probability space $(\Omega, \mathcal{F}, \mathbf{P})$. We will always assume that our random variables live in this probability space. Given a random variable $X\colon \Omega \to \mathbb{R}^n$ and a probability measure $Q$ on $\mathbb{R}^n$, we write $X \sim Q$ to indicate that $Q$ is the probability distribution of $X$, namely, for any $A \in \mathcal{B}(\mathbb{R}^n)$,

$$\mathbf{P}(X \in A) = \int_{\mathbb{R}^n} \mathbf{1}_A(x)\, \mathrm{d}Q(x).$$

We write $\mathbf{E}$ and $\mathbf{Var}$ to denote the expectation and the variance with respect to the probability $\mathbf{P}$, respectively.

Given a sequence of random variables $(X_n)_{n\in\mathbb{N}}$ and a random variable $Y$ we write

$$X_n \xrightarrow[n\to\infty]{d} Y, \qquad X_n \xrightarrow[n\to\infty]{\mathbf{P}} Y, \qquad X_n \xrightarrow[n\to\infty]{\text{a.s.}} Y,$$

to indicate that $(X_n)_{n\in\mathbb{N}}$ converges to $Y$ in distribution, probability or almost surely, respectively, as $n \to \infty$.

We write $N \sim \mathcal{N}(0, \Sigma)$ and say that $N$ is a centred Gaussian random vector in $\mathbb{R}^n$ with covariance matrix $\Sigma$, i.e., its density function w.r.t. the Lebesgue measure is given by

$$f(x) = \frac{1}{\sqrt{(2\pi)^n \det \Sigma}} \exp\Big(-\frac{1}{2}\big\langle x, \Sigma^{-1}x\big\rangle\Big), \qquad x \in \mathbb{R}^n.$$

For $\alpha, \theta > 0$, we write $X \sim \Gamma(\alpha, \vartheta)$ (resp. $X \sim \beta(\alpha, \vartheta)$) and say that $X$ has a Gamma distribution (resp. a Beta distribution) with parameters $\alpha$ and $\vartheta$ if the probability density function of $X$ w.r.t. to the Lebesgue measure is proportional to $x \mapsto x^{\alpha-1}e^{-\vartheta x}\mathbf{1}_{[0,\infty)}(x)$ (resp. $x \mapsto x^{\alpha-1}(1-x)^{\vartheta-1}\mathbf{1}_{[0,1]}(x)$). We also say that $X$ has a uniform distribution on $[0, 1]$ if $X \sim \mathrm{Unif}([0, 1]) := \beta(1, 1)$ or an exponential distribution with parameter 1 if $X \sim \exp(\vartheta) := \Gamma(1, \vartheta)$.

The following properties of the aforementioned distributions are of interest and easy to verify by direct computation:

$$\text{if } X \sim \Gamma(\alpha, \vartheta) \text{ and } Y \sim \Gamma(\tilde{\alpha}, \vartheta) \text{ are independent, then } \frac{X}{X+Y} \sim \beta(\alpha, \tilde{\alpha}), \tag{9.1}$$

$$\text{if } X \sim \mathrm{Unif}([0,1]), \text{ then } X^k \sim \beta(1/k, 1), \tag{9.2}$$

for any $\alpha, \tilde{\alpha}, \vartheta, k \in (0, \infty)$.

Given a real sequence $(a_n)_{n\in\mathbb{N}}$, we write $a_n \equiv a$ if $a_n = a$ for every $n \in \mathbb{N}$. If $(b_n)_{n\in\mathbb{N}}$ is a positive sequence, we write $a_n = \mathcal{O}(b_n)$ if there exists $C \in (0, \infty)$ such that $|a_n| \leq C b_n$ for every $n \in \mathbb{N}$, and $a_n = o(b_n)$ if $\lim_{n\to\infty}(a_n/b_n) = 0$.

### 9.2.2 The $\ell_p^n$-Balls

For $n \in \mathbb{N}$, let $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ and define the $p$-norm of $x$ via

$$\|x\|_p := \begin{cases} \Big(\sum_{i=1}^n |x_i|^p\Big)^{1/p} & \text{if } p \in [1, \infty), \\ \max_{1\leq i\leq n} |x_i| & \text{if } p = \infty. \end{cases}$$

The unit ball $\mathbb{B}_p^n$ and sphere $\mathbb{S}_p^{n-1}$ with respect to this norm are defined as

$$\mathbb{B}_p^n := \{x \in \mathbb{R}^n : \|x\|_p \leq 1\} \qquad \text{and} \qquad \mathbb{S}_p^{n-1} := \{x \in \mathbb{R}^n : \|x\|_p = 1\} = \partial\mathbb{B}_p^n.$$

As usual, we shall write $\ell_p^n$ for the Banach space $(\mathbb{R}^n, \|\cdot\|_p)$. The exact value of $|\mathbb{B}_p^n|$ is known since Dirichlet [8] and is given by

$$|\mathbb{B}_p^n| = \frac{(2\Gamma(1+1/p))^n}{\Gamma(1+n/p)}.$$

The interested reader may consult [19] for a modern computation. The volume-normalized ball shall be denoted by $\mathbb{D}_p^n$ and is given by

$$\mathbb{D}_p^n = \frac{\mathbb{B}_p^n}{|\mathbb{B}_p^n|^{1/n}}.$$

For convenience, in what follows we will use the convention that in the case $p = \infty$, $1/p := 0$. It is worth noticing that the restriction on the domain of $p$ is due to the fact that an analogous definition of $\|\cdot\|_p$ for $p < 1$ does only result in a quasi-norm, meaning that the triangle inequality does not hold. As a consequence, $\mathbb{B}_p^n$ is convex if and only if $p \geq 1$. Although a priori many arguments of this survey do not rely on $\|\cdot\|_p$ being a norm, we restrict our presentation to the case $p \geq 1$, since it is necessary in some of the theorems.

### *9.2.3 Grassmannian Manifolds*

The group of $(n \times n)$-orthogonal matrices is denoted by $\mathbb{O}(n)$ and we let $\mathbb{SO}(n)$ be the subgroup of orthogonal $n \times n$ matrices with determinant 1. As subsets of $\mathbb{R}^{n^2}$, $\mathbb{O}(n)$ and $\mathbb{SO}(n)$ can be equipped with the trace $\sigma$-field of $\mathcal{B}(\mathbb{R}^{n^2})$. Moreover, both compact groups $\mathbb{O}(n)$ and $\mathbb{SO}(n)$ carry a unique Haar probability measure which we denote by $\eta$ and $\tilde{\eta}$, respectively. Since $\mathbb{O}(n)$ consists of two copies of $\mathbb{SO}(n)$, the measure $\eta$ can easily be derived from $\tilde{\eta}$ and vice versa. Given $k \in \{0, 1, \dots, n\}$, we use the symbol $\mathbb{G}_k^n$ to denote the Grassmannian of $k$-dimensional linear subspaces of $\mathbb{R}^n$. We supply $\mathbb{G}_k^n$ with the metric

$$d(E, F) := \max\Big\{ \sup_{x\in B_E} \inf_{y\in B_F} \|x - y\|_2, \sup_{y\in B_F} \inf_{x\in B_E} \|x - y\|_2 \Big\}, \qquad E, F \in \mathbb{G}_k^n,$$

where $B_E$ and $B_F$ stand for the Euclidean unit balls in $E$ and $F$, respectively. The Borel $\sigma$-field on $\mathbb{G}_k^n$ induced by this metric is denoted by $\mathcal{B}(\mathbb{G}_k^n)$ and we supply the arising measurable space $\mathbb{G}_k^n$ with the unique Haar probability measure $\eta_k^n$. It can be identified with the image measure of the Haar probability measure $\tilde{\eta}$ on $\mathbb{SO}(n)$ under the mapping $\mathbb{SO}(n) \to \mathbb{G}_k^n$, $T \mapsto TE_0$ with $E_0 := \mathrm{span}(\{e_1, \dots, e_k\})$. Here, we write $e_1 := (1, 0, \dots, 0)$, $e_2 := (0, 1, 0, \dots, 0)$, $\dots$, $e_n := (0, \dots, 0, 1) \in \mathbb{R}^n$ for the standard orthonormal basis in $\mathbb{R}^n$ and $\mathrm{span}(\{e_1, \dots, e_k\}) \in \mathbb{G}_k^n$, $k \in \{1, \dots, n\}$, for the $k$-dimensional linear subspace spanned by the first $k$ vectors of this basis.

### *9.2.4 Large Deviation Principles*

Consider a sequence $(X_n)_{n\in\mathbb{N}}$ of i.i.d. integrable real random variables and let

$$S_n := \frac{1}{n} \sum_{i=1}^{n} X_i$$

be the empirical average of the first $n$ random variables of the sequence. It is well known that the law of large numbers provides the asymptotic behaviour of $S_n$, as $n$ tends to infinity. In particular, the strong law of large numbers says that

$$S_n \xrightarrow[n\to\infty]{\text{a.s.}} \mathbf{E}[X_1].$$

If $X_1$ has also positive and finite variance, then the classical central limit theorem states that the fluctuations of $S_n$ around $\mathbf{E}[X_1]$ are normal and of scale $1/\sqrt{n}$. More precisely,

$$\sqrt{n}(S_n - \mathbf{E}[X_1]) \xrightarrow[n\to\infty]{d} \mathcal{N}(0, \mathbf{Var}[X_1]).$$

One of the important features of the central limit theorem is its universality, i.e., that the limiting distribution is normal independently of the precise distribution of the summands $X_1, X_2, \ldots$. This allows to have a good estimate for probabilities of the kind

$$\mathbf{P}(S_n > x), \qquad x \in \mathbb{R},$$

when $n$ is large, but fixed. However, such estimate can be quite imprecise if $x$ is much larger than $\mathbf{E}[X_1]$. Moreover, it does not provide any rate of convergence for such tail probabilities as $n$ tends to infinity for fixed $x$.

In typical situations, if $S_n$ arises as a sum of $n$ independent random variables $X_1, \ldots, X_n$ with finite exponential moments, say, one has that

$$\mathbf{P}(S_n > x) \approx e^{-n\mathcal{I}(x)}, \qquad x > \mathbf{E}[X_1]$$

if $n \to \infty$, where $\mathcal{I}$ is the so-called rate function. Here $\approx$ expresses an asymptotic equivalence up to sub-exponential functions of $n$. For concreteness, let us consider two examples. If $\mathbf{P}(X_1 = 1) = \mathbf{P}(X_1 = 0) = 1/2$, then

$$\mathcal{I}(x) = \begin{cases} x\log x + (1-x)\log(1-x) + \log 2 & \text{if } x \in [0,1], \\ +\infty & \text{otherwise,} \end{cases}$$

which describes the upper large deviations. If on the other hand $X_1 \sim \mathcal{N}(0, \sigma^2)$, then the rate function is given by

$$\mathcal{I}(x) = \frac{x^2}{2\sigma^2}, \qquad x \in \mathbb{R}.$$

Contrarily to the universality shown in the central limit theorem, these two examples already underline that the function $\mathcal{I}$ and thus the decay of the tail probabilities is much more sensitive and specific to the distribution of $X_1$.

The study of the atypical situations (in contrast to the typical ones described in the laws of large numbers and the central limit theorem) is called Large Deviations Theory. The concept expressed heuristically in the examples above can be made formal in the following way. Let $\mathbf{X} := (X_n)_{n\in\mathbb{N}}$ be a sequence of random vectors taking values in $\mathbb{R}^d$. Further, let $s\colon \mathbb{N} \to [0,\infty]$ be a non-negative sequence such that $s(n) \uparrow \infty$ and assume that $\mathcal{I}\colon \mathbb{R}^d \to [0,\infty]$ is a lower semi-continuous function, i.e., all of its lower level sets $\{x \in \mathbb{R}^d : \mathcal{I}(x) \leq \ell\}$, $\ell \in [0, \infty]$, are closed. We say that $\mathbf{X}$ satisfies a large deviation principle (or simply LDP) with speed $s(n)$ and rate function $\mathcal{I}$ if and only if

$$-\inf_{x \in A^\circ} \mathcal{I}(x) \leq \liminf_{n\to\infty} \frac{1}{s(n)} \log \mathbf{P}(X_n \in A) \leq \limsup_{n\to\infty} \frac{1}{s(n)} \log \mathbf{P}(X_n \in A) \leq -\inf_{x \in \overline{A}} \mathcal{I}(x)$$

for all $A \in \mathcal{B}(\mathbb{R}^d)$. Moreover, $\mathcal{I}$ is said to be a good rate function if all of its lower level sets are compact. The latter property is essential to guarantee the so-called exponential tightness of the sequence of measures.

The following result, known as Cramér's Theorem, guarantees an LDP for the empirical average of a sequence of i.i.d. random vectors, provided that their common distribution is sufficiently nice (see, e.g. [14, Theorem 27.5]).

**Theorem 9.2.1 (Cramér's Theorem)** *Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of i.i.d. random vectors in $\mathbb{R}^d$ such that the cumulant generating function of $X_1$,*

$$\Lambda(u) := \log \mathbf{E}\big[\exp X_1 u\big], \qquad u \in \mathbb{R}^d,$$

*is finite in a neighbourhood of $0 \in \mathbb{R}^d$. Let $\mathbf{S} := (\frac{1}{n}\sum_{i=1}^n X_i)_{n\in\mathbb{N}}$ be the sequence of the sample means. Then $\mathbf{S}$ satisfies an LDP with speed $n$ and good rate function $\mathcal{I} = \Lambda^*$, where*

$$\Lambda^*(x) := \sup_{u\in\mathbb{R}^d} \big(xu - \Lambda(u)\big), \qquad x \in \mathbb{R}^d,$$

*is the Fenchel-Legendre transform of $\Lambda$.*

Cramér's Theorem is a fundamental tool that allows to prove an LDP if the random variables of interest can be transformed into a sum of independent random variables.

Sometimes there is the need to 'transport' a large deviation principle from one space to another by means of a continuous function. This can be done with a device known as the contraction principle and we refer to [6, Theorem 4.2.1] or [14, Theorem 27.11(i)].

**Proposition 9.2.2 (Contraction Principle)** *Let $d_1, d_2 \in \mathbb{N}$ and let $F : \mathbb{R}^{d_1} \to \mathbb{R}^{d_2}$ be a continuous function. Further, let $\mathbf{X} := (X_n)_{n\in\mathbb{N}}$ be a sequence of $\mathbb{R}^{d_1}$-valued random vectors that satisfies an LDP with speed $s(n)$ and rate function $\mathcal{I}_\mathbf{X}$. Then the sequence $\mathbf{Y} := (F(X_n))_{n\in\mathbb{N}}$ of $\mathbb{R}^{d_2}$-valued random vectors satisfies an LDP with the same speed and with good rate function $\mathcal{I}_\mathbf{Y} = \mathcal{I}_\mathbf{X} \circ F^{-1}$, i.e., $\mathcal{I}_\mathbf{Y}(y) := \inf\{\mathcal{I}_\mathbf{X}(x) : F(x) = y\}$, $y \in \mathbb{R}^{d_2}$, with the convention that $\mathcal{I}_\mathbf{Y}(y) = +\infty$ if $F^{-1}(\{y\}) = \emptyset$.*

While this form of the contraction principle is sufficient to analyse the large deviation behavior for one-dimensional random projections of $\ell_p^n$-balls, a refinement to treat the higher-dimensional cases is needed. To handle this situation, the classical contraction principle can be extended to allow a dependency on $n$ of the continuous function $F$. We refer the interested reader to [6, Corollary 4.2.21] for the precise statement.

## 9.3 Probability Measures on Convex Bodies

There is a variety of probability measures that can be defined on the family of $\ell_p^n$-balls or spheres. We shall present some of them and their key properties below.

### 9.3.1 Probability Measures on an $\ell_p^n$-Ball

One can endow $\mathbb{B}_p^n$ with a natural volume probability measure. This is defined as follows,

$$\nu_p^n(A) := \frac{|A \cap \mathbb{B}_p^n|}{|\mathbb{B}_p^n|}, \tag{9.3}$$

for any $A \in \mathcal{B}(\mathbb{R}^n)$. We also refer to $\nu_p^n$ as the uniform distribution on $\mathbb{B}_p^n$.

As far as $\mathbb{S}_p^{n-1}$ is concerned, there are two probability measures that are of particular interest. The first is the so-called surface measure, which we denote by $\sigma_p^n$, and which is defined as the normalised $(n-1)$-dimensional Hausdorff measure. The second, $\mu_p^n$, is the so-called cone (probability) measure and is defined via

$$\mu_p^n(A) := \frac{|[0,1]A|}{|\mathbb{B}_p^n|}, \qquad A \in \mathcal{B}(\mathbb{S}_p^{n-1}). \tag{9.4}$$

In other words, $\mu_p^n(A)$ is the normalised volume of the cone that intersects $\mathbb{S}_p^{n-1}$ in $A$, intersected with $\mathbb{B}_p^n$. The cone measure is known to be the unique measure that satisfies the following polar integration formula for any integrable function $f$ on $\mathbb{R}^n$ (see, e.g., [18, Proposition 1])

$$\int_{\mathbb{R}^n} f(x)\,\mathrm{d}x = n\,|\mathbb{B}_p^n| \int_0^\infty r^{n-1} \int_{\mathbb{S}_p^{n-1}} f(rz)\,\mathrm{d}\mu_p^n(z)\,\mathrm{d}r. \tag{9.5}$$

In particular, whenever $f$ is $p$-radial, i.e., there exists a function $g$ defined on $\mathbb{R}^+$ such that $f(x) = g(\|x\|_p)$, then

$$\int_{\mathbb{R}^n} g(\|x\|_p)\,\mathrm{d}x = n\,|\mathbb{B}_p^n| \int_0^\infty r^{n-1} g(r)\,\mathrm{d}r. \tag{9.6}$$

The relation between $\sigma_p^n$ and $\mu_p^n$ has been deeply investigated. It is known, for example, that they coincide whenever $p \in \{1, 2, \infty\}$ (see, e.g., [20]). In the other cases, Naor [17] provided a bound on the total variation distance of these two measures.

**Proposition 9.3.1** *Let $\sigma_p^n$ and $\mu_p^n$ be the surface probability and cone probability measure on $\mathbb{S}_p^{n-1}$, respectively. Then*

$$d_{\mathrm{TV}}(\sigma_p^n, \mu_p^n) := \sup\left\{|\sigma_p^n(A) - \mu_p^n(A)| : A \in \mathcal{B}(\mathbb{S}_p^{n-1})\right\} \le C\left(1 - \frac{1}{p}\right)\left|1 - \frac{2}{p}\right|\frac{\sqrt{np}}{n+p},$$

*where $C \in (0, \infty)$ is an absolute constant.*

In particular, the above proposition ensures that for $p$ fixed, such a distance decreases to 0 not slower than $n^{-1/2}$.

An important feature of the cone measure is described by the following probabilistic representation, due to Schechtman and Zinn [22] (independently discovered by Rachev and Rüschendorf [20]). We will below present a proof in a more general set-up.

**Theorem 9.3.2** *Let $n \in \mathbb{N}$ and $p \in [1, \infty]$. Let $(Z_i)_{i \in \mathbb{N}}$ be independent and $p$-generalized Gaussian random variables, meaning absolutely continuous w.r.t. to the Lebesgue measure on $\mathbb{R}$ with density*

$$f_p(x) := \begin{cases} \dfrac{1}{2p^{1/p}\Gamma(1+1/p)} e^{-|x|^p/p} & \text{if } p \in [1, \infty), \\ \dfrac{1}{2}\mathbf{1}_{[0,1]}(|x|) & \text{if } p = \infty. \end{cases} \tag{9.7}$$

*Consider the random vector $Z := (Z_1, \dots, Z_n) \in \mathbb{R}^n$ and let $U \sim \mathrm{Unif}([0, 1])$ be independent of $Z_1, \dots, Z_n$. Then*

$$\frac{Z}{\|Z\|_p} \sim \mu_p^n \qquad \text{and} \qquad U^{1/n}\frac{Z}{\|Z\|_p} \sim \nu_p^n.$$

*Moreover, $Z/\|Z\|_p$ is independent of $\|Z\|_p$.*

It is worth noticing that in [22] the density used by the authors for $Z_1$ is actually proportional to $x \mapsto \exp(-|x|^p)$. As will become clear later, this difference is irrelevant as far as the conclusion of the theorem is concerned.

Indeed, although the statement of Theorem 9.3.2 reflects the focus of this survey on the $\ell_p^n$-balls and the literature on the topic, its result is not strictly dependent on the particular choice of $f_p$ in Eq. (9.7). In fact, it is not even a prerogative of the $\ell_p^n$-balls, as subsequently explained in Proposition 9.3.3.

### 9.3.2 The Cone Measure on a Symmetric Convex Body

Consider a symmetric convex body $K \subseteq \mathbb{R}^n$, meaning that if $x \in K$ then also $-x \in K$. Define the functional $\|\cdot\|_K : \mathbb{R}^n \to [0, \infty)$ by

$$\|x\|_K := \inf\{r > 0 : x \in rK\}.$$

The functional $\|\cdot\|_K$ is known as the Minkowski functional associated with $K$ and, under the aforementioned conditions on $K$, defines a norm on $\mathbb{R}^n$. We will also say that $\|x\|_K$ is the $K$-norm of the vector $x \in \mathbb{R}^n$. Whenever a function on $\mathbb{R}^n$ is dependent only on $\|\cdot\|_K$, we say that it is a $K$-radial function. Analogously, we call a probability measure $K$-radial when its distribution function is $K$-radial. We will also write $p$-radial meaning $\mathbb{B}_p^n$-radial.

In analogy with Eqs. (9.3) and (9.4), it is possible to define a uniform probability measure $\nu_K$ on $K$ and a cone measure $\mu_K$ on $\partial K$, respectively, as

$$\nu_K(A) := \frac{|A \cap K|}{|K|} \qquad \text{and} \qquad \mu_K(B) := \frac{|[0,1]B|}{|K|},$$

for any $A \in \mathcal{B}(\mathbb{R}^n)$ and $B \in \mathcal{B}(\partial K)$.

Note that $\mu_K$, as a ratio of volumes, is invariant under a simultaneous transformation of both the numerator and the denominator. In particular, for any $I \in \mathcal{B}(\mathbb{R}^+)$, such that $|I| > 0$, it holds

$$\mu_K(B) = \frac{|IB|}{|I\partial K|}, \tag{9.8}$$

for any $B \in \mathcal{B}(\partial K)$ (note that $K = [0,1]\partial K$). This fact will be used in the proof of the following generalization of Theorem 9.3.2 to arbitrary symmetric convex bodies.

**Proposition 9.3.3** *Let $K \subseteq \mathbb{R}^n$ be a symmetric convex body. Suppose that there exists a continuous function $f : [0, \infty) \to [0, \infty)$ with the property $\int_{\mathbb{R}^n} f(\|x\|_K)\,\mathrm{d}x = 1$ such that the distribution of a random vector $Z$ on $\mathbb{R}^n$ is given by*

$$\mathbf{P}(Z \in A) = \int_A f(\|x\|_K)\,\mathrm{d}x,$$

*for any $A \in \mathcal{B}(\mathbb{R}^n)$. Also, let $U \sim \mathrm{Unif}([0,1])$ be independent of $Z$. Then,*

$$\frac{Z}{\|Z\|_K} \sim \mu_K \qquad \text{and} \qquad U^{1/n}\frac{Z}{\|Z\|_K} \sim \nu_K. \tag{9.9}$$

*In addition, $Z/\|Z\|_K$ is independent of $\|Z\|_K$.*

The proof of Proposition 9.3.3 is based on the following polar integration formula, which generalizes Eq. (9.5). It says that for measurable functions $h : \mathbb{R}^n \to [0, \infty)$,

$$\int_{\mathbb{R}^n} h(x)\,\mathrm{d}x = n|K| \int_0^\infty r^{n-1} \int_{\partial K} h(rz)\;\mathrm{d}\mu_K(z)\,\mathrm{d}r. \tag{9.10}$$

By the usual measure-theoretic standard procedure to prove Eq. (9.10) it is sufficient to consider functions $h$ of the form $h(x) = \mathbf{1}_A(x)$, where $A = (a,b)E$ with $0 < a < b < \infty$ and $E$ a Borel subset of $\partial K$. However, in this case, the left-hand side is just $|A|$, while for the right-hand side we obtain, by definition of the cone measure $\mu_K$,

$$\begin{aligned} n|K| \int_0^\infty r^{n-1}\mathbf{1}_{(a,b)}(r) \int_{\partial K} \mathbf{1}_E(z)\;\mathrm{d}\mu_K(z)\,\mathrm{d}r &= n|K| \int_a^b r^{n-1}\,\mathrm{d}r\, \frac{|[0,1]E|}{|K|} \\ &= (b^n - a^n)|[0,1]E|, \end{aligned}$$

which is clearly also equal to $|A|$.

*Proof of Proposition 9.3.3* Let $\varphi : \mathbb{R}^n \to \mathbb{R}$ and $\psi : \mathbb{R} \to \mathbb{R}$ be non-negative measurable functions. Applying the polar integration formula, Eq. (9.10), yields

$$\begin{aligned} \mathbf{E}\Big[\varphi\Big(\frac{Z}{\|Z\|_K}\Big)\psi(\|Z\|_K)\Big] &= \int_{\mathbb{R}^n} \varphi\Big(\frac{x}{\|x\|_K}\Big)\psi(\|x\|_K) f(\|x\|_K)\,\mathrm{d}x \\ &= n|K| \int_0^\infty \psi(r) f(r) r^{n-1}\,\mathrm{d}r \int_{\partial K} \varphi(z)\,\mathrm{d}\mu_K(z). \end{aligned}$$

By the product structure of the last expression this first shows the independence of $Z/\|Z\|_K$ and $\|Z\|_K$. Moreover, choosing $\psi \equiv 1$ we see that

$$\mathbf{E}\,\varphi\Big(\frac{Z}{\|Z\|_K}\Big) = n|K| \int_0^\infty f(r) r^{n-1}\,\mathrm{d}r \int_{\partial K} \varphi(z)\,\mathrm{d}\mu_K(z) = \int_{\partial K} \varphi(z)\,\mathrm{d}\mu_K(z)$$

by definition of $f$. This proves that $Z/\|Z\|_K \sim \mu_K$. That $U^{1/n}\frac{Z}{\|Z\|_K} \sim \nu_K$ finally follows from the fact that $U^{1/n} \sim \beta(n,1)$, which has density $r \mapsto nr^{n-1}$ for $r \in (0,1)$. □

The main reason why the theory treated in this survey is restricted to $\ell_p^n$-balls, and not to more general convex bodies $K$, is that $\ell_p^n$-balls are a class of convex bodies whose Minkowski functional is of the form

$$\|x\|_K = F\Big(\sum_{i=1}^n f_i(x_i)\Big) \tag{9.11}$$

for certain functions $f_1, \dots, f_n$ and invertible positive function $F$. This is necessary for $Z$ to have independent coordinates. Indeed, in this case one can assign a joint density on $Z$ that factorizes into its components, like for example (omitting the normalizing constant),

$$e^{-F^{-1}(\|x\|_K)} = e^{-\sum_{i=1}^n f_i(x_i)} = \prod_{i=1}^n e^{-f_i(x_i)},$$

which ensures the independence of the coordinates $Z_i$ of $Z$.

Already for slightly more complicated convex bodies than $\ell_p^n$-balls, Eq. (9.11) no longer holds. For example, considering the convex body defined as

$$\mathbb{B}_{1,2}^2 := \{x \in \mathbb{R}^2 : |x_1| + x_2^2 \le 1\}.$$

It can be computed that $\|x\|_{\mathbb{B}_{1,2}^2} = |x_1|/2 + \sqrt{x_1^2/4 + x_2^2}$, which is not of the form (9.11).

On the other hand, the coordinate-wise representation of the density of $Z$ in the precise form given by Eq. (9.7), is also convenient to explicitly compute the distribution of some functionals of $Z$, as we will see in the following section.

### *9.3.3 A Different Probabilistic Representation for p-Radial Probability Measures*

Another probabilistic representation for a $p$-symmetric probability measure on $\mathbb{B}_p^n$ has been given by Barthe et al. [4] in the following way,

**Theorem 9.3.4** *Let $Z$ be a random vector in $\mathbb{R}^n$ defined as in Theorem 9.3.2. Let $W$ be a non-negative random variable with probability distribution $\mathbf{P}_W$ and independent of $Z$. Then*

$$\frac{Z}{(Z^p + W)^{1/p}} \sim \mathbf{P}_W(\{0\})\, \mu_p^n + \mathrm{H}_W(\cdot)\, \nu_p^n,$$

*where* $\mathrm{H}_W : \mathbb{B}_p^n \to \mathbb{R}$, $\mathrm{H}_W(x) = h(x)$, *with*

$$h(r) = \frac{1}{\Gamma(1+n/p)(1-r^p)^{1+n/p}} \int_{(0,\infty)} s^{n/p} e^{sr^p/(r^p-1)}\, \mathrm{d}\mathbf{P}_W(s).$$

*Remark* Note that all the distributions obtainable from Theorem 9.3.4 are $p$-radial, especially the $p$-norm of $Z/(Z^p + W)^{1/p}$ is

$$R = \Big(\frac{Z^p}{Z^p + W}\Big)^{1/p}.$$

Moreover, some particular choices of $W$ in Theorem 9.3.4 lead to interesting distributions:

1. When $W \equiv 0$ we recover the cone measure of Theorem 9.3.2;
2. For $\alpha > 0$, choosing $W \sim \Gamma(\alpha, 1)$ results in the density proportional to $x \mapsto (1 - x^p)^{\alpha-1}$ for $x \leq 1$.
3. As a particular case of the previous one, when $W \sim \exp(1) = \Gamma(1, 1)$, then $H_W \equiv 1$ and

$$\frac{Z}{(\|Z\|_p^p + W)^{1/p}} \sim \nu_p^n.$$

This is not in contrast with Theorem 9.3.2. Indeed, it is easy to compute that

$$\|Z\|_p^p \sim \Gamma(n/p, 1).$$

In view of the properties (9.1) and (9.2), this implies

$$\frac{\|Z\|_p^p}{\|Z\|_p^p + W} \sim \beta(n/p, 1) \sim U^{p/n}.$$

As a consequence of this fact, the orthogonal projection of the cone measure $\mu_p^{n+p}$ on $\partial\mathbb{B}_p^{n+p}$ onto the first n coordinates is $\nu_p^n$. Indeed, if $W = \sum_{i=n+1}^{n+p} |Z_i|^p$, then $W \sim \exp(1)$, while

$$\frac{Z}{(\|Z\|_p^p + W)^{1/p}} = \frac{(Z_1, \ldots, Z_n)}{(\sum_{i=1}^{n+p} |Z_i|^p)^{1/p}}$$

is the required projection. We refer to [4, Corollaries 3-4] for more details in this direction.

## 9.4 Central Limit Theorems and Laws of Large Numbers

The law of large numbers and the central limit theorem are arguably among the most prominent limit theorems in probability theory. Thanks to the probabilistic representation for the various geometric measures on $\ell_p^n$-balls described in Sect. 9.3.1, both of these limit theorems can successfully applied to deduce information about the geometry of $\ell_p^n$-balls. This—by now classical—approach will be described here, but we will also consider some more recent developments in this direction as well as several generalizations of known results.

### 9.4.1 Classical Results: Limit Theorems à la Schechtman-Schmuckenschläger

The following result on the absolute moments of a $p$-generalized Gaussian random variable is easy to derive by direct computation, and therefore we omit its proof, which the reader can find in [13, Lemma 4.1]

**Lemma 9.4.1** *Let $p \in (0, \infty]$ and let $Z_0$ be a $p$-generalized Gaussian random variable (i.e., its density is given by Eq. (9.7)). Then, for any $q \in [0, \infty]$,*

$$\mathbf{E}\big[|Z_0|^q\big] = \begin{cases} \dfrac{p^{q/p}}{q+1} \dfrac{\Gamma\big(1+\frac{q+1}{p}\big)}{\Gamma\big(1+\frac{1}{p}\big)} =: M_p(q) & \text{if } p < \infty, \\ \dfrac{1}{q+1} =: M_\infty(q) & \text{if } p = \infty. \end{cases}$$

For convenience, we will also indicate $m_{p,q} := M_p(q)^{1/q}$ and

$$C_p(q,r) := \mathbf{Cov}(|Z_0|^q, |Z_0|^r) = M_p(q+r) - M_p(q)M_p(r).$$

We use the convention that $M_\infty(\infty) = C_\infty(\infty, \infty) = C_\infty(\infty, q) = 0$. The next theorem is a version of the central limit theorem in [24, Proposition 2.4].

**Theorem 9.4.2** *Let $0 < p, q < \infty$, $p \neq q$ and $X \sim \nu_p^n$. Then*

$$\sqrt{n}\Big(n^{1/p-1/q}\frac{\|X\|_q}{m_{p,q}} - 1\Big) \xrightarrow[n\to\infty]{d} N,$$

*where $N \sim \mathcal{N}\big(0, \sigma_{p,q}^2\big)$ and*

$$\sigma_{p,q}^2 := \frac{C_p(q,q)}{q^2 M_p(q)^2} - \frac{2C_p(p,q)}{pq M_p(q)} + \frac{C_p(p,p)}{p^2}$$

Note that, since $M_p(p) = 1$, then $\sigma_{p,p}^2 = 0$. In fact, in such a case

$$\sqrt{n}(\|X\|_p - 1) \xrightarrow[n\to\infty]{d} 0,$$

and a different normalization than $\sqrt{n}$ is needed to obtain a non-degenerate limit distribution. Moreover, $\sigma_{p,q}^2 > 0$ whenever $p \neq q$.

For our purposes, it is convenient to define the following quantities

$$k_{p,n} := n^{1/p}|\mathbb{B}_p^n|^{1/n}, \qquad k_{q,n} := n^{1/q}|\mathbb{B}_q^n|^{1/n}$$

and

$$A_{p,q,n} := \frac{k_{p,n}}{m_{p,q} k_{q,n}}.$$

It is easy to verify with Sterling's approximation that, for any $p, q > 0$, $A_{p,q,n} = A_{p,q} + \mathcal{O}(1/n)$ for $A_{p,q} \in (0, \infty)$, as $n \to \infty$.

With this definition in mind, we exploit Theorem 9.4.2 to prove a result on the volume of the intersection of $\ell_p^n$-balls. This can be regarded as a generalization of the main results in Schechtman and Schmuckenschläger [21], and Schmuckenschläger [23, 24].

**Corollary 9.4.3** *Let $0 < p, q < \infty$ and $p \neq q$. Let $r \in [0, 1]$ and $(t_n)_{n \in \mathbb{N}} \subseteq \mathbb{R}^+$ be such that*

$$\lim_{n\to\infty} \sqrt{n}(t_n A_{p,q} - 1) = \Phi_{p,q}^{-1}(r),$$

*where $\Phi_{p,q} : [-\infty, +\infty] \to [0, 1]$ is the distribution function of $N \sim \mathcal{N}(0, \sigma_{p,q}^2)$ and $\sigma_{p,q}^2$ is defined in Theorem 9.4.2, i.e.,*

$$\Phi_{p,q}(x) := \frac{1}{\sqrt{2\pi\sigma_{p,q}^2}} \int_{-\infty}^{x} e^{-s^2/(2\sigma_{p,q}^2)}\, \mathrm{d}s.$$

*Then*

$$\lim_{n\to\infty} \left|\mathbb{D}_p^n \cap t_n \mathbb{D}_q^n\right| = r.$$

*In particular, when $t_n \equiv t$, then*

$$\lim_{n\to\infty} \left|\mathbb{D}_p^n \cap t\, \mathbb{D}_q^n\right| = \begin{cases} 0 & \text{if } t < 1/A_{p,q}, \\ 1/2 & \text{if } t = 1/A_{p,q}, \\ 1 & \text{if } t > 1/A_{p,q}. \end{cases}$$

*Proof* First of all, note that, since $A_{p,q,n} = A_{p,q} + \mathcal{O}(1/n)$, then

$$\lim_{n\to\infty} \sqrt{n}(t_n A_{p,q,n} - 1) = \lim_{n\to\infty} \sqrt{n}(t_n A_{p,q} - 1),$$

provided that the latter exists in $[-\infty, \infty]$, as per assumption. In particular, taking the limit on both sides of the following equality,

$$\mathbf{P}\big(\|X\|_q \le t_n k_{p,n} k_{q,n}^{-1} n^{1/p-1/q}\big) = \mathbf{P}\big(\sqrt{n}(n^{1/p-1/q} m_{p,q}^{-1} \|X\|_q - 1) \le \sqrt{n}(t_n A_{p,q,n} - 1)\big),$$

we get, because of Theorem 9.4.2,

$$\lim_{n\to\infty} \mathbf{P}\big(\|X\|_q \le t_n k_{p,n} k_{q,n}^{-1} n^{1/p-1/q}\big) = \mathbf{P}\big(N \le \Phi_{p,q}^{-1}(r)\big) = r.$$

On the other hand, it is true that the following chain of equalities hold:

$$\begin{aligned}
\mathbf{P}\big(\|X\|_q \le t_n k_{p,n} k_{q,n}^{-1} n^{1/p-1/q}\big) &= \frac{|z \in \mathbb{B}_p^n : z \in t_n k_{p,n} k_{q,n}^{-1} n^{1/p-1/q} \mathbb{B}_q^n|}{|\mathbb{B}_p^n|} \\
&= \big|z \in |\mathbb{B}_p^n|^{-1/n} \mathbb{B}_p^n : z \in t_n k_{p,n} k_{q,n}^{-1} n^{1/p-1/q} |\mathbb{B}_p^n|^{-1/n} \mathbb{B}_q^n\big| \\
&= |z \in \mathbb{D}_p^n : z \in t_n \mathbb{D}_q^n| \\
&= |\mathbb{D}_p^n \cap t_n \mathbb{D}_q^n|,
\end{aligned}$$

which concludes the main part of proof. For the last observation, note that for any $t$ constant, either $\sqrt{n}(tA_{p,q} - 1) \equiv 0$ or it diverges. □

## *9.4.2 Recent Developments*

### 9.4.2.1 The Multivariate CLT

We present here a multivariate central limit theorem that recently appeared in [13]. It constitutes the multivariate generalization of Theorem 9.4.2. Similar to the classical results of Schechtman and Schmuckenschläger [21], and Schmuckenschläger [23, 24] this was used to study intersections of (this time multiple) $\ell_p^n$-balls. In part 1, we replace the original assumption $X \sim \nu_p^n$ of [13] to a more general one, that appears naturally from the proof. Part 2 is substantially different and cannot be generalized with the same assumption.

**Theorem 9.4.4** *Let $n, k \in \mathbb{N}$ and $p \in [1, \infty]$.*

*1. Let $X$ be a continuous $p$-radial random vector in $\mathbb{R}^n$ such that*

$$\sqrt{n}\big(1 - \|X\|_p\big) \xrightarrow[n\to\infty]{\mathbf{P}} 0. \tag{9.12}$$

*Fix a $k$-tuple $(q_1, \dots, q_k) \in ([1,\infty) \setminus \{p\})^k$. We have the multivariate central limit theorem*

$$\sqrt{n}\Big(n^{1/p-1/q_1} \frac{\|X\|_{q_1}}{m_{p,q_1}} - 1, \dots, n^{1/p-1/q_k} \frac{\|X\|_{q_k}}{m_{p,q_k}} - 1\Big) \xrightarrow[n\to\infty]{d} N,$$

*where* $N = (N_1, \ldots, N_k) \sim \mathcal{N}(0, \Sigma)$*, with covariance matrix* $\Sigma = (c_{i,j})_{i,j=1}^k$ *whose entries are given by*

$$c_{i,j} := \begin{cases} \dfrac{1}{q_i q_j}\left(\dfrac{\Gamma(\frac{1}{p})\Gamma(\frac{q_i+q_j+1}{p})}{\Gamma(\frac{q_i+1}{p})\Gamma(\frac{q_j+1}{p})} - 1\right) - \dfrac{1}{p} & \text{if } p < \infty, \\ \dfrac{1}{q_i + q_j + 1} & \text{if } p = \infty. \end{cases} \tag{9.13}$$

2. *Let* $X \sim \nu_p^n$*. If* $p < \infty$*, then we have the non-central limit theorem*

$$\frac{n^{1/p}}{(p \log n)^{1/p-1}} \|X\|_\infty - A_n^{(p)} \xrightarrow[n\to\infty]{d} G,$$

*where*

$$A_n^{(p)} := p \log n - \frac{1-p}{p} \log(p \log n) + \log(p^{1/p}\Gamma(1 + 1/p))$$

*and* $G$ *is a Gumbel random variable with distribution function* $\mathbb{R} \ni t \mapsto e^{-e^{-t}}$.

*Remark* Note that the assumptions of Theorem 9.4.4 include the cases $X \sim \nu_p^n$ and $X \sim \mu_p^n$. In fact, condition (9.12) is just the quantitative version of the following concept: to have Gaussian fluctuations it is necessary that the bigger $n$ gets, the more the distribution of $X$ is concentrated in near $\partial\mathbb{B}_p^n$. It is relevant to note that (9.12) also keeps open the possibility for a non-trivial limit distribution when rescaling $(1 - \|X\|_p)$ with a sequence that grows faster than $\sqrt{n}$. This would yield a limit-theorem for $\|X\|_p$. For example, when $X \sim \nu_p^n$, we already noted that $\|X\|_p \overset{d}{=} U^{1/n}$, so that

$$n(1 - \|X\|_p) \xrightarrow[n\to\infty]{d} E \sim \exp(1).$$

On the other hand, when $X \sim \mu_p^n$, then $1 - \|X\|_p \equiv 0$.

*Proof* We only give a proof for the first part of the theorem, the second one can be found in [13].

Let first $p \in [1, \infty)$. Consider a sequence of independent $p$-generalized Gaussian random variables $(Z_j)_{j\in\mathbb{N}}$, also independent from every $X$. Set $Z = (Z_1, \ldots, Z_n)$. For any $n \in \mathbb{N}$ and $i \in \{1, \ldots, k\}$, consider the random variables

$$\xi_n^{(i)} := \frac{1}{\sqrt{n}} \sum_{j=1}^n \big(|Z_j|^{q_i} - M_p(q_i)\big) \qquad \text{and} \qquad \eta_n := \frac{1}{\sqrt{n}} \sum_{j=1}^n \big(|Z_j|^p - 1\big).$$

According to the classical multivariate central limit theorem, we get

$$(\xi_n^{(1)}, \dots, \xi_n^{(k)}, \eta_n) \xrightarrow[n\to\infty]{d} (\xi^{(1)}, \dots, \xi^{(k)}, \eta) \sim \mathcal{N}(0, \widetilde{\Sigma})$$

with covariance matrix given by

$$\widetilde{\Sigma} = \begin{pmatrix} C_p(q_1, q_1) & \cdots & C_p(q_1, q_k) & C_p(q_1, p) \\ \vdots & \ddots & \vdots & \vdots \\ C_p(q_k, q_1) & \cdots & C_p(q_k, q_k) & C_p(q_k, p) \\ \\ C_p(p, q_1) & \cdots & C_p(p, q_k) & C_p(p, p) \end{pmatrix}$$

Using Theorem 9.3.2 and the aforementioned definitions we can write, for $i \in \{1, \dots, k\}$,

$$\begin{aligned} \|X\|_{q_i} &\overset{d}{=} \frac{X\|Z\|_{q_i}}{\|Z\|_p} \\ &= \|X\|_p \frac{(nM_p(q_i) + \sqrt{n}\xi_n^{(i)})^{1/q_i}}{(n + \sqrt{n}\eta_n)^{1/p}} \\ &= \|X\|_p \frac{(nM_p(q_i))^{1/q_i}}{n^{1/p}} F_i\Big(\frac{\xi_n^{(i)}}{\sqrt{n}}, \frac{\eta_n}{\sqrt{n}}\Big) \\ &= \|X\|_p \, n^{1/q_i - 1/p} m_{p,q} F_i\Big(\frac{\xi_n^{(i)}}{\sqrt{n}}, \frac{\eta_n}{\sqrt{n}}\Big) \\ &= (\|X\|_p - 1) n^{1/q_i - 1/p} m_{p,q} F_i\Big(\frac{\xi_n^{(i)}}{\sqrt{n}}, \frac{\eta_n}{\sqrt{n}}\Big) + n^{1/q_i - 1/p} m_{p,q} F_i\Big(\frac{\xi_n^{(i)}}{\sqrt{n}}, \frac{\eta_n}{\sqrt{n}}\Big) \end{aligned}$$

where we defined the function $F_i : \mathbb{R} \times (\mathbb{R} \setminus \{-1\}) \to \mathbb{R}$ as

$$F_i(x, y) := \frac{(1 + x/M_p(q_i))^{1/q_i}}{(1 + y)^{1/p}}.$$

Note that $F_i$ is continuously differentiable around $(0, 0)$ with Taylor expansion given by

$$F_i(x, y) = 1 + \frac{x}{q_i M_p(q_i)} - \frac{y}{p} + \mathcal{O}(x^2 + y^2).$$

Since, for the law of large numbers, $\xi_n^{(i)}/\sqrt{n} \xrightarrow[n\to\infty]{\text{a.s.}} 0$ and $\eta_n/\sqrt{n} \xrightarrow[n\to\infty]{\text{a.s.}} 0$, the previous equation means that there exists a random variable $C$, independent of $n$, such that

$$\Big|F_i\Big(\frac{\xi_n^{(i)}}{\sqrt{n}},\frac{\eta_n}{\sqrt{n}}\Big)-\Big(1+\frac{1}{q_iM_p(q_i)}\frac{\xi_n^{(i)}}{\sqrt{n}}-\frac{1}{p}\frac{\eta_n}{\sqrt{n}}\Big)\Big|\le C\frac{(\xi_n^{(i)})^2+\eta_n^2}{n}.$$

In particular,

$$\begin{aligned}\sqrt{n}(\|X\|_p-1)&\Big(1+\frac{1}{q_iM_p(q_i)}\frac{\xi_n^{(i)}}{\sqrt{n}}-\frac{1}{p}\frac{\eta_n}{\sqrt{n}}-C\frac{(\xi_n^{(i)})^2+\eta_n^2}{n}\Big)\\&+\Big(\frac{1}{q_iM_p(q_i)}\xi_n^{(i)}-\frac{1}{p}\eta_n-C\frac{(\xi_n^{(i)})^2+\eta_n^2}{\sqrt{n}}\Big)\\&\le\sqrt{n}\Big(n^{1/p-1/q_i}\frac{\|X\|_{q_i}}{m_{p,q_i}}-1\Big)\\&\le\sqrt{n}(\|X\|_p-1)\Big(1+\frac{1}{q_iM_p(q_i)}\frac{\xi_n^{(i)}}{\sqrt{n}}-\frac{1}{p}\frac{\eta_n}{\sqrt{n}}+C\frac{(\xi_n^{(i)})^2+\eta_n^2}{n}\Big)\\&+\Big(\frac{1}{q_iM_p(q_i)}\xi_n^{(i)}-\frac{1}{p}\eta_n+C\frac{(\xi_n^{(i)})^2+\eta_n^2}{\sqrt{n}}\Big)\end{aligned}$$

Note that the first summand of both bounding expressions tends to 0 in distribution by assumption (9.12), while the second converges in distribution to $\frac{1}{q_iM_p(q_i)}\xi^{(i)}-\frac{1}{p}\eta$. This implies that

$$\sqrt{n}\Big(n^{1/p-1/q_i}\frac{\|X\|_{q_i}}{m_{p,q_i}}-1\Big)\xrightarrow[n\to\infty]{d}\frac{1}{q_iM_p(q_i)}\xi^{(i)}-\frac{1}{p}\eta=:N_i,$$

where $N_i$ is a centered Gaussian random variable. To obtain the final multivariate central limit theorem, we only have to compute the covariance matrix $\Sigma$. For $\{i,j\}\subseteq\{1,\dots,k\}$, its entries are given by

$$\begin{aligned}c_{i,j}&=\mathbf{Cov}\Big(\frac{\xi^{(i)}}{q_iM_p(q_i)}-\frac{\eta}{p},\frac{\xi^{(j)}}{q_jM_p(q_j)}-\frac{\eta}{p}\Big)\\&=\frac{\mathbf{Cov}(\xi^{(i)},\xi^{(j)})}{q_iq_jM_p(q_i)M_p(q_j)}-\frac{1}{p}\Big(\frac{\mathbf{Cov}(\xi^{(i)},\eta)}{q_iM_p(q_i)}+\frac{\mathbf{Cov}(\eta,\xi^{(j)})}{q_jM_p(q_j)}\Big)+\frac{\mathbf{Cov}(\eta,\eta)}{p^2}\\&=\frac{C_p(q_i,q_j)}{q_iq_jM_p(q_i)M_p(q_j)}-\frac{1}{p}\Big(\frac{C_p(q_i,p)}{q_iM_p(q_i)}+\frac{C_p(q_j,p)}{q_jM_p(q_j)}\Big)+\frac{C_p(p,p)}{p^2},\end{aligned}$$

and this can be made explicit to get Eq. (9.13). The remaining case of $p=\infty$ can be repeated using the aforementioned conventions on the quantities $M_\infty$ and $C_\infty$. □

*Remark* From the proof is evident that in the case when $\sqrt{n}(X-1)$ converges in distribution to a random variable $F$, independence yields, for every $i\in\{1,\dots,k\}$, the convergence in distribution

$$\sqrt{n}\Big(n^{1/p-1/q_i}\frac{\|X\|_{q_i}}{m_{p,q_i}}-1\Big)\xrightarrow[n\to\infty]{d}F+N_i$$

in which case the limiting random variable is not normal in general. Analogously, if there exists a sequence $(a_n)_{n\in\mathbb{N}}$, $a_n=o(\sqrt{n})$ and a random variable $F$ such that

$$a_n(\|X\|_p-1)\xrightarrow[n\to\infty]{d}F,$$

then the previous proof, with just a change of normalization, yields the limit theorem

$$a_n\Big(n^{1/p-1/q}\frac{\|X\|_q}{m_{p,q}}-1\Big)\xrightarrow[n\to\infty]{d}F$$

for every $q\in[1,\infty)$, as $n\to\infty$.

In analogy to Corollary 9.4.3, one can prove in a similar way the following result concerning the simultaneous intersection of several dilated $\ell_p$-balls. In particular, we emphasize that the volume of the simultaneous intersection of three balls $\mathbb{D}_p^n\cap t_1\mathbb{D}_{q_1}^n\cap t_2\mathbb{D}_{q_2}$ is *not* equal to $1/4$ if these balls are in 'critical' position, as one might conjecture in view of Corollary 9.4.3.

**Corollary 9.4.5** *Let $n,k\in\mathbb{N}$ and $p\in[1,\infty]$. Fix a $k$-uple $(q_1,\dots,q_k)\in([1,\infty)\setminus\{p\})^k$. Let $t_1,\dots,t_k$ be positive constants and define the sets $I_\star:=\{i\in\{1,\dots,k\}:A_{p,q_i}t_i\star 1\}$, where $\star$ is one of the symbols $>$, $=$ or $<$. Then,*

$$\lim_{n\to\infty}|\mathbb{D}_p^n\cap t_1\mathbb{D}_{q_1}^n\cap\cdots\cap t_k\mathbb{D}_{q_k}^n|=\begin{cases}1 & \text{if } \#I_>=k,\\ \mathbf{P}(N_i\le 0: i\in I_=) & \text{if } \#I_=\ge 1 \text{ and } \#I_<=0,\\ 0 & \text{if } \#I_<\ge 1,\end{cases}$$

*where $N=(N_1,\dots,N_k)$ is as in Theorem 9.4.4.*

#### 9.4.2.2 Outlook: The Non-commutative Setting

Very recently, Kabluchko et al. obtained in [11] a non-commutative analogue of the classical result by Schechtman and Schmuckenschläger [21]. Instead of considering the family of $\ell_p^n$-balls, they studied the volumetric properties of unit balls in classes of classical matrix ensembles.

More precisely, we let $\beta\in\{1,2,4\}$ and consider the collection $\mathscr{H}_n(\mathbb{F}_\beta)$ of all self-adjoint $n\times n$ matrices with entries from the (skew) field $\mathbb{F}_\beta$, where $\mathbb{F}_1=\mathbb{R}$, $\mathbb{F}_2=\mathbb{C}$ or $\mathbb{F}_4=\mathbb{H}$ (the set of Hamiltonian quaternions). By $\lambda_1(A),\dots,\lambda_n(A)$ we denote the (real) eigenvalues of a matrix $A$ from $\mathscr{H}_n(\mathbb{F}_\beta)$ and consider the following matrix analogues of the classical $\ell_p^n$-balls discussed above:

$$\mathbb{B}^n_{p,\beta} := \Big\{A \in \mathscr{H}_n(\mathbb{F}_\beta) : \sum_{j=1}^n |\lambda_j(A)|^p \le 1\Big\}, \qquad \beta \in \{1,2,4\} \quad \text{and} \quad p \in [1,\infty],$$

where we interpret the sum in brackets as $\max\{\lambda_j(A) : j = 1, \dots, n\}$ if $p = \infty$. As in the case of the classical $\ell_p^n$-balls we denote by $\mathbb{D}^n_{p,\beta}$, $\beta \in \{1,2,4\}$ the volume normalized versions of these matrix unit balls. Here the volume can be identified with the $(\beta\frac{n(n-1)}{2} + \beta n)$-dimensional Hausdorff measure on $\mathscr{H}_n(\mathbb{F}_\beta)$.

**Theorem 9.4.6** *Let* $1 \le p, q < \infty$ *with* $p \ne q$ *and* $\beta \in \{1,2,4\}$. *Then*

$$\lim_{n\to\infty} |\mathbb{D}^n_{p,\beta} \cap t\,\mathbb{D}^n_{q,\beta}| = \begin{cases} 0 & \text{if } t < e^{\frac{1}{2p}-\frac{1}{2q}}\big(\frac{2p}{p+q}\big)^{1/q}, \\ 1 & \text{if } t > e^{\frac{1}{2p}-\frac{1}{2q}}\big(\frac{2p}{p+q}\big)^{1/q}. \end{cases}$$

To obtain this result, one first needs to study the asymptotic volume of the unit balls of $\mathscr{H}_n(\mathbb{F}_\beta)$. This is done by resorting to ideas from the theory of logarithmic potentials with external fields. The second ingredient is a weak law of large numbers for the eigenvalues of a matrix chosen uniformly at random from $\mathbb{B}^n_{p,\beta}$. For details we refer the interested reader to [11].

## 9.5 Large Deviations vs. Large Deviation Principles

The final section is devoted to large deviations and large deviation principles for geometric characteristics of $\ell_p^n$-balls. We start by presenting some classical results on large deviations related to the geometry of $\ell_p^n$-balls due to Schechtman and Zinn. Its LDP counterpart has entered the stage of asymptotic geometry analysis only recently in [13]. We then continue by presenting a large deviation principle for one-dimensional random projections of $\ell_p^n$-balls of Gantert et al. [9]. Finally, we present a similar result for higher-dimensional projections as well.

### *9.5.1 Classical Results: Large Deviations à la Schechtman-Zinn*

We start by rephrasing the large deviation inequality of Schechtman and Zinn [22]. It is concerned with the $\ell_q$-norm of a random vector in an $\ell_p^n$-balls. The proof that we present follows the argument of [17].

**Theorem 9.5.1** *Let $1 \le p < q \le \infty$ and $X \sim \nu_p^n$ or $X \sim \mu_p^n$. Then there exists a constant $c \in (0, \infty)$, depending only on $p$ and $q$, such that*

$$\mathbf{P}(n^{1/p-1/q}\|X\|_q > z) \le \exp(-c\, n^{p/q} z^p),$$

*for every $z > 1/c$.*

*Proof* We sketch the proof for the case that $X \sim \mu_p^n$. Let $Z_1, \dots, Z_n$ be $p$-generalized Gaussian random variables and put $S_r := |Z_1|^r + \dots + |Z_n|^r$ for $r \ge 1$. Now observe that by the exponential Markov inequality and Theorem 9.3.2, for $t > 0$,

$$\begin{aligned}\mathbf{P}(n^{1/p-1/q}\|X\|_q > z) &= \mathbf{P}\Big(\frac{S_q^{p/q}}{S_p} > \frac{z^p}{n^{1-p/q}}\Big) \\ &\le \exp\Big(-\frac{tz^p}{n^{1-p/q}}\Big)\,\mathbf{E}\exp\Big(t\frac{S_q^{p/q}}{S_p}\Big) \\ &\le \exp\Big(-\frac{tz^p}{n^{1-p/q}}\Big)\,\mathbf{E}\exp\Big(t\frac{S_q^{p/q}}{\mathbf{E}\,S_p}\Big),\end{aligned}$$

where we also used the independence property in Theorem 9.3.2 in the last step. Next, we observe that $\mathbf{E}\,S_p = n$ by Lemma 9.4.1. Moreover from [17, Corollary 3] it is known that there exists a constant $c \in (0, \infty)$ only depending on $p$ and $q$ such that

$$\mathbf{E}\exp\big(tS_q^{p/q}\big) \le n^{1-p/q}\big(1-ct\big)^{-n^{p/q}}$$

as long as $0 < t < 1/c$. Thus, choosing $t = \frac{n}{c} - \frac{n}{z^p}$ we arrive at

$$\mathbf{P}(n^{1/p-1/q}\|X\|_q > z) \le n^{1-p/q}\Big(\frac{ez^p}{c}\Big)^{n^{p/q}} \exp(-cn^{p/q}z^p).$$

This implies the result. □

## 9.5.2 Recent Developments

### 9.5.2.1 The LDP Counterpart to Schechtman-Zinn

After having presented the classical Schechtman-Zinn large deviation inequality, we turn now to a LDP counterpart. The next result is a summary of the results presented in from [13, Theorems from 1.2 to 1.5]. The speed and the rate function in its part 4 resembles the right hand side of the inequality in Theorem 9.5.1.

**Theorem 9.5.2** *Let $n \in \mathbb{N}$, $p \in [1,\infty]$, $q \in [1,\infty)$ and $X \sim \nu_p^n$. Define the sequence*

$$\|\mathbf{X}\| := (n^{1/p-1/q} X)_{n\in\mathbb{N}}.$$

1. *If $q < p < \infty$, then $\|\mathbf{X}\|$ satisfies an LDP with speed $n$ and good rate function*

$$\mathcal{I}_{\|\mathbf{X}\|}(x) = \begin{cases} \inf\{\mathcal{I}_1(x_1) + \mathcal{I}_2(x_2) : x = x_1 x_2, x_1 \geq 0, x_2 \geq 0\} & \text{if } x \geq 0, \\ +\infty & \text{otherwise.} \end{cases}$$

*Here*

$$\mathcal{I}_1(x) = \begin{cases} -\log(x) & \text{if } x \in (0,1], \\ +\infty & \text{otherwise,} \end{cases} \tag{9.14}$$

*and*

$$\mathcal{I}_2(x) = \begin{cases} \inf\{\Lambda^*(y,z) : x = y^{1/q} z^{-1/p}, y \geq 0, z \geq 0\} & \text{if } x \geq 0 \\ +\infty & \text{otherwise,} \end{cases}$$

*where $\Lambda^*$ is the Fenchel-Legendre transform of the function*

$$\Lambda(t_1,t_2) := \log \int_0^{+\infty} \frac{1}{p^{1/p}\Gamma(1+1/p)} e^{t_1 s^q + (t_2 - 1/p)s^p}\,\mathrm{d}s, \qquad (t_1,t_2) \in \mathbb{R} \times \Big(-\infty, \frac{1}{p}\Big).$$

2. *If $q < p = \infty$, then $\|\mathbf{X}\|$ satisfies an LDP with speed $n$ and good rate function*

$$\mathcal{I}_{\|\mathbf{X}\|}(x) = \begin{cases} \Psi^*(x) & \text{if } x \geq 0, \\ +\infty & \text{otherwise,} \end{cases}$$

*where $\Psi^*$ is the Fenchel-Legendre transform of the function*

$$\Psi(t) := \int_0^1 e^{ts^q}\,\mathrm{d}s, \qquad t \in \mathbb{R}.$$

3. *If $p = q$, then $\|\mathbf{X}\|$ satisfies an LDP with speed $n$ and good rate function $\mathcal{I}_1$ defined in Eq. (9.14).*
4. *If $p < q$, then $\|\mathbf{X}\|$ satisfies an LDP with speed $n^{p/q}$ and good rate function*

$$\mathcal{I}_{\|\mathbf{X}\|}(x) = \begin{cases} \frac{1}{p}\big(x^q - m_{p,q}^q\big)^{p/q} & \text{if } x \geq m_{p,q}, \\ +\infty & \text{otherwise.} \end{cases}$$

#### 9.5.2.2 LDPs for Projections of $\ell_p^n$-Balls: One-Dimensional Projections

We turn now to a different type of large deviation principles. More precisely, we consider random projections of points uniformly distributed in an $\ell_p^n$-ball or distributed according to the corresponding cone probability measure onto a uniform random direction. The following result is a summary of from [9, Theorems 2.2,2.3]. The proof of the first part follows rather directly from Cramér's theorem (Theorem 9.2.1) and the contraction principle (Proposition 9.2.2), the second part is based on large deviation theory for sums of stretched exponentials.

**Theorem 9.5.3** *Let $n \in \mathbb{N}$ and $p \in [1, \infty)$. Let $X \sim \nu_p^n$ or $X \sim \mu_p^n$ and $\Theta \sim \sigma_2^n$ be independent random vectors. Consider the sequence*

$$\mathbf{W} := (n^{1/p-1/2} X\Theta)_{n\in\mathbb{N}}.$$

*1. If $p \geq 2$, then $\mathbf{W}$ satisfies an LDP with speed $n$ and good rate function*

$$\mathcal{I}_{\mathbf{W}}(w) = \inf\{\Phi^*(\tau_0, \tau_1, \tau_2) : w = \tau_0^{-1/2}\tau_1\tau_2^{-1/p}, \tau_0 > 0, \tau_1 \in \mathbb{R}, \tau_2 > 0\},$$

*where $\Phi^*$ is the Fenchel-Legendre transform of*

$$\Phi(t_0, t_1, t_2) := \log \int_{\mathbb{R}}\int_{\mathbb{R}} e^{t_0 z^2 + t_1 zy + t_2 |z|^p} f_2(z) f_p(y)\, \mathrm{d}z\, \mathrm{d}y, \qquad t_0, t_1, t_2 \in \mathbb{R}.$$

*2. If $p < 2$, then $\mathbf{W}$ satisfies an LDP with speed $n^{2p/(2+p)}$ and good rate function*

$$\mathcal{I}_{\mathbf{W}}(w) = \frac{2+p}{2p}|w|^{2p/(2+p)}.$$

*Proof* Let us sketch the proof for the case that $p > 2$, by leaving out any technical details. For this, let $Z_1, \ldots, Z_n$ be $p$-generalized Gaussian random variables, $G_1, \ldots, G_n$ be Gaussian random variables and $U$ be a uniform random variable over $[0, 1]$. Also assume that all the aforementioned random variables are independent. Also put $Z := (Z_1, \ldots, Z_n)$ and $G := (G_1, \ldots, G_n)$. When $X \sim \mu_p^n$, by Theorem 9.3.2, we can state that for each $n \in \mathbb{N}$ the target random variable $n^{1/p-1/2} X\Theta$ has the same distribution as

$$n^{1/p-1/2} \frac{\sum_{i=1}^n G_i Z_i}{\|G\|_2 \|Z\|_p} = \frac{\frac{1}{n}\sum_{i=1}^n G_i Z_i}{\left(\frac{1}{n}\sum_{i=1}^n |G_i|^2\right)^{1/2}\left(\frac{1}{n}\sum_{i=1}^n |Z_i|^p\right)^{1/p}}. \qquad (9.15)$$

Note that $\Phi$ is finite whenever $p < 2$, $t_0 < 1/2$, $t_1 \in \mathbb{R}$ and $t_2 < 1/p$. Then, Cramér's theorem (Theorem 9.2.1) shows that the $\mathbb{R}^3$-valued sum

$$\frac{1}{n}\sum_{i=1}^{n}\big(|G_i|^2, G_i Z_i, |Z_i|^p\big)$$

satisfies an LDP with speed $n$ and rate function $\Phi^*$. Applying the contraction principle (Proposition 9.2.2) to the function $F(x, y, z) = x^{-1/2} y z^{-1/p}$ yields the LDP for $\mathbf{W}$ with speed $n$ and the desired rate function $\mathcal{I}_{\mathbf{W}}$. Once the LDP is proven for the cone measure, it can be pushed to the case of the uniform measure. By Theorem 9.3.2, multiplying the expression in Eq. (9.15) by $U^{1/n}$, we obtain a random variable distributed according to $\nu_p^n$. It is proven in [9, Lemma 3.2] that multiplying by $U^{1/n}$ every element of the sequence $\mathbf{W}$, we obtain a new sequence of random variables that also satisfies an LDP with the same speed and the same rate function as $\mathbf{W}$. On the other hand, when $p < 2$, $\Phi(t_0, t_1, t_2) = \infty$ for any $t_1 \neq 0$, hence suggesting that in this case the LDP could only occur at a lower speed than $n$. □

#### 9.5.2.3 LDPs for Projections of $\ell_p^n$-Balls: The Grassmannian Setting

Finally, let us discuss projections to higher dimensional subspaces, generalizing thereby the set-up from the previous section. We adopt the Grassmannian setting and consider the 2-norm of the projection to a uniformly distributed random subspace in the Grassmannian $\mathbb{G}_n^k$ of $k$-dimensional subspaces of $\mathbb{R}^n$ of a point uniformly distributed in the $\ell_p^n$-unit ball. Since we are interested in the asymptotic regime where $n \to \infty$, we also allow the subspace dimension $k$ to vary with $n$. However, in order to keep our notation transparent, we shall nevertheless write $k$ instead of $k(n)$. The next result is the collection of [1, Theorems 1.1,1.2].

**Theorem 9.5.4** *Let $n \in \mathbb{N}$. Fix $p \in [1, \infty]$ and a sequence $k = k(n) \in \{1, \dots, n-1\}$ such that the limit $\lambda := \lim_{n\to\infty}(k/n)$ exists. Let $P_E X$ be the orthogonal projection of a random vector $X \sim \nu_p^n$ onto a random independent linear subspace $E \sim \eta_k^n$. Consider the sequence*

$$\|\boldsymbol{P_E X}\| := (n^{1/p-1/2}\|P_E X\|_2)_{n\in\mathbb{N}}.$$

1. *If* $p \geq 2$*, then* $\|P_E X\|$ *satisfies an LDP with speed n and good rate function*

$$\mathcal{I}_{\|P_E X\|}(y) := \begin{cases} \inf\limits_{x>y} \Big[\frac{\lambda}{2}\log\Big(\frac{\lambda x^2}{y^2}\Big) + \frac{1-\lambda}{2}\log\Big(\frac{1-\lambda}{1-y^2x^{-2}}\Big) + \mathcal{J}_p(x)\Big] & \text{if } y > 0, \\ \mathcal{J}_p(0) & \text{if } y = 0,\ \lambda \in (0,1], \\ \inf\limits_{x\geq 0} \mathcal{J}_p(x) & \text{if } y = 0,\ \lambda = 0, \\ +\infty & \text{if } y < 0\,, \end{cases}$$

*where we use the convention* $0\log 0 := 0$ *and for* $p \neq \infty$ *we have*

$$\mathcal{J}_p(y) := \inf_{\substack{x_1,x_2>0\\ x_1^{1/2}x_2^{-1/p}=y}} \mathcal{I}_p^*(x_1,x_2), \qquad y \in \mathbb{R}\,,$$

*and* $\mathcal{I}_p^*(x_1,x_2)$ *is the Fenchel-Legendre transform of*

$$\mathcal{I}_p(t_1,t_2) := \log\int_{\mathbb{R}} e^{t_1x^2+t_2|x|^p} f_p(x)\,\mathrm{d}x, \qquad (t_1,t_2) \in \mathbb{R}\times\Big(-\infty, \frac{1}{p}\Big).$$

*For* $p = \infty$*, we write* $\mathcal{J}_\infty(y) := \mathcal{I}_\infty^*(y^2)$ *with* $\mathcal{I}_\infty^*$ *being the Fenchel-Legendre transform of* $\mathcal{I}_\infty(t) := \log\int_0^1 e^{tx^2}\,\mathrm{d}x$*.*

2. *If* $p < 2$ *and* $\lambda > 0$*, then* $\|P_E X\|$ *satisfies and LDP with speed* $n^{p/2}$ *and good rate function*

$$\mathcal{I}_{\|P_E X\|}(y) := \begin{cases} \frac{1}{p}\Big(\frac{y^2}{\lambda} - m\Big)^{p/2} & \text{if } y \geq \sqrt{\lambda m_p}\,, \\ +\infty & \text{otherwise}, \end{cases}$$

*where* $m_p := p^{p/2}\Gamma(1+3/p)/(3\Gamma(1+1/p))$*.*

Let us emphasize that the proof of this theorem is in some sense similar to its one-dimensional counterpart that we have discussed in the previous section. However, there are a number of technicalities that need to be overcome when projections to high-dimensional subspaces are considered. Among others, one needs a new probabilistic representation of the target random variables. In fact, the previous theorem heavily relies on the following probabilistic representation, proved in [1, Theorem 3.1] for the case $X \sim \nu_n^p$. We shall give a proof here for a more general set-up, which might be of independent interest.

**Theorem 9.5.5** *Let* $n \in \mathbb{N}$*,* $k \in \{1,\dots,n\}$ *and* $p \in [1,\infty]$*. Let X be a continuous p-radial random vector in* $\mathbb{R}^n$ *and* $E \sim \eta_k^n$ *be a random k-dimensional linear subspace. Let* $Z = (Z_1,\dots,Z_n)$ *and* $G = (G_1,\dots,G_n)$ *having i.i.d. coordinates, distributed according to the densities* $f_p$ *and* $f_2$*, respectively. Moreover, let X, E,* *Z and G be independent. Then*

$$\|P_E X\|_2 \overset{d}{=} \|X\|_p \frac{\|Z\|_2}{\|Z\|_p} \frac{\|(G_1,\dots,G_k)\|_2}{\|G\|_2}.$$

*Proof* *Fix a vector $x \in \mathbb{R}^n$. By construction of the Haar measure $\eta_k^n$ on $\mathbb{G}_k^n$ and uniqueness of the Haar measure $\eta$ on $\mathbb{O}(n)$, we have that, for any $t \in \mathbb{R}$,*

$$\begin{aligned}\eta_k^n(E \in \mathbb{G}_k^n : \|P_E x\|_2 \ge t) &= \eta(T \in \mathbb{O}(n) : \|P_{TE_0} x\|_2 \ge t)\\ &= \eta(T \in \mathbb{O}(n) : \|P_{E_0} T x\|_2 \ge t)\\ &= \eta\big(T \in \mathbb{O}(n) : \|x\|_2 \big\|P_{E_0} T(x/\|x\|_2)\big\|_2 \ge t\big),\end{aligned}$$

*where $E_0 := \operatorname{span}(\{e_1,\dots,e_k\})$. Again, by the uniqueness of the Haar measure $\sigma_2^n$ on $\mathbb{S}_2^{n-1}$, $T(x/\|x\|_2) \sim \sigma_2^n$, provided that $T \in \mathbb{O}(n)$ has distribution $\eta$. Thus,*

$$\eta\Big(T \in \mathbb{O}(n) : \|x\|_2 \Big\|P_{E_0} T\Big(\frac{x}{\|x\|_2}\Big)\Big\|_2 \ge t\Big) = \sigma_2^n(u \in \mathbb{S}_2^{n-1} : \|x\|_2 \|P_{E_0} u\|_2 \ge t).$$

*By Theorem 9.3.2, $G/\|G\|_2 \sim \sigma_2^n$. Thus,*

$$\sigma_2^n(u \in \mathbb{S}_2^{n-1} : \|x\|_2 \|P_{E_0} T u\|_2 \ge t) = \mathbf{P}\Big(\|x\|_2 \frac{\|P_{E_0} G\|_2}{\|G\|_2} \ge t\Big).$$

*Therefore, if $E \in \mathbb{G}_k^n$ is a random subspace independent of $X$ having distribution $\eta_k^n$, and $G$ is a standard Gaussian random vector in $\mathbb{R}^n$ that is independent of $X$ and $E$, we have that*

$$\mathbf{P}_{(X,E)}\big((x,F) \in \mathbb{R}^n \times \mathbb{G}_k^n : \|P_F x\|_2 \ge t\big) = \mathbf{P}_{(X,G)}\Big((x,g) \in \mathbb{R}^n \times \mathbb{R}^n : \|x\|_2 \frac{\|P_{E_0} g\|_2}{\|g\|_2} \ge t\Big).$$

*Here, $\mathbf{P}_{(X,E)}$ denotes the joint distribution of the random vector $(X,E) \in \mathbb{R}^n \times \mathbb{G}_k^n$, while $\mathbf{P}_{(X,G)}$ stands for that of $(X,G) \in \mathbb{R}^n \times \mathbb{R}^n$. By Proposition 9.3.3, $X$ has the same distribution as $XZ/Z$. Therefore,*

$$\begin{aligned}&\mathbf{P}_{(X,G)}\Big((x,g) \in \mathbb{R}^n \times \mathbb{R}^n : \|x\|_2 \frac{\|P_{E_0} g\|_2}{\|g\|_2} \ge t\Big)\\ &= \mathbf{P}_{(X,Z,G)}\Big((x,z,g) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n : \|x\|_p \frac{\|z\|_2}{\|z\|_p} \frac{\|P_{E_0} g\|_2}{\|g\|_2} \ge t\Big)\end{aligned}$$

*with $\mathbf{P}_{(X,Z,G)}$ being the joint distribution of the random vector $(X,Z,G) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n$. Consequently, we conclude that the two random variables $\|P_E X\|_2$ and $\|X\|_p \frac{\|Z\|_2}{\|Z\|_p} \frac{\|P_{E_0} G\|_2}{\|G\|_2}$ have the same distribution.* □

*Remark* Let us remark that in his PhD thesis, Kim [15] was recently able to extend the results from [1] and [9] to more general classes of random vectors under an asymptotic thin-shell-type condition in the spirit of [2] (see [15, Assumption 5.1.2]). For instance, this condition is satisfied by random vectors chosen uniformly at random from an Orlicz ball.

#### 9.5.2.4 Outlook: The Non-commutative Setting

The body of research on large deviation principles in asymptotic geometric analysis, which we have just described above, is complemented by another paper of Kim and Ramanan [16], in which they proved an LDP for the empirical measure of an $n^{1/p}$ multiple of a point drawn from an $\ell_p^n$-sphere with respect to the cone or surface measure. The rate function identified is essentially the so-called relative entropy perturbed by some $p$-th moment penalty (see [16, Equation (3.4)]).

While this result is again in the commutative setting of the $\ell_p^n$-balls, Kabluchko et al. [12] recently studied principles of large deviations in the non-commutative framework of self-adjoint and classical Schatten $p$-classes. The self-adjoint setting is the one of the classical matrix ensembles which has already been introduced in Sect. 9.4.2.2 (to avoid introducing further notation, for the case of Schatten trace classes we refer the reader to [12] directly). In the spirit of [16], they proved a so-called Sanov-type large deviations principles for the spectral measure of $n^{1/p}$ multiples of random matrices chosen uniformly (or with respect to the cone measure on the boundary) from the unit balls of self-adjoint and non self-adjoint Schatten $p$-classes where $0 < p \leq +\infty$. The good rate function identified and the speed are quite different in the non-commutative setting and the rate is essentially given by the logarithmic energy (which is the negative of Voiculescu's free entropy introduced in [25]). Interestingly also a perturbation by a constant connected to the famous Ullman distribution appears. This constant already made an appearance in the recent works [10, 11], where the precise asymptotic volume of unit balls in classical matrix ensembles and Schatten trace classes were computed using ideas from the theory of logarithmic potentials with external fields.

The main result of [12] for the self-adjoint case is the following theorem, where we denote by $\mathcal{M}(\mathbb{R})$ the space of Borel probability measures on $\mathbb{R}$ equipped with the topology of weak convergence. On this topological space we consider the Borel $\sigma$-algebra, denoted by $\mathcal{B}(\mathcal{M}(\mathbb{R}))$.

**Theorem 9.5.6** *Fix $p \in (0, \infty)$ and $\beta \in \{1, 2, 4\}$. For every $n \in \mathbb{N}$, let $Z_n$ be a random matrix chosen according to the uniform distribution on $\mathbb{B}^n_{p,\beta}$ or the cone measure on its boundary. Then the sequence of random probability measures*

$$\mu_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{n^{1/p}\lambda_i(Z_n)}, \qquad n \in \mathbb{N},$$

*satisfies an LDP on* $\mathcal{M}(\mathbb{R})$ *with speed* $n^2$ *and good rate function* $\mathcal{I} : \mathcal{M}(\mathbb{R}) \to [0, +\infty]$ *defined by*

$$\mathcal{I}(\mu) = \begin{cases} -\frac{\beta}{2}\int_{\mathbb{R}}\int_{\mathbb{R}} \log|x-y|\,\mu(\mathrm{d}x)\,\mu(\mathrm{d}y) + \frac{\beta}{2p}\log\Big(\frac{\sqrt{\pi}\,p\Gamma(\frac{\beta}{2})}{2^p\sqrt{e}\,\Gamma(\frac{p+1}{2})}\Big) & \text{if } \int_{\mathbb{R}}|x|^p\mu(\mathrm{d}x) \le 1, \\ +\infty & \text{if } \int_{\mathbb{R}}|x|^p\mu(\mathrm{d}x) > 1. \end{cases}$$

Let us note that the case $p = +\infty$ as well as the case of Schatten trace classes is also covered in that paper (see [12, Theorems 1.3 and 1.5]). The proof of Theorem 9.5.6 requires to control *simultaneously* the deviations of the empirical measures and their $p$-th moments towards arbitrary small balls in the product topology of the weak topology on the space of probability measures and the standard topology on $\mathbb{R}$. It is then completed by proving exponential tightness. Moreover, they also use the probabilistic representation for random points in the unit balls of classical matrix ensembles which they have recently obtained in [10]. We close this survey by saying that as a consequence of the LDP in Theorem 9.5.6, they obtained that the spectral measure of $n^{1/p}Z_n$ converges weakly almost surely to a non-random limiting measure given by the Ullman distribution, as $n \to \infty$ (see [12, Corollary 1.4] for the self-adjoint case and [12, Corollary 1.6] for the non-self-adjoint case).

## References


1. D. Alonso-Gutiérrez, J. Prochno, C. Thäle, Large deviations for high-dimensional random projections of $\ell_p^n$-balls. Adv. Appl. Math. **99**, 1–35 (2018)
2. M. Anttila, K. Ball, I. Perissinaki, The central limit problem for convex bodies. English. Trans. Am. Math. Soc. **355**(12), 4723–4735 (2003)
3. S. Artstein-Avidan, A. Giannopoulos, V. Milman, *Asymptotic Geometric Analysis. Part I*. Mathematical Surveys and Monographs, vol. 202 (American Mathematical Society, Providence, 2015), pp. xx+451
4. F. Barthe, O. Guédon, S. Mendelson, A. Naor, A probabilistic approach to the geometry of the $l_p^n$-ball. Ann. Probab. **33**(2), 480–513 (2005)
5. S. Brazitikos, A. Giannopoulos, P. Valettas, B.-H. Vritsiou, *Geometry of Isotropic Convex Bodies*. Mathematical Surveys and Monographs, vol. 196 (American Mathematical Society, Providence, 2014), pp. xx+594
6. A. Dembo, O. Zeitouni, *Large Deviations. Techniques and Applications*. Stochastic Modelling and Applied Probability, vol. 38. Corrected reprint of the second (1998) edition (Springer, Berlin, 2010), pp. xvi+396
7. F. den Hollander, *Large Deviations*. Fields Institute Monographs, vol. 14 (American Mathematical Society, Providence, 2000), pp. x+143
8. P. Dirichlet, Sur une nouvelle méthode pour la détermination des intégrales multiples. J. Math. Pures Appl. **4**, 164–168 (1839)
9. N. Gantert, S. Kim, K. Ramanan, Large deviations for random projections of $\ell^p$ balls. Ann. Probab. **45**(6B), 4419–4476 (2017)
10. Z. Kabluchko, J. Prochno, C. Thäle, Exact asymptotic volume and volume ratio of Schatten unit balls. ArXiv e-prints (Apr. 2018). arXiv: 1804.03467 [math.FA]
11. Z. Kabluchko, J. Prochno, C. Thäle, Intersection of unit balls in classical matrix ensembles. Israel J. Math. (to appear). ArXiv e-prints (Apr. 2018). arXiv: 1804.03466 [math.FA]

12. Z. Kabluchko, J. Prochno, C. Thäle, Sanov-type large deviations in Schatten classes. Ann. Inst. H. Poincaré Probab. Statist. (to appear). ArXiv e-prints (Aug. 2018). arXiv: 1808.04862 [math.PR]
13. Z. Kabluchko, J. Prochno, C. Thäle, High-dimensional limit theorems for random vectors in $\ell_p^n$-balls. Commun. Contemp. Math. **21**, 1750092 (2019)
14. O. Kallenberg, *Foundations of Modern Probability*, Second. Probability and Its Applications (Springer, New York, 2002), pp. xx+638
15. S. Kim, Problems at the interface of probability and convex geometry: random projections and constrained processes. Ph.D. thesis. Brown University, 2017
16. S. Kim, K. Ramanan, A conditional limit theorem for high-dimensional $\ell^p$ spheres. ArXiv e-prints (Sept. 2015). arXiv: 1509.05442 [math.PR]
17. A. Naor, The surface measure and cone measure on the sphere of $\ell_p^n$. Trans. Am. Math. Soc. **359**(3), 1045–1079 (2007)
18. A. Naor, D. Romik, Projecting the surface measure of the sphere of $\ell_p^n$. English. Ann. Inst. Henri Poincaré, Probab. Stat. **39**(2), 241–261 (2003)
19. G. Pisier, *The Volume of Convex Bodies and Banach Space Geometry* (Cambridge University Press, Cambridge, 1999)
20. S. Rachev, L. Rüschendorf, Approximate independence of distributions on spheres and their stability properties. Ann. Probab. **19**(3), 1311–1337 (1991)
21. G. Schechtman, M. Schmuckenschläger, Another remark on the volume of the inter-section of two $L_p^n$ balls, in *Geometric Aspects of Functional Analysis (1989–90)*, vol. 1469. Lecture Notes in Mathematics (Springer, Berlin, 1991), pp. 174–178
22. G. Schechtman, J. Zinn, On the volume of the intersection of two $L_p^n$balls. Proc. Am. Math. Soc. **110**(1), 217–224 (1990)
23. M. Schmuckenschläger, Volume of intersections and sections of the unit ball of $\ell_p^n$. Proc. Am. Math. Soc. **126**(5), 1527–1530 (1998)
24. M. Schmuckenschläger, CLT and the volume of intersections of $\ell_p^n$-balls. English. Geom. Dedicata **85**(1–3), 189–195 (2001)
25. D. Voiculescu, The analogues of entropy and of Fisher's information measure in free probability theory. I. Commun. Math. Phys. **155**(1), 71–92 (1993)

# Chapter 10
# Remarks on Superconcentration and Gamma Calculus: Applications to Spin Glasses

**Kevin Tanguy**

**Abstract** This note is concerned with the so-called superconcentration phenomenon. It shows that the Bakry-Émery's Gamma calculus can provide relevant bound on the variance of function satisfying a inverse, integrated, curvature criterion. As an illustration, we present some variance bounds for the Free Energy in different models from Spin Glasses Theory.

## 10.1 Introduction

Superconcentration phenomenon has been introduced by Chatterjee in [7] and has given birth to a lot a work (cf. [15] for a survey). Each of these works, used various ad-hoc methods to improve upon sub-optimal bounds given by classical concentration of measure (cf. [4, 10]). In this note, we want to show that the celebrated Gamma calculus from Bakry and Émery's Theory is relevant to such improvements. To this task, we introduce an inverse, integrated, $\Gamma_2$ criterion which provides a useful bound on the variance of a particular function. As far as we know, this criterion seems to be new. We give below a sample of our modest achievement.

Denote by $\gamma_n$ the standard Gaussian measure on $\mathbb{R}^n$ and by $(P_t)_{t\geq 0}$ the standard Ornstein–Uhlenbeck semigroup. $\Gamma$ will stand for the so-called "carré du champ" operator, associated to the infinitesimal generator $L = \Delta - x\cdot\nabla$ of $(P_t)_{t\geq 0}$, and $\Gamma_2$ its iterated operator. We refer to Sect. 10.2 for more details about this topic.

**Theorem 10.1.1** *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a regular function and assume that there exists $\psi : \mathbb{R}_+ \to \mathbb{R}$ such that*

*(1) for any $t \geq 0$,*

$$\int_{\mathbb{R}^n} \Gamma_2(P_t f)d\gamma_n \leq \int_{\mathbb{R}^n} \Gamma(P_t f)d\gamma_n + \psi(t), \tag{10.1}$$

K. Tanguy (✉)
University of Angers, Angers, France
e-mail: kevin.tanguy@univ-angers.fr

 
N. Gozlan et al. (eds.), *High Dimensional Probability VIII*,
Progress in Probability 74, https://doi.org/10.1007/978-3-030-26391-1_10

*(2)*

$$\int_0^\infty e^{-2t} \int_t^\infty e^{2s} \psi(s) ds dt < \infty.$$

*Then the following holds*

$$\mathrm{Var}_{\gamma_n}(f) \leq \left| \int_{\mathbb{R}^n} \nabla f d\gamma_n \right|^2 + 4 \int_0^\infty e^{-2t} \int_t^\infty e^{2s} \psi(s) ds dt.$$

*with* $|\cdot|$ *the standard Euclidean norm.*

*Remark* Equation (10.14) can be seen as an inverse, integrated, curvature inequality for the function $f$.

As an application of Theorem 10.1.1, we show that some results due to Chatterjee can be expressed in terms of such criterion. From our point of view, this expression seems to ease the original scheme of proof and could possibly lead to various extensions. It also permits to easily recover some known variance bounds in Spin Glass Theory (cf. [5, 6, 11, 12]). Therefore let us present a short introduction to this theory.

Most of the time, in Spin Glasses Theory, it is customary to consider a centered Gaussian field $\left(H_n(\sigma)\right)_{\sigma \in \{-1,1\}^n}$ on the discrete cube $\{-1, 1\}^n$ (the map $\sigma \mapsto H_n(\sigma)$ is called the Hamiltonian of the system) and to focus on $\max_{\sigma \in \{-1,1\}^n} H_n(\sigma)$ (or $\min_{\sigma \in \{-1,1\}^n} H_n(\sigma)$). In general, this quantity is rather complex and presents a lack of regularity. Therefore, one focusses on a smooth approximation of the maximum (or the minimum) called the Free Energy $F_{n,\beta}$. This function is defined as follow

$$F_{n,\beta} = \pm \frac{1}{\beta} \log \left( \sum_{\sigma \in \{-1,1\}^n} e^{\pm \beta H_n(\sigma)} \right)$$

where $\beta > 0$ corresponds to (the inverse of) the temperature and its sign depends on whether you want to study the maximum or the minimum of $H_n$ over the discrete cube.

For instance, for the Random Energy Model (REM in short), we have

$$H_n(\sigma) = \sqrt{n} X_\sigma, \quad \sigma \in \{-1, 1\}^n$$

where $(X_\sigma)_{\sigma \in \{-1,1\}^n}$ is a sequence of i.i.d. standard Gaussian random variables.

For the Sherrington and Kirkpatrick's model (SK model in short), the Hamiltonian is more complex,

$$H_n(\sigma) = -\frac{1}{\sqrt{n}} \sum_{i,j=1}^{n} X_{ij}\sigma_i\sigma_j, \quad \sigma \in \{-1, 1\}^n$$

with $(X_{ij})_{1\leq i,j\leq n}$ a sequence of i.i.d. standard Gaussian random variables.

As an application of our methodology (cf. Sect. 10.4), we prove the following two Propositions.

**Proposition 10.1.1** *The following holds for the SK model. Let* $0 < \beta < \frac{1}{2}$*, then*

$$\mathrm{Var}(F_{n,\beta}) \leq C_\beta, \quad n \geq 1 \tag{10.2}$$

*where* $C_\beta > 0$ *is a constant depending only on* $\beta$*.*

*Remark* Talagrand obtained (cf. [11, 12]) such upper bound on the variance, for $0 < \beta < 1$, as a consequence of precise (and much harder to prove than our variance bounds) concentration inequalities for the Free Energy together with second moment method. As far as we know, it is the first time that such bound is obtained through semigroups arguments.

The methodology can also be used for the Random Energy Model (REM in short) (cf. Sect. 10.4 for more details) and provides the following bounds.

**Proposition 10.1.2** *The following holds in the REM.*

*High temperature regime: for* $0 < \beta < \frac{1}{\sqrt{2n}}$*, we have*

$$\mathrm{Var}_{\gamma_{2^n}}(F_{n,\beta}) \leq \frac{n}{2^n}\left(\frac{1-n\beta^2}{1-2n\beta^2}\right), \quad n \geq 1$$

*with* $C > 0$ *a universal constant.*

*Remark*

(1) The preceding bound has to be compared with the results exposed in [6, 7] (be careful with the different renormalization, in [6] the free energy is $\frac{F_{n,\beta}}{n}$). In [6], it is shown that

$$\mathrm{Var}_{\gamma_{2^n}}(F_{n,\beta}) \sim \frac{1}{2^n} \times \frac{e^{n\beta^2}}{\beta^2}, \quad \beta < \sqrt{\frac{\log 2}{2}}.$$

The dependance (in $n$ and $\beta$) is clearly not optimal in this regime but, as presented in Proposition 10.1.1, the scheme of proof of our method is robust enough to treat more complicated models. It seems natural that it can fail to capture precise behaviour such as the one obtained in [6]. Notice also that in [6], the authors obtained various (according to the temperature $\beta$) asymptotic convergence results for the (renormalized) Free Energy. Therefore, their results only indicate the correct order of the variance of this functional. However, to our best knowledge, this is the first time that such non-asymptotic bounds on the variance of the Free energy is obtained for the high temperature regime temperature.

(2) In [6] the low temperature regime was also investigated. Non-asymptotic variance bound, in accordance with the convergence results from Bovier et al., was already obtained in [7] and is presented and commented in Sect. 10.4 (Proposition 10.4.4) for the sake of completeness.

(3) As we will see latter in this note, it is easier to do the proof (of the preceding result) with the standard Gaussian measure on $\mathbb{R}^n$ and then to perform the following substitutions

$$n \longleftrightarrow 2^n \quad \text{and} \quad \beta \longleftrightarrow \sqrt{n}\beta$$

to fit the framework of [6].

This note is organized as follows. In Sect. 10.2, we recall some facts about superconcentration and Gamma calculus. In Sect. 10.3, we will prove our main results. Finally, in Sect. 10.4, we will give some applications in Spin Glass Theory.

## 10.2 Framework and Tools

In this section, we briefly recall some notions about superconcentration, Gamma calculus and interpolation methods by semigroups. General references about these topics could be, respectively, [1, 7].

### *10.2.1 Superconcentration*

It is well known (cf. [4, 10]), that concentration of measure of phenomenon is useful in various mathematical contexts. Such phenomenon can be obtained through functional inequalities. For instance, the standard Gaussian measure, on $\mathbb{R}^n$, $\gamma_n$ satisfies a Poincaré's inequality:

**Proposition 10.2.1** *For any function $f : \mathbb{R}^n \to \mathbb{R}$ smooth enough, the following holds*

$$\mathrm{Var}_{\gamma_n}(f) \leq \int_{\mathbb{R}^n} |\nabla f|^2 d\gamma_n \tag{10.3}$$

*where $|\cdot|$ stands for the Euclidean norm.*

Although this inequality holds for a large class of function, it could lead to sub-optimal bounds. A classical example is the function $f(x) = \max_{i=1,\ldots,n} x_i$. For such function, Poincaré's inequality implies that

$$\mathrm{Var}_{\gamma_n}(f) \leq 1$$

but it is known that $\mathrm{Var}_{\gamma_n}(f) \sim \frac{C}{\log n}$ for some constant $C > 0$. In Chatterjee's terminology, in this Gaussian framework, a function $f$ is said to be superconcentrated when Poincaré's inequality (10.3) is sub-optimal.

As we have said in the introduction, this phenomenon has been studied in various manner: semigroup interpolation [14], Renyi's representation of order statistics [3], Optimal Transport [15], Ehrard's inequality [17],...(cf. the Thesis [16] for a recent survey about superconcentration). In this note, we want to show that some differential inequalities between the operator $\Gamma$ and $\Gamma_2$ from Bakry and Émery's Theory could provide superconcentration.

### *10.2.2 Semigroups Interpolation and Gamma Calculus*

For more details about semigroups interpolation and $\Gamma$ calculus, we refer to [1, 9]. Although our work can easily be extended to a more general framework, we will focus on a Gaussian setting.

The Ornstein–Uhlenbeck process $(X_t)_{t\geq 0}$ is defined as follow:

$$X_t = e^{-t}X + \sqrt{1-e^{-2t}}Y, \quad t \geq 0,$$

with $X$ and $Y$ i.i.d. standard Gaussian vectors in $\mathbb{R}^n$. The semigroup $(P_t)_{t\geq 0}$, associated to this process, acts on a class of smooth function $\mathcal{A}$ (due to the integrability of Gaussian densities, one can choose here for $\mathcal{A}$ the class of $C^\infty$ functions whose derivatives are rapidly decreasing) and admits an explicit representation formula:

$$P_t f(x) = \int_{\mathbb{R}^n} f\big(xe^{-t} + \sqrt{1-e^{-2t}}y\big)d\gamma_n(y), \quad x \in \mathbb{R}^n,\ t \geq 0$$

Its infinitesimal generator is given by

$$L = \Delta - x \cdot \nabla$$

Furthermore, $\gamma_n$ is the invariant and reversible measure of $(P_t)_{t\geq 0}$. That is to say, for any function $f$ and $g$ belonging to $\mathcal{A}$,

$$\int_{\mathbb{R}^n} P_t f d\gamma_n = \int_{\mathbb{R}^n} f d\gamma_n \quad \text{and} \quad \int_{\mathbb{R}^n} f P_t g d\gamma_n = \int_{\mathbb{R}^n} g P_t f d\gamma_n.$$

Now, let us recall some properties satisfied by $(P_t)_{t\geq 0}$ which will be useful in the sequel.

**Proposition 10.2.2** *The Ornstein–Uhlenbeck semigroup $(P_t)_{t\geq 0}$ satisfies the following properties*

- *$P_t(f)$ is a solution of the heat equation associated to $L$*

$$\text{i.e.}\quad \partial_t(P_t f) = P_t(Lf) = L(P_t f). \tag{10.4}$$

- *$(P_t)_{t\geq 0}$ is ergodic, that is to say, for $f \in \mathcal{A}$*

$$\lim_{t\to+\infty} P_t(f) = \int_{\mathbb{R}^n} f\, d\gamma_n = \mathbb{E}_{\gamma_n}[f] \tag{10.5}$$

- *$(P_t)_{t\geq 0}$ commutes with the gradient $\nabla$. More precisely, for any function $f \in \mathcal{A}$,*

$$\nabla P_t(f) = e^{-t} P_t(\nabla f), \quad t \geq 0. \tag{10.6}$$

- *$(P_t)_{t\geq 0}$ is a contraction in $L^p(\gamma_n)$, for any function $f \in L^p(\gamma_n)$ and every $t \geq 0$,*

$$\|P_t(f)\|_p \leq \|f\|_p. \tag{10.7}$$

As it is exposed in [1], it is possible to give a dynamical representation of the variance of a function $f$ along the semigroup $(P_t)_{t\geq 0}$:

$$\mathrm{Var}_{\gamma_n}(f) = 2\int_0^\infty \int_{\mathbb{R}^n} |\nabla P_s(f)|^2 d\gamma_n ds = 2\int_0^\infty e^{-2s} \int_{\mathbb{R}^n} |P_s(\nabla f)|^2 d\gamma_n ds \tag{10.8}$$

### 10.2.3 Gamma Calculus and Poincaré's Inequality

Let us introduce the fundamental operator $\Gamma_2$ and $\Gamma$ from Bakry and Emery's Theory. Given an infinitesimal generator $L$ set, for $f$ and $g$, two smooth functions,

$$\Gamma(f,g) = \frac{1}{2}\big[L(fg) - fLg - Lfg\big] \quad \text{and} \quad \Gamma_2(f,g) = \frac{1}{2}\big[L\Gamma(f,g) - \Gamma(f,Lg) - \Gamma(Lf,g)\big]$$

In the case of the Ornstein–Uhlenbeck's infinitesimal generator $L = \Delta - x\cdot\nabla$, it is easily seen that

$$\Gamma(f) = |\nabla f|^2 \quad \text{and} \quad \Gamma_2(f) = \|\mathrm{Hess} f\|_2^2 + |\nabla f|^2 \tag{10.9}$$

where $\|\text{Hess} f\|_2 = \left(\sum_{i,j=1}^n \left(\frac{\partial^2 f}{\partial x_i \partial x_j}\right)^2\right)^{1/2}$ is the Hilbert–Schmidt norm of the tensor of the second derivatives of $f$.

Now, let us briefly recall how a relationship between $\Gamma$ and $\Gamma_2$ can be used to give a elementary proof of Poincaré's inequality (10.3).

First, notice that the representation formula of the variance (10.8) can be expressed in terms of $\Gamma$:

$$\text{Var}_{\gamma_n}(f) = 2\int_0^\infty \int_{\mathbb{R}^n} \Gamma(P_t f) d\gamma_n ds. \tag{10.10}$$

Then, observe that (10.9) implies the celebrated curvature-dimension criterion $CD(1, +\infty)$ (cf. [1])

$$\Gamma_2 \geq \Gamma. \tag{10.11}$$

Set $I(t) = \int_{\mathbb{R}^n} \Gamma(P_t f) d\gamma_n$. It is classical that

$$I'(t) = -2\int_{\mathbb{R}^n} \Gamma_2(P_t f) d\gamma_n, \quad t \geq 0$$

Thus, the inequality (10.11) leads to a differential inequality

$$\int_{\mathbb{R}^n} \Gamma_2(P_t f) d\gamma_n \geq \int_{\mathbb{R}^n} \Gamma(P_t f) d\gamma_n \Leftrightarrow 2I + I' \leq 0, \quad t \geq 0 \tag{10.12}$$

which can be easily integrated between $s$ and $t$ (with $0 \leq s \leq t$). That is

$$I(t)e^{2t} \leq I(s)e^{2s}.$$

It is now classical to let $s \to 0$ to easily recover Poincaré's inequality (10.3) for the measure $\gamma_n$. As we will see in the next section, we will show that a differential inequality of the form

$$I' \geq -2(I + \psi), \tag{10.13}$$

for some function $\psi$, can be used to obtain relevant bound (with respect to superconcentration phenomenon) on the variance of the function $f$ (being fixed) by letting $s$ fixed and $t \to +\infty$.

*Remark* Let us make few remarks.

(1) As it is proved in [1], the integrated curvature dimension inequality (10.12) is, in fact, equivalent to the Poincaré's inequality (10.3).
(2) As we will see in the next section, the inequality $I' \geq -2(I+\psi)$ is equivalent to an inverse, integrated, curvature dimension inequality which seems to be new. However, notice that the major difference between (10.12) and (10.13) is that

the first one holds for a large class of function whereas the second is only true for a particular function $f$ (and $\psi$ depends on $f$).

## 10.3 Inverse, Integrated, Curvature Inequality

In this section, we will use the methodology exposed in the preceding section to obtain variance bounds for a (fixed) function $f$ satisfying an inverse, integrated, curvature inequality $IC_{\gamma_n}(1, \psi)$.

First, let us state a definition. We want to highlight the fact that this definition will be stated in a Gaussian framework $(\mathbb{R}^n, \Gamma, \gamma_n)$ with $\Gamma$ associated to the infinitesimal generator $L = \Delta - x \cdot \nabla$ and the Ornstein–Uhlenbeck's semigroup $(P_t)_{t \geq 0}$. The next definition can be extended, mutatis mutandis, to fit the general framework of [1].

**Definition 10.3.1** Let $f : \mathbb{R}^n \to \mathbb{R}$ be a smooth function. We say that $f$ satisfy an inverse, integrated, curvature criterion with function $\psi : \mathbb{R}_+ \to \mathbb{R}$ if

$$\int_{\mathbb{R}^n} \Gamma_2(P_t f) d\gamma_n \leq \int_{\mathbb{R}^n} \Gamma(P_t f) d\gamma_n + \psi(t), \quad t \geq 0 \tag{10.14}$$

When the previous inequality is satisfied we denote it by $f \in IC_{\gamma_n}(1, \psi)$.

*Remark*

(1) Notice, again, that the inequality (10.14) holds, a priori, only for the function $f$.
(2) More generally, as it will be needed in the sequel, if $\mu$ is a Gaussian measure we will say that $f \in IC_\mu(1, \psi)$ if Eq. (10.14) is satisfied with $\mu$ instead of $\gamma_n$ and with the operators $\Gamma$ and $\Gamma_2$ associated to the Markov Triple $(\mathbb{R}^n, L, \mu)$.

Now, let us prove our main result Theorem 10.1.1.

*Proof (of Theorem 10.1.1)* Assume that $f \in IC_{\gamma_n}(1, \psi)$ (cf. Eq. (10.14)) holds. This is equivalent to the following differential inequality:

$$I' \geq -2(I + \psi), \tag{10.15}$$

where $I(t) = \int_{\mathbb{R}^n} |\nabla P_t f|^2 d\gamma_n$, $t \geq 0$. Set $I(t) = K(t)e^{-2t}$, inequality (10.15) becomes

$$K'(t) \geq -2e^{2t}\psi(t), \quad t \geq 0 \tag{10.16}$$

Now, integrate inequality (10.16) between $s$ and $t$. That is

$$K(t) - K(s) \geq -2 \int_s^t e^{2u}\psi(u)du, \quad \text{for all} \quad 0 \leq s \leq t.$$

Then, let $t \to \infty$, this yields

$$K(s) \le \big[\lim_{t\to\infty} K(t)\big] + 2\int_s^{\infty} e^{2u}\psi(u)du, \quad s \ge 0,$$

To conclude, observe that

$$K(t) = I(t)e^{2t} \to_{t\to\infty} \left|\int_{\mathbb{R}^n} \nabla f d\gamma_n\right|^2$$

by ergodicity of $(P_t)_{t\ge 0}$. Finally, we have, for every $t \ge 0$,

$$I(t) = \int_{\mathbb{R}_n} \Gamma(P_t f)d\gamma_n \le e^{-2t}\left(\left|\int_{\mathbb{R}^n} \nabla f d\gamma_n\right|^2 + 2\int_t^{\infty} e^{2s}\psi(s)ds\right). \tag{10.17}$$

It suffices to use the dynamical representation of the variance (10.8) with elementary calculus to end the proof. □

*Remark* This method of interpolation, between $t$ and $+\infty$, has also been used in [13] in order to obtain Talagrand's inequality of higher order.

### *10.3.1 Another Variance Bound*

As we will see in the last section, it is sometimes useful to restrict an $IC_\mu(1,\psi)$, for some probability measure $\mu$, up to a time $T$ in order to improve the dependance with respect to some parameter.

In other words, the setting is the following: assume that an $IC_\mu(1,\psi)$ holds and that we are able to produce some $T > 0$ such that the bound of $I(T)$ (given by Eq. (10.17)) is particularly nice (with respect to some parameter). Now, we have to bound the variance in a different manner in order to use the information on $I(T)$. To this task, we will prove the next proposition.

**Proposition 10.3.1** *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function smooth enough. Then, for any $T > 0$*

$$\mathrm{Var}_{\gamma_n}(f) \le \frac{2TI(0)}{1-e^{-2T}}\left[\frac{1}{\log a} - \frac{1}{a\log a}\right]$$

*with $a = \frac{I(0)}{I(T)}$ and $I(t) = \int_{\mathbb{R}^n}\Gamma(P_t f)d\gamma_n$.*

*Remark* This proposition will be used to show that the Free Energy is superconcentrated for some Spin Glasses models. Although we stated the preceding Proposition 10.3.1 for the standard Gaussian measure $\gamma_n$, it will also hold (up to obvious renormalization) for $\mu$ the law of a centered Gaussian vector with covariance matrix $M$.

To prove the preceding theorem, we will need two further arguments.

First, we present an inequality due to Cordero-Erausquin and Ledoux [8]. The proof of this inequality rests on the fact that the Poincaré's inequality satisfied by $\gamma_n$ implies an exponential decay of the variance along the semigroup $(P_t)_{t\geq 0}$.

**Lemma 10.3.1 (Cordero-Erausquin–Ledoux)** *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function smooth enough. Then, for any $T > 0$, the following holds*

$$\mathrm{Var}_{\gamma_n}(f) \leq \frac{2}{1-e^{-2T}} \int_0^T I(t)dt \tag{10.18}$$

*with $I(t) = \int_{\mathbb{R}^n} \Gamma(P_t f) d\gamma_n$.*

*Proof* For the sake of completeness we give the proof of the preceding Lemma.

$$\begin{aligned}
\mathrm{Var}_{\gamma_n}(f) &= \mathbb{E}_{\gamma_n}[f^2] - \mathbb{E}_{\gamma_n}[(P_T f)^2] + \mathbb{E}_{\gamma_n}[(P_T f)^2] - \mathbb{E}_{\gamma_n}[P_T f]^2 \\
&= -\int_0^T \frac{d}{ds}\mathbb{E}_{\gamma_n}[(P_s f)^2]ds + \mathrm{Var}_{\gamma_n}(P_T f) \\
&\leq 2\int_0^T I(s)ds + e^{-2T}\mathrm{Var}_{\gamma_n}(f).
\end{aligned}$$

□

Secondly, we will use the fact that the infinitesimal generator $(-L)$ of the Ornstein–Uhlenbeck process $(X_t)_{t\geq 0}$ admits a (discrete) spectral decomposition. Then, denote by $dE_\lambda$ the spectral resolution of $(-L)$. According to [1], this leads to a different representation of $t \mapsto I(t)$. With $f : \mathbb{R}^n \to \mathbb{R}$ being fixed, we have:

$$I(t) = \int_{\mathbb{R}^n} |\nabla P_t f|^2 d\gamma_n = \int_0^\infty \lambda e^{-2\lambda t} dE_\lambda(f), \quad t \geq 0$$

As it is proven in [2] (cf. Corollary 5.6), $t \mapsto I(t)$ satisfies, with the preceding representation, an Hölder-type inequality. That is to say, for every $T > 0$,

**Lemma 10.3.2 (Baudoin–Wang)**

$$I(s) \leq I(0)^{1-s/T} I(T)^{s/T}, \quad 0 \leq s \leq T \tag{10.19}$$

Now, we can prove Proposition 10.3.1 with the help of preceding Lemma.

*Proof (of Proposition 10.3.1)* First use Lemma 10.3.1 to get

$$\mathrm{Var}_{\gamma_n}(f) \leq \frac{2}{1-e^{-2T}} \int_0^T I(t)dt.$$

Then, use Lemma 10.3.2. This yields

$$\begin{aligned}\mathrm{Var}_{\gamma_n}(f) &\leq \frac{2}{1-e^{-2T}} \int_0^T I(0)^{1-t/T} I(T)^{t/T} dt \\ &= \frac{2I(0)}{1-e^{-2T}} \int_0^T e^{-\frac{t}{T}\log a} dt\end{aligned}$$

where $a = \frac{I(0)}{I(T)} \geq 1$ and $I(t) = \int_{\mathbb{R}^n} \Gamma(P_t f) d\gamma_n$. Finally, elementary calculus ends the proof. □

## 10.4 Application in Spin Glasses's Theory

In the remaining of this section, we will show how Theorem 10.1.1 can be used to provide relevant bounds on the variance of $F_{n,\beta}$. We will focus on the REM and the SK Model. For the remaining of this note we will denote by $f_\beta$, for $\beta > 0$, the following function

$$f_\beta(x) = \frac{1}{\beta} \log\Big(\sum_{i=1}^n e^{\beta x_i}\Big), \quad x = (x_1, \ldots, x_n) \in \mathbb{R}^n$$

### *10.4.1 Random Energy Model*

In this section we will show how Theorem 10.1.1 is useful to obtain relevant bound on the variance of the Free Energy $F_{n,\beta}$ (with $\beta$ close to 0) for the REM.

**Proposition 10.4.1** *For any $\beta > 0$, $f_\beta \in IC_{\gamma_n}(1, \psi)$ with*

$$\psi(t) = 2\beta^2 e^{-2t} I(t)$$

*where, let us recall it, $I(t) = \int_{\mathbb{R}^n} \Gamma(P_t f_\beta) d\gamma_n$ and $\Gamma$ is the standard "carré du champ" operator.*

We will need the following Lemma to prove the preceding Proposition.

**Lemma 10.4.1** *Let $(u_i)_{i=1,\ldots,n}$ be a family of functions, with $u_i : \mathbb{R}^n \to \mathbb{R}$ for any $i = 1, \ldots, n$, satisfying the following condition*

$$\sum_{i=1}^n u_i^2(x) \leq 1 \quad \textit{for all} \quad x \in \mathbb{R}^n$$

*Then, for any function $v : \mathbb{R}^n \to \mathbb{R}_+$ and any probability measure $\mu$, we have*

$$\sum_{i=1}^n \left( \int_{\mathbb{R}^n} u_i(x)v(x)d\mu(x) \right)^2 \le \left( \int_{\mathbb{R}^n} v d\mu \right)^2$$

*Proof* Consider the vector $U = (u_1 v, \dots, u_n v) \in \mathbb{R}^n$ and recall that $|\cdot|$ stands for the Euclidean norm. Then, it holds

$$\left[ \sum_{i=1}^n \left( \int_{\mathbb{R}^n} u_i(x)v(x)d\mu \right)^2 \right]^{1/2} = \left| \int_{\mathbb{R}^n} U d\mu \right| \le \int_{\mathbb{R}^n} |U| d\mu = \int_{\mathbb{R}^n} \left[ \sum_{i=1}^n u_i^2(x) \right]^{1/2} v(x)d\mu$$
$$\le \int_{\mathbb{R}^n} v(x)d\mu$$

where the first upper bound comes from Jensen's inequality. □

Now we turn to the proof of Proposition 10.4.1.

*Proof (of Proposition 10.4.1)* First, observe that the condition $IC_{\gamma_n}(1, \psi)$ is equivalent to

$$\int_{\mathbb{R}^n} \Gamma_2\big(P_t(f_\beta)\big)d\gamma_n \le (1 + 2\beta^2 e^{-2t}) \int_{\mathbb{R}^n} \Gamma\Big(P_t(f_\beta)\Big)d\gamma_n, \quad t \ge 0.$$

That is (since $\Gamma_2(f) = \|\mathrm{Hess} f\|_2^2 + |\nabla f|^2$ and $\Gamma(f) = |\nabla f|^2$)

$$\int_{\mathbb{R}_n} \|\mathrm{Hess} P_t(f_\beta)\|_2^2 d\gamma_n \le 2\beta^2 e^{-2t} \int_{\mathbb{R}^n} |\nabla P_t(f_\beta)|^2 d\gamma_n, \quad t \ge 0. \tag{10.20}$$

Now, observe that, pointwise, Eq.(10.20) is equivalent to (thanks to the commutation property between $\nabla$ and $(P_t)_{t\ge 0}$)

$$\sum_{i,j=1}^n [P_t(\partial_{ij}^2 f_\beta)]^2 \le 2\beta^2 \sum_{i=1}^n [P_t(\partial_i f_\beta)]^2, \quad \forall t \ge 0$$

Elementary calculus yields, for every $i = 1, \dots, n$, and every $\beta > 0$,

$$\partial_i f_\beta = \frac{e^{\beta x_i}}{\sum_{k=1}^n e^{\beta x_k}}$$

and, for every $j = 1, \dots, n$,

$$\partial_j \partial_i f_\beta = \beta(\partial_i f_\beta \delta_{ij} - \partial_i f_\beta \partial_j f_\beta).$$

Thus, for every $t \geq 0$,

$$\sum_{i,j=1}^{n} [P_t(\partial_{ij}^2 f_\beta)]^2 = \beta^2 \sum_{i=1}^{n} \big[P_t(\partial_i f_\beta)\big]^2 - 2\beta \sum_{i=1}^{n} P_t(\partial_i f_\beta) P_t\big[(\partial_i f_\beta)^2\big] + \beta^2 \sum_{i,j=1}^{n} \big[P_t(\partial_i f_\beta \partial_j f_\beta)\big]^2.$$

First ignore the crossed terms (which are always non positive), then apply Lemma 10.4.1 to the third term.

Indeed, let $i \in \{1, \dots, n\}$ be fixed and set $u_j = \partial_j f_\beta$ and $v = \partial_i f_\beta$. Thus, Lemma 10.4.1 implies

$$\sum_{j=1}^{n} \big[P_t(\partial_i f_\beta \partial_j f_\beta)\big]^2 \leq P_t^2(\partial_i f_\beta).$$

This inequality finally yields,

$$\sum_{i,j=1}^{n} [P_t(\partial_{ij}^2 f_\beta)]^2 \leq \beta^2 \sum_{i=1}^{n} \big[P_t(\partial_i f_\beta)\big]^2 + \beta^2 \sum_{i,j=1}^{n} \big[P_t(\partial_i f_\beta \partial_j f_\beta)\big]^2 \leq 2\beta^2 \sum_{i=1}^{n} \big[P_t(\partial_i f_\beta)\big]^2.$$

□

Now, the criterion $IC_{\gamma_n}(1, \psi)$ can be used gives to provide relevant bound on the variance of $F_{n,\beta}$ as stated in Proposition 10.1.2.

*Proof (of Proposition 10.1.2)* As mentioned earlier, the proof will be done for the standard Gaussian measure on $\mathbb{R}^n$ and then it will be enough to perform a change of variable. As it will be useful in the sequel, observe that (by symmetry) the following holds

$$\int_{\mathbb{R}^n} \partial_i f_\beta d\gamma_n = \frac{1}{n}, \quad \forall i = 1, \dots, n.$$

Now, let $\beta > 0$ and use Theorem 10.1.1 which implies that

$$\mathrm{Var}_{\gamma_n}(F_{n,\beta}) \leq \frac{1}{n} + 4\beta^2 \int_0^\infty e^{-2s}(1 - e^{-2s}) \sum_{i=1}^{n} \int_{\mathbb{R}^n} P_s^2(\partial_i f_\beta) d\gamma_n ds \tag{10.21}$$

where we used Fubini's Theorem and the commutation property between $\nabla$ and $P_s$.

For the first bound, when $\beta \in \big(0, \frac{\sqrt{2}}{2}\big)$, it is possible to rewrite (thanks to the dynamical representation of the variance (10.21)) the integral in the right hand side as

$$2\beta^2 \mathrm{Var}_{\gamma_n}(F_{n,\beta}) - 4\beta^2 \int_0^\infty e^{-4s} \sum_{i=1}^{n} \int_{\mathbb{R}^n} P_s^2(\partial_i f_\beta) d\gamma_n ds \tag{10.22}$$

Furthermore, by Jensen's inequality and the invariance of $(P_t)_{t\geq 0}$ with respect to $\gamma_n$, we have

$$\int_{\mathbb{R}^n} P_s^2(\partial_i f_\beta) d\gamma_n \geq \left( \int_{\mathbb{R}^n} P_s(\partial_i f_\beta) d\gamma_n \right)^2 = \frac{1}{n^2}, \quad \forall i = 1, \dots, n, \quad \forall s > 0$$

Thus, $\mathrm{Var}_{\gamma_n}(F_{n,\beta}) \leq \left(\frac{1-\beta^2}{1-2\beta^2}\right)\frac{1}{n}$.

To conclude, as announced, it is enough to substitute $n$ by $2^n$ and $\beta$ by $\sqrt{n}\beta$ to get the result. □

*Remark* Incidentally, the preceding proof can be used to get a lower bound on the variance of the Free energy. More precisely, it is possible to deduce from (10.21) and (10.22) the following lower bound

$$\mathrm{Var}_{\gamma_{2^n}}(F_{n,\beta}) \geq \frac{n}{2^n} \frac{(1 - n\beta^2)}{(1 - 2\beta^2 n)}, \quad \text{for} \quad \beta > \frac{1}{\sqrt{2n}}$$

### *10.4.2 SK Model*

In this section we show how some work of Chatterjee (from [7]) can be rewritten in term of an inverse, integrated, curvature criterion. Then, it allows us to easily recover a bound, obtained by Talagrand (cf. [11, 12]), on the variance of the Free Energy for the SK model at high temperature.

First, we need to express the $\Gamma$ and $\Gamma_2$ operator when $\gamma_n$ is replaced by $\mu$ the law of a centered Gaussian vector, in $\mathbb{R}^n$, with covariance matrix $M$.

Let $X$ be a random Gaussian vector with $\mathcal{L}(X) = \mu$ and consider $Y$ an independent copy of $X$. It is then possible to define the generalized Ornstein–Uhlenbeck process, which we will still denote by $(X_t)_{t\geq 0}$, as follow

$$X_t = e^{-t} X + \sqrt{1 - e^{-2t}}\, Y, \quad t \geq 0$$

Similarly, we also denote by $(P_t)_{t\geq 0}$ the associated semigroup. Then, it is known (cf. [7, 14, 16]) that, for any smooth function $f : \mathbb{R}^n \to \mathbb{R}$,

$$I(t) = \int_{\mathbb{R}^n} \Gamma(P_t f) d\mu = 2 \int_{\mathbb{R}^n} e^{-2t} \sum_{i,j} M_{ij} (\partial_i f) P_t(\partial_j f) d\mu, \quad t \geq 0$$

As we will see latter, it will be more convenient to work with

$$I_r(t) = 2 \int_{\mathbb{R}^n} e^{-2t} \sum_{i,j} (M_{ij})^r (\partial_i f) P_t(\partial_j f) d\mu, \quad t \geq 0$$

where $r$ is a positive integer. In the rest of this section, we choose $f = f_\beta$.

**Proposition 10.4.2 (Chatterjee)** *Assume that* $M_{ij} \geq 0$ *for all* $(i, j) \in \{1, \dots, n\}^2$. *Then, for any* $t \geq 0$, *the following holds*

$$I_r'(t) \geq -2\big[I_r(t) + 2\beta^2 e^{-2t} J_{r+1}(t)\big] \tag{10.23}$$

*with* $J_r(t) = e^{2t} I_r(t)$.

*Remark*

(1) In [7], Chatterjee proved that $J_r'(t) \geq -4\beta^2 e^{-2t} J_{r+1}(t)$ for any $r \in \mathbb{N}^*$. The proof is similar the proof of Lemma 10.4.1 with the additional use of Hölder's inequality.
(2) In particular, when $r = 1$, Chatterjee's proposition amounts of saying that

$$f_\beta \in IC_\mu(1, \psi)$$

with $\psi(t) = 2\beta^2 e^{-2t} J_2(t)$. Unfortunately, it remains hard to upper bound this quantity by something relevant.

As observed in the preceding remark, the inverse, integrated, curvature criterion can not be used in the present form. However, it is possible to recycle the arguments of Sect. 10.3. That is, use $l$ times, with $l \in \mathbb{N}$, the fundamental Theorem of analysis (on $t \mapsto I_r(t)$) together with the inequality (10.23) and let $l \to +\infty$. This leads to a useful bound on the function $t \mapsto I_r(t)$ for any $r \in \mathbb{N}^*$.

**Theorem 10.4.1 (Chatterjee)** *Assume that* $M_{ij} \geq 0$ *for all* $(i, j) \in \{1, \dots, n\}^2$. *Then, for any* $t \geq 0$, *the following holds*

$$I_r(t) \leq e^{-2t} \sum_{i,j=1}^{n} (M_{ij})^r e^{2\beta^2 e^{-2t} M_{ij}} \nu_i \nu_j, \quad \forall r \geq 1 \tag{10.24}$$

*where* $\nu_i = \int_{\mathbb{R}^n} \partial_i f_\beta d\mu$ *for all* $i = 1, \dots, n$.

*Remark* When $r = 1$, the main step of Chatterjee's proof is equivalent to show that $f_\beta \in IC_\mu(1, \psi)$ with $\psi(t) = 2\beta^2 e^{-2t} \sum_{i,j=1} M_{ij} e^{2\beta^2 e^{-2t} M_{ij}} \nu_i \nu_j$. The proof of this result can be found in [7, pp. 108–110] .

Unfortunately, the repeated use of the differential inequality (10.23) degrades the upper bound on $t \mapsto I_r(t)$. As we will briefly see in the next subsection, Chatterjee used Eq. (10.24) only for a fixed $T > 0$ (large enough). We show, in the next Proposition, that this bound (for $r = 1$) is still relevant to recover some work of Talagrand on the variance of $F_{n,\beta}$, with small $\beta$, for the SK model (cf. [11, 12]).

Now, let us prove Proposition 10.1.1.

*Proof (of Proposition 10.1.1)* First we show that inequality (10.24) leads to a general upper bound on the variance of $F_{n,\beta}$ which might be of independent interest. Then, we choose $M$ to be the covariance structure of the $SK$ model and proved inequality (10.2).

When $r = 1$, Eq. (10.24) combined with Eq. (10.10) implies that, for any $\beta > 0$,

$$\begin{aligned}\mathrm{Var}_\mu(F_{n,\beta}) &\leq 2\int_0^\infty e^{-2t} \sum_{i,j=1}^n M_{ij} e^{2\beta^2 e^{-2t} M_{ij}} \nu_i \nu_j dt \\ &\leq \frac{1}{2\beta^2} \sum_{i,j=1}^n e^{2\beta^2 M_{ij}} \nu_i \nu_j\end{aligned}$$

Following Chatterjee (cf. [7]), choose $M$ to be the covariance structure of the SK model. That is,

$$M_{\sigma\sigma'} = \left(\frac{1}{\sqrt{n}}\sum_{i=1}^n \sigma_i \sigma_i'\right)^2, \quad \forall \sigma, \sigma' \in \{-1, 1\}^n.$$

Besides, observe (by symmetry) that, for each $\sigma \in \{-1, 1\}^n$,

$$\nu_\sigma = \mathbb{E}_\mu\big[\partial_\sigma F_{n,\beta}\big] = \frac{1}{2^n}.$$

Thus,

$$\mathrm{Var}_\mu(F_{n,\beta}) \leq \frac{1}{2\beta^2}\mathbb{E}_{\sigma,\sigma'}\Big[e^{2\beta^2\left(\frac{1}{\sqrt{n}}\sigma_i\sigma_i'\right)^2}\Big]$$

where $\mathbb{E}_{\sigma'\sigma}$ stands for the expectation under the product measure induced by the Rademacher random variables $\sigma_i, \sigma_i', \ i = 1, \ldots, n$.

Finally, if $\beta \in \left(0, \frac{1}{2}\right)$ we have $\mathbb{E}_{\sigma,\sigma'}\Big[e^{2\beta^2\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n \sigma_i\sigma_i'\right)^2}\Big] = C(\beta)$. Indeed, observe first that $\sum_{i=1}^n \sigma_i\sigma_i'$ has the same distribution as $\sum_{i=1}^n \sigma_i$. Then, it is enough to use Hoeffding's inequality (cf.[4]), which gives the following deviation inequality

$$\mathbb{P}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n \sigma_i > t\right) \leq e^{-t^2/2} \quad t \geq 0,$$

to conclude. □

### 10.4.3 Improvements of Variance Bounds with Respect to the Parameter β

Let us collect some results of Chatterjee and briefly explain how Proposition 10.3.1 can be used to improve the dependence of the variance bounds with respect to $\beta$. However, the dependance in $n$ will be worse.

Chatterjee used, in [7], a Theorem of Bernstein about completely monotone function. As far as we are concerned, the spectral framework exposed in Sect. 10.3 seems to be more natural to work with and provides equivalent results.

The arguments, in order to improve the dependance in $\beta$, can be summarize as follow: choose $T$ such that $I(T)$ can be bounded by a relevant quantity and apply Proposition 10.3.1.

**Proposition 10.4.3 (Chatterjee)** *In the $SK$ model the following holds*

$$\mathrm{Var}_\mu(F_{n,\beta}) \leq \frac{C_1 n \log(2 + C_2\beta)}{\log n}, \quad \forall \beta > 0$$

*with $C_1, C_2 > 0$ two numerical constants.*

*Remark* Here $T > 0$ is chosen such that

$$\mathbb{E}_{\sigma,\sigma'}\left[M_{\sigma\sigma'} e^{2\beta^2 e^{-2T} M_{\sigma\sigma'}}\right] = C_\beta, \quad \forall \beta > 0$$

where $M_{\sigma\sigma'} = \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \sigma_i \sigma_i'\right)^2$ and $C_\beta > 0$ is a constant that does not depend on $n$. That is $T = \frac{1}{2} \log\left(\frac{2\beta^2}{\gamma}\right)$ for some sufficiently small constant $\gamma > 0$ (cf. [7]).

**Proposition 10.4.4 (Chatterjee)** *In the REM, the following holds for $\beta > 2\sqrt{\log 2}$,*

$$\mathrm{Var}_\mu(F_{n,\beta}) \leq C_\beta$$

*where $C_\beta > 0$ is a constant that does not depend on $n$.*

*Remark* Here $T$ is chosen as $T = \frac{1}{2}\log(2\beta^2)$ so that $I(T) \leq \frac{n}{2^n} e^{-2T} e^n$ and the upper bound is relevant in the low temperature regime (cf. [6, 7]). Again, notice the difference of renormalization with Proposition 10.1.2 (one has to replace the number of random variables $n$ by $2^n$ and the i.i.d. standard Gaussian random variables $(X_i)_{i=1,\ldots,2^n}$ by $\sqrt{n}X_i$ in the Proposition). In [7], Chatterjee also proved that the upper bound is tight.

In fact, it also possible to use hypercontractive arguments instead of Theorem 10.4.1 to achieve the upper bound of Proposition 10.4.4. Indeed, one can use the inequality (10.21) together with hypercontractive estimates of $(P_t)_{t\geq 0}$ (cf. [7, 8, 15, 16]). More precisely, we have

$$\|P_s(\partial_i f_\beta)\|_2^2 \leq \|\partial_i f_\beta\|_{1+e^{-2s}}^2, \quad \forall i = 1, \ldots, n, \quad \forall s > 0$$

It is then standard, cf. Section 4 in [16] for instance, to prove that

$$\int_0^\infty e^{-2s}(1 - e^{-2s})\|\partial_i f_\beta\|_{1+e^{-2s}}^2 ds \leq \frac{C\|\partial_i f_\beta\|_2^2}{\left[1 + \log \frac{\|\partial_i f_\beta\|_2}{\|\partial_i f_\beta\|_1}\right]^2}$$

where $C > 0$ is a numerical constant. Then, it is elementary to conclude. Notice that such estimates are already implicit in the celebrated $L^1/L^2$ Talagrand's inequality (presented in [7, 8] for instance), which one can also be directly used to recover the content of Proposition 10.4.4.

**Acknowledgements** I thank M. Ledoux for fruitful discussions on this topic. I also warmly thank the referee for helpful comments in improving the exposition and the simplification of the proof of Lemma 10.4.1.

## References

1. D. Bakry, I. Gentil, M. Ledoux, *Analysis and Geometry of Markov Diffusion Operators*. Grundlehren der Mathematischen Wissenschaften, vol. 348 (Springer, Berlin, 2014)
2. F. Baudoin, J. Wang, Curvature dimension inequalities and subelliptic heat kernel gradient bounds on contact manifolds. Potential Anal. **40**, 163–193 (2014)
3. S. Boucheron, M. Thomas, Concentration inequalities for order statistics. Electron. Commun. Probab. **17**, 1–12 (2012)
4. S. Boucheron, G. Lugosi, P. Massart, *Concentration Inequalities: A Nonasymptotic Theory of Independance* (Oxford University Press, Oxford, 2013)
5. A. Bovier, *Gaussian Processes on Trees: From Spin Glasses to Branching Brownian Motion*. Cambridge Studies in Advanced Mathematics, vol. 163 (Cambridge University Press, Cambridge, 2016)
6. A. Bovier, I. Kurkova, M. Löwe, Fluctuations of the free energy in the REM and the $p$-spin SK models. Ann. Probab. **30**(2), 605–651 (2002)
7. S. Chatterjee, *Superconcentration and Related Topics* (Springer, Berlin, 2014)
8. D. Cordero-Erausquin, M. Ledoux, Hypercontractive measures, Talagrand's inequality, and influences, in *Geometric Aspects of Functional Analysis*. Lectures Notes in Mathematics, vol. 2050 (Springer, Berlin, 2012), pp.169–189
9. M. Ledoux, The geometry of Markov diffusions operators. Ann. Fac. Sci. Toulouse Math. 6 **9**(2), 305–366 (2000)
10. M. Ledoux, *The Concentration of Measure Phenomenon*. Mathematical Surveys and Monographs, vol. 89 (American Mathematical Society, Providence, 2001)
11. M. Talagrand, *Mean Field Models for Spin Glasses. Volume I*. Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge. A Series of Modern Surveys in Mathematics (Springer, Berlin, 2011)
12. M. Talagrand, *Mean Field Models for Spin Glasses. Volume II: Advanced Replica-Symmetry and Low Temperature*, Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge. A Series of Modern Surveys in Mathematics, vol. 55 (Springer, Heidelberg, 2011)
13. K. Tanguy, Talagrand's inequality at higher order and application to Boolean analysis (2019). https://arxiv.org/pdf/1801.08931.pdf
14. K. Tanguy, Some superconcentration inequalities for extrema of stationary Gaussian processes. Stat. Probab. Lett. **106**, 239–246 (2015)
15. K. Tanguy, Non asymptotic variance bounds and deviation inequalities by optimal transport (2017). Preprint, http://arxiv.org/abs/1708.08620
16. K. Tanguy, *Quelques inégalités de concentration: théorie et applications (in French)*. Ph.D. thesis, Institute of Mathematics of Toulouse, 2017
17. P. Valettas, On the tightness of Gaussian concentration for convex functions. J. Anal. Math. (to appear)

# Chapter 11
# Asymptotic Behavior of Rényi Entropy in the Central Limit Theorem

**Sergey G. Bobkov and Arnaud Marsiglietti**

**Abstract** We explore an asymptotic behavior of Rényi entropy along convolutions in the central limit theorem with respect to the increasing number of i.i.d. summands. In particular, the problem of monotonicity is addressed under suitable moment hypotheses.

**Keywords** Rényi entropy · Central limit theorem

**2010 Mathematics Subject Classification** Primary 60E, 60F

## 11.1 Introduction

Given a (continuous) random variable $X$ with density $p$, the associated Rényi entropy and Rényi entropy power of index $r$ $(1 < r < \infty)$ are defined by

$$h_r(X) = -\frac{1}{r-1} \log \int_{-\infty}^{\infty} p(x)^r \, dx, \qquad N_r(X) = e^{2h_r(X)} = \left( \int_{-\infty}^{\infty} p(x)^r \, dx \right)^{-\frac{2}{r-1}}.$$

Being translation invariant and homogeneous of order 2, the functional $N_r$ is similar to the variance and is often interpreted as measure of uncertainty hidden in the distribution of $X$. Another representation

$$N_r(X)^{-\frac{1}{2}} = \left(\mathbb{E}\, p(X)^{r-1}\right)^{\frac{1}{r-1}}$$

S. G. Bobkov
School of Mathematics, University of Minnesota, Minneapolis, MN, USA
e-mail: bobko001@umn.edu

A. Marsiglietti (✉)
Department of Mathematics, University of Florida, Gainesville, FL, USA
e-mail: a.marsiglietti@ufl.edu

 
N. Gozlan et al. (eds.), *High Dimensional Probability VIII*,
Progress in Probability 74, https://doi.org/10.1007/978-3-030-26391-1_11

shows that $N_r$ is non-increasing in $r$, so that $0 \leq N_\infty \leq N_r \leq N_1 \leq \infty$. Here, for the extreme indexes, the Rényi entropy power is defined by the monotonicity,

$$N_\infty(X) = \lim_{r \uparrow \infty} N_r(X) = \|p\|_\infty^{-2}, \qquad N_1(X) = \lim_{r \downarrow 1} N_r(X) = e^{2h_1(X)},$$

where $\|p\|_\infty$ is the essential supremum of $p(x)$. In the case $r = 1$, we arrive at the Shannon differential entropy $h_1(X) = h(X) = -\int p(x) \log p(x)\, dx$ with entropy power $N_1 = N = e^{2h}$ (provided that $N_r(X) > 0$ for some $r > 1$).

Much of the analysis about the Shannon and Rényi entropies is focused on the behavior of these functionals on convolutions, i.e., for sums $S_n = X_1 + \cdots + X_n$ of independent random variables (including a multidimensional setting). First, let us recall a fundamental entropy power inequality, which may be written in terms of the normalized sums $Z_n = S_n/\sqrt{n}$ as

$$N(Z_n) \geq \frac{1}{n} \sum_{k=1}^{n} N(X_k). \tag{11.1}$$

There are also some extensions of this relation to the Rényi case (cf. [4, 5, 9, 10]).

When $X_k$'s are independent and identically distributed (i.i.d.), with mean zero and variance one, the central limit theorem (CLT) asserts that $Z_n \Rightarrow Z$ with weak convergence in distribution to the Gaussian limit $Z \sim N(0, 1)$. In this case, the right-hand side of (11.1) is constant, while the sequence on the left is monotone, as was shown by Artstein, Ball, Barthe and Naor [1], cf. also [12] (the inequality (11.1) itself ensures that $N(Z_n)$ are non-decreasing along the values $n = 2^l$). Moreover, by another important result due to Barron [2], we have the entropic CLT: $N(Z_n)$ are convergent to the entropy power $N(Z)$, as long as $N(Z_{n_0}) > 0$ for some $n_0$.

These results give rise to a number of natural questions about an asymptotic behavior of the Rényi entropy powers $N_r(Z_n)$. In particular, when do they converge to $N_r(Z)$, and if so, what is the rate of convergence? Is the monotonicity still true? As we will see, such questions may be studied, at least partially, under suitable moment assumptions.

Let us state a few observations in these directions, assuming throughout that $X, X_1, X_2, \dots$ are i.i.d. random variables with $\mathbb{E}X = 0$ and $\mathrm{Var}(X) = 1$. Put $\beta_s = \mathbb{E}\,|X|^s$ for real $s \geq 2$. In order to describe necessary and sufficient conditions for the convergence of the Rényi entropies in the CLT, we also introduce the common characteristic function

$$f(t) = \mathbb{E}\, e^{itX} \qquad (t \in \mathbb{R}).$$

**Theorem 11.1.1** *Given $1 < r \leq \infty$, we have the convergence $N_r(Z_n) \to N_r(Z)$ or equivalently $h_r(Z_n) \to h_r(Z)$ as $n \to \infty$, if and only if*

$$\int_{-\infty}^{\infty} |f(t)|^\nu \, dt < \infty \quad for\ some\ \ \nu \geq 1. \tag{11.2}$$

*Equivalently, this holds if and only if $Z_n$ have bounded densities for all* (*some*) $n$ *large enough.*

This characterization coincides with the one for the uniform local limit theorem due to Gnedenko, cf. [11]. Since (11.2) is equivalent to the property that $Z_n$ have bounded and hence bounded $C^k$-smooth densities for any fixed $k$ and all $n$ large enough, it is often referred to as the smoothing condition. In general, (11.2) is stronger than what is needed in the entropic case $r = 1$. In this connection, let us note that there is still no explicit description such as (11.2) for the validity of the entropic CLT in terms of the characteristic function $f(t)$.

Once (11.2) is fulfilled, one may ask about the rate of convergence in Theorem 11.1.1, which may be guaranteed assuming that the absolute moment $\beta_s$ is finite for some $s > 2$. Moreover, in this case one may obtain asymptotic expansions for $N_r(Z_n)$ in powers of $1/n$ similarly to the entropic expansions derived in [8]. They involve the moments of $X$ up to order $m = [s]$, or equivalently—the cumulants

$$\gamma_k = i^{-k}\,(\log f)^{(k)}(0), \qquad k = 1, \dots, m.$$

In the Gaussian case $X \sim N(0,1)$, all cumulants are vanishing, starting with $k = 2$. In the general case, they indicate how close a given distribution to the normal. As for the asymptotic behavior of Rényi's entropies, it turns out that a special role is played by the quantity

$$b = b(r) = -\frac{1}{r}\left[\frac{2-r}{12}\,\gamma_3^2 + \frac{r-1}{8}\,\gamma_4\right].$$

Here, $\gamma_3 = \mathbb{E}X^3$ and $\gamma_4 = \mathbb{E}X^4 - 3$, while for the extreme indexes, one may just put

$$b(1) = \lim_{r \to 1} b(r) = -\frac{1}{12}\,\gamma_3^2, \qquad b(\infty) = \lim_{r \to \infty} b(r) = \frac{1}{12}\,\gamma_3^2 - \frac{1}{8}\,\gamma_4.$$

This can be seen from the following assertion.

**Theorem 11.1.2** *Suppose that the smoothing condition* (11.2) *is fulfilled. If* $\beta_s$ *is finite for* $2 \le s < 4$, *then for any* $1 < r < \infty$,

$$h_r(Z_n) = h_r(Z) + o(n^{-\frac{s-2}{2}}), \qquad N_r(Z_n) = N_r(Z) + o(n^{-\frac{s-2}{2}}). \tag{11.3}$$

*Moreover, in case* $4 \le s < 6$,

$$h_r(Z_n) = h_r(Z) + b\,n^{-1} + o(n^{-\frac{s-2}{2}}), \tag{11.4}$$
$$N_r(Z_n) = N_r(Z)\left(1 + 2b\,n^{-1}\right) + o(n^{-\frac{s-2}{2}}).$$

This assertion remains valid in the entropic case $r = 1$ as well (with a slight logarithmic improvement in the remainder $o$-term, cf. [8]). In case $s = 6$, the remainder term may be improved to $O(n^{-2})$, and in fact, one may add quadratic terms to get an expansion

$$h_r(Z_n) = h_r(Z) + b\,n^{-1} + b_2 n^{-2} + o(n^{-2}) \tag{11.5}$$

with some functional $b_2 = b_2(r)$ depending also on $\gamma_5$ and $\gamma_6$. Regardless of its value, one may therefore conclude about an eventual monotonicity of $N_r(Z_n)$ based on the sign of $b$. Moreover, the above expansions continue to hold for $r = \infty$, so that this case may be included as well.

**Theorem 11.1.3** *Suppose that the smoothing condition* (11.2) *is fulfilled, and let $\beta_6$ be finite. Given $1 < r \le \infty$, there exists $n_0 \ge 1$ such that the sequence $N_r(Z_n)$ is increasing for $n \ge n_0$, whenever $b(r) < 0$, that is, if*

$$\frac{2-r}{3}\,\gamma_3^2 + \frac{r-1}{2}\,\gamma_4 > 0 \quad (1 < r < \infty), \qquad \gamma_4 > \frac{2}{3}\,\gamma_3^2 \quad (r = \infty).$$

*This sequence is decreasing for $n \ge n_0$, if $b(r) > 0$.*

In particular, under the last condition $\gamma_4 > \frac{2}{3}\gamma_3^2$, the sequence $N_r(Z_n)$ is eventually increasing for any fixed $r \ge 1$. For example, this holds for $X = \frac{\xi-\alpha}{\sqrt{\alpha}}$, where the random variable $\xi$ has a Gamma distribution with $\alpha$ degrees of freedom (in which case $\gamma_3 = 2/\sqrt{\alpha}$ and $\gamma_4 = 6/\alpha$).

On the other hand, if $X$ is uniformly distributed in the interval $(-\sqrt{3}, \sqrt{3})$, then $\gamma_3 = 0$, $\gamma_4 = -6/5$, so $N_r(Z_n)$ is eventually decreasing for any $r > 1$, although the opposite property takes place for $r = 1$.

The paper is organized as follows. We start with the proof of Theorem 11.1.1 (Sect. 11.2), and then collect together basic results on Edgeworth expansions for densities $p_n$ of $Z_n$ (Sect. 11.3). They are used in Sects. 11.4–11.5 to construct a formal asymptotic expansion for $L^r$-norms of $p_n$ in powers of $1/n$ up to order $[\frac{m-2}{2}]$ with remainder term as in (11.3)–(11.4). One particular case, where the first moments of $X$ agree with those of $Z \sim N(0,1)$, is discussed separately in Sect. 11.6, while the range $4 \le s \le 8$ in such expansion is treated in Sect. 11.7. The transition to the Rényi entropy is performed in Sect. 11.8, where Theorem 11.1.2 is proved. Some comparison with the entropic CLT is given in Sect. 11.9, with remarks leading to Theorem 11.1.3 for finite $r$. Finally, the index $r = \infty$ is treated separately in Sect. 11.10. We refer to [6] for an extended version of the article where more computational details are provided.

## 11.2 Proof of Theorem 11.1.1

From now on, let $X, X_1, X_2, \dots$ be i.i.d. random variables with $\mathbb{E}X = 0$ and $\mathrm{Var}(X) = 1$, for which we define the normalized sums

$$Z_n = \frac{X_1 + \cdots + X_n}{\sqrt{n}}, \qquad n = 1, 2, \dots$$

First, let us recall Gnedenko's uniform local limit theorem. Assuming the smoothing condition (11.2), it asserts that, for all $n$ large enough, the random variables $Z_n$ have bounded densities $p_n$, and moreover, in that case as $n \to \infty$,

$$\sup_x |p_n(x) - \varphi(x)| \to 0. \tag{11.6}$$

Here, as usual,

$$\varphi(x) = \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2} \qquad (x \in \mathbb{R})$$

denotes the density of the standard normal random variable $Z$. Clearly, the property (11.6) is also necessary for the uniform boundedness of $p_n$'s.

Let us explain the equivalence of the two conditions—in terms of the characteristic function as in (11.2), and in terms of densities (via the existence of a bounded density). Since $|f(t)| \le 1$ for all $t$, the property (11.2) is getting weaker for growing $\nu$, so it is sufficient to consider integer values of $\nu$. Since $Z_n$ has characteristic function

$$f_n(t) = \mathbb{E}\, e^{itZ_n} = f(t/\sqrt{n})^n,$$

(11.2) implies that $Z_n$ has a bounded, continuous density $p_n$ for $n = \nu$, by the Fourier inversion formula. Hence the same is true for all $n \ge \nu$, by the convolution character of the distributions of $Z_n$. Conversely, suppose that $Z_n$ has a bounded density $p_n$ for $n = n_0$. This implies that $p_n \in L^r(\mathbb{R})$ for any $r \ge 1$, with norm

$$\|p_n\|_r = \left(\int_{-\infty}^{\infty} p_n(x)^r\, dx\right)^{1/r},$$

and in particular $p_n \in L^2(\mathbb{R})$. By Plancherel's theorem, the characteristic function $f_n$ is also in $L^2(\mathbb{R})$. But this means that (11.2) is fulfilled with $\nu = 2n_0$.

Also note that, under the condition (11.2), we have $f_\nu(t) \to 0$ as $t \to \infty$ (the Riemann-Lebesgue lemma), and thus $f(t) \to 0$. Hence, (11.2) represents a sharpening of the Cramér condition $\limsup_{t\to\infty} |f(t)| < 1$, which is used to establish a number of asymptotic results related to the CLT. In particular, using the Fourier inversion formula, one can easily obtain (11.6) and actually a sharper statement such as

$$\sup_x \,(1+x^2)\,|p_n(x)-\varphi(x)| \to 0 \qquad (n\to\infty). \tag{11.7}$$

*Proof of Theorem 11.1.1* First, let $r=\infty$. As explained, the smoothing condition (11.2) implies the uniform local limit theorem (11.6). In turn, the latter yields $\|p_n\|_\infty \to \|\varphi\|_\infty$, that is, $N_\infty(Z_n)\to N_\infty(Z)$ as $n\to\infty$. Conversely, this convergence ensures that $N_\infty(Z_n)>0$ for all $n$ large enough, that is, $\|p_n\|_\infty<\infty$. As was also emphasized, this implies (11.2).

Now, let $1<r<\infty$. If $N_r(Z_n)\to N_r(Z)$ as $n\to\infty$, then $N_r(Z_n)>0$ for all $n$ large enough, say $n\ge n_0$. Equivalently, for such $n$, $Z_n$ have densities $p_n$ with $\|p_n\|_r<\infty$. If $r\ge 2$, then $\|p_n\|_2\le 1+\|p_n\|_r<\infty$, so that $p_n$ and therefore $f_n$ are in $L^2(\mathbb{R})$. This means that (11.2) is fulfilled for $\nu=2n_0$. In the case $1<r<2$, one may apply the Hausdorff-Young inequality

$$\|\hat u\|_{r'}\le\|u\|_r, \qquad \text{where} \quad \hat u(t)=\int_{-\infty}^{\infty} e^{2\pi itx}\,u(x)\,dx, \quad r'=\frac{r}{r-1}.$$

It implies that $\|f_n\|_{r'}\le\sqrt{2\pi}\,\|p_n\|_r<\infty$, which means that (11.2) is fulfilled for $\nu=r'n_0$.

Thus, the smoothing condition (11.2) is indeed necessary. To argue in the other direction, we apply the uniform local limit theorem: For all $n\ge n_0$ large enough, $Z_n$ have densities $p_n$, bounded by a constant $M$ and moreover, the relation (11.6) holds true, i.e.,

$$\sup_x \big|p_n(x)^r-\varphi(x)^r\big|\le\varepsilon_n\to 0 \qquad (n\to\infty). \tag{11.8}$$

For a given $\varepsilon>0$, applying the usual central limit theorem, one may pick up $T>0$ such that

$$\mathbb{P}\{|Z_n|>T\}+\mathbb{P}\{|Z|>T\}<\varepsilon, \qquad n\ge n_1\ge n_0.$$

Hence

$$\int_{|x|>T} p_n(x)^r\,dx \;\le\; M^{r-1}\int_{|x|>T} p_n(x)\,dx \;=\; M^{r-1}\,\mathbb{P}\{|Z_n|>T\} \;<\; M^{r-1}\varepsilon,$$

and similarly for $\varphi(x)$. Hence

$$\bigg|\int_{|x|>T} p_n(x)^r\,dx-\int_{|x|>T}\varphi(x)^r\,dx\bigg|<M^{r-1}\varepsilon. \tag{11.9}$$

On the other hand, by (11.8),

$$\left| \int_{|x|\le T} p_n(x)^r\,dx - \int_{|x|\le T} \varphi(x)^r\,dx \right| \le \int_{|x|\le T} |p_n(x)^r - \varphi(x)^r|\,dx \le 2T\varepsilon_n \le \varepsilon,$$

where the last inequality holds for all $n \ge n_2$ with some $n_2 \ge n_1$. Together with (11.9), we get

$$\big|\, \|p_n\|_r^r - \|\varphi\|_r^r \big| < (M^{r-1}+1)\,\varepsilon, \qquad n \ge n_2.$$

That is, $\|p_n\|_r^r \to \|\varphi\|_r^r$ as $n \to \infty$, thus proving the theorem. □

## 11.3 Limit Theorems About Edgeworth Expansions

As is well-known, in case of the finite 3-rd absolute moment $\beta_3 = \mathbb{E}\,|X|^3$, and assuming the smoothness condition (11.2), the local limit theorems (11.6)–(11.7) can be sharpened to

$$\sup_x\,(1+|x|^3)\,|p_n(x) - \varphi(x)| = o\Big(\frac{1}{\sqrt{n}}\Big) \qquad (n \to \infty). \tag{11.10}$$

Here, the rate cannot be improved in general. However, under higher order moment assumptions, the limit normal density may slightly be modified, which leads to the sharpening of the right-hand side of (11.10). Namely, if $\beta_m = \mathbb{E}\,|X|^m$ is finite for an integer $m \ge 2$, one may introduce the cumulants

$$\gamma_k = i^{-k}\,(\log f)^{(k)}(0), \qquad k = 1, \dots, m.$$

They represent certain polynomials in the moments $\alpha_i = \mathbb{E}X^i$ up to order $k$, namely,

$$\gamma_k = k! \sum (-1)^{j-1}\,(j-1)!\,\frac{1}{r_1! \dots r_k!}\,\Big(\frac{\alpha_1}{1!}\Big)^{r_1} \dots \Big(\frac{\alpha_k}{k!}\Big)^{r_k},$$

where $j = r_1 + \dots + r_k$ and where the summation is running over all tuples $(r_1, \dots, r_k)$ of non-negative integers such that $r_1 + 2r_2 + \dots + kr_k = k$.

For example, with our moment assumptions $\mathbb{E}X = 0$, $\mathrm{Var}(X) = 1$, we have $\gamma_1 = 0$, $\gamma_2 = 1$, $\gamma_3 = \alpha_3$, $\gamma_4 = \alpha_4 - 3$.

**Definition 11.3.1** An Edgeworth correction of the standard normal law of order $m$ for the distribution of $Z_n$ is a finite signed measure $\nu_m$ with density

$$\varphi_m(x) = \varphi(x) + \varphi(x) \sum_{k=1}^{m-2} Q_k(x)\,n^{-k/2}, \tag{11.11}$$

where

$$Q_k(x) = \sum \frac{1}{r_1! \dots r_k!} \left(\frac{\gamma_3}{3!}\right)^{r_1} \dots \left(\frac{\gamma_{k+2}}{(k+2)!}\right)^{r_k} H_{k+2j}(x). \tag{11.12}$$

Here, the summation is running over all collections of non-negative integers $r_1, \dots, r_k$ such that $r_1 + 2r_2 + \dots + kr_k = k$, with notation $j = r_1 + \dots + r_k$.

As usual, $H_k$ denotes the Chebyshev-Hermite polynomial of degree $k$ with leading term $x^k$. The polynomial $Q_k$ in (11.11) has degree at most $3(m-2)$ in the variable $x$. The index $m$ for $\varphi_m$ indicates that the cumulants up to $\gamma_m$ participate in the construction. The sum in (11.11) may also be viewed as a polynomial in $1/\sqrt{n}$ of degree at most $m-2$.

For example, $\varphi_2 = \varphi$, and there are no terms in the sum (11.11). For $m = 3, 4, 5, 6$, in (11.12) we correspondingly have

$$Q_1(x) = \frac{\gamma_3}{3!} H_3(x),$$

$$Q_2(x) = \frac{\gamma_3^2}{2!\,3!^2} H_6(x) + \frac{\gamma_4}{4!} H_4(x),$$

$$Q_3(x) = \frac{\gamma_3^3}{3!^4} H_9(x) + \frac{\gamma_3\gamma_4}{3!\,4!} H_7(x) + \frac{\gamma_5}{5!} H_5(x),$$

$$Q_4(x) = \frac{\gamma_3^4}{4!\,3!^4} H_{12}(x) + \frac{\gamma_3^2\gamma_4}{2!\,3!^2\,4!} H_{10}(x) + \frac{\gamma_3\gamma_5}{3!\,5!} H_8(x) + \frac{\gamma_4^2}{2!\,4!^2} H_8(x) + \frac{\gamma_6}{6!} H_6(x).$$

Moreover, if the first $m-1$ moments of $X$ coincide with those of $Z \sim N(0,1)$, then the first $m-1$ cumulants of $X$ are vanishing, and (11.11) is simplified to

$$\varphi_m(x) = \varphi(x)\Big(1 + \frac{\gamma_m}{m!} H_m(x)\, n^{-\frac{m-2}{2}}\Big), \qquad \gamma_m = \mathbb{E}X^m - \mathbb{E}Z^m. \tag{11.13}$$

The following observation, generalizing and refining the non-uniform local limit theorems (11.7) and (11.10), is due to Petrov [14], cf. also [3, 15]. From now on, we always assume that the smoothing condition (11.2) is fulfilled.

**Lemma 11.3.2** *If $\beta_m < \infty$ for an integer $m \geq 2$, then as $n \to \infty$*

$$\sup_x \, (1 + |x|^m)\, |p_n(x) - \varphi_m(x)| \; = \; o\big(n^{-\frac{m-2}{2}}\big). \tag{11.14}$$

Without the polynomial weight $1 + |x|^m$, a similar result was earlier obtained by Gnedenko. However, in some applications the appearance of this weight turns out to be crucial.

If $m \geq 3$, one may also take $\varphi_{m-1}$ as an approximation of $p_n$, and then (11.14) together with Definition 11.3.1 imply that

$$\sup_x \, (1 + |x|^m) \, |p_n(x) - \varphi_{m-1}(x)| \, = \, O\big(n^{-\frac{m-2}{2}}\big). \tag{11.15}$$

A further generalization was given in [7] to employ the case of fractional moments.

**Lemma 11.3.3** *Let $\beta_s < \infty$ for some real $s \geq 2$, and $m = [s]$. Then uniformly over all $x$, as $n \to \infty$,*

$$(1+|x|^s)\,(p_n(x)-\varphi_m(x)) \, = \, o\big(n^{-\frac{s-2}{2}}\big)+(1+|x|^{s-m})\,\Big(O\big(n^{-\frac{m-1}{2}}\big)+o\big(n^{-(s-2)}\big)\Big).$$

*In particular, for some constant $\alpha > 0$ depending on $s$,*

$$\sup_{|x| \leq n^\alpha} \, (1 + |x|^s) \, |p_n(x) - \varphi_m(x)| \, = \, o\big(n^{-\frac{s-2}{2}}\big). \tag{11.16}$$

Thus, (11.16) extends (11.14) when taking the supremum over relatively large interval.

There are also similar results about the distribution functions $F_n(x) = \mathbb{P}\{Z_n \leq x\}$, which may be approximated by

$$\Phi_m(x) \, = \, \nu_m((-\infty, x]) \, = \, \int_{-\infty}^{x} \varphi_m(y) \, dy \, = \, \Phi(x) - \varphi(x) \sum_{k=1}^{m-2} R_k(x) \, n^{-k/2}, \tag{11.17}$$

where

$$R_k(x) = \sum \frac{1}{r_1! \dots r_k!} \left(\frac{\gamma_3}{3!}\right)^{r_1} \dots \left(\frac{\gamma_{k+2}}{(k+2)!}\right)^{r_k} H_{k+2j-1}(x)$$

with summation as in Definition 11.3.1. The next result is due to Osipov and Petrov [13].

**Lemma 11.3.4** *Suppose that $\beta_s < \infty$ for some real $s \geq 2$, and let $m = [s]$. Then, as $n \to \infty$,*

$$\sup_x \, (1 + |x|^s) \, |F_n(x) - \Phi_m(x)| \, = \, o\big(n^{-\frac{s-2}{2}}\big).$$

*In particular, when $s = m \geq 3$ is integer, we have*

$$\sup_x \, (1 + |x|^s) \, |F_n(x) - \Phi_{m-1}(x)| \, = \, O\big(n^{-\frac{s-2}{2}}\big).$$

This statement holds under the weaker assumption in comparison with (11.2): nothing should be required in case $2 \le s < 3$, while for $s \ge 3$ the Cramér condition is sufficient.

*Remark 11.3.5* Since the densities $p_n$ can properly be approximated by the functions $\varphi_m$, it makes sense to isolate the leading term in the sum (11.11), by rewriting the definition as

$$\varphi_m(x) = \varphi(x) + \varphi(x) \frac{\gamma_{k+2}}{(k+2)!} H_{k+2}(x)\, n^{-k/2} + \varphi(x) \sum_{j=k+1}^{m-2} Q_j(x)\, n^{-j/2} \tag{11.18}$$

for some unique $1 \le k \le m-2$. The value of $k$ is the maximal one in the interval $[1, m-2]$ such that $\gamma_3 = \cdots = \gamma_{k+1} = 0$, which means that the first moments of $X$ up to order $k+1$ coincide with those of $Z \sim N(0,1)$. In this case, necessarily $\gamma_{k+2} = \mathbb{E}X^{k+2} - \mathbb{E}Z^{k+2}$.

Of course, if $m = 2$, there are no terms on the right-hand side of (11.18) except for $\varphi$.

## 11.4 Approximation for $L^r$-Norm of Densities $p_n$

Lemmas 11.3.2–11.3.4 can be applied to explore an asymptotic behavior of the functionals

$$I(p) = \|p\|_r^r = \int_{-\infty}^{\infty} p(x)^r\, dx \qquad (r > 1)$$

with $p = p_n$. Since the densities $p_n$ are approximated by $\varphi_m$, we may expect that $I(p_n) \sim I(\varphi_m)$ for large $n$. However, $\varphi_m$ do not need to be positive on the whole real line, and it is more natural to consider the integrals

$$I_T(p) = \int_{|x| \le T} p(x)^r\, dx, \qquad T > 0,$$

over relatively long intervals. Actually, one may take $T = T_n = \sqrt{(s-2)\log n}$ $(s > 2)$. We have with some constants depending on the first $m$ absolute moments of $X$ that

$$\sum_{k=1}^{m-2} |Q_k(x)|\, n^{-k/2} \le C\,(1+|x|)^{3(m-2)} \frac{1}{\sqrt{n}} \le C' \frac{(\log n)^{3(m-2)/2}}{\sqrt{n}} \le \frac{1}{2}, \qquad |x| \le T_n,$$

for all $n$ large enough in the last inequality. Hence, by Definition 11.3.1, for all $n$ large enough,

$$|\varphi_m(x) - \varphi(x)| \leq \frac{1}{2}\,\varphi(x), \qquad |x| \leq T_n, \tag{11.19}$$

so $\varphi_m$ is positive on $[-T_n, T_n]$. On these intervals and for large $n$, consider the functions

$$\varepsilon_n(x) = \frac{p_n(x) - \varphi_m(x)}{\varphi_m(x)}.$$

By (11.16) and (11.19), for $|x| \leq T_n$, we have

$$|\varepsilon_n(x)| \leq 2\delta_n \frac{n^{-\frac{s-2}{2}}}{\varphi(x)} \leq 2\sqrt{2\pi}\,\delta_n,$$

for some positive sequence $\delta_n \to 0$. Thus, for large $n$, $p_n(x) = \varphi_m(x)(1 + \varepsilon_n(x))$ with $|\varepsilon_n(x)| \leq \frac{1}{2}$. Hence, by Taylor's formula, and using (11.19) together with the non-uniform bound (11.16), we get

$$\begin{aligned} |p_n(x)^r - \varphi_m(x)^r| &\leq c\,\varphi(x)^r\,|\varepsilon_n(x)| \\ &\leq 2c\,\varphi(x)^{r-1}\,|p_n(x) - \varphi_m(x)| \,\leq\, \delta_n\,\frac{\varphi(x)^{r-1}}{1+|x|^s}\,n^{-\frac{s-2}{2}} \end{aligned}$$

with some constant $c$ which does not depend on $x$ and $n \geq n_0$ and some positive sequence $\delta_n \to 0$. After integration over $[-T_n, T_n]$, this gives

$$I_T(p_n) = I_T(\varphi_m) + o(n^{-\frac{s-2}{2}}). \tag{11.20}$$

In case $s = m \geq 3$ is integer, by a similar argument based on (11.15), we also have

$$I_T(p_n) = I_T(\varphi_{m-1}) + O(n^{-\frac{s-2}{2}}). \tag{11.21}$$

The remaining part of the integral,

$$J_T(p) = \int_{|x|>T} p(x)^r\,dx,$$

can be shown to be sufficiently small for $p = p_n$ on the basis of Lemma 11.3.4. Indeed, first

$$\mathbb{P}\{|Z| > T_n\} \leq \frac{1}{T_n}\,e^{-T_n^2/2} = o\big(n^{-\frac{s-2}{2}}\big), \qquad Z \sim N(0,1).$$

On the other hand, by Definition 11.3.1, using polynomial bounds $|Q_k(x)| \le c_k\,(1+|x|^N)$ with $N = 3(m-2)$ and some constants $c_k$ which do not depend on $x$, we have

$$|\varphi_m(x)| \le \varphi(x) + \frac{c}{\sqrt{n}}\,(1+|x|^N)\,\varphi(x)$$

with some $c$ independent of $x$ and $n$. In addition,

$$\int_{|x|>T_n} |x|^N\,\varphi(x)\,dx \ \le\ c_N'\,(1+T_n^N)\,e^{-T_n^2/2} \ \le\ c_N''\ \log(n)^{\frac{N}{2}}\,n^{-\frac{s-2}{2}}$$

with constants $c_N'$ and $c_N''$ independent of $n$. This gives

$$\begin{aligned}\big|\nu_m\{|x|>T_n\}\big| &\le \int_{|x|>T_n} |\varphi_m(x)|\,dx \\ &\le \int_{|x|>T_n} \varphi(x)\,dx + \frac{c}{\sqrt{n}}\int_{|x|>T_n} (1+|x|^N)\,\varphi(x)\,dx \\ &\le \mathbb{P}\{|Z|>T_n\} + \frac{c_N''}{\sqrt{n}}\ \log(n)^{\frac{N}{2}}\,n^{-\frac{(s-2)}{2}},\end{aligned}$$

and thus

$$\big|\nu_m\{|x|>T_n\}\big| \ =\ o\big(n^{-\frac{s-2}{2}}\big).$$

Since we assume the smoothness condition (11.2), the densities $p_n$ are uniformly bounded by some constant $M$ for all $n \ge n_0$. Therefore, by Lemma 11.3.4, for all $n$ large enough,

$$\begin{aligned}J_T(p_n) &\le M^{r-1}\int_{|x|>T_n} p_n(x)\,dx \ =\ M^{r-1}\,\mathbb{P}\{|Z_n|>T_n\} \\ &\le M^{r-1}\,\big|\nu_m\{x: |x|>T_n\}\big| + T_n^{-s}\,o\big(n^{-\frac{s-2}{2}}\big) \ =\ o\big(n^{-\frac{s-2}{2}}\big).\end{aligned}$$

Combining this relation with (11.20) and (11.21), we arrive at:

**Lemma 11.4.1** *Suppose that $\beta_s < \infty$ for $s \ge 2$. Then for all $n$ large enough, $Z_n$ have bounded densities $p_n$. Moreover, for any $r > 1$, as $n \to \infty$,*

$$\int_{-\infty}^{\infty} p_n(x)^r\,dx \ =\ \int_{-T_n}^{T_n} \varphi_m(x)^r\,dx + o\big(n^{-\frac{s-2}{2}}\big), \qquad m = [s], \tag{11.22}$$

*where $T_n = \sqrt{(s-2)\log n}$. In particular, if $s = m \ge 3$ is integer, we also have*

$$\int_{-\infty}^{\infty} p_n(x)^r\,dx \ =\ \int_{-T_n}^{T_n} \varphi_{m-1}(x)^r\,dx + O\big(n^{-\frac{s-2}{2}}\big). \tag{11.23}$$

## 11.5 Truncated $L^r$-Norm of Approximating Densities $\varphi_m$

Let us now find an explicit expression for the second integral in (11.22), by applying the Edgeworth approximation

$$\varphi_m(x) = \varphi(x) \Big(1 + \sum_{k=1}^{m-2} Q_k(x)\, n^{-k/2}\Big), \qquad m = [s]. \tag{11.24}$$

In the case $2 < s < 3$, when $\varphi_m = \varphi_2 = \varphi$, one may extend the integration in (11.22) to the whole real line at the expense of the error

$$\int_{|x|>T_n} \varphi(x)^r\, dx \ < \int_{|x|>T_n} \varphi(x)\, dx \ = \ \mathbb{P}\{|Z| > T_n\} \ = \ o\big(n^{-\frac{s-2}{2}}\big),$$

where $T_n = \sqrt{(s-2)\log n}$ as before. Hence, (11.22) yields

$$\int_{-\infty}^{\infty} p_n(x)^r\, dx \ = \int_{-\infty}^{\infty} \varphi(x)^r\, dx + o\big(n^{-\frac{s-2}{2}}\big), \qquad 2 < s < 3. \tag{11.25}$$

This assertion remains to hold for $s = 2$ as well (Theorem 11.1.1).

Next, assume that $s \geq 3$. As we know, when $n$ is large enough, $\varphi_m(x)$ is positive for $|x| \leq T_n$, so the second integral in (11.22) makes sense, cf. (11.19). Moreover, in order to raise $\varphi_m(x)$ to the power $r$ on the basis of (11.24), one may apply the Taylor expansion

$$(1+\varepsilon)^r \ = \ 1 + \sum_{k=1}^{N} \frac{(r)_k}{k!}\, \varepsilon^k + O(\varepsilon^{N+1}), \qquad N = 1, 2, \dots, \ \ (\varepsilon \to 0),$$

where the constant in $O$ depends on $N$ only, as long as $|\varepsilon| \leq \frac{1}{2}$. Here we used the standard notation $(r)_k = r(r-1)\dots(r-k+1)$, with convention $(r)_0 = 1$ to be used later on. Choosing

$$\varepsilon = \sum_{k=1}^{m-2} Q_k(x)\, n^{-k/2}, \qquad |x| \leq T_n,$$

for all $n$ large enough the above Taylor expansion is thus valid. Hence, uniformly over all $x \in [-T_n, T_n]$, as $n \to \infty$,

$$(1+\varepsilon)^r \ = \ 1 + \sum_{k=1}^{N} \frac{(r)_k}{k!}\, \varepsilon^k + \varepsilon_n(x) \tag{11.26}$$

with

$$\varepsilon_n(x) = O\Big((1+|x|)^{3(m-2)(N+1)}\, n^{-(N+1)/2}\Big).$$

Furthermore, by the polynomial formula,

$$\varepsilon^k = \sum \frac{k!}{k_1! \dots k_{m-2}!}\, Q_1^{k_1}(x) \dots Q_{m-2}^{k_{m-2}}(x)\, n^{-\frac{1}{2}\,(k_1+2k_2+\dots+(m-2)\,k_{m-2})},$$

where the summation is running over all non-negative integers $k_1, \dots, k_{m-2}$ such that $k_1 + \dots + k_{m-2} = k$. Inserting this in (11.26) and recalling (11.24), we can represent $\varphi_m(x)^r$ as

$$\varphi(x)^r \sum \frac{(r)_{k_1+\dots+k_{m-2}}}{k_1! \dots k_{m-2}!}\, Q_1^{k_1}(x) \dots Q_{m-2}^{k_{m-2}}(x)\, n^{-\frac{1}{2}\,(k_1+2k_2+\dots+(m-2)\,k_{m-2})} + \varphi(x)^r\, \varepsilon_n(x)$$

with summation over all non-negative integers $k_1, \dots, k_{m-2}$ such that $k_1 + \dots + k_{m-2} \le N$. One may now note that

$$\int_{-T_n}^{T_n} \varphi(x)^r\, \varepsilon_n(x)\, dx = O\big(n^{-\frac{N+1}{2}}\big),$$

where the constant in $O$ depends on $N$ only, as long as $n$ is large enough.

Let us then choose $N = m-2$. Integrating the above expression for $\varphi_m(x)^r$ over the interval $[-T_n, T_n]$, we can represent $\int_{-T_n}^{T_n} \varphi_m(x)^r\, dx$ as

$$\sum \frac{(r)_{k_1+\dots+k_{m-2}}}{k_1! \dots k_{m-2}!} \int_{-T_n}^{T_n} \varphi(x)^r\, Q_1^{k_1}(x) \dots Q_{m-2}^{k_{m-2}}(x)\, dx\, \frac{1}{n^{\frac{1}{2}\,(k_1+2k_2+\dots+(m-2)\,k_{m-2})}}$$

at the expense of an error $O(n^{-\frac{m-1}{2}})$. Moreover, using the property

$$\int_{|x| \ge T_n} x^N \varphi(x)^r\, dx = o(n^{-\frac{s-2}{2}}),$$

the above integration may be extended to the whole real line. Hence, $\int_{-T_n}^{T_n} \varphi_m(x)^r\, dx$ is represented as

$$\sum \frac{(r)_{k_1+\dots+k_{m-2}}}{k_1! \dots k_{m-2}!} \int_{-\infty}^{\infty} \varphi(x)^r\, Q_1^{k_1}(x) \dots Q_{m-2}^{k_{m-2}}(x)\, dx\, \frac{1}{n^{\frac{1}{2}\,(k_1+2k_2+\dots+(m-2)\,k_{m-2})}} + o\big(n^{-\frac{s-2}{2}}\big).$$

Here, it is sufficient to keep only the powers of $1/n$ not exceeding $(m-2)/2$. But in that case, for any fixed value of

$$j = k_1 + 2k_2 + \dots + (m-2)\,k_{m-2},$$

the constraint $j \le m-2$ implies that $k_{j+1} = \cdots = k_{m-2} = 0$. That is, we only need to consider the collections $k_1, \dots, k_j$ of length $j$. Thus, the above representation is simplified to

$$\int_{-T_n}^{T_n} \varphi_m(x)^r \, dx = \int_{-\infty}^{\infty} \varphi(x)^r \, dx$$
$$+ \sum \frac{(r)_{k_1+\dots+k_j}}{k_1! \dots k_j!} \int_{-\infty}^{\infty} \varphi(x)^r \, Q_1^{k_1}(x) \dots Q_j^{k_j}(x) \, dx \, n^{-j/2} + o\big(n^{-\frac{s-2}{2}}\big) \quad (11.27)$$

with summation over all $j = 1, \dots, m-2$ and over all non-negative integers $k_1, \dots, k_j$ such that $k_1 + 2k_2 + \cdots + j\,k_j = j$.

As the last simplifying step, we note that $Q_{2k-1}(x)$ represents a linear combination of the polynomials $H_{2i-1}(x)$ and has a leading term $x^{3(2k-1)}$ up to a constant. In particular, it is an odd function. On the other hand, $Q_{2k}(x)$ represents a linear combination of $H_{2i}(x)$'s and has a leading term $x^{6k}$, so it is an even function. It follows that any function of the form

$$Q = Q_1^{k_1}(x) \dots Q_j^{k_j}(x) \qquad (k_1 + 2k_2 + \cdots + j\,k_j = j) \tag{11.28}$$

is either odd or even, depending on whether $j$ is odd or even. Indeed, for polynomials of the class 1, defined by

$$P(x) = c_0 + c_2 x^2 + \cdots + c_{2N} x^{2N},$$

let us put $\mathrm{Ev}(P) = 2N \pmod 2 = 0$, and for the class 2, defined by

$$P(x) = c_1 x + \cdots + c_{2N-1}\, x^{2N-1},$$

let us put $\mathrm{Ev}(P) = 2N - 1 \pmod 2 = 1$. The products of such polynomials belong to one of the classes, and we have the property $\mathrm{Ev}(P_1 P_2) = (\mathrm{Ev}(P_1) + \mathrm{Ev}(P_2)) \pmod 2$. Therefore, using $\mathrm{Ev}(Q_i) = 3i \pmod 2 = i \pmod 2$ and the summation in the group $\mathbb{Z}_2$, we have

$$\begin{aligned}\mathrm{Ev}(Q) &= k_1\, \mathrm{Ev}(Q_1) + \cdots + k_j\, \mathrm{Ev}(Q_j) \\ &= k_1 \cdot 1 \pmod 2 + \cdots + k_j \cdot j \pmod 2 \;=\; (k_1 + \cdots + jk_j) \pmod 2 \;=\; j \pmod 2.\end{aligned}$$

Thus, $Q$ is an odd function in (11.28), as long as $j$ is odd, and then the corresponding integral in (11.27) is vanishing. As a result, (11.22) and (11.27) yield the following asymptotic expansion, which also holds for $2 \le s < 3$, in view of (11.25).

**Proposition 11.5.1** *Suppose that $\beta_s < \infty$ for $s \geq 2$. Then, with $m = [s]$, for any $r > 1$,*

$$\int_{-\infty}^{\infty} p_n(x)^r \, dx = \int_{-\infty}^{\infty} \varphi(x)^r \, dx \left(1 + \sum_{j=1}^{[\frac{m-2}{2}]} \frac{a_j}{n^j}\right) + o\big(n^{-\frac{s-2}{2}}\big) \tag{11.29}$$

*with coefficients defined by*

$$a_j \int_{-\infty}^{\infty} \varphi(x)^r \, dx = \sum \frac{(r)_{k_1+\dots+k_{2j}}}{k_1! \dots k_{2j}!} \int_{-\infty}^{\infty} Q_1^{k_1}(x) \dots Q_{2j}^{k_{2j}}(x) \, \varphi(x)^r \, dx. \tag{11.30}$$

*Here, the summation runs over all integers $k_1, \dots, k_{2j} \geq 0$ such that $k_1 + 2k_2 + \dots + 2j\, k_{2j} = 2j$ with notation $(r)_k = r(r-1)\dots(r-k+1)$.*

It follows from Definition 11.3.1 that each polynomial $Q_k$ is determined by the moments of $X$ up to order $k+2$. Hence, each $a_j$ in (11.30) is only determined by $r$ and by the moments, hence, by the cumulants of $X$ up to order $2j+2$. Moreover, $a_j = 0$ if these cumulants are vanishing.

## 11.6 The Case Where the First Cumulants Are Vanishing

For $2 \leq s < 4$, we necessarily have $m \leq 3$, so that the sum in (11.29) has no term, and then

$$\int_{-\infty}^{\infty} p_n(x)^r \, dx = \int_{-\infty}^{\infty} \varphi(x)^r \, dx + o\big(n^{-\frac{s-2}{2}}\big). \tag{11.31}$$

In the more interesting case $s \geq 4$, the leading term in the Edgeworth expansion (11.24) may be written explicitly, as was already done in the representation (11.18). It implies that, for some unique $1 \leq k \leq m-2$,

$$\varphi_m(x) = \varphi(x) + \varphi(x) \frac{\gamma_{k+2}}{(k+2)!} H_{k+2}(x) \, n^{-k/2} + C(x)\varphi(x) \, (1 + |x|^{3(m-2)}) \, n^{-(k+1)/2} \tag{11.32}$$

with some function $C(x)$ bounded by a constant which does not depend on $x$ and large $n \geq n_0$.

To study an asymptotic behavior of the truncated $L^r$-norm of $\varphi_m$, one may repeat computations of the previous section in this simple particular case, or alternatively, one may just refer to the general result described in Proposition 11.5.1. Indeed, (11.32) is equivalent to saying that the first moments of $X$ up to order $k+1$ coincide with those of $Z \sim N(0, 1)$ for some $1 \leq k \leq m-2$. Therefore, as emphasized after Proposition 11.5.1, $a_j = 0$ whenever $2j + 2 \le k + 1$, that is, $j \le \frac{k-1}{2}$. Then also $Q_j = 0$. In case $2j + 2 = k + 2$, that is, $j = k/2$ with even $k$, all terms in the sum (11.30) are vanishing, except (potentially) for the term corresponding to the collection with $k_1 = \cdots = k_{2j-1} = 0$, $k_{2j} = 1$. Then the right-hand side of (11.30) becomes

$$r \int_{-\infty}^{\infty} Q_{2j}(x)\, \varphi(x)^r\, dx = r \int_{-\infty}^{\infty} Q_k(x)\, \varphi(x)^r\, dx = r\, \frac{\gamma_{k+2}}{(k+2)!} \int_{-\infty}^{\infty} H_{k+2}(x)\, \varphi(x)^r\, dx,$$

and hence (11.29) yields

$$\int_{-\infty}^{\infty} p_n(x)^r\, dx \;=\; \int_{-\infty}^{\infty} \varphi(x)^r\, dx + A n^{-k/2} + O(n^{-\frac{k+1}{2}}) + o\big(n^{-\frac{s-2}{2}}\big), \qquad (11.33)$$

where

$$A = r\, \frac{\gamma_{k+2}}{(k+2)!} \int_{-\infty}^{\infty} H_{k+2}(x)\, \varphi(x)^r\, dx, \qquad \gamma_{k+2} = \mathbb{E}X^{k+2} - \mathbb{E}Z^{k+2}.$$

In particular, $A = 0$ for odd $k$, since then the Chebyshev-Hermite polynomial $H_{k+2}(x)$ is odd.

To proceed, we focus on the integrals $I(k,r) = \int_{-\infty}^{\infty} H_k(x)\, \varphi(x)^r\, dx$ with even $k$.

**Lemma 11.6.1** *For any* $k = 1, 2, \dots,$

$$I(2k,r) \;=\; \frac{(2k-1)!!}{r^{\frac{2k+1}{2}}\, (2\pi)^{\frac{r-1}{2}}}\, (1-r)^k. \qquad (11.34)$$

*Proof* The $k$-th Chebyshev-Hermite polynomial

$$H_k(x) \;=\; (-1)^k \left(e^{-x^2/2}\right)^{(k)} e^{x^2/2} \;=\; \mathbb{E}\,(x + iZ)^k, \qquad Z \sim N(0,1), \qquad (11.35)$$

has generating function

$$\sum_{k=0}^{\infty} H_k(x)\, \frac{z^k}{k!} = e^{xz - z^2/2}, \qquad z \in \mathbb{C},$$

from which one can find the generating function for the sequence $c_k = I(k,r)$. Namely,

$$\sum_{k=0}^{\infty} c_k\, \frac{z^k}{k!} \;=\; \int_{-\infty}^{\infty} e^{xz - z^2/2}\, \varphi(x)^r\, dx \;=\; \frac{1}{(2\pi)^{\frac{r-1}{2}} \sqrt{r}}\, e^{-\frac{1}{2}\,(1-\frac{1}{r})\, z^2}.$$

Differentiating this equality $2k$ times and applying the definition (11.35), we arrive at

$$c_{2k} = \frac{1}{(2\pi)^{\frac{r-1}{2}}\sqrt{r}} \Big(1 - \frac{1}{r}\Big)^k H_{2k}(0).$$

It remains to apply the equality (11.35), which gives $H_{2k}(0) = (-1)^k\, \mathbb{E}Z^{2k} = (-1)^k\,(2k-1)!!$ □

For the first three even values $k = 2, 4, 6$, we thus have

$$I(2,r) = -\frac{1}{r^{3/2}\,(2\pi)^{\frac{r-1}{2}}}\,(r-1), \qquad I(4,r) = \frac{3}{r^{5/2}\,(2\pi)^{\frac{r-1}{2}}}\,(r-1)^2,$$
$$I(6,r) = -\frac{15}{r^{7/2}\,(2\pi)^{\frac{r-1}{2}}}\,(r-1)^3. \tag{11.36}$$

One may also evaluate the integrals $\int_{-\infty}^{\infty} H_k(x)^2\,\varphi(x)^r\,dx$. For example,

$$\int_{-\infty}^{\infty} H_3(x)^2\,\varphi(x)^r\,dx = \frac{1}{\sqrt{r}\,(2\pi)^{\frac{r-1}{2}}}\,\mathbb{E}\bigg(\Big(\frac{Z}{\sqrt{r}}\Big)^3 - 3\Big(\frac{Z}{\sqrt{r}}\Big)\bigg)^2 = \frac{3\,(5-6r+3r^2)}{r^{7/2}\,(2\pi)^{\frac{r-1}{2}}}. \tag{11.37}$$

Thus, the formula (11.34) may be used in the asymptotic representation (11.33). The particular case $k = [s]-2$ should be mentioned separately.

**Corollary 11.6.2** *Suppose that $\mathbb{E}X^l = \mathbb{E}Z^l$ for $l = 1, \dots, m-1$ ($m \geq 3$), where $Z \sim N(0,1)$. If $\beta_s < \infty$ for some $s \in [m, m+1)$, then for all $n$ large enough, $Z_n$ have bounded densities $p_n$. Moreover,*

$$\int_{-\infty}^{\infty} p_n(x)^r\,dx = \int_{-\infty}^{\infty} \varphi(x)^r\,dx + An^{-\frac{m-2}{2}} + o\big(n^{-\frac{s-2}{2}}\big) \tag{11.38}$$

*with $A = 0$ in the case $m = 2k-1$ is odd, while in the case where $m = 2k$ is even, we have*

$$A = \frac{\gamma_{2k}}{2^k k!}\,\frac{(1-r)^k}{(2\pi)^{\frac{r-1}{2}}\,r^{\frac{2k-1}{2}}}, \qquad \gamma_{2k} = \mathbb{E}X^{2k} - \mathbb{E}Z^{2k}.$$

*If $\beta_s < \infty$ for $s = m+1$, then $o$-term in (11.38) may be replaced with $O$-term.*

For example, if $\gamma_3 = \mathbb{E}X^3 = 0$, so that $m = 4$, $4 \leq s < 5$, we have

$$A = \frac{\gamma_4}{8}\,\frac{1}{(2\pi)^{\frac{r-1}{2}}}\,\frac{(1-r)^2}{r^{\frac{3}{2}}}, \qquad \gamma_4 = \mathbb{E}X^4 - \mathbb{E}Z^4 = \mathbb{E}X^4 - 3,$$

and (11.38) becomes

$$\int_{-\infty}^{\infty} p_n(x)^r\,dx \;=\; \int_{-\infty}^{\infty} \varphi(x)^r\,dx + An^{-1} + o\big(n^{-\frac{s-2}{2}}\big). \tag{11.39}$$

By (11.33), a similar formula remains to hold in the case $5 \le s < 6$, but then the $o$-term should be replaced with $O(n^{-3/2})$.

## 11.7 Moments of Order $4 \le s \le 8$

Returning to the general expansion (11.29) in Proposition 11.5.1 with coefficients $a_j$ described in (11.30), let us now derive formulas similar to (11.39) for two regions of the values of $s$ without additional assumptions on the first cumulants. To evaluate the integrals in that definition, we will use the formulas for the polynomials $Q_j$ described in Sect. 11.3 for the indexes $j \le 4$.

If $4 \le s < 6$, the expansion (11.29) contains only one term, namely, we get

$$\int_{-\infty}^{\infty} p_n(x)^r\,dx \;=\; \int_{-\infty}^{\infty} \varphi(x)^r\,dx + \frac{a_1}{n}\int_{-\infty}^{\infty} \varphi(x)^r\,dx + o\big(n^{-\frac{s-2}{2}}\big) \tag{11.40}$$

with the coefficient for $j = 1$ in front of $1/n$, i.e.,

$$\begin{aligned} A_1 \;\equiv\; a_1 \int_{-\infty}^{\infty} \varphi(x)^r\,dx &= \frac{(r)_1}{1!}\int_{-\infty}^{\infty} Q_2(x)\,\varphi(x)^r\,dx + \frac{(r)_2}{2!}\int_{-\infty}^{\infty} Q_1^2(x)\,\varphi(x)^r\,dx \\ &= r\int_{-\infty}^{\infty} \Big(\frac{\gamma_4}{4!}\,H_4(x) + \frac{1}{2!}\Big(\frac{\gamma_3}{3!}\Big)^2 H_6(x)\Big)\,\varphi(x)^r\,dx \\ &\quad + \frac{r(r-1)}{2}\int_{-\infty}^{\infty} \Big(\frac{\gamma_3}{3!}\,H_3(x)\Big)^2 \varphi(x)^r\,dx. \end{aligned}$$

Applying the formulas (11.36)–(11.37), we find that

$$\begin{aligned} A_1 &= r\,\frac{\gamma_3^2}{2!\,3!^2}\,I(6,r) + r\,\frac{\gamma_4}{4!}\,I(4,r) + \frac{r(r-1)}{2}\Big(\frac{\gamma_3}{3!}\Big)^2 \int_{-\infty}^{\infty} H_3(x)^2\,\varphi(x)^r\,dx \\ &= -r\,\frac{\gamma_3^2}{72}\,\frac{15}{r^{7/2}\,(2\pi)^{\frac{r-1}{2}}}\,(r-1)^3 + r\,\frac{\gamma_4}{24}\,\frac{3}{r^{5/2}\,(2\pi)^{\frac{r-1}{2}}}\,(r-1)^2 + r(r-1)\,\frac{\gamma_3^2}{24}\,\frac{5-6r+3r^2}{r^{7/2}\,(2\pi)^{\frac{r-1}{2}}}. \end{aligned}$$

Equivalently,

$$(2\pi)^{\frac{r-1}{2}}\,\frac{r^{5/2}}{r-1}\,A_1 = -\frac{5}{24}\,(r-1)^2\,\gamma_3^2 + \frac{1}{8}\,r(r-1)\,\gamma_4 + \frac{1}{24}\,(5-6r+3r^2)\,\gamma_3^2.$$

Collecting the coefficients in front of $\gamma_3^2$, we arrive at the following refinement of (11.40).

**Proposition 11.7.1** *Suppose that* $\beta_s < \infty$ *for* $4 \le s < 6$. *Then, for any* $r > 1$,

$$\int_{-\infty}^{\infty} p_n(x)^r \, dx \ = \ \int_{-\infty}^{\infty} \varphi(x)^r \, dx + A_1 n^{-1} + o\big(n^{-\frac{s-2}{2}}\big), \tag{11.41}$$

*where the constant* $A_1 = A_1(r)$ *is given by*

$$(2\pi)^{\frac{r-1}{2}} \frac{r^{3/2}}{r-1} \, A_1(r) \ = \ \frac{2-r}{12} \, \gamma_3^2 + \frac{r-1}{8} \, \gamma_4. \tag{11.42}$$

*In the case* $s = 6$, *the formula* (11.41) *remains valid with the remainder term* $O(n^{-2})$.

Note that $\lim_{r \to 1} \frac{A_1(r)}{r-1} = \frac{1}{12} \gamma_3^2$. Also, if $\gamma_3 = 0$, then (11.42) is simplified and defines exactly the constant $A$ in the equality (11.39).

For the region $6 \le s < 8$, the sum in (11.29) contains two terms, proportional to $\frac{1}{n}$ and $\frac{1}{n^2}$. The coefficient $a_1$ is as before, while according to (11.30), we arrive at the following refinement.

**Proposition 11.7.2** *Suppose that* $\beta_s < \infty$ *for* $6 \le s < 8$. *Then, for any* $r > 1$,

$$\int_{-\infty}^{\infty} p_n(x)^r \, dx \ = \ \int_{-\infty}^{\infty} \varphi(x)^r \, dx + A_1 n^{-1} + A_2 n^{-2} + o\big(n^{-\frac{s-2}{2}}\big), \tag{11.43}$$

*where* $A_1$ *is given in* (11.42) *and*

$$A_2 = r \int_{-\infty}^{\infty} Q_4(x) \, \varphi(x)^r \, dx + \frac{(r)_2}{2} \int_{-\infty}^{\infty} \big(Q_2^2(x) + 2\, Q_1(x) Q_3(x)\big) \, \varphi(x)^r \, dx$$
$$+ \frac{(r)_3}{2} \int_{-\infty}^{\infty} Q_1^2(x) Q_2(x) \, \varphi(x)^r \, dx + \frac{(r)_4}{24} \int_{-\infty}^{\infty} Q_1^4(x) \, \varphi(x)^r \, dx.$$

*In the case* $s = 8$, *the formula* (11.43) *remains valid with the remainder term* $O(n^{-3})$.

One can rewrite $A_2$ explicitly in terms of the cumulants of $X$, cf. [6]. In the case $\gamma_3 = 0$, a long expression for this constant is simplified to

$$A_2 \ = \ \frac{\gamma_6 \, r}{6!} \int_{-\infty}^{\infty} H_6(x) \, \varphi(x)^r \, dx + \frac{\gamma_4^2 \, r}{2! \, 4!^2} \int_{-\infty}^{\infty} H_8(x) \, \varphi(x)^r \, dx + \frac{\gamma_4^2 \, r(r-1)}{2! \, 4!^2} \int_{-\infty}^{\infty} H_4(x)^2 \, \varphi(x)^r \, dx.$$

## 11.8 Expansions for Rényi Entropies

Let us now reformulate the asymptotic results about the integrals $\int_{-\infty}^{\infty} p_n(x)^r\,dx$ in terms of the Rényi entropies and entropy powers

$$h_r(Z_n) = -\frac{1}{r-1}\log\int_{-\infty}^{\infty} p_n(x)^r\,dx, \qquad N_r(Z_n) = \Big(\int_{-\infty}^{\infty} p_n(x)^r\,dx\Big)^{-\frac{2}{r-1}}.$$

Since these functionals represent smooth functions of the $L^r$-norm, from Proposition 11.5.1 together with Taylor's formulas

$$\begin{aligned}\log(a+b+c) &= \log a + a^{-1}b + O(b^2+|c|),\\ (a+b+c)^q &= a^q + qa^{q-1}b + O(b^2+|c|),\end{aligned} \tag{11.44}$$

holding with $a > 0$, $q \neq 0$, and $b, c \to 0$, we immediately obtain:

**Proposition 11.8.1** *Let $\mathbb{E}\,|X|^s < \infty$ for some $s \geq 2$, and $m = [s]$. Then, for any $r > 1$,*

$$h_r(Z_n) = h_r(Z) + \sum_{j=1}^{[\frac{m-2}{2}]} \frac{b_j}{n^j} + o\big(n^{-\frac{s-2}{2}}\big), \tag{11.45}$$

$$N_r(Z_n) = N_r(Z)\Bigg(1 + \sum_{j=1}^{[\frac{m-2}{2}]} \frac{c_j}{n^j}\Bigg) + o\big(n^{-\frac{s-2}{2}}\big), \tag{11.46}$$

*with coefficients $b_j$ and $c_j$ that are determined by $r$ and by the moments of $X$ up to order $2j+2$.*

*Proof of Theorem 11.1.2* To evaluate the first coefficients in the expansions (11.45)–(11.46), we apply Taylor's formulas (11.44). For $q = -\frac{2}{r-1}$, the last equality in (11.44) reads

$$(a+b+c)^{-\frac{2}{r-1}} = a^{-\frac{2}{r-1}} - \frac{2}{r-1}\,a^{-\frac{r+1}{r-1}}\,b + O(b^2+|c|). \tag{11.47}$$

In particular (with $b = 0$), the expansion of the form

$$\int_{-\infty}^{\infty} p_n(x)^r\,dx = \int_{-\infty}^{\infty} \varphi(x)^r\,dx + o\big(n^{-\frac{s-2}{2}}\big),$$

which corresponds to Proposition 11.5.1 to the region $2 < s < 4$, implies

$$\log \int_{-\infty}^{\infty} p_n(x)^r \, dx = \log \int_{-\infty}^{\infty} \varphi(x)^r \, dx + o\big(n^{-\frac{s-2}{2}}\big).$$

Equivalently, $h_r(Z_n) = h_r(Z) + o(n^{-\frac{s-2}{2}})$ or $N_r(Z_n) = N_r(Z) + o(n^{-\frac{s-2}{2}})$ for $Z \sim N(0,1)$.

More generally, applying (11.44)–(11.47) to the expansion

$$\int_{-\infty}^{\infty} p_n(x)^r \, dx \ = \ \int_{-\infty}^{\infty} \varphi(x)^r \, dx + A_1 \, n^{-1} + o\big(n^{-\frac{s-2}{2}}\big),$$

corresponding to Proposition 11.7.1 with its region $4 \le s < 6$, we get

$$\log \int_{-\infty}^{\infty} p_n(x)^r \, dx \ = \ \log \int_{-\infty}^{\infty} \varphi(x)^r \, dx + A_1 \, n^{-1} \bigg(\int_{-\infty}^{\infty} \varphi(x)^r \, dx\bigg)^{-1} + o\big(n^{-\frac{s-2}{2}}\big),$$

and

$$\bigg(\int_{-\infty}^{\infty} p_n(x)^r \, dx\bigg)^{-\frac{2}{r-1}} = \bigg(\int_{-\infty}^{\infty} \varphi(x)^r \, dx\bigg)^{-\frac{2}{r-1}}$$
$$- \frac{2A_1}{r-1} \, n^{-1} \bigg(\int_{-\infty}^{\infty} \varphi(x)^r \, dx\bigg)^{-\frac{r+1}{r-1}} + o\big(n^{-\frac{s-2}{2}}\big).$$

Thus,

$$h_r(Z_n) \ = \ h_r(Z) - \frac{A_1}{r-1} \, N_r(Z)^{\frac{r-1}{2}} \, n^{-1} + o(n^{-\frac{s-2}{2}}), \tag{11.48}$$

and (equivalently)

$$N_r(Z_n) \ = \ N_r(Z) \Big[1 - \frac{2A_1}{r-1} \, N_r(Z)^{\frac{r-1}{2}} \, n^{-1}\Big] + o(n^{-\frac{s-2}{2}}). \tag{11.49}$$

Recall that $A_1 = A_1(r)$ is determined by $r$ and the cumulants $\gamma_3 = \mathbb{E}X^3$ and $\gamma_4 = \mathbb{E}X^4 - 3$. More precisely, according to the formula (11.42) of Proposition 11.7.1,

$$\frac{A_1}{r-1} \ = \ \frac{1}{(2\pi)^{\frac{r-1}{2}} \, r^{3/2}} \Big[\frac{2-r}{12} \, \gamma_3^2 + \frac{r-1}{8} \, \gamma_4\Big].$$

Since also

$$N_r(Z)^{\frac{r-1}{2}} = \bigg(\int_{-\infty}^{\infty} \varphi(x)^r \, dx\bigg)^{-1} = (2\pi)^{\frac{r-1}{2}} \, r^{1/2},$$

the coefficients $b_1$ and $c_1$ in (11.45)–(11.46) in front of $n^{-1}$ are simplified according to (11.48)–(11.49) as

$$b_1 = -\frac{A_1}{r-1}\, N_r(Z)^{\frac{r-1}{2}} = -\frac{1}{r}\left[\frac{2-r}{12}\,\gamma_3^2 + \frac{r-1}{8}\,\gamma_4\right], \qquad c_1 = 2b_1.$$

□

Let us complement the expansions of Theorem 11.1.2 with similar assertions corresponding to the scenario from Corollary 11.6.2, where the first $m-1$ moments of $X$ coincide with those of $Z \sim N(0,1)$, for some integer $m \geq 3$. If $\beta_s$ is finite for $s \in [m, m+1)$, in that case we have an expansion of the form

$$\int_{-\infty}^{\infty} p_n(x)^r\,dx \;=\; \int_{-\infty}^{\infty} \varphi(x)^r\,dx + An^{-\frac{m-2}{2}} + o\big(n^{-\frac{s-2}{2}}\big).$$

Hence, by (11.44)–(11.47),

$$\log\int_{-\infty}^{\infty} p_n(x)^r\,dx = \log\int_{-\infty}^{\infty} \varphi(x)^r\,dx + A\,n^{-\frac{m-2}{2}}\bigg(\int_{-\infty}^{\infty} \varphi(x)^r\,dx\bigg)^{-1} + O(n^{-(m-2)}) + o\big(n^{-\frac{s-2}{2}}\big),$$

and

$$\bigg(\int_{-\infty}^{\infty} p_n(x)^r\,dx\bigg)^{-\frac{2}{r-1}} = \bigg(\int_{-\infty}^{\infty} \varphi(x)^r\,dx\bigg)^{-\frac{2}{r-1}} - \frac{2A}{r-1}\,n^{-\frac{m-2}{2}}\bigg(\int_{-\infty}^{\infty} \varphi(x)^r\,dx\bigg)^{-\frac{r+1}{r-1}} + O(n^{-(m-2)}) + o\big(n^{-\frac{s-2}{2}}\big).$$

Since $m-2 > \frac{s-2}{2}$, here $O$-term may be removed. In addition, as before, the last integral with its power can be written as $N_r(Z)^{\frac{r+1}{2}}$. Therefore, we obtain the asymptotic relations

$$h_r(Z_n) = h_r(Z) - \frac{A}{r-1}\,N_r(Z)^{\frac{r-1}{2}}\,n^{-\frac{m-2}{2}} + o(n^{-\frac{s-2}{2}})$$

and

$$N_r(Z_n) \;=\; N_r(Z)\left[1 - \frac{2A}{r-1}\,N_r(Z)^{\frac{r-1}{2}}\,n^{-\frac{m-2}{2}}\right] + o(n^{-\frac{s-2}{2}})$$

in full analogy with (11.48)–(11.49). The only difference is that we have a different formula for the constant $A = A(r)$. As stated in Corollary 11.6.2, here $A = 0$ in the case $m = 2k - 1$ is odd, while in the case $m = 2k$ is even, we have

$$A = \frac{\gamma_{2k}}{2^k k!} \, \frac{(1-r)^k}{(2\pi)^{\frac{r-1}{2}} r^{\frac{2k-1}{2}}}, \qquad \gamma_{2k} = \mathbb{E}X^{2k} - \mathbb{E}Z^{2k}.$$

Using again $N_r(Z)^{\frac{r-1}{2}} = (2\pi)^{\frac{r-1}{2}} r^{1/2}$, the coefficients $b_{k-1}$ and $c_{k-1}$ in (11.45)–(11.46) in front of $n^{-\frac{m-2}{2}} = n^{-(k-1)}$ are simplified to

$$b_{k-1} = -\frac{A}{r-1} N_r(Z)^{\frac{r-1}{2}} = \frac{\gamma_{2k}}{2^k k!} \, \frac{(1-r)^{k-1}}{r^{k-1}}, \qquad c_{k-1} = 2b_{k-1}.$$

Let us also remind that, if $\beta_s < \infty$ for $s = m+1$, then $o$-term may be replaced with $O(n^{-\frac{m-1}{2}})$. We are thus ready to make a corresponding statement.

**Proposition 11.8.2** *Suppose that* $\mathbb{E}X^l = \mathbb{E}Z^l$ *for* $l = 3, \dots, m-1$ $(m \geq 3)$. *If* $\beta_s < \infty$ *for some* $s \in [m, m+1)$, *then for any* $r > 1$,

$$h_r(Z_n) = h_r(Z) + bn^{-\frac{m-2}{2}} + o(n^{-\frac{s-2}{2}}),$$
$$N_r(Z_n) = N_r(Z)\left(1 + 2b\,n^{-\frac{m-2}{2}}\right) + o(n^{-\frac{s-2}{2}})$$

*with constant* $b = 0$ *in the case* $m = 2k - 1$ *is odd, while in the case* $m = 2k$ *is even,*

$$b = b_{k-1} = \frac{\gamma_{2k}}{2^k k!} \left(\frac{1}{r} - 1\right)^{k-1}, \qquad \gamma_{2k} = \mathbb{E}X^{2k} - \mathbb{E}Z^{2k}.$$

*If* $\beta_s < \infty$ *for* $s = m + 1$, *then o-term may be replaced with* $O(n^{-\frac{m-1}{2}})$.

For example, if $\gamma_3 = \mathbb{E}X^3 = 0$, we return to the equality (11.4) from Theorem 11.1.2.

## 11.9 Comparison with the Entropic CLT: Monotonicity

Put

$$\Delta_n(r) = h_r(Z) - h_r(Z_n), \qquad \Delta_n = \Delta_n(1).$$

The latter quantity, which may also be written as $D(Z_n||Z) = \int_{-\infty}^{\infty} p_n(x) \log \frac{p_n(x)}{\varphi(x)} \, dx$, represents the Kullback-Leibler distance from the distribution of $Z_n$ to the standard normal law (or, the relative entropy). As was mentioned, the sequence $\Delta_n$ is always non-negative and non-increasing. Moreover, the entropic CLT asserts that $\Delta_n \to 0$ as $n \to \infty$, as long as $\Delta_n$ is finite for some $n$ (in general, it is a weaker condition in comparison with (11.2)). The basic references for these results are [1, 2, 12].

The rate of convergence of $\Delta_n$ to zero was studied in [8], and here we recall a few asymptotic results, assuming that $\Delta_n < \infty$ for some $n$, and that $\beta_s = \mathbb{E}\,|X|^s < \infty$ for a real number $s \geq 2$. Namely, we have

$$\Delta_n = o\bigg(\frac{1}{(n\log n)^{\frac{s-2}{2}}}\bigg), \qquad 2 \leq s < 4.$$

Modulo a logarithmic term, it is the same rate as for $\Delta_n(r)$ indicated in Theorem 11.1.2. Nevertheless, it is not yet clear, if one can similarly improve Theorem 11.1.2. On the other hand, for any prescribed $\eta > 1$, it may occur that, for all $n$ large enough,

$$\Delta_n \geq \frac{c}{(n\log n)^{\frac{s-2}{2}}\,(\log n)^{\eta}}$$

with some constant $c = c(\eta, s) > 0$ depending on $\eta$ and $s$ only ([8], Theorem 11.1.3).

The range $s \geq 4$ is more interesting, since then one may control the speed of $\Delta_n$. In particular,

$$\Delta_n = \frac{\gamma_3^2}{12}\,n^{-1} + o\bigg(\frac{1}{(n\log n)^{\frac{s-2}{2}}}\bigg), \qquad 4 \leq s < 6,$$

$$\Delta_n = \frac{\gamma_3^2}{12}\,n^{-1} + O\bigg(\frac{1}{(n\log n)^{2}}\bigg), \qquad s = 6.$$

Thus, if $\gamma_3 \neq 0$, then $\Delta_n$ is equivalent to a decreasing sequence, which decreases at rate $n^{-1}$. (Strictly speaking, this property does not imply the monotonicity itself.)

Let us compare this asymptotic with what is given in Theorem 11.1.2. Namely, for any $r > 1$, we have

$$\Delta_n(r) = B_1\,n^{-1} + o\big(n^{-\frac{s-2}{2}}\big), \qquad 4 \leq s < 6, \tag{11.50}$$

$$\Delta_n(r) = B_1\,n^{-1} + O\big(n^{-2}\big), \qquad s = 6, \tag{11.51}$$

where

$$B_1 = B_1(r) = -b = \frac{1}{4r}\left[\frac{2-r}{3}\,\gamma_3^2 + \frac{r-1}{2}\,\gamma_4\right].$$

We see that $B(r) \to \frac{1}{12}\,\gamma_3^2$ as $r \to 1$, so that we recover the main term in the asymptotic for $\Delta_n$, and at the same rate modulo a logarithmic factor.

However, what can one say about the sign of $B_1(r)$ with fixed $r > 1$? First suppose that $\gamma_3 \neq 0$. When $r$ is sufficiently close to 1, then $B_1(r) > 0$, so that $\Delta_n(r)$ is equivalent to a decreasing sequence like for $r = 1$. More precisely, this is true for all $r > 1$, whenever $\gamma_4 \geq \frac{2}{3}\gamma_3^2$. But, if $\gamma_4 < \frac{2}{3}\gamma_3^2$, then $B_1(r) < 0$ for all

$$r > r_0 = \frac{4\gamma_3^2 - 3\gamma_4}{2\gamma_3^2 - 3\gamma_4}.$$

Hence $\Delta_n(r)$ becomes to be equivalent to an increasing sequence. In that case, necessarily $h_r(Z_n) > h_r(Z)$ for all $n$ large enough, which is impossible in the Shannon case $r = 1$. This shows that $\Delta_n(r)$ may not serve as distance!

If $\gamma_3 = 0$ (as in case of symmetric distributions), the constant is simplified to

$$B_1 = B_1(r) = \frac{r-1}{8r}\gamma_4, \qquad \gamma_4 = \mathbb{E}X^4 - 3,$$

and then the sign of $B_1$ coincides with the sign of $\gamma_4$. Both cases, $\gamma_4 > 0$ or $\gamma_4 < 0$, are typical, and one can make a similar conclusion as before, but for the whole range $r > 1$. Namely, if $\gamma_4 > 0$, then $\Delta_n(r)$ is equivalent to a decreasing sequence, which decreases at rate $n^{-1}$, and if $\gamma_4 < 0$, then $\Delta_n(r)$ is equivalent to an increasing sequence, which increases also at rate $n^{-1}$.

*Proof of Theorem 11.1.3 in Case $r < \infty$* In order to make a more rigorous conclusion about the monotonicity of $\Delta_n(r)$ for large $n$, the expansions for Rényi entropy $h_r(Z_n)$ such as (11.50)–(11.51) are insufficient. We need to use more terms in the general Proposition 11.8.1 involving the quadratic terms $b_2/n^2$ and $c_2/n^2$. This is possible under stronger moment assumptions, corresponding to the range $6 \leq s < 8$. Indeed, in that case, Proposition 11.8.1 provides the expansion (11.5) in which the coefficient $b_1 = b$ is as before, and we also know that the coefficient $b_2$ is only determined by $r$ and by the moments of $X$ up to order 6. In fact, one may evaluate $b_2$ on the basis of equality (11.43) of Proposition 11.7.2, which specializes Proposition 11.5.1 to the range $6 \leq s < 8$. Since the formula for the coefficient $A_2 = A_2(r)$ is somewhat complicated, we will not go into tedious computations.

Now, from (11.5) it follows that

$$h_r(Z_{n+1}) - h_r(Z_n) = \frac{B_1}{n(n+1)} + o(n^{-2}),$$

which thus proves Theorem 11.1.3 in case of finite $r$. □

## 11.10 Maximum of Density (the Case $r = \infty$)

Recall that $N_\infty(X) = \|p\|_\infty^{-2}$ when a random variable $X$ has density $p$. An expansion similar to the one of Proposition 11.5.1 can also be obtained for $\|p_n\|_\infty$ and hence for $N_\infty(Z_n)$. In order to deduce monotonicity, let us assume that $\beta_6 < \infty$.

From the non-uniform local limit theorem it follows that $\|p_n - \varphi_6\|_\infty = o(n^{-2})$ as $n \to \infty$, where $\varphi_6$ is the Edgeworth expansion of order 6. Hence

$$\|p_n\|_\infty = \|\varphi_6\|_\infty + o(n^{-2}). \tag{11.52}$$

Here

$$\varphi_6(x) = \varphi(x)\Big(1 + Q_1(x)\frac{1}{\sqrt{n}} + Q_2(x)\frac{1}{n} + Q_3(x)\frac{1}{n^{\frac{3}{2}}} + Q_4(x)\frac{1}{n^2}\Big),$$

where the polynomials $Q_k(x)$ are the same as in Sect. 11.3.

Let us find an asymptotic expansion for $\|\varphi_6\|_\infty$ (we refer to [6] for more computational details). Since $\varphi_6(x)$ is vanishing at infinity, there exists a point $x_6(n)$ such that $\|\varphi_6\|_\infty = |\varphi_6(x_6(n))|$. Since also the functions $\varphi(x)\ Q_k(x)$ are bounded, we have $|\varphi_6(x)| = O(\frac{1}{\sqrt{n}})$ uniformly in the region $|x| \geq \sqrt{\log n}$. On the other hand,

$$\varphi_6(0) = \varphi(0) + \varphi(0)\sum_{k=1}^{4} Q_k(0)\, n^{-\frac{k}{2}} \geq \frac{1}{2}\,\varphi(0)$$

for $n$ large. Therefore, $\varphi_6(0) > |\varphi_6(x)|$ for all $n$ large enough, as long as $|x| \geq \sqrt{\log n}$, and we conclude that

$$\|\varphi_6\|_\infty = \sup_{|x| \leq \sqrt{\log n}} |\varphi_6(x)| \qquad \text{and} \qquad |x_6(n)| \leq \sqrt{\log n}. \tag{11.53}$$

Since $x = x_6(n)$ is the point of local extremum, we have $\varphi_6'(x) = 0$, that is,

$$x = \frac{Q_1'(x) - xQ_1(x)}{\sqrt{n}} + \frac{Q_2'(x) - xQ_2(x)}{n} + \frac{Q_3'(x) - xQ_3(x)}{n^{\frac{3}{2}}} + \frac{Q_4'(x) - xQ_4(x)}{n^2}. \tag{11.54}$$

Using (11.53), we deduce from (11.54) that $x_6(n) = O\big(\frac{1}{\sqrt{n}}\,(\log n)^{\frac{13}{2}}\big)$ and hence $|x_6(n)| \leq 1$ for all $n$ large enough. But then, from (11.54) again, $x_6(n) = O(\frac{1}{\sqrt{n}})$. For $x = x_6(n)$, we thus have

$$\frac{xQ_3(x)}{n^{\frac{3}{2}}} = O\big(n^{-5/2}\big), \qquad \frac{Q_4'(x)}{n^2} = O\big(n^{-5/2}\big), \qquad \frac{xQ_4(x)}{n^2} = O\big(n^{-5/2}\big),$$

and (11.54) is simplified to

$$x = \frac{Q_1'(x) - xQ_1(x)}{\sqrt{n}} + \frac{Q_2'(x) - xQ_2(x)}{n} + \frac{Q_3'(x)}{n^{\frac{3}{2}}} + O\big(n^{-5/2}\big).$$

The Chebyshev-Hermite polynomials satisfy the relation $H_k'(x) - xH_k(x) = -H_{k+1}(x)$, so

$$\begin{aligned} H_3'(x) - xH_3(x) = -H_4(x) \ &= \ -3 + 6x^2 - x^4 \\ H_4'(x) - xH_4(x) = -H_5(x) \ &= \ -15x + 10x^3 - x^5 \\ H_6'(x) - xH_6(x) = -H_7(x) \ &= \ 105\,x - 105\,x^3 + 21\,x^5 - x^7. \end{aligned}$$

Using these identities in the formulas for $Q_k$'s, we easily find for $x = O(\frac{1}{\sqrt{n}})$ that

$$\begin{aligned} \frac{Q_1'(x) - xQ_1(x)}{\sqrt{n}} &= -\frac{\gamma_3}{2\sqrt{n}} + \gamma_3 \frac{x^2}{\sqrt{n}} + O\big(n^{-5/2}\big), \\ \frac{Q_2'(x) - xQ_2(x)}{n} &= \Big(\frac{105}{2!\,3!^2}\,\gamma_3^2 - \frac{15}{4!}\,\gamma_4\Big)\,\frac{x}{n} + O\big(n^{-5/2}\big), \\ \frac{Q_3'(x)}{n^{\frac{3}{2}}} &= \Big(\frac{945}{3!^4}\,\gamma_3^3 - \frac{105}{3!\,4!}\,\gamma_3\gamma_4 + \frac{15}{5!}\,\gamma_5\Big)\,\frac{1}{n^{\frac{3}{2}}} + O\big(n^{-5/2}\big). \end{aligned}$$

As a result,

$$\begin{aligned} x \ = \ x_6(n) = -\frac{\gamma_3}{2\sqrt{n}} + \gamma_3 \frac{x^2}{\sqrt{n}} + \Big(\frac{105}{2\cdot 3!^2}\,\gamma_3^2 - \frac{15}{4!}\,\gamma_4\Big)\frac{x}{n} \\ + \Big(\frac{945}{3!^4}\,\gamma_3^3 - \frac{105}{3!4!}\,\gamma_3\gamma_4 + \frac{15}{5!}\,\gamma_5\Big)\frac{1}{n^{\frac{3}{2}}} + O\big(n^{-5/2}\big). \quad (11.55) \end{aligned}$$

One may use this asymptotic equation to find an expansion for $x_6(n)$ in powers of $1/\sqrt{n}$. Indeed, first we immediately obtain that

$$x = x_6(n) = -\frac{\gamma_3}{2\sqrt{n}} + O\big(n^{-\frac{3}{2}}\big),$$

implying

$$\frac{x^2}{\sqrt{n}} \ = \ \frac{\gamma_3^2}{4}\frac{1}{n^{\frac{3}{2}}} + O\big(n^{-5/2}\big), \qquad \frac{x}{n} \ = \ -\frac{\gamma_3}{2}\frac{1}{n^{\frac{3}{2}}} + O\big(n^{-5/2}\big).$$

Inserting the above to (11.55), we deduce that

$$x \,=\, x_6(n) \,=\, \frac{a_1}{\sqrt{n}} + \frac{a_2}{n^{\frac{3}{2}}} + O\big(n^{-5/2}\big)$$

with coefficients

$$a_1 \,=\, -\frac{1}{2}\,\gamma_3, \qquad a_2 \,=\, \frac{1}{4}\,\gamma_3^3 - \frac{5}{12}\,\gamma_3\gamma_4 + \frac{1}{8}\,\gamma_5.$$

In particular, $a_1 = a_2 = 0$ and therefore $x = x_6(n) = O\big(n^{-5/2}\big)$, as long as the distribution of $X$ is symmetric about the origin (in which case $\gamma_3 = \gamma_5 = 0$).

Still in the general case, keeping these coefficients, we deduce for $x = x_6(n)$ that

$$x = \frac{1}{\sqrt{n}}\Big(a_1 + a_2\,\frac{1}{n}\Big) + O\big(n^{-5/2}\big), \quad x^2 \,=\, \frac{1}{n}\Big(a_1^2 + 2a_1a_2\,\frac{1}{n}\Big) + O\big(n^{-5/2}\big),$$

$$x^3 = \frac{1}{n^{\frac{3}{2}}}\,a_1^3 + O\big(n^{-5/2}\big), \quad x^4 \,=\, \frac{1}{n^2}\,a_1^4 + O\big(n^{-5/2}\big), \quad x^p \,=\, O\big(n^{-5/2}\big) \quad (p \geq 5).$$

Hence

$$\frac{Q_1(x)}{\sqrt{n}} \,=\, \frac{\gamma_3}{6\sqrt{n}}\,(x^3 - 3x) \,=\, \frac{\gamma_3^2}{4n} + \frac{b_1}{n^2} + O\big(n^{-5/2}\big), \qquad b_1 = \frac{\gamma_3}{3!}\,(a_1^3 - 3a_2).$$

Similarly,

$$\frac{Q_2(x)}{n} = \Big(\frac{3}{4!}\,\gamma_4 - \frac{15}{2!\cdot 3!^2}\,\gamma_3^2\Big)\,\frac{1}{n} + \frac{b_2}{n^2} + O\big(n^{-5/2}\big),$$

$$\frac{Q_3(x)}{n^{\frac{3}{2}}} = \frac{b_3}{n^2} + O\big(n^{-5/2}\big),$$

$$\frac{Q_4(x)}{n^2} = \frac{b_4}{n^2} + O\big(n^{-5/2}\big)$$

with

$$b_2 = \Big(\frac{45}{2\cdot 3!^2}\,\gamma_3^2 - \frac{6}{4!}\,\gamma_4\Big)\,a_1^2, \qquad b_3 = \Big(\frac{945}{3!^4}\,\gamma_3^3 - \frac{105}{3!4!}\,\gamma_3\gamma_4 + \frac{15}{5!}\,\gamma_5\Big)\,a_1,$$

and

$$b_4 = \frac{10\,395}{4!\cdot 3!^4}\gamma_3^4 - \frac{945}{2\cdot 3!^2 4!}\,\gamma_3^2\gamma_4 + \frac{105}{3!\cdot 5!}\,\gamma_3\gamma_5 + \frac{105}{2\cdot 4!^2}\,\gamma_4^2 - \frac{15}{6!}\,\gamma_6.$$

Note that in the case of symmetric distributions, $b_1 = b_2 = b_3 = 0$, while $b_4 = \frac{105}{2\cdot 4!^2}\,\gamma_4^2 - \frac{15}{6!}\,\gamma_6$.

Now, as $x \to 0$,

$$\frac{\varphi(x)}{\|\varphi\|_\infty} = 1 - \frac{1}{2}\,x^2 + \frac{1}{8}\,x^4 + O(x^6),$$

and recall that, for $x = x_6(n)$, we have $x^2 = \frac{1}{n}\,(a_1^2 + 2a_1a_2\,\frac{1}{n}) + O(n^{-5/2})$ and $x^4 = \frac{1}{n^2}\,a_1^4 + O(n^{-5/2})$. Thus,

$$\frac{\varphi(x)}{\|\varphi\|_\infty} = 1 - \frac{a_1^2}{2n} + \Big(\frac{a_1^4}{8} - a_1a_2\Big)\frac{1}{n^2} + O\big(n^{-5/2}\big).$$

Therefore, denoting $b = b_1 + b_2 + b_3 + b_4$, we get

$$\begin{aligned}\frac{\|\varphi_6\|_\infty}{\|\varphi\|_\infty} = \frac{\varphi_6(x)}{\|\varphi\|_\infty} &= \frac{\varphi(x)}{\|\varphi\|_\infty}\Big(1 + \frac{Q_1(x)}{\sqrt{n}} + \frac{Q_2(x)}{n} + \frac{Q_3(x)}{n^{\frac{3}{2}}} + \frac{Q_4(x)}{n^2}\Big)\\ &= 1 + \Big(-\frac{1}{2}\,a_1^2 + \frac{1}{4}\,\gamma_3^2 + \frac{3}{4!}\,\gamma_4 - \frac{15}{2!\cdot 3!^2}\,\gamma_3^2\Big)\frac{1}{n}\\ &\quad + \bigg(b + \frac{1}{8}\,a_1^4 - a_1a_2 - \frac{1}{2}\Big(\frac{1}{4}\,\gamma_3^2 + \frac{3}{4!}\,\gamma_4 - \frac{15}{2!\cdot 3!^2}\,\gamma_3^2\Big)\,a_1^2\bigg)\,\frac{1}{n^2} + O\big(n^{-5/2}\big).\end{aligned}$$

Simplifying the term in front of $1/n$, we arrive at

$$\|\varphi_6\|_\infty \;=\; \|\varphi\|_\infty + \frac{\|\varphi\|_\infty}{n}\,A + \frac{\|\varphi\|_\infty}{n^2}\,B + O\big(n^{-5/2}\big),$$

where

$$A \;=\; \frac{1}{8}\Big(\gamma_4 - \frac{2}{3}\,\gamma_3^2\Big), \qquad B \;=\; b + \frac{1}{8}\,a_1^4 - a_1a_2 - \frac{1}{2}\Big(\frac{1}{4}\,\gamma_3^2 + \frac{3}{4!}\,\gamma_4 - \frac{15}{2!\cdot 3!^2}\,\gamma_3^2\Big)\,a_1^2. \tag{11.56}$$

Using our assumptions, let us summarize by recalling the assertion (11.52). We then get

$$\|p_n\|_\infty = \|\varphi\|_\infty\Big(1 + \frac{1}{n}\,A + \frac{1}{n^2}\,B\Big) + o\big(n^{-2}\big), \tag{11.57}$$

where $A$ and $B$ are as above with $a_1 = -\frac{1}{2}\,\gamma_3$ and $a_2 = \frac{1}{4}\,\gamma_3^3 - \frac{5}{12}\,\gamma_3\gamma_4 + \frac{1}{8}\,\gamma_5$.

One can now reformulate this result in terms of the Rényi entropy of index $r = \infty$. Since $N_\infty(Z_n) = \|p_n\|_\infty^{-2}$ and $N_\infty(Z) = \|\varphi\|_\infty^{-2}$ for $Z \sim N(0,1)$, the expansion (11.57) yields:

**Proposition 11.10.1** *If $\beta_6$ is finite, then as $n \to \infty$,*

$$N_\infty(Z_n) \;=\; N_\infty(Z)\Big(1 - \frac{\widetilde{A}}{n} + \frac{\widetilde{B}}{n^2}\Big) + o\Big(\frac{1}{n^2}\Big)$$

*with* $\widetilde{A} = \frac{1}{4}\,(\gamma_4 - \frac{2}{3}\,\gamma_3^2)$*,* $\widetilde{B} = 3A^2 - 2B$*, where the constants* $A$ *and* $B$ *are given in* (11.56).

*Proof of Theorem 11.1.3 in Case* $r = \infty$ Denoting $\Delta_n = N_\infty(Z) - N_\infty(Z_n)$, from (11.57) we get $\Delta_{n+1} - \Delta_n = -\frac{\widetilde{A}}{n(n+1)} + o(\frac{1}{n^2})$. □

In the case $\gamma_3 = \gamma_5 = 0$, for example when $X$ is symmetric, the coefficients in Proposition 11.10.1 are simplified. Indeed, recalling the formula for $b_4$ in such a case, we have

$$A = \frac{1}{8}\,\gamma_4, \qquad B = b_4 = \frac{105}{2 \cdot 4!^2}\,\gamma_4^2 - \frac{15}{6!}\,\gamma_6,$$

and therefore,

$$\widetilde{A} = \frac{1}{4}\,\gamma_4, \qquad \widetilde{B} = 3A^2 - 2B = \frac{1}{24}\,\gamma_6 - \frac{13}{96}\,\gamma_4^2.$$

As a consequence, the eventual monotonicity of $N_\infty(Z_n)$ can be deduced based on the sign of $\gamma_4$. However, if also $\gamma_4 = 0$, we need to look at the sign of $\gamma_6$.

**Acknowledgements** This research was partially supported by NSF grant DMS-1855575 and BSF grant 2016050. Also, this research was supported in part by the Walter S. Baer and Jeri Weiss CMI Postdoctoral Fellowship.

## References


1. S. Artstein, K.M. Ball, F. Barthe, A. Naor, Solution of Shannon's problem on the monotonicity of entropy. J. Am. Math. Soc. **17**(4), 975–982 (2004)
2. A.R. Barron, Entropy and the central limit theorem. Ann. Probab. **14**(1), 336–342 (1986)
3. R.N. Bhattacharya, R. Ranga Rao, *Normal Approximation and Asymptotic Expansions* (Wiley, Hoboken, 1976). Also: Soc. for Industrial and Appl. Math., Philadelphia, 2010
4. S.G. Bobkov, G.P. Chistyakov, Entropy power inequality for the Rényi entropy. IEEE Trans. Inform. Theory **61**(2), 708–714 (2015)
5. S.G. Bobkov, A. Marsiglietti, Variants of the entropy power inequality. IEEE Trans. Inform. Theory **63**(12), 7747–7752 (2017)
6. S.G. Bobkov, A. Marsiglietti, Asymptotic behavior of Rényi entropy in the central limit theorem (extended version) (2018). Preprint. arXiv:1802.10212
7. S.G. Bobkov, G.P. Chistyakov, F. Götze, Non-uniform bounds in local limit theorems in case of fractional moments. I. Math. Methods Stat. **20**(3), 171–191 (2011); **20**(4), 269–287 (2011)
8. S.G. Bobkov, G.P. Chistyakov, F. Götze, Rate of convergence and Edgeworth-type expansion in the entropic central limit theorem. Ann. Probab. **41**(4), 2479–2512 (2013)
9. T.M. Cover, J.A. Thomas, *Elements of Information Theory*, 2nd edn. (Wiley-Interscience [Wiley], Hoboken, 2006) xxiv+748 pp.
10. A. Dembo, T.M. Cover, J.A. Thomas, Information-theoretic inequalities. IEEE Trans. Inform. Theory **37**(6), 1501–1518 (1991)

11. B.V. Gnedenko, A.N. Kolmogorov, *Limit Distributions for Sums of Independent Random Variables*. Translated and annotated by K. L. Chung. With an Appendix by J. L. Doob (Addison-Wesley Publishing Company, Cambridge, 1954), ix+264 pp.
12. M. Madiman, A. Barron, Generalized entropy power inequalities and monotonicity properties of information. IEEE Trans. Inform. Theory **53**(7), 2317–2329 (2007)
13. L.V. Osipov, V.V. Petrov, On the estimation of the remainder term in the central limit theorem. Teor. Veroyatn. Primen. **12**, 322–329 (1967)
14. V.V. Petrov, Local limit theorems for sums of independent random variables (Russian). Teor. Verojatnost. Primenen. **9**, 343–352 (1964)
15. V.V. Petrov, *Sums of Independent Random Variables*. Translated from the Russian by A. A. Brown. Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 82 (Springer, New York, 1975), x+346 pp. Russian ed.: Moscow, Nauka, 1972, 414 pp.

# Chapter 12
# Uniform-in-Bandwidth Functional Limit Laws for Multivariate Empirical Processes

**Paul Deheuvels**

**Abstract** We provide uniform-in-bandwidth functional limit laws for multivariate local empirical processes. Statistical applications to kernel density estimation are given to motivate these results.

**Keywords** Functional limit laws · Kernel density estimation · Weak laws

**AMS 2000 Subject Classification** Primary 60F15, 60F17; Secondary 60G07

## 12.1 Introduction and Motivation

We establish uniform-in-bandwidth functional limit laws for local empirical processes in $\mathbb{R}^d$. Our main result, stated in Theorem 12.2.1, is motivated by statistical applications presented in Theorem 12.1.1. Let $\mathbf{X}^* = (\mathbf{X}, Y) \in \mathbb{R}^{d+1}$, with $\mathbf{X} := (X(1), \ldots, X(d)) \in \mathbb{R}^d$ and $Y \in \mathbb{R}$, denote a random vector [rv], with continuous density $g_{\mathbf{X},Y}(\cdot, \cdot)$ on $\mathbb{R}^{d+1} = \mathbb{R}^d \times \mathbb{R}$, and support in $\mathbf{J} \times L$, where $\mathbf{J}$ and $L$ are bounded open subsets of $\mathbb{R}^d$ and $\mathbb{R}$, respectively. Under these assumptions, the marginal density $f(\cdot)$ of $\mathbf{X}$ is continuous on $\mathbb{R}^d$, with $f(\mathbf{x}) = 0$ for $\mathbf{x} \notin \mathbf{J}$, and

$$f(\mathbf{x}) := \int_L g_{\mathbf{X},Y}(\mathbf{x}, y)\mathrm{d}y \quad \text{for} \quad \mathbf{x} \in \mathbb{R}^d. \tag{12.1}$$

Let $\mathcal{K}$ denote a family of *kernels* on $\mathbb{R}^d$, namely, of mappings $\mathbf{K} : \mathbb{R}^d \to \mathbb{R}$, fulfilling conditions $(K.1)$–$(K.4)$ below. For $\mathbf{u} := (u_1, \ldots, u_d) \in \mathbb{R}^d$ and $\mathbf{v} := (v_1, \ldots, v_d) \in \mathbb{R}^d$, we write $\mathbf{u} \le \mathbf{v}$ when $u_j \le v_j$ for $j = 1, \ldots, d$. When this condition holds, we set $(\mathbf{u}, \mathbf{v}] := \prod_{j=1}^d (u_j, v_j]$, and define likewise, with obvious notation, $[\mathbf{u}, \mathbf{v}]$ and $(\mathbf{u}, \mathbf{v})$. In general, by an *interval* in $[r, s]^d$ will be

P. Deheuvels (✉)
LPSM, Sorbonne Université, Bourg-la-Reine, France
e-mail: paul.deheuvels@upmc.fr



N. Gozlan et al. (eds.), *High Dimensional Probability VIII*,
Progress in Probability 74, https://doi.org/10.1007/978-3-030-26391-1_12

meant a product of $d$ subintervals of $[r, s]$. We set $\mathbf{0} := (0, \ldots, 0) \in \mathbb{R}^d$ and $\mathbf{1} := (1, \ldots, 1) \in \mathbb{R}^d$, and adopt a similar notation for $\boldsymbol{\infty} := (\infty, \ldots, \infty)$.

$(K.1)$ There exist an $A < \infty$, such that, for each $\mathbf{K} \in \mathcal{K}$, $\mathbf{K}(\mathbf{t}) = 0$ when $|\mathbf{t}| \geq A$ (with $|\cdot|$ denoting the Euclidian norm in $\mathbb{R}^d$);
$(K.2)$ There exists a $B < \infty$ such that each $\mathbf{K} \in \mathcal{K}$ has a Hardy-Krause variation $\mathcal{V}_{\mathrm{HK}}(\mathbf{K})$ in $\mathbb{R}^d$, fulfilling $\mathcal{V}_{\mathrm{HK}}(\mathbf{K}) \leq B$ (see Sect. 12.2.3 below for details);
$(K.3)$ Each $\mathbf{K}(\mathbf{t}) \in \mathcal{K}$ is a right-continuous function of $\mathbf{t} = (t_1, \ldots, t_d)$;
$(K.4)$ For all $\mathbf{K} \in \mathcal{K}$, $\int_{\mathbb{R}^d} \mathbf{K}(\mathbf{t}) \mathrm{d}\mathbf{t} = 1$ (where $\mathrm{d}\mathbf{t}$ denotes Lebesgue measure).

Let $\psi : \mathbb{R} \to \mathbb{R}$ denote a right-continuous function of bounded variation $\|\mathrm{d}\psi\|_L$ on $L$. We will denote by $\|\mathrm{d}\psi\| := \|\mathrm{d}\psi\|_{\mathbb{R}}$ the total variation of $\psi$ on $\mathbb{R}$. In most of our examples, $\psi$ will be a linear combination of the identity mapping, $\mathcal{I}(y) = y$, and of the unit function, $\mathbb{I}(y) = 1$, for $y \in \mathbb{R}$. Consider a sequence of independent and identically distributed [iid] random replicæ $\mathbf{X}_i^* = (\mathbf{X}_i, Y_i)$, $i = 1, 2, \ldots$, of $\mathbf{X}^* = (\mathbf{X}, Y)$. Introduce the *kernel statistic* indexed by $\mathbf{K} \in \mathcal{K}$,

$$f_{\psi;n;h;\mathbf{K}}(\mathbf{x}) := (nh)^{-1} \sum_{i=1}^{n} \psi(Y_i) \mathbf{K}\left(h^{-1/d}(\mathbf{X}_i - \mathbf{x})\right) \quad \text{for} \quad \mathbf{x} \in \mathbb{R}^d, \tag{12.2}$$

where $h > 0$ is a *bandwidth* parameter. In particular, $f_{n;h;\mathbf{K}}(\mathbf{x}) := f_{\mathbb{I};n;h;\mathbf{K}}(\mathbf{x})$ is the Parzen-Rosenblatt [29, 30] kernel estimator of $f(\mathbf{x})$, which, under $(K.1)$–$(K.4)$, fulfills $\int_{\mathbb{R}^d} f_{\mathbb{I};n;h;\mathbf{K}}(\mathbf{x}) \mathrm{d}\mathbf{x} = 1$.

Let $\mathbf{I} := \prod_{j=1}^{d} [u_j, v_j] \subset \mathbf{J}$ with $-\infty < u_j < v_j < \infty$ for $j = 1, \ldots, d$, be such that $f(\mathbf{x}) > 0$ for all $\mathbf{x} \in \mathbf{I}$. The conditional expectation (or *regression*) of $\psi(Y)$, given that $\mathbf{X} = \mathbf{x}$, is continuous over $\mathbf{x} \in \mathbf{I}$, and defined by

$$\begin{aligned} m_\psi(\mathbf{x}) &:= \mathbb{E}(\psi(Y)|\mathbf{X} = \mathbf{x}) = \frac{f_\psi(\mathbf{x})}{f(\mathbf{x})} = \frac{f_\psi(\mathbf{x})}{f_{\mathbb{I}}(\mathbf{x})} \\ &= \frac{1}{f(\mathbf{x})} \int_L \psi(y) g_{\mathbf{X},Y}(\mathbf{x}, y) \mathrm{d}y \quad \text{for} \quad \mathbf{x} \in \mathbf{I}, \end{aligned} \tag{12.3}$$

where, for each measurable $\phi : \mathbb{R} \to \mathbb{R}$, rendering meaningful the expression below, we set

$$f_\phi(\mathbf{x}) := \int_L \phi(y) g_{\mathbf{X},Y}(\mathbf{x}, y) \mathrm{d}y \quad \text{for} \quad \mathbf{x} \in \mathbf{I}. \tag{12.4}$$

In view of (12.1) and (12.4), for $\phi = \mathbb{I}$, (12.4) reduces to $f_{\mathbb{I}}(\mathbf{x}) = f(\mathbf{x})$. Under the above assumptions, the conditional variance of $\psi(Y)$, given $\mathbf{X} = \mathbf{x}$, is continuous over $\mathbf{x} \in \mathbf{I}$, and given by

$$\begin{aligned} \sigma_\psi^2(\mathbf{x}) &:= \mathrm{Var}\,(\psi(Y)|\mathbf{X} = \mathbf{x}) \\ &= \frac{1}{f(\mathbf{x})} \int_L \left(\psi(y) - m_\psi(\mathbf{x})\right)^2 g_{\mathbf{X},Y}(\mathbf{x}, y) \mathrm{d}y \quad \text{for} \quad \mathbf{x} \in \mathbf{I}. \end{aligned} \tag{12.5}$$

The kernel estimator of the *regression function* $m_\psi(\mathbf{x}) = \mathbb{E}(\psi(Y)|\mathbf{X} = \mathbf{x})$ [25, 40], is then defined, for $\mathbf{x} \in \mathbf{I}$, by

$$m_{\psi;n;h;\mathbf{K}}(\mathbf{x}) := \begin{cases} \dfrac{f_{\psi;n;h;\mathbf{K}}(\mathbf{x})}{f_{\mathrm{II};n;h;\mathbf{K}}(\mathbf{x})} & \text{when} \quad f_{\mathrm{II};n;h;\mathbf{K}}(\mathbf{x}) > 0, \\ \overline{Y} := n^{-1}\sum_{i=1}^n Y_i & \text{when} \quad f_{\mathrm{II};n;h;\mathbf{K}}(\mathbf{x}) \leq 0. \end{cases} \tag{12.6}$$

Introduce, whenever properly defined, the centering factor

$$\widehat{\mathbb{E}}\left(m_{\psi;n;h;\mathbf{K}}(\mathbf{x})\right) := \frac{\mathbb{E}\Big(\psi(Y)\mathbf{K}(h^{-1/d}(\mathbf{X}-\mathbf{x}))\Big)}{\mathbb{E}\Big(\mathbf{K}(h^{-1/d}(\mathbf{X}-\mathbf{x}))\Big)}. \tag{12.7}$$

*Remark 12.1.1* Under $(K.1)$–$(K.4)$, for $\mathbf{x} \in \mathbf{I}$, we have $\mathbb{E}\left(f_{n;h;\mathbf{K}}(\mathbf{x})\right) \to f(\mathbf{x})$ and $\widehat{\mathbb{E}}\left(m_{\psi;n;h;\mathbf{K}}(\mathbf{x})\right) \to m_\psi(\mathbf{x})$, as $h \to 0$. (see, e.g., [9]). Thus, in the study of the consistency of $f_{n;h;\mathbf{K}}(\mathbf{x})$ and $m_{\psi;n;h;\mathbf{K}}(\mathbf{x})$, we will limit ourselves to the evaluation of the limiting behavior of the *random components* $f_{n;h;\mathbf{K}}(\mathbf{x}) - \mathbb{E}\left(f_{n;h;\mathbf{K}}(\mathbf{x})\right)$ and $m_{\psi;n;h;\mathbf{K}}(\mathbf{x}) - \widehat{\mathbb{E}}\left(m_{\psi;n;h;\mathbf{K}}(\mathbf{x})\right)$ of the estimators.

Let $0 < a_n \leq b_n$, for $n \geq 1$, be sequences of real constants, and set $\log_+ x := \log(x \vee e)$ for $x \in \mathbb{R}$. We have the following theorem.

**Theorem 12.1.1** *Assume $(K.1)$–$(K.4)$, and let $0 < a_n \leq b_n$ be such that, as $n \to \infty$,*

$$na_n/\log n \to \infty \quad \textit{and} \quad b_n \to 0. \tag{12.8}$$

*Then, with $\mathcal{H}_n := [a_n, b_n]$, we have, as $n \to \infty$,*

$$\sup_{\mathbf{K}\in\mathcal{K}}\left(\sup_{h\in\mathcal{H}_n}\left|\left\{\frac{nh}{2\log_+(1/h)}\right\}^{1/2}\sup_{\mathbf{x}\in\mathbf{I}}\pm\Big\{f_{n;h;\mathbf{K}}(\mathbf{x}) - \mathbb{E}\left(f_{n;h;\mathbf{K}}(\mathbf{x})\right)\Big\} - \left\{\sup_{\mathbf{x}\in\mathbf{I}} f(\mathbf{x})\int_{\mathbb{R}^d}\mathbf{K}(\mathbf{t})^2 d\mathbf{t}\right\}^{1/2}\right|\right) = o_{\mathbb{P}}(1), \tag{12.9}$$

*and*

$$\sup_{\mathbf{K}\in\mathcal{K}}\left(\sup_{h\in\mathcal{H}_n}\left|\left\{\frac{nh}{2\log_+(1/h)}\right\}^{1/2}\sup_{\mathbf{x}\in\mathbf{I}}\pm\Big\{m_{\psi;n;h;\mathbf{K}}(\mathbf{x}) - \widehat{\mathbb{E}}\left(m_{\psi;n;h;\mathbf{K}}(\mathbf{x})\right)\Big\} - \left\{\sup_{\mathbf{x}\in\mathbf{I}}\frac{\sigma_\psi^2(\mathbf{x})}{f(\mathbf{x})}\int_{\mathbb{R}^d}\mathbf{K}(\mathbf{t})^2 d\mathbf{t}\right\}^{1/2}\right|\right) = o_{\mathbb{P}}(1). \tag{12.10}$$

*Remark 12.1.2*

1°) When $\mathcal{K} = \{\mathbf{K}\}$ and $d = 1$, (12.9) in Theorem 12.1.1 reduces to Theorem 2 of Deheuvels and Ouadah [10]. This property does not hold for an arbitrary $f(\cdot)$, when (12.8) is not fulfilled (see Remark 1 in [10]).

2°) By Theorem 12.1.1, taken with $\mathcal{K} = \{\mathbf{K}\}$ and $h_n := a_n = b_n$, the condition

$$h_n \to 0 \quad \text{and} \quad nh_n/\log n \to \infty, \tag{12.11}$$

implies that, as $n \to \infty$,

$$\left\{\frac{nh_n}{2\log_+(1/h_n)}\right\}^{1/2} \sup_{\mathbf{x}\in\mathbf{I}} \pm\left\{f_{n;h_n;\mathbf{K}}(\mathbf{x}) - \mathbb{E}\left(f_{n;h_n;\mathbf{K}}(\mathbf{x})\right)\right\} \tag{12.12}$$
$$\xrightarrow{\mathbb{P}} \left\{\sup_{\mathbf{x}\in\mathbf{I}} f(\mathbf{x}) \int_{\mathbb{R}^d} \mathbf{K}(\mathbf{t})^2 d\mathbf{t}\right\}^{1/2},$$

and

$$\left\{\frac{nh_n}{2\log_+(1/h_n)}\right\}^{1/2} \sup_{\mathbf{x}\in\mathbf{I}} \pm\{m_{\psi;n;h_n;\mathbf{K}}(\mathbf{x}) - \widehat{\mathbb{E}}\left(m_{\psi;n;h_n;\mathbf{K}}(\mathbf{x})\right)\} \tag{12.13}$$
$$\xrightarrow{\mathbb{P}} \left\{\sup_{\mathbf{x}\in\mathbf{I}} \frac{\sigma_\psi^2(\mathbf{x})}{f(\mathbf{x})} \int_{\mathbb{R}^d} \mathbf{K}(\mathbf{t})^2 d\mathbf{t}\right\}^{1/2}.$$

The limiting statement (12.12) is due to Deheuvels [8] for $d = 1$, and [6] for $d \geq 1$ (see, e.g., Deheuvels and Einmahl [5], Deheuvels and Mason [9]). Earlier, Silverman [32] had established (12.12) for $d = 1$, under more stringent assumptions. Equation (12.13) is a particular case of Theorem 1.1 in Deheuvels and Mason [9] for $d = 1$, and of Theorem 1.2 in Deheuvels [7] for $d \geq 2$. The case where the rv $Y$ has an unbounded support, will be considered elsewhere.

3°) The conclusion of Theorem 12.1.1 remains valid when $a_n \leq b_n$ are random sequences such that (12.8) holds in probability. As follows from the results of Deheuvels and Mason [8] and Deheuvels [5], additional conditions are required to obtain an almost sure [a.s.] version of this theorem.

4°) The properties of the estimators (12.2) and (12.6) have been extensively investigated since the seminal work of Rosenblatt [30], Parzen [29], Nadaraya [25] and Watson [40]. To allow data-dependent bandwidths, several authors (see, e.g., Mason et al. [24], Nolan and Marron [27], Deheuvels [4], Deheuvels and Mason [9]) have provided *uniform-in-bandwidth* limit laws for $f_{n,h}(\cdot)$, in the spirit of (12.9) and (12.10). Einmahl and Mason [16, 17] initiated the use of empirical processes indexed by functions to investigate this problem. For example, Theorem 1 of [17] shows that, for each $r > 0$,

$$\limsup_{n\to\infty} \Bigg( \sup_{\frac{r\log n}{n} \leq h \leq 1} \left\{\frac{nh}{\log(1/h) \vee \log\log n}\right\}^{1/2} \tag{12.14}$$
$$\sup_{\mathbf{x}\in\mathbf{I}} |f_{n;h;\mathbf{K}}(\mathbf{x}) - \mathbb{E}\left(f_{n;h;\mathbf{K}}(\mathbf{x})\right)| \Bigg) =: \mathcal{K}(I, r) < \infty,$$

a.s. for some $\mathcal{K}(I,r)$. We refer to Mason [22], Mason and Swanepoel [23], Dony [11, 13], Dony and Einmahl [12, 13], Dony et al. [15], Mason [21], Viallon [38], Varron [36, 37] and van Keilegom and Varron [35], for details on this methodology. In particular, an adaptation of the arguments of [16, 17] should allow us to prove that, under (12.8), as $n \to \infty$

$$\sup_{h\in\mathcal{H}_n}\left\{\frac{nh}{2\log_+(1/h)}\right\}^{1/2}\sup_{\mathbf{x}\in\mathbf{I}}\left|f_{n;h;\mathbf{K}}(\mathbf{x})-\mathbb{E}\left(f_{n;h;\mathbf{K}}(\mathbf{x})\right)\right| \qquad (12.15)$$
$$-\left\{\sup_{\mathbf{x}\in\mathbf{I}} f(\mathbf{x})\int_{\mathbb{R}^d}\mathbf{K}(\mathbf{t})^2 d\mathbf{t}\right\} = o_{\mathbb{P}}(1).$$

It is not clear whether a proof of (12.9) (which is a stronger statement that (12.15)) can be achieved or not by these methods. Here, we make use of a different argument, based on the ideas of Deheuvels and Mason [8] and Deheuvels [5]. Further references are that of Dony and Mason [14] and Mason [20].

An outline of the remainder of our paper is as follows. We establish, in Theorem 12.2.1 below, a functional limit law for multivariate increments of a *non-uniform* empirical process (which is new, even for $d = 1$). To prove this theorem, we rely on classical arguments, to obtain, in the forthcoming Sect. 12.3.1, rough upper bounds for the modulus of continuity of multivariate empirical processes. Our proof then reduces to show that, for each fixed $M \geq 1$, the $N := M^d$ properly rescaled increments of the multivariate empirical process over sets of the form $\prod_{j=1}^d(\frac{k_j}{M},\frac{k_j+1}{M}]$, cluster onto the unit ball of $\mathbb{R}^N$. To establish this property, we extend arguments of Deheuvels and Ouadah [10] to an dimension-free framework. The proof of Theorem 12.1.1 given Theorem 12.2.1 is captured in Sect. 12.2.4 below. The proofs being quite lengthy, we limit ourselves to the main arguments.

## 12.2 Functional Limit Laws

### 12.2.1 Main Result

For $d \geq 1$, let $(B([0,1]^d),\mathcal{U})$ denote the set $B([0,1]^d)$ of bounded functions on $[0,1]^d$, endowed with the topology $\mathcal{U}$, induced by the sup-norm $\|g\| := \sup_{\mathbf{u}\in[0,1]^d}|g(\mathbf{u})|$. Let $AC([0,1]^d)$ denote the set of absolutely continuous (with respect to the Lebesgue measure) functions on $[0,1]^d$, and set $AC_{\mathbf{0}}([0,1]^d) := \{g \in AC([0,1]^d) : g(\mathbf{0}) = 0\}$, with $\mathbf{0} := (0,\dots,0) \in \mathbb{R}^d$. For each $\varepsilon > 0$ and $g \in B([0,1]^d)$, set $\mathcal{N}_\varepsilon(g) := \left\{\phi \in B([0,1]^d) : \|\phi - g\| < \varepsilon\right\}$, and for each

$A \subseteq [0,1]^d$, set $A^\varepsilon := \bigcup_{g \in A} \mathcal{N}_\varepsilon(g)$, with the convention that $\bigcup_{\emptyset}(\cdot) := \emptyset$. Define the sup-norm Hausdorff set-distance of $A, B \subseteq B([0,1]^d)$ by

$$\Delta(A, B) := \inf\{\theta > 0 : A \subseteq B^\theta \text{ and } B \subseteq A^\theta\},$$
$$\text{whenever such a } \theta \text{ exists, and}$$
$$\Delta(A, B) := \infty \quad \text{otherwise.}$$

Let $\dot{g}$ denote the Lebesgue derivative of $g \in AC([0,1]^d)$, and consider the Hilbert norm, defined on $B([0,1]^d)$ by

$$|g|_{\mathbb{H}} := \left\{ \int_{[0,1]^d} \dot{g}(\mathbf{t})^2 d\mathbf{t} \right\}^{1/2} \text{ when } g \in AC_{\mathbf{0}}([0,1]^d),$$
$$|g|_{\mathbb{H}} := \infty \quad \text{otherwise.}$$

Set $\mathbb{S}_d = \{g \in B([0,1]^d) : |g|_{\mathbb{H}} \leq 1\}$. For $d = 1$, we will use this notation with subscripts omitted, and write, e.g., $\mathbb{S}$ for $\mathbb{S}_1$. The following relations follow readily from the Schwarz inequality and the definitions of $|\cdot|_{\mathbb{H}}$ and $\mathbb{S}_d$. For any $\psi \in B([0,1]^d)$, we have

$$\|\psi\| \leq |\psi|_{\mathbb{H}} \quad \text{and} \quad \sup_{g \in \mathbb{S}_d} \|g\| = 1. \tag{12.16}$$

Letting $\mathbf{X} := \mathbf{X}_1, \mathbf{X}_2, \ldots$ be as in Sect. 12.1, we denote the distribution function [df] of $\mathbf{X}$ by $\mathbb{F}(\mathbf{x}) := \mathbb{P}(\mathbf{X} \leq \mathbf{x})$ for $\mathbf{x} \in \mathbb{R}^d$. Here, we write $\mathbf{x} \leq \mathbf{y}$, for $\mathbf{x} = (x(1), \ldots, x(d)) \in \mathbb{R}^d$ and $\mathbf{y} = (y(1), \ldots, y(d)) \in \mathbb{R}^d$, whenever $x(j) \leq y(j)$ for $j = 1, \ldots, d$. Denote the empirical df based upon $\mathbf{X}_1, \ldots, \mathbf{X}_n$, by

$$\mathbb{F}_n(\mathbf{x}) := n^{-1} \#\{\mathbf{X}_i \leq \mathbf{x} : 1 \leq i \leq n\} \quad \text{for} \quad \mathbf{x} \in \mathbb{R}^d, \tag{12.17}$$

where # denotes cardinality. Introduce the empirical process

$$a_n(\mathbf{x}) := n^{1/2}(\mathbb{F}_n(\mathbf{x}) - \mathbb{F}(\mathbf{x})) \quad \text{for} \quad \mathbf{x} \in \mathbb{R}^d. \tag{12.18}$$

Let $\mathbf{I} \subset \mathbf{J}$, with $\mathbf{I} = \prod_{j=1}^d [u_j, v_j]$, and $-\infty < u_j < v_j < \infty$ for $j = 1, \ldots, d$, be as in Sect. 12.1. We assume that the density $f(\cdot)$ of $\mathbf{X}$ is defined and continuous on $\mathbf{J}$, and bounded away from 0 on $\mathbf{I} \subset \mathbf{J}$. For $a > 0$, and $\mathbf{x} \in \mathbf{I}$, we consider the increment functions

$$\upsilon_n(a; \mathbf{x}; \mathbf{u}) := \{a_n(\mathbf{x} + a^{1/d}\mathbf{u}) - a_n(\mathbf{x})\} / \sqrt{f(\mathbf{x})}, \tag{12.19}$$
$$\text{for} \quad \mathbf{u} \in [0,1]^d,$$

and set, for each $a > 0$, and $\mathbf{L} \subseteq \mathbf{I}$,

$$\mathcal{F}_{n;a;\mathbf{L}} = \left\{ \frac{\upsilon_n(a; \mathbf{x}; \cdot)}{\sqrt{2a \log_+(1/a)}} : \mathbf{x} \in \mathbf{L} \right\}. \tag{12.20}$$

Our main theorem may now be stated as follows.

**Theorem 12.2.1** *Let* $0 < a_n \leq b_n$ *be such that, as* $n \to \infty$,

$$b_n \to 0 \quad \text{and} \quad na_n / \log n \to \infty. \tag{12.21}$$

*Then, with* $\mathcal{H}_n = [a_n, b_n]$, *we have, as* $n \to \infty$,

$$\sup_{a \in \mathcal{H}_n} \Delta \left( \mathcal{F}_{n;a;\mathbf{I}}, \mathbb{S}_d \right) = o_{\mathbb{P}}(1). \tag{12.22}$$

*Remark 12.2.1*

1°) It will become obvious from our proofs that the conclusion of Theorem 12.2.1 remains valid if, in the definition (12.19) of $\upsilon(a; \mathbf{x}; \mathbf{u})$, $\mathbf{u}$ is assumed to vary in $\left[-\frac{1}{2}, \frac{1}{2}\right]$ (or in any specified bounded interval $[r, s]$, with $r < s$) instead of $[0, 1]$.

2°) To our best knowledge, the only version of Theorem 12.2.1 available up to now correspond to $d = 1$, and under the assumption that $\mathbf{X}$ uniformly distributed on $(0, 1)$ (see, e.g., Theorem 1(1) of Deheuvels and Ouadah [10]). When $a_n = b_n$ the problem has been considered by Deheuvels and Mason [8] and Deheuvels [5]) for $d = 1$, and by Mason [21] for $d \geq 1$. We note that the methods of [10] cannot be extended to $d \geq 2$, since the proofs rely on invariance principles for empirical processes, which are not presently available with the proper approximation rates.

The proof of Theorem 12.2.1 is postponed until Sect. 12.3. In the forthcoming Sect. 12.2.4, we shall provide a proof of Theorem 12.1.1 given Theorem 12.2.1.

### *12.2.2 A Limit Law for Local Empirical Processes Indexed by Functions*

Let $\mathcal{K}$ denote a class of measurable functions defined on $\mathbb{R}^d$, with support in $\left[-\frac{1}{2}, \frac{1}{2}\right]^d$, and fulfilling $(K.1)$–$(K.3)$. Following (2.3)–(2.4) in Mason [21], for each $n \geq 1$, $h > 0$ and $\mathbf{x} \in \mathbb{R}^d$, denote the local empirical process at $\mathbf{x}$ indexed by $\mathbf{K} \in \mathcal{K}$ by

$$\mathcal{E}_n(h; \mathbf{x}; K) := (nh)^{-1/2} \sum_{i=1}^{n} \Big\{ \mathbf{K}(h^{-1/d}(\mathbf{x} - \mathbf{X}_i)) - \mathbb{E}\mathbf{K}(h^{-1/d}(\mathbf{x} - \mathbf{X}_i)) \Big\} = \sqrt{nh} \left\{ f_{n;h;\mathbf{K}}(\mathbf{x}) - \mathbb{E}\left( f_{n;h;\mathbf{K}}(\mathbf{x}) \right) \right\}, \tag{12.23}$$

and set, for $\mathbf{x} \in \mathbf{I}$,

$$\mathcal{L}_n(a; \mathbf{x}; K) = \frac{\mathcal{E}_n(a; \mathbf{x}; \mathbf{K})}{\sqrt{2 \log_+(1/a) f(\mathbf{x})}}. \tag{12.24}$$

*Remark 12.2.2* Mason [21] make use of different conditions imposed upon $\mathcal{K}$. He assumes, namely that

$$\lim_{\|\mathbf{t}\| \to 0} \sup_{\mathbf{K} \in \mathcal{K}} \int_{\mathbb{R}^d} [\mathbf{K}(\mathbf{x} + \mathbf{t}) - \mathbf{K}(\mathbf{x})]^2 \, d\mathbf{x} = 0,$$

$$\lim_{\lambda \to 1} \sup_{\mathbf{K} \in \mathcal{K}} \int_{\mathbb{R}^d} [\mathbf{K}(\lambda \mathbf{x}) - \mathbf{K}(\mathbf{x})]^2 \, d\mathbf{x} = 0,$$

### *12.2.3 Properties of Kernels*

We discuss here $(K.1)$–$(K.4)$. In $(K.1)$, the choice of the interval $[-A, A]^d \subset \mathbb{R}^d$ supporting the kernels $\mathbf{K} \in \mathcal{K}$, is a matter of convenience, so that we will work, without loss of generality, under the following variant of this assumption, for some $0 < \epsilon < \frac{1}{2}$.

$(K.1)^*$ Each $\mathbf{K} \in \mathcal{K}$ is such that $\mathbf{K}(\mathbf{t}) = 0$ for all $\mathbf{t} \notin \mathbf{I}_\epsilon := [\epsilon, 1 - \epsilon]^d$.

The condition $(K.2)$, requires each $\mathbf{K} \in \mathcal{K}$ to be of *Hardy-Krause bounded variation*. For functions of several variables, this notion is involved (see, e.g., Adams and Clarkson [1, 3], Niederreiter [26]), and some details must be given. The most common forms of variation [18, 19, 39], are as follows (see, e.g., Niederreiter [26, p. 22]). Set $\mathbf{I}_0 = [0, 1]^d$, and, for $1 \leq k \leq 1$ and $1 \leq i_1 < \ldots < i_k \leq d$, define a *face* of $\mathbf{I}_0$, by $\mathbf{I}_0(i_1, \ldots, i_k) := \{\mathbf{t} = (t_1, \ldots, t_d) \in \mathbf{I}_0 : t_j = 1 \text{ for } j \notin \{i_1, \ldots, i_k\}\}$. By an *interval* $\mathcal{J} \subseteq \mathbf{I}_0$, will be meant a product of $d$ subintervals of [0, 1]. Denote the lower endpoint of $\mathcal{J}$ by $\mathbf{t}(\mathcal{J})$. For any function $\boldsymbol{\kappa}$ defined on $\mathbf{I}_0$, let $\Delta(\boldsymbol{\kappa}; \mathcal{J})$

denote the alternating sum of values of $\boldsymbol{\kappa}$ at vertices of $\mathcal{J}$, where $\boldsymbol{\kappa}(\mathbf{t}(\mathcal{J}))$ has coefficient 1. The *Vitali variation* of $\boldsymbol{\kappa}$ on $\mathbf{I}_0$ is then given by

$$\mathcal{V}_{\mathrm{V}}(\boldsymbol{\kappa};\mathbf{I}_0) := \sup_{\mathcal{P}(\mathbf{I}_0)} \sum_{\mathcal{J}\in\mathcal{P}(\mathbf{I}_0)} |\Delta(\boldsymbol{\kappa};\mathcal{J})|,$$

where the supremum is taken over all partitions $\mathcal{P}(\mathbf{I}_0)$ of $\mathbf{I}_0$ into subintervals $\mathcal{J} \subseteq \mathbf{I}_0$. The *Hardy-Krause variation* of $\boldsymbol{\kappa}$ on $\mathbf{I}_0$ is, in turn, defined by

$$\mathcal{V}_{\mathrm{HK}}(\boldsymbol{\kappa};\mathbf{I}_0) := \sum_{k=1}^{d}\Bigg\{ \sum_{1\le i_1<\ldots<i_k\le d} \mathcal{V}_{\mathrm{V}}(\kappa;\mathbf{I}_0(i_1,\ldots,i_k))\Bigg\},$$

which sums, over all faces $\mathbf{I}_0(i_1,\ldots,i_k)$ of $\mathbf{I}_0$, the Vitali variation of the restriction of $\boldsymbol{\kappa}$ to $\mathbf{I}_0(i_1,\ldots,i_k)$. For $d = 1$, the Vitali and Hardy-Krause variations coincide with the usual *total variation*. In these definitions, we may replace $\mathbf{I}_0$ by other intervals of $\mathbb{R}^d$, via book-keeping arguments. In particular, we set, in $(K.2)$, $\mathcal{V}_{\mathrm{HK}}(\boldsymbol{\kappa}) := \mathcal{V}_{\mathrm{HK}}(\boldsymbol{\kappa};\mathbb{R}^d) := \sup_{m\ge1}\mathcal{V}_{\mathrm{HK}}(\boldsymbol{\kappa};[-m,m]^d)$.

Subject to the existence of continuous partial derivatives of $\boldsymbol{\kappa}$, the Vitali and Hardy-Krause variations of $\boldsymbol{\kappa}$ on $\mathbf{I}_0$ are given, respectively, by

$$\mathcal{V}_{\mathrm{V}}(\boldsymbol{\kappa};\mathbf{I}_0) = \int_{\mathbf{I}_0}\left|\frac{\partial^d\boldsymbol{\kappa}(\mathbf{t})}{\partial t_1\ldots\partial t_d}\right| d\mathbf{t},$$

$$\mathcal{V}_{\mathrm{HK}}(\boldsymbol{\kappa};\mathbf{I}_0) = \sum_{k=1}^{d}\Bigg\{ \sum_{1\le i_1<\ldots<i_k\le d} \int_{\mathbf{I}_0(i_1,\ldots,i_k)}\left|\frac{\partial^k\boldsymbol{\kappa}(\mathbf{t})}{\partial t_{i_1}\ldots\partial t_{i_k}}\right| dt_{i_1}\ldots dt_{i_k}\Bigg\}.$$

In this case, an induction on $d$ allows us to write, for each $\mathbf{0} \le \mathbf{u} \le \mathbf{v} \le \mathbf{1}$,

$$\begin{aligned}\boldsymbol{\kappa}(\mathbf{v}) - \boldsymbol{\kappa}(\mathbf{u}) = \sum_{k=1}^{d}\Bigg\{ \sum_{1\le i_1<\ldots<i_k\le d} &\int_{\mathbf{t}\in\mathbf{I}_0(i_1,\ldots,i_k),\ \mathbf{u}<\mathbf{t}\le\mathbf{v}} \\ &(-1)^{k-d}\,\frac{\partial^k\boldsymbol{\kappa}(\mathbf{t})}{\partial t_{i_1}\ldots\partial t_{i_k}}\,\mathrm{d}t_{i_1}\ldots\mathrm{d}t_{i_k}\Bigg\},\end{aligned} \tag{12.25}$$

In general, subject to $\mathcal{V}_{\mathrm{HK}}(\boldsymbol{\kappa};\mathbf{I}_0) < \infty$, the totally bounded Lebesgue-Stieltjes signed measure $\boldsymbol{\nu} = \mathrm{d}\boldsymbol{\kappa}(\cdot)$, associated with $\boldsymbol{\kappa}$ and supported by $\mathbf{I}_0$, is defined by setting, for each continuous function $\phi$ on $\mathbf{I}_0$,

$$\begin{aligned}\int_{\mathbf{I}_0}\phi(\mathbf{t})\mathrm{d}\boldsymbol{\kappa}(\mathbf{t}) = \sum_{k=1}^{d}\Bigg\{ \sum_{1\le i_1<\ldots<i_k\le d} &\lim_{|\mathcal{P}(\mathbf{I}_0(i_1,\ldots,i_k))|\to0} \\ &\sum_{\mathcal{J}\in\mathcal{P}(\mathbf{I}_0(i_1,\ldots,i_k))} (-1)^{k-d}\phi(\mathbf{t}(\mathcal{J}))\Delta(\boldsymbol{\kappa};\mathcal{J})\Bigg\}.\end{aligned} \tag{12.26}$$

Here, we set $|\mathcal{P}(\mathbf{I}_0(i_1,\ldots,i_k))| \to 0$, when the supremum vertice length of the intervals $\mathcal{J} \in \mathcal{P}(\mathbf{I}_0(i_1,\ldots,i_k))$ tends to 0. The kernel functions we consider have simple expressions in terms of $\boldsymbol{\nu} = \mathrm{d}\boldsymbol{\kappa}$. When $\boldsymbol{\kappa}$ is right-continuous, with $\boldsymbol{\kappa}(\mathbf{t}) = 0$ for $\mathbf{t} \notin \mathbf{I}_\varepsilon = [\epsilon, 1-\epsilon]^d$, $\boldsymbol{\kappa}(\mathbf{0}) = \boldsymbol{\kappa}(\mathbf{1}) = 0$, so that, by (12.26),

$$\boldsymbol{\kappa}(\mathbf{t}) = -\boldsymbol{\nu}((\mathbf{t},\mathbf{1}]) = \boldsymbol{\nu}((\mathbf{0},\mathbf{t}]) \quad \text{for} \quad \mathbf{t} \in \mathbf{I}_0. \tag{12.27}$$

Observe that $\boldsymbol{\nu} = \mathrm{d}\boldsymbol{\kappa}(\cdot)$ in (12.27) is a totally bounded signed measure with support in $\mathbf{I}_\varepsilon$. Letting $\boldsymbol{\nu} = \boldsymbol{\nu}_+ - \boldsymbol{\nu}_-$ denote the Hahn-Jordan decomposition of $\boldsymbol{\nu}$ into the difference of nonnegative bounded measures with supports in $\mathbf{I}_\varepsilon \subset \mathbf{I}_0$, we infer from (12.27) that these component measures fulfill

$$\boldsymbol{\kappa}(\mathbf{0}) = -\boldsymbol{\nu}((\mathbf{0},\mathbf{1}]) = -\boldsymbol{\nu}(\mathbf{I}_0) = -\boldsymbol{\nu}(\mathbf{I}_\varepsilon) = \boldsymbol{\nu}_-(\mathbf{I}_\varepsilon) - \boldsymbol{\nu}_+(\mathbf{I}_\varepsilon) = 0,$$

so that $0 \le \boldsymbol{\nu}_+(\mathbf{I}_\varepsilon) = \boldsymbol{\nu}_-(\mathbf{I}_\varepsilon) < \infty$. Following Bouleau [2] (see, e.g., p. 166 in Pagès and Xiao [28]), we define the *measure variation* of $\boldsymbol{\kappa}$ on $\mathbf{I}_0$, by

$$\mathcal{V}_{\mathrm{M}}(\boldsymbol{\kappa};\mathbf{I}_0) = \|\mathrm{d}\boldsymbol{\kappa}\|_{\mathrm{M}} := |\boldsymbol{\nu}|(\mathbf{I}_0) := \boldsymbol{\nu}_+(\mathbf{I}_0) + \boldsymbol{\nu}_-(\mathbf{I}_0). \tag{12.28}$$

The above-defined variations are related through the inequalities

$$\mathcal{V}_{\mathrm{V}}(\boldsymbol{\kappa};\mathbf{I}_0) \le \mathcal{V}_{\mathrm{M}}(\boldsymbol{\kappa};\mathbf{I}_0) \le \mathcal{V}_{\mathrm{HK}}(\boldsymbol{\kappa};\mathbf{I}_0) \le (2^d-1)\mathcal{V}_{\mathrm{M}}(\boldsymbol{\kappa};\mathbf{I}_0), \tag{12.29}$$

where $2^d - 1$ stands for the number of faces $\mathbf{I}_0(i_1,\ldots,i_k)$ of $\mathbf{I}_0$. In view of (12.29), under $(K.1)^*$–$(K.3)$, the assumption $(K.2)$ is equivalent to:

$(K.2)^*$ There exists a $B^* < \infty$ such that each $\mathbf{K} \in \mathcal{K}$ has a measure variation in $\mathbf{I}_0$ fulfilling $\mathcal{V}_{\mathrm{M}}(\mathbf{K};\mathbf{I}_0) \le B^*$.

Armed with these arguments, we establish, in Lemma 12.2.1 below, a useful integration by parts formula. We consider nonnegative bounded measures $\boldsymbol{\mu}_i$, $i = 1, 2$ and $\boldsymbol{\nu}_i$, $i = 1, 2$, with supports in $\mathbf{I}_\epsilon := [\epsilon, 1-\epsilon]^d$, and such that $\boldsymbol{\mu}_1(\mathbf{I}_\epsilon) = \boldsymbol{\mu}_2(\mathbf{I}_\epsilon)$, and $\boldsymbol{\nu}_1(\mathbf{I}_\epsilon) = \boldsymbol{\nu}_2(\mathbf{I}_\epsilon)$. Set, for $\mathbf{0} \le \mathbf{s} \le \mathbf{t} \le \mathbf{1}$,

$$\mathbf{M}_1(\mathbf{s},\mathbf{t}) = \Big\{\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\Big\}((\mathbf{s},\mathbf{t}]) \quad \text{and} \quad \mathbf{M}_2(\mathbf{s},\mathbf{t}) = \Big\{\boldsymbol{\nu}_1 - \boldsymbol{\nu}_2\Big\}((\mathbf{s},\mathbf{t}]) .$$

By (12.27), taken with $\boldsymbol{\nu} = d\{-\mathbf{M}_2(\mathbf{t},\mathbf{1})\}$ and $\boldsymbol{\kappa}(\mathbf{t}) = -\mathbf{M}_2(\mathbf{t},\mathbf{1})$, we see that $\boldsymbol{\nu}_1 - \boldsymbol{\nu}_2 = \mathrm{d}\{-\mathbf{M}_2(\mathbf{t},\mathbf{1})\}$ coincides with the Lebesgue-Stieltjes measure $\boldsymbol{\nu}$ induced by $-\mathbf{M}_2(\mathbf{t},\mathbf{1})$. Likewise, by (12.27), taken with $\boldsymbol{\nu} = \mathrm{d}\mathbf{M}_2(\mathbf{0},\mathbf{t})$ and $\boldsymbol{\kappa}(\mathbf{t}) = \mathbf{M}_1(\mathbf{0},\mathbf{t})$, we see that $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = \mathrm{d}\mathbf{M}_1(\mathbf{0},\mathbf{t})$ coincides with the Lebesgue-Stieltjes measure $\boldsymbol{\nu}$ induced by $\mathbf{M}_1(\mathbf{0},\mathbf{t})$.

**Lemma 12.2.1** *Under the assumptions above, we have the integration by parts formula*

$$\int_{\mathbf{I}_0} \mathbf{M}_1(\mathbf{0},\mathbf{t})\mathrm{d}\mathbf{M}_2(\mathbf{t},\mathbf{1}) = \int_{\mathbf{I}_0} \mathbf{M}_2(\mathbf{t},\mathbf{1})\mathrm{d}\mathbf{M}_1(\mathbf{0},\mathbf{t}). \tag{12.30}$$

*Proof* We limit ourselves to the case where $\mathbf{M}_1(\mathbf{0},\mathbf{t})$ and $\mathbf{M}_2(\mathbf{t},\mathbf{1})$ have continuous partial derivatives of order $d$ over $\mathbf{t} \in \mathbb{R}^d$. The proof in the general case is achieved by a smoothing argument which we omit. Observe that, for all $1 \le k < d$ and $1 \le i_1 < \ldots < i_k \le d$, we have $\mathbf{M}_2(\mathbf{t},\mathbf{1}) = 0$ for $\mathbf{t} \in \mathbf{I}_0(i_1,\ldots,i_k)$. Therefore, we may rewrite (12.25) into

$$\mathbf{M}_2(\mathbf{t},\mathbf{1}) = (-1)^d \int_{\mathbf{s}\in\mathbf{I}_0,\ \mathbf{t}<\mathbf{s}\le\mathbf{1}} \frac{\partial^d \mathbf{M}_2(\mathbf{s},\mathbf{1})}{\partial s_1 \ldots \partial s_d}\,\mathrm{d}\mathbf{s}. \tag{12.31}$$

By a similar argument, with the formal replacement of $\mathbf{M}_2(\mathbf{t},\mathbf{1})$ by $\mathbf{M}_1(\mathbf{0},\mathbf{t})$, we may rewrite (12.25) into

$$\mathbf{M}_1(\mathbf{0},\mathbf{t}) = \int_{\mathbf{s}\in\mathbf{I}_0,\ \mathbf{0}<\mathbf{s}\le\mathbf{t}} \frac{\partial^d \mathbf{M}_1(\mathbf{0},\mathbf{s})}{\partial s_1 \ldots \partial s_d}\,\mathrm{d}\mathbf{s}. \tag{12.32}$$

This shows that the signed measures $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = \mathrm{d}\mathbf{M}_1(\mathbf{0},\mathbf{t})$ and $-\{\boldsymbol{\nu}_1 - \boldsymbol{\nu}_2\} = \mathrm{d}\mathbf{M}_2(\mathbf{t},\mathbf{1})$ are absolutely continuous with respect to the Lebesgue measure in $\mathbb{R}^d$, with densities given, respectively, by

$$\mathbf{m}(\mathbf{t}) := \frac{\mathrm{d}\mathbf{M}_1(\mathbf{0},\mathbf{t})}{\mathrm{d}\mathbf{t}} = \frac{\partial^d \mathbf{M}_1(\mathbf{0},\mathbf{t})}{\partial t_1 \ldots \partial t_d},$$

and

$$\mathbf{n}(\mathbf{t}) := \frac{\mathrm{d}\mathbf{M}_2(\mathbf{t},\mathbf{1})}{\mathrm{d}\mathbf{t}} = (-1)^d \frac{\partial^d \mathbf{M}_2(\mathbf{t},\mathbf{1})}{\partial t_1 \ldots \partial t_d}.$$

Set $\mathbf{M}_{1;0}(\mathbf{t}) = \mathbf{m}(\mathbf{t})$, $\mathbf{M}_{2;0}(\mathbf{t}) = \mathbf{n}(\mathbf{t})$, and, for $1 \le k \le d$,

$$\mathbf{M}_{1;k}(\mathbf{t}) = \int_0^{t_1} \ldots \int_0^{t_k} \mathbf{m}(\mathbf{s})\mathrm{d}s_1 \ldots \mathrm{d}s_k$$

and

$$\mathbf{M}_{2;k}(\mathbf{t}) = \int_{t_1}^{1} \ldots \int_{t_k}^{1} \mathbf{n}(\mathbf{s})\mathrm{d}s_1 \ldots \mathrm{d}s_k.$$

Observe that $\mathbf{M}_{1;d}(\mathbf{t}) = \mathbf{M}_1(\mathbf{0},\mathbf{t})$, $\mathbf{M}_{2;d}(\mathbf{t}) = \mathbf{M}_2(\mathbf{t},\mathbf{1})$, and, for $1 \le k \le d$, $\frac{\partial}{dt_k}\mathbf{M}_{1;k}(\mathbf{t}) = \mathbf{M}_{1;k-1}(\mathbf{t})$ and $\frac{\partial}{dt_k}\mathbf{M}_{2;k}(\mathbf{t}) = -\mathbf{M}_{2;k-1}(\mathbf{t})$. In addition, for $1 \le k \le d$,

$\mathbf{M}_{1;k}(\mathbf{t}) = 0$ when $t_k = 0$ and $\mathbf{M}_{2;k}(\mathbf{t}) = 0$ when $t_k = 1$. We may therefore write the chain of equalities

$$\int_{[0,1]^d} \mathbf{M}_1(\mathbf{0}, \mathbf{t}) d\mathbf{M}_2(\mathbf{t}, \mathbf{1}) = (-1)^d \int_{[0,1]^d} \mathbf{M}_{1;d}(\mathbf{t})\mathbf{n}(\mathbf{t})\mathrm{d}\mathbf{t}$$

$$= (-1)^d \int_{[0,1]^d} \mathbf{M}_{1;d}(\mathbf{t})\mathbf{M}_{2;0}(\mathbf{t})d\mathbf{t} = (-1)^d \int_{[0,1]^d} \mathbf{M}_{1;d}(\mathbf{t})\tfrac{\partial}{\partial t_1}\mathbf{M}_{2;1}(\mathbf{t})\mathrm{d}\mathbf{t}$$

$$= (-1)^d \int_{[0,1]^{d-1}} \Big\{\Big[\mathbf{M}_{1;d}(\mathbf{t})\mathbf{M}_{2;1}(\mathbf{t})\Big]_{t_1=0}^{t_1=1}$$

$$- \int_0^1 \tfrac{\partial}{\partial t_1}\mathbf{M}_{1;d}(\mathbf{t})\mathbf{M}_{2;1}(\mathbf{t})\mathrm{d}t_1\Big\}\mathrm{d}t_2 \dots \mathrm{d}t_d$$

$$= (-1)^{d-1} \int_{[0,1]^d} \mathbf{M}_{1;d-1}(\mathbf{t})\mathbf{M}_{2;1}(\mathbf{t})\mathrm{d}\mathbf{t} = \dots = \int_{[0,1]^d} \mathbf{M}_{1;0}(\mathbf{t})\mathbf{M}_{2;d}(\mathbf{t})\mathrm{d}\mathbf{t}$$

$$= \int_{[0,1]^d} \mathbf{M}_{2;d}(\mathbf{t})\mathbf{m}(\mathbf{t})\mathrm{d}\mathbf{t} = \int_{[0,1]^d} \mathbf{M}_2(\mathbf{t}, \mathbf{1})\mathrm{d}\mathbf{M}_1(\mathbf{0}, \mathbf{t}),$$

which is (12.30). □

*Remark 12.2.3* The version of (12.30) corresponding to $d = 1$, is readily checked, when $\mathbf{m}(\cdot)$ and $\mathbf{n}(\cdot)$ are continuous on [0, 1]. We obtain the relations

$$\int_0^1 \Big\{\int_0^t \mathbf{m}(s)ds\Big\} \mathrm{d}\Big\{\int_t^1 \mathbf{n}(s)ds\Big\} = \Big[\Big\{\int_0^t \mathbf{m}(s)ds\Big\}\Big\{\int_t^1 \mathbf{n}(s)ds\Big\}\Big]_{t=0}^{t=1}$$

$$- \int_0^1 \Big\{\int_t^1 \mathbf{n}(s)ds\Big\} \mathrm{d}\Big\{\int_0^t \mathbf{m}(s)ds\Big\} = - \int_0^1 \Big\{\int_t^1 \mathbf{n}(s)ds\Big\} \mathbf{m}(t)\mathrm{d}t.$$

### *12.2.4 Proof of Theorem 12.1.1*

For each $\mathbf{K} \in \mathcal{K}$, set $\widetilde{\mathbf{K}}(\mathbf{u}) = \mathbf{K}(-\mathbf{u})$, and let $\widetilde{\mathcal{K}} = \{\widetilde{\mathbf{K}} : \mathbf{K} \in \mathcal{K}\}$. Following the arguments pp. 1278–1281 of [8], we may reduce the proof of (12.9) to the case where $\widetilde{K}$ fulfills $(K.1)^*$–$(K.2)^*$ and $(K.3)$, so that $\widetilde{\mathbf{K}}(\mathbf{u}) := \mathbf{K}(-\mathbf{u}) = 0$ for $\mathbf{u} \notin (0, 1)^d$. In view of (12.27), let $\mathrm{d}\widetilde{\mathbf{K}}(\cdot)$ be the Lebesgue-Stieltjes measure induced by $\widetilde{\mathbf{K}}$, in such a way that

$$-\widetilde{\mathbf{K}}(\mathbf{t}) = \int_{(\mathbf{t},\mathbf{1}]} \mathrm{d}\widetilde{\mathbf{K}}(\mathbf{u}).$$

Let $h_0 > 0$ be so small that $\mathbf{I}+h_0^{1/d}[0,1]^d \subset \mathbf{J}$. By an application of Lemma 12.2.1, and making use of the definition (12.19) of $\upsilon_n(h;\mathbf{x};\mathbf{u})$, we see that, for each $\mathbf{x} \in \mathbf{I}$, and $0 < h \le h_0$,

$$\begin{aligned}
&\left\{\frac{nh}{2\log_+(1/h)}\right\}^{1/2} \left(f_{n,h}(\mathbf{x}) - \mathbb{E}\left(f_{n,h}(\mathbf{x})\right)\right) \qquad (12.33)\\
&= \int_{[0,1]^d} \widetilde{\mathbf{K}}(\mathbf{u}) \left\{\frac{\mathrm{d}\{a_n(\mathbf{x}+h^{1/d}\mathbf{u}) - a_n(\mathbf{x})\}}{\sqrt{2h\log_+(1/h)}}\right\}\\
&= -\int_{[0,1]^d} \left\{\frac{a_n(\mathbf{x}+h^{1/d}\mathbf{u}) - a_n(\mathbf{x})}{\sqrt{2h\log_+(1/h)}}\right\} \mathrm{d}\widetilde{\mathbf{K}}(\mathbf{u})\\
&= -\sqrt{f(\mathbf{x})} \int_{[0,1]^d} \frac{\upsilon_n(h;\mathbf{x};\mathbf{u})}{\sqrt{2h\log_+(1/h)}} \mathrm{d}\widetilde{\mathbf{K}}(\mathbf{u}).
\end{aligned}$$

We will need the following analytical result (see, e.g., Lemma 1 in [10]). Let $\mathcal{M}$ denote a subset of $B([0,1]^d)$, such that $\mathbb{S}_d \subseteq \mathcal{M} \subseteq B([0,1]^d)$, and let $\mathcal{T}$ denote a non-empty class of mappings $\Theta : \mathcal{M} \to \mathbb{R}$, continuous with respect to the uniform topology on $\mathcal{M}$. We assume that $\mathcal{T}$ has the following equicontinuity property. For each $\epsilon > 0$, there exists an $\eta(\epsilon) > 0$ such that, for each $\phi \in \mathcal{M}$ and $g \in \mathbb{S}_d$, we have

$$\|\phi - g\| < \eta(\epsilon) \;\Rightarrow\; \sup_{\Theta\in\mathcal{T}} |\Theta(\phi) - \Theta(g)| < \epsilon. \tag{12.34}$$

**Lemma 12.2.2** *Under the assumptions above, for each $\varepsilon > 0$, there exists a $\zeta(\varepsilon) > 0$, such that, for any $\mathcal{F} \subseteq \mathcal{M}$, we have*

$$\Delta(\mathcal{F},\mathbb{S}) < \zeta(\varepsilon) \;\Rightarrow\; \sup_{\Theta\in\mathcal{T}} \left| \sup_{\phi\in\mathcal{F}} \Theta(\phi) - \sup_{g\in\mathbb{S}_d} \Theta(g) \right| < \varepsilon. \tag{12.35}$$

Consider an arbitrary $\Theta \in \mathcal{T}$. By compactness of $\mathbb{S}_d$ and continuity of $\Theta$, there exists a $g_\Theta \in \mathbb{S}_d$ such that $\Theta(g_\Theta) = \sup_{g\in\mathbb{S}_d} \Theta(g)$. Letting $\eta(\varepsilon)$ be as in (12.34), we see that, for each $\varepsilon > 0$, and $\phi \in \mathcal{M}$ such that $\|\phi - g_\Theta\| \le \eta(\varepsilon)$, we have $\sup_{\Theta\in\mathcal{T}} |\Theta(\phi) - \Theta(g_\Theta)| < \varepsilon$. In view of the implication $\Delta(\mathcal{F},\mathbb{S}_d) \le \eta(\varepsilon) \;\Rightarrow\; \mathbb{S}_d \subseteq \mathcal{F}^{\eta(\varepsilon)}$, we see that $\Delta(\mathcal{F},\mathbb{S}_d) \le \eta(\varepsilon)$ implies the existence of a $\phi_\Theta \in \mathcal{F}$ such that $\|\phi_\Theta - g_\Theta\| < \eta(\varepsilon)$. By an application of (12.34), we obtain therefore, that, whenever $\Delta(\mathcal{F},\mathbb{S}_d) \le \eta(\varepsilon)$,

$$\forall\, \Theta \in \mathcal{T} : \sup_{\phi\in\mathcal{F}} \Theta(\phi) - \sup_{g\in\mathbb{S}_d} \Theta(g) \ge \Theta(\phi_\Theta) - \Theta(g_\Theta) \ge -\varepsilon. \tag{12.36}$$

Consider now the assumption

$$(H): \quad \left\{\forall\, \eta > 0, \exists\, \phi \in \mathcal{M} \cap \mathbb{S}_d^{\eta} : \sup_{\Theta \in \mathcal{T}} \left\{\Theta(\phi) - \sup_{g \in \mathbb{S}_d} \Theta(g)\right\} \geq \varepsilon\right\}.$$

Under $(H)$, there exists a sequence $(\phi_n, \Theta_n) \in (\mathcal{M} \cap \mathbb{S}_d^{1/n}, \mathcal{T})$, $n = 1, 2, \ldots$, such that $\phi_n \in \mathcal{M} \cap \mathbb{S}_d^{1/n}$, and $\Theta_n(\phi_n) \geq \sup_{g \in \mathbb{S}_d} \Theta_n(g) + \varepsilon$, for all $n \geq 1$. The condition $\phi_n \in \mathbb{S}_d^{1/n}$ implies the existence, for each $n \geq 1$, of a $\psi_n \in \mathbb{S}$, such that $\|\phi_n - \psi_n\| \leq 1/n$. The compactness of $\mathbb{S}$ implies the existence of a convergent subsequence $\psi_{n_k} \to \psi \in \mathbb{S}_d$ as $k \to \infty$. Since then, $\|\phi_{n_k} - \psi\| \to 0$, as $k \to \infty$, an application of (12.34) shows that, as $k \to \infty$, $\sup_{\Theta \in \mathcal{T}} \left|\Theta(\phi_{n_k}) - \Theta(\psi)\right| \to 0$. This entails that, for all $k$ sufficiently large,

$$\Theta_{n_k}(\phi_{n_k}) < \Theta_{n_k}(\psi) + \varepsilon \leq \sup_{g \in \mathbb{S}_d} \Theta_{n_k}(g) + \varepsilon,$$

which contradicts $(H)$. The impossibility of $(H)$ implies the existence of an $\eta_1(\varepsilon)$ such that whenever $\mathcal{F} \subseteq \mathcal{M}$ fulfills $\Delta(\mathcal{F}, \mathbb{S}_d) \leq \eta_1(\varepsilon)$, and hence, $\mathcal{F} \subseteq \mathbb{S}_d^{\eta_1(\varepsilon)}$, we have

$$\forall\, \Theta \in \mathcal{T} : \sup_{\phi \in \mathcal{F}} \Theta(\phi) - \sup_{g \in \mathbb{S}_d} \Theta(g) \leq \varepsilon. \tag{12.37}$$

The conclusion (12.35) follows from (12.36) to (12.37), with $\zeta(\varepsilon) := \eta(\varepsilon) \wedge \eta_1(\varepsilon)$. □

*Example 12.2.1*

1°) Let $\mathcal{M} = B([0, 1]^d)$, and $\mathcal{T} = \{\Theta_0\}$, with $\Theta_0(g) := \|g\|$. Since $\sup_{\Theta \in \mathcal{T}} |\Theta(\phi) - \Theta(g)| = \|\phi - g\|$, we see that (12.34) holds with $\eta(\varepsilon) = \varepsilon$, so that the assumptions of Lemma 12.2.2 are fulfilled.

2°) Let $\mathcal{K}$, where $\mathcal{K}$ fulfill $(K.1)^*$–$(K.2)^*$–$(K.3)$, and choose $\mathcal{M}$ as the set of all bounded measurable functions on $[0, 1]^d$. The inclusions $\mathbb{S}_d \subseteq \mathcal{M} \subseteq B([0, 1]^d)$ are then straightforward. Consider the functionals

$$g \in BV_{\mathbf{0};\mathrm{HK}}([0, 1]^d) \to \Theta_K(g) = \int_{[0,1]^d} g(\mathbf{u})\mathrm{d}K(\mathbf{u}),$$

for $\mathbf{K} \in \mathcal{K}$. In view of the obvious inequality, for $g_1, g_2 \in BV_0([0, 1]^d)$,

$$|\Theta_K(g_1) - \Theta_K(g_2)| \leq \|g_1 - g_2\| \times \mathcal{V}_{\mathrm{M}}(\mathbf{K}, \mathbf{I}_0) \leq B^* \|g_1 - g_2\|,$$

we see that (12.34) is fulfilled, with $\eta(\varepsilon) = \varepsilon / B^*$.

By a *rectangle* in $\mathbb{R}^d$ will be meant a product of $d$ subintervals of $\mathbb{R}$. Below, we will denote by $|A|$ the Lebesgue measure of a measurable $A \subset \mathbb{R}^d$. Since $f(\cdot)$ is continuous on $\mathbf{J} \supset \mathbf{I}$, for each $0 < \epsilon < \nu := \inf_{\mathbf{x}\in\mathbf{I}} \sqrt{f(\mathbf{x})}$, we may partition the rectangle $\mathbf{I}$ into $\mathbf{I} = \mathbf{I}_1 \cup \ldots \cup \mathbf{I}_M$, where $\mathbf{I}_1, \ldots, \mathbf{I}_M \subset \mathbf{I}$ are disjoint rectangles in $\mathbb{R}^d$ such that, for $j = 1, \ldots, M$, $|\mathbf{I}_j| > 0$ and

$$m_j := \sup_{\mathbf{x}\in\mathbf{I}_j} \sqrt{f(\mathbf{x})} \geq \inf_{\mathbf{x}\in\mathbf{I}_j} \sqrt{f(\mathbf{x})} > m_j - \epsilon \geq \nu - \epsilon > 0.$$

By setting $\mathbf{L} = \mathbf{I}_j$, for $j = 1, \ldots, M$, and $a = h$ in (12.20), we may therefore write, for each $j = 1, \ldots, M$ and $0 < h \leq h_0$, the relations

$$\sup_{\mathbf{x}\in\mathbf{I}_j} \left| \left\{ m_j - \sqrt{f(\mathbf{x})} \right\} \int_{[0,1]^d} \frac{\upsilon_n(h;\mathbf{x};\mathbf{u})}{\sqrt{2h\log_+(1/h)}} \, d\widetilde{\mathbf{K}}(\mathbf{u}) \right| \tag{12.38}$$

$$\leq \epsilon \left\{ \sup_{g\in\mathcal{F}_{n;h;\mathbf{I}_j}} \|g\| \right\} \int_{[0,1]^d} |d\widetilde{\mathbf{K}}(\mathbf{u})| = \epsilon \left\{ \sup_{g\in\mathcal{F}_{n;h;\mathbf{I}_j}} \|g\| \right\} \|d\mathbf{K}\|,$$

where $\|d\mathbf{K}\| < \infty$ denotes the total variation of $\mathbf{K}(\cdot)$ on $\mathbb{R}^d$. Set now $\Theta(g) = \Theta_0(g) := \|g\|$ and $\mathcal{F} = \mathcal{F}_{n;h;\mathbf{I}_j}$. In view of (12.16) and (12.20), and by a repeated application of Theorem 12.2.1 with the formal replacement of $\mathbf{I}$ by $\mathbf{I}_j$, for $j = 1, \ldots, M$, we infer from (12.22) that, whenever $\mathcal{H}_n = [a_n, b_n]$ fulfills (12.21), we have, as $n \to \infty$,

$$\sup_{h\in\mathcal{H}_n} \left| \sup_{g\in\mathcal{F}_{n;h;\mathbf{I}_j}} \|g\| - \sup_{g\in\mathbb{S}_d} \|g\| \right| = \sup_{h\in\mathcal{H}_n} \left| \sup_{g\in\mathcal{F}_{n;h;\mathbf{I}_j}} \|g\| - 1 \right| = o_{\mathbb{P}}(1). \tag{12.39}$$

We infer readily from (12.38) and (12.39) that, as $n \to \infty$,

$$\mathbb{P}\Bigg( \max_{1\leq j\leq M} \sup_{h\in\mathcal{H}_n} \Bigg| \Bigg\{ \sup_{\mathbf{x}\in\mathbf{I}_j} \pm \left\{ \frac{nh}{2\log_+(1/h)} \right\}^{1/2} \left( f_{n,h}(\mathbf{x}) - \mathbb{E}\left( f_{n,h}(\mathbf{x}) \right) \right) \Bigg\}$$

$$- m_j \sup_{\mathbf{x}\in\mathbf{I}_j} \left\{ \pm (-1)^d \int_{[0,1]^d} \frac{\upsilon_n(h;\mathbf{x};\mathbf{u})}{\sqrt{2h\log_+(1/h)}} \, d\widetilde{\mathbf{K}}(\mathbf{u}) \right\} \Bigg| \geq 2\epsilon \|d\mathbf{K}\| \Bigg)$$

$$\leq \mathbb{P}\Bigg( \max_{1\leq j\leq M} \sup_{h\in\mathcal{H}_n} \Bigg\{ \sup_{\mathbf{x}\in\mathbf{I}_j} \Bigg| \left\{ m_j - \sqrt{f(\mathbf{x})} \right\}$$

$$\int_{[0,1]^d} \frac{\upsilon_n(h;\mathbf{x};\mathbf{u})}{\sqrt{2h\log_+(1/h)}} \, d\widetilde{\mathbf{K}}(\mathbf{u}) \Bigg| \Bigg\} \geq 2\epsilon \|d\mathbf{K}\| \Bigg) \to 0. \tag{12.40}$$

Set now

$$\Theta(g) = \Theta_1(g) := \pm \int_{[0,1]^d} g(\mathbf{u})\widetilde{\mathbf{K}}(\mathbf{u})d\mathbf{u}.$$

We may rewrite (12.40) into

$$\mathbb{P}\Big( \max_{1\leq j\leq M} \sup_{h\in\mathcal{H}_n} \Big| \Big\{ \sup_{\mathbf{x}\in\mathbf{I}_j} \pm \Big\{ \frac{nh}{2\log_+(1/h)} \Big\}^{1/2} \big(f_{n,h}(\mathbf{x}) - \mathbb{E}\big(f_{n,h}(\mathbf{x})\big)\big) \Big\} - m_j \sup_{g\in\mathcal{F}_{n;h;\mathbf{I}_j}} \Theta(g) \Big| \geq 2\epsilon\|d\mathbf{K}\| \Big) \to 0. \tag{12.41}$$

After integrating by parts, we combine the definition of $\mathbb{S}_d$ with the Schwarz inequality, to obtain that

$$\sup_{g\in\mathbb{S}_d} \Theta(g) = \sup_{g\in\mathbb{S}_d} \Big\{ \mp \int_{[0,1]^d} g(\mathbf{u}) d\widetilde{\mathbf{K}}(\mathbf{u}) \Big\} \tag{12.42}$$

$$= \sup_{g\in\mathbb{S}_d} \Big\{ \pm \int_{[0,1]^d} \dot{g}(\mathbf{u})\widetilde{\mathbf{K}}(\mathbf{u})d\mathbf{u} \Big\} = \Big\{ \int_{[0,1]^d} \mathbf{K}(\mathbf{u})^2 d\mathbf{u} \Big\}^{1/2}.$$

For $j = 1,\ldots,M$, set $\mathcal{F} = \mathcal{F}_{n;h;\mathbf{I}_j}$. In view of (12.16)–(12.20), and by an application of Theorem 12.2.1, with $\mathbf{I} = \mathbf{I}_j$, for $j = 1,\ldots,M$, we infer from (12.22) that, whenever $\mathcal{H}_n = [a_n, b_n]$ fulfills (12.21), we have, as $n \to \infty$,

$$\max_{1\leq j\leq M} \sup_{h\in\mathcal{H}_n} \Big| \sup_{g\in\mathcal{F}_{n;h;\mathbf{I}_j}} \Theta(g) - \sup_{g\in\mathbb{S}_d} \Theta(g) \Big|$$

$$= \max_{1\leq j\leq M} \sup_{h\in\mathcal{H}_n} \Big| \sup_{g\in\mathcal{F}_{n;h;\mathbf{I}_j}} \Theta(g) - \Big\{ \int_{[0,1]^d} \mathbf{K}(\mathbf{u})^2 d\mathbf{u} \Big\}^{1/2} \Big| = o_{\mathbb{P}}(1).$$

This, when combined with (12.41), implies that, as $n \to \infty$,

$$\mathbb{P}\Big( \sup_{h\in\mathcal{H}_n} \Big| \Big\{ \sup_{\mathbf{x}\in\mathbf{I}} \pm \Big\{ \frac{nh}{2\log_+(1/h)} \Big\}^{1/2} \big(f_{n,h}(\mathbf{x}) - \mathbb{E}(f_{n,h}(\mathbf{x}))\big) \Big\} - \Big\{ \sup_{\mathbf{x}\in\mathbf{I}} \sqrt{f(\mathbf{x})} \Big\} \Big\{ \int_{[0,1]^d} \mathbf{K}(\mathbf{u})^2 d\mathbf{u} \Big\}^{1/2} \Big| \geq \epsilon + 2\epsilon\|d\mathbf{K}\| \Big) \to 0. \tag{12.43}$$

Since $\epsilon \in (0, h_0]$ in (12.43) may be chosen arbitrarily small, we infer (12.9) from (12.43). This, together with routine arguments completes the proof of (12.9), given Theorem 12.2.1.

## 12.3 Proof of Theorem 12.2.1

### 12.3.1 A Bound for the Oscillation Modulus

In Proposition 12.3.1 below, we establish a rough bound for the oscillation modulus of the multivariate empirical process. This result will be instrumental in the proof of Theorem 12.2.1. We will work under the assumption that the support of the distribution of $\mathbf{X}$ is equal to $[0, 1]^d$, and that the density $f(\cdot)$ of $\mathbf{X}$ is continuous and bounded away from 0 on $[0, 1]^d$. This implies the existence of constants $C_1, C_2$, such that

$$0 < C_1 \leq f(\mathbf{x}) \leq C_2 < \infty \quad \text{for} \quad \mathbf{x} \in [0, 1]^d. \tag{12.44}$$

The assumption that $\int_{[0,1]^d} f(\mathbf{x})d\mathbf{x} = 1$, implies that $C_1, C_2$ in (12.44) fulfill

$$0 < C_1 \leq 1 \leq C_2 < \infty. \tag{12.45}$$

Moreover, we may extend the definition of $f(\cdot)$ to $\overline{\mathbb{R}}^d := [-\infty, \infty]^d$, by setting

$$f(\mathbf{x}) = 0 \quad \text{for} \quad \mathbf{x} \notin [0, 1]^d. \tag{12.46}$$

This entails that the *distribution function* [df] $\mathbb{F}(\mathbf{x}) := \mathbb{P}(\mathbf{X} \leq \mathbf{x})$ of $\mathbf{X} = (X(1), \ldots, X(d)) \in \mathbb{R}^d$, is continuous on $\overline{\mathbb{R}}^d$. For each $j = 1, \ldots, d$, set $\mathbf{x}^{[j]} := (x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_d)$ and $d\mathbf{x}^{[j]} := dx_1 \ldots dx_{j-1}dx_{j+1} \ldots dx_d$. As follows from (12.44)–(12.46), for each $j = 1, \ldots, d$, the $j$-th coordinate $X(j)$ of $\mathbf{X}$ has a continuous density $f^{[j]}(\cdot)$ on $[0, 1]$, fulfilling, for all $x_j \in [0, 1]$,

$$C_1 \leq f^{[j]}(x_j) = \int_{\mathbf{x}^{[j]} \in [0,1]^{d-1}} f(\mathbf{x})d\mathbf{x}^{[j]} \leq C_2. \tag{12.47}$$

This, in turn, implies that for each $j = 1, \ldots, d$, the $j$-th marginal df of $\mathbb{F}(\cdot)$, denoted by $F^{[j]}(x) := \mathbb{P}(X(j) \leq x)$, $x \in \overline{\mathbb{R}}$, is continuous on $\overline{\mathbb{R}}$, and such that $U(j) := F^{[j]}(X(j))$ is uniformly distributed on $[0, 1]$. For $j = 1, \ldots, d$, let $Q^{[j]}(t) := \inf\{x : F^{[j]}(x) \geq t\}$, $0 < t < 1$, $Q^{[j]}(0) := \inf\{x : F^{[j]}(x) > 0\}$, $Q^{[j]}(1) := \sup\{x : F^{[j]}(x) < 1\}$, denote the *quantile function* pertaining to $F^{[j]}(\cdot)$. For $j = 1, \ldots, d$, we have, almost surely [a.s.], $X(j) = Q^{[j]}(U(j))$. Without loss of generality, will therefore work on the set of probability 1 on which these relations hold. It is noteworthy that, unless $f(\mathbf{x}) = \prod_{j=1}^d f^{[j]}(x_j)$ for all $\mathbf{x} = (x_1, \ldots, x_d) \in [0, 1]^d$, the components $U(1), \ldots, U(d)$ of $\mathbf{U} := (U(1), \ldots, U(d))$ are not independent. Their joint df, $\mathbb{C}(\mathbf{u}) := \mathbb{P}(\mathbf{U} \leq \mathbf{u})$, $\mathbf{u} \in \mathbb{R}^d$, is the *copula function* of $F(\cdot)$ (see, e.g., Schweizer and Wolff [31]). We have the reciprocal relations

$$\mathbb{F}(\mathbf{x}) = \mathbb{C}(F^{[1]}(x_1), \ldots, F^{[d]}(x_d)) \quad \text{for} \quad \mathbf{x} = (x_1, \ldots, x_d) \in \overline{\mathbb{R}}^d, \tag{12.48}$$

and

$$\mathbb{C}(\mathbf{u}) = \mathbb{F}(Q^{[1]}(u_1), \ldots, Q^{[d]}(u_d)) \quad \text{for} \quad \mathbf{u} = (u_1, \ldots, u_d) \in [0, 1]^d. \quad (12.49)$$

We infer from (12.47) that, for each $j = 1, \ldots, d$, the $j$-th quantile density function $q^{[j]}(t) := \frac{d}{dt} Q^{[j]}(t)$, $t \in (0, 1)$, is defined and continuous on $(0, 1)$, and fulfills, for $0 < t < 1$,

$$0 < \frac{1}{C_2} \le q^{[j]}(t) = \frac{d}{dt} Q^{[j]}(t) = \frac{1}{f_j(Q^{[j]}(t))} \le \frac{1}{C_1} < \infty. \quad (12.50)$$

The relations (12.44), (12.47)–(12.50), readily imply that the copula function $\mathbb{C}(\cdot)$ has a density $c(\cdot)$ on $(0, 1)^d$, fulfilling the relations, for $\mathbf{x} = (x_1, \ldots, x_d) \in (0, 1)^d$ and $\mathbf{u} = (u_1, \ldots, u_d) \in (0, 1)^d$

$$\begin{aligned} 0 < C_1 \le f(\mathbf{x}) &= \frac{\partial^d}{\partial x_1 \ldots \partial x_d} \mathbb{F}(x_1, \ldots, x_d) \\ &= c(F^{[1]}(x_1), \ldots, F^{[d]}(x_d)) \prod_{j=1}^{d} f^{[j]}(x_j) \le C_2 < \infty, \quad (12.51) \end{aligned}$$

$$\begin{aligned} 0 < \frac{C_1}{C_2^d} \le c(\mathbf{u}) &= \frac{\partial^d}{\partial u_1 \ldots \partial u_d} \mathbb{C}(u_1, \ldots, u_d) \\ &= f(Q^{[1]}(u_1), \ldots, Q^{[d]}(u_d)) \prod_{j=1}^{d} q^{[j]}(u_j) \le \frac{C_2}{C_1^d} < \infty. \quad (12.52) \end{aligned}$$

Let now $\mathbf{X}_i = (X_i(1), \ldots, X_i(d))$, $i \ge 1$, be iid random copies of $\mathbf{X}$, and set $\mathbf{U}_i = (U_i(1), \ldots, U_i(d)) := (F^{[1]}(X_i(1)), \ldots, F^{[d]}(X_i(d)))$, $i \ge 1$. In agreement with the notation of Sect. 12.2.1, the empirical df's based, respectively, upon $\mathbf{U}_1, \ldots, \mathbf{U}_n$ and $\mathbf{X}_1, \ldots, \mathbf{X}_n$, are denoted by

$$\mathbb{C}_n(\mathbf{u}) := n^{-1}\#\{\mathbf{U}_i \le \mathbf{u} : 1 \le i \le n\}, \quad \mathbf{u} \in \mathbb{R}^d,$$

and

$$\mathbb{F}_n(\mathbf{x}) := n^{-1}\#\{\mathbf{X}_i \le \mathbf{x} : 1 \le i \le n\}, \quad \mathbf{x} \in \mathbb{R}^d.$$

The corresponding empirical processes are denoted by

$$a_{n;\mathbb{C}}(\mathbf{u}) := n^{1/2}\{\mathbb{C}_n(\mathbf{u}) - \mathbb{C}(\mathbf{u})\}, \quad \mathbf{u} \in \overline{\mathbb{R}}^d,$$

and

$$a_{n;\mathbb{F}}(\mathbf{x}) := n^{1/2}\{\mathbb{F}_n(\mathbf{x}) - \mathbb{F}(\mathbf{x})\}, \quad \mathbf{x} \in \overline{\mathbb{R}}^d.$$

Denote the set of all rectangles in $[0, 1]^d$ by $\mathcal{R}_d$. The empirical measures indexed by $\mathcal{R}_d$, based, respectively, upon $\mathbf{U}_1, \dots, \mathbf{U}_n$ and $\mathbf{X}_1, \dots, \mathbf{X}_n$, are denoted by

$$\mu_{n;\mathbb{C}}(I) = n^{-1}\#\{\mathbf{U}_i \in I : 1 \le i \le n\}, \quad I \in \mathcal{R}_d,$$

and

$$\mu_{n;\mathbb{F}}(I) = n^{-1}\#\{\mathbf{X}_i \in I : 1 \le i \le n\}, \quad I \in \mathcal{R}_d,$$

with expectations, given, respectively, by

$$\mu_{\mathbb{C}}(I) = \int_I c(\mathbf{u})d\mathbf{u} \quad \text{and} \quad \mu_{\mathbb{F}}(I) = \int_I f(\mathbf{x})d\mathbf{x} \quad \text{for} \quad I \in \mathcal{R}_d.$$

The corresponding empirical processes indexed by $\mathcal{R}_d$ are denoted by

$$a_{n;\mathbb{C}}(I) := n^{1/2}\left\{\mu_{n;\mathbb{C}}(I) - \mu_{\mathbb{C}}(I)\right\} \quad \text{for} \quad I \in \mathcal{R}_d,$$

and

$$a_{n;\mathbb{F}}(I) := n^{1/2}\left\{\mu_{n;\mathbb{F}}(I) - \mu_{\mathbb{F}}(I)\right\} \quad \text{for} \quad I \in \mathcal{R}_d.$$

For $0 \le u, v \le 1$, consider the modulus of continuity of $a_{n;\mathbb{C}}$ and $a_{n;\mathbb{F}}$, defined, respectively, by

$$\omega_{n;\mathbb{C}}(v) = \sup\Big\{|a_{n;\mathbb{C}}(\mathbf{t} + vI)| : I \in \mathcal{R}_d, \\ \mathbf{t} \in [0, 1]^d, \mathbf{t} + vI \subseteq [0, 1]^d\Big\}, \tag{12.53}$$

$$\omega_{n;\mathbb{F}}(u) = \sup\Big\{|a_{n;\mathbb{F}}(\mathbf{x} + uI)| : I \in \mathcal{R}_d, \mathbf{x} \in \mathbb{R}^d\Big\}. \tag{12.54}$$

Recall the definition (12.44) of the constant $C_2$.

**Lemma 12.3.1** *For all* $0 \le u \le 1/C_2$, *we have the inequality*

$$\omega_{n;\mathbb{F}}(u) \le \omega_{n;\mathbb{C}}(C_2 u). \tag{12.55}$$

*Proof* Denote by $\overline{\mathcal{R}}_d$ the set of all closed rectangles of $\mathcal{R}_d$. Since (12.55) is trivial for $u = 0$, we assume that $0 < u \le 1$, and set, for $\mathbf{x} := (x_1, \dots, x_d) \in [0, 1]^d$ and $I := \prod_{j=1}^d \left[y_j, z_j\right] \subseteq [0, 1]^d$, $I \in \overline{\mathcal{R}}_d$, such that $\mathbf{x} + uI \in [0, 1]^d$,

$$\mathbf{x} + uI = \prod_{j=1}^d [r_j(u, \mathbf{x}), s_j(u, \mathbf{x})] \subseteq [0, 1]^d, \tag{12.56}$$

where, for $j = 1, \dots, d$, $r_j(u, \mathbf{x})$ and $s_j(u, \mathbf{x})$ are such that

$$0 \le r_j(u, \mathbf{x}) := x_j + uy_j \le s_j(u, \mathbf{x}) := x_j + uz_j \le 1, \tag{12.57}$$

and

$$0 \le s_j(u,\mathbf{x}) - r_j(u,\mathbf{x}) = u(z_j - y_j) \le u \le 1. \tag{12.58}$$

It is noteworthy that the mappings $\mathcal{F}$ and $\mathcal{Q}$, defined by

$$\mathbf{x} = (x(1),\ldots,x(d)) \in [0,1]^d \tag{12.59}$$
$$\to \mathcal{F}(\mathbf{x}) := (F^{[1]}(x(1)),\ldots,F^{[d]}(x(d)) \in [0,1]^d,$$
$$\mathbf{u} = (u(1),\ldots,u(d)) \in [0,1]^d \tag{12.60}$$
$$\to \mathcal{Q}(\mathbf{u}) := (Q^{[1]}(u(1)),\ldots,Q^{[d]}(u(d)) \in [0,1]^d,$$

are continuous mappings of $[0,1]^d$ onto itself, fulfilling $\mathcal{F} \circ \mathcal{Q} = \mathcal{Q} \circ \mathcal{F} = \mathbb{I}$, where $\mathbb{I}$ denotes identity. Therefore, for each $i \ge 1$, and $I \in \overline{\mathcal{R}}_d$, the event $\{\mathbf{X}_i \in \mathbf{x} + uI\}$ is identical to the event $\{\mathcal{F}(\mathbf{X}_i) = \mathbf{U}_i \in \mathcal{F}(\mathbf{x} + uI)\}$. Now we infer from (12.56), (12.57)–(12.58) and (12.59)–(12.60), that, with $\mathbf{x}$, $u$ and $I$ as above,

$$\mathcal{F}(\mathbf{x} + uI) = \prod_{j=1}^{d} \Big[F^{[j]}(r_j(u,\mathbf{x})), F^{[j]}(s_j(u,\mathbf{x}))\Big] = \mathbf{t} + vJ,$$

where $\mathbf{t} \in [0,1]^d$, $v \in (0,1]$ and $J \in \overline{\mathcal{R}}_d$ are such that

$$\mathbf{t} := \Big(F^{[1]}(r_1(u,\mathbf{x})),\ldots,F^{[d]}(r_d(u,\mathbf{x}))\Big),$$
$$vJ := \prod_{j=1}^{d} \Big[0, F^{[j]}(s_j(u,\mathbf{x})) - F^{[j]}(r_j(u,\mathbf{x}))\Big],$$

with

$$v := C_2 u \quad \text{and} \quad J := \prod_{j=1}^{d} \left[0, \frac{F^{[j]}(s_j(u,\mathbf{x})) - F^{[j]}(r_j(u,\mathbf{x}))}{C_2 u}\right].$$

By (12.47) and (12.57)–(12.58), we see that, for $j = 1,\ldots,d$ and $0 < u \le 1$,

$$0 \le F^{[j]}(s_j(u,\mathbf{x})) - F^{[j]}(r_j(u,\mathbf{x}))$$
$$\le \left\{\sup_{0 \le x \le 1} f^{[j]}(x)\right\} \big(s_j(u,\mathbf{x}) - r_j(u,\mathbf{x})\big) \le C_2 u.$$

Thus, we see that $J \subseteq [0,1]^d$, whereas the inequality $0 < v \le 1$ is implied by the assumption $0 < u \le 1/C_2$. By all this, whenever $\mathbf{x} \in [0,1]^d$, $I \in \overline{\mathcal{R}}_d$ and $0 < u \le 1/C_2$ are such that $\mathbf{x} + uI \subseteq [0,1]^d$, then $\mathcal{F}(\mathbf{x} + uI) \subseteq [0,1]^d$ is of the form $\mathbf{t} + vJ$, for some $\mathbf{t} \in [0,1]^d$, $J \in \overline{\mathcal{R}}_d$, and $0 < v = C_2 u \leq 1$. In view of the definitions (12.53)–(12.54) of $\omega_{n;\mathbb{F}}(\cdot)$ and $\omega_{n;\mathbb{C}}(\cdot)$, and, making use of a similar argument for non-closed rectangles of $\mathcal{R}^d$, we readily obtain (12.55). □

The following fact is a special case of Theorem 1.5 in Stute [34].

**Fact 12.3.1** *For each* $0 < \delta < \frac{1}{2}$, *there exist constants* $0 < c_1(\delta), c_2(\delta) < \infty$ *and* $C_3(\delta) > 0$, *such that, for all*

$$u^d \leq c_2(\delta) \quad \textit{and} \quad 0 < t \leq c_1(\delta)\sqrt{\frac{nu^d}{2\log(1/u^d)}},$$

*we have*

$$\mathbb{P}\left(\frac{\omega_{n;\mathbb{C}}(u)}{\sqrt{2u^d \log_+(1/u^d)}} \geq t\sqrt{\sup_{\mathbf{x}\in[0,1]^d} c(\mathbf{x})}\right) \leq C_3(\delta) u^{d((1-\delta)t^2-1)}. \tag{12.61}$$

**Lemma 12.3.2** *There exist constants* $c_3 > 0$, $c_4 > 0$, $C_4 > 0$ *and* $C_5 > 0$, *such that, whenever* $0 < a_n \leq b_n < \infty$ *fulfill*

$$\frac{na_n^d}{\log(1/a_n^d)} \geq c_3 \quad \textit{and} \quad b_n \leq c_4, \tag{12.62}$$

*we have, with* $\mathcal{H}_n = [a_n, b_n]$, *as* $n \to \infty$,

$$\mathbb{P}\left(\sup_{a\in\mathcal{H}_n} \frac{\omega_{n;\mathbb{C}}(a)}{\sqrt{2a^d \log_+(1/a^d)}} \geq C_4\right) \leq C_5 b_n^{2d}. \tag{12.63}$$

*Proof* First, we observe that, for any $\frac{1}{2} \leq \lambda \leq 1$ and $h^d \leq 1/e$, $\log_+(1/h^d) = \log(1/h^d) \leq \log_+(1/(\lambda h)^d) = \log(1/(\lambda h)^d)$, and, therefore,

$$\frac{\omega_{n;\mathbb{C}}(\lambda h)}{\sqrt{2(\lambda h)^d \log_+(1/(\lambda h)^d)}} \tag{12.64}$$

$$\leq \frac{\omega_{n;\mathbb{C}}(h)}{\sqrt{2h^d \log(1/h^d)}} \times \frac{\sqrt{\log(1/h^d)}}{\sqrt{\lambda^d \log(1/(\lambda h)^d)}} \leq \frac{2^{d/2}\omega_{n;\mathbb{C}}(h)}{\sqrt{2h^d \log(1/h^d)}}.$$

Let now $0 < a \leq 1$ be such that $a^d \leq 1/e$ and select any $N \geq 0$. By a repeated application of (12.64) for $h = 2^{-k}a$ for $k = 0, \ldots, N$, we readily obtain that, for each $N \geq 0$,

$$A_N := \mathbb{P}\left( \sup_{2^{-N-1}a \leq h \leq a} \frac{\omega_{n;\mathbb{C}}(h)}{\sqrt{2h^d \log_+(1/h^d)}} \geq 2^{1+d/2} \sqrt{\sup_{\mathbf{x}\in[0,1]^d} c(\mathbf{x})} \right) \quad (12.65)$$

$$\mathbb{P}\left( \bigcup_{k=0}^{N} \left\{ \sup_{2^{-k-1}a \leq h \leq 2^{-k}a} \frac{2^{-d/2}\omega_{n;\mathbb{C}}(h)}{\sqrt{2h^d \log_+(1/h^d)}} \geq 2\sqrt{\sup_{\mathbf{x}\in[0,1]^d} c(\mathbf{x})} \right\} \right)$$

$$\leq \sum_{k=0}^{N} \mathbb{P}\left( \sup_{\frac{1}{2}\leq\lambda\leq 1} \frac{2^{-d/2}\omega_{n;\mathbb{C}}(\lambda 2^{-k}a)}{\sqrt{2(\lambda 2^{-k}a)^d \log_+(1/(\lambda 2^{-k}a)^d)}} \geq 2\sqrt{\sup_{\mathbf{x}\in[0,1]^d} c(\mathbf{x})} \right)$$

$$\leq \sum_{k=0}^{N} \mathbb{P}\left( \frac{\omega_{n;\mathbb{C}}(2^{-k}a)}{\sqrt{2(2^{-k}a)^d \log_+(1/(2^{-k}a)^d)}} \geq 2\sqrt{\sup_{\mathbf{x}\in[0,1]^d} c(\mathbf{x})} \right).$$

Let now $0 < a \leq 1$ and $N \geq 0$ be such that

$$a^d \leq c_2(\tfrac{1}{4}) \wedge \{1/e\} \quad \text{and} \quad 2 \leq c_1(\tfrac{1}{4})\sqrt{\frac{n(2^{-N}a)^d}{2\log(1/((2^{-N}a)^d))}}.$$

By combining (12.65) with a repeated application of Fact 12.3.1, taken with $\delta = \frac{1}{4}$, $t = 2$ (so that $(1-\delta)t^2 - 1 = 2$) and $u = 2^{-k}a$ for $k = 0, \ldots, N$, we readily obtain that

$$A_N \leq C_3(\tfrac{1}{4}) \sum_{k=0}^{N} \left(2^{-k}a\right)^{2d} \quad (12.66)$$

$$\leq C_3(\tfrac{1}{4})a^{2d} \sum_{k=0}^{\infty} \left(2^{-2d}\right)^k \leq \tfrac{4}{3}C_3(\tfrac{1}{4})a^{2d},$$

where we have used the fact that, independently of $d \geq 1$,

$$\sum_{k=0}^{\infty} \left(2^{-2d}\right)^k = \frac{1}{1-2^{-2d}} \leq \tfrac{4}{3}.$$

We now set $a = b_n$ and choose $N \geq 0$ in such a way that

$$2^{-N-1}a \leq a_n \leq 2^{-N}a, \quad (12.67)$$

so that $a_n \le 2^{-N}a < 2a_n$. Next, we observe that the function $\psi(t) := t/\log(1/t)$ is increasing on $(0, e]$. Thus, if we assume that $(2b_n)^d \le e$, we obtain that $(2a_n)^d \le e$, and $(a_n)^d \le (2^{-N}a)^d \le e$. We get therefore

$$\frac{na_n^d}{2\log(1/a_n^d)} \le \frac{n(2^{-N}a)^d}{2\log(1/((2^{-N}a)^d))}.$$

By setting $\mathcal{H}_n = [a_n, b_n]$, we infer from (12.65)–(12.67) that, whenever $a_n \le b_n$ fulfill

$$b_n^d \le c_2(\tfrac{1}{4}) \wedge \{1/e\} \wedge \left\{2^{-d}e\right\} \quad \text{and} \quad \frac{na_n^d}{2\log(1/a_n^d)} \ge \frac{4}{c_1(\frac{1}{4})^2},$$

we have

$$\mathbb{P}\left(\sup_{a\in\mathcal{H}_n} \frac{\omega_{n;\mathbb{C}}(a)}{\sqrt{2a^d\log_+(1/a^d)}} \ge 2^{1+1/d}\sqrt{\sup_{\mathbf{x}\in[0,1]^d} c(\mathbf{x})}\right) \tag{12.68}$$
$$\le A_N \le \tfrac{4}{3}C_3(\tfrac{1}{4})b_n^{2d}.$$

Recalling (12.52), we set

$$C_4 := 2^{1+1/d}C_2^{1/2}C_1^{-d/2} \ge 2^{1+1/d}\sqrt{\sup_{\mathbf{x}\in[0,1]^d} c(\mathbf{x})}.$$

We therefore infer from (12.68) that (12.63) holds under (12.62), when the constants $c_3$, $c_4$ and $C_5$ are defined by

$$c_3 := 8/c_1(\tfrac{1}{4})^2,$$
$$c_4 := \left(c_2(\tfrac{1}{4}) \wedge \{1/e\} \wedge \left\{2^{-d}e\right\}\right)^{1/d},$$

and $C_5 := \frac{4}{3}C_3(\frac{1}{4})$. □

**Proposition 12.3.1** *There exist constants $c_5 > 0$, $c_6 > 0$, $C_6 > 0$ and $C_7 > 0$, such that, whenever $0 < a_n \le b_n < \infty$ fulfill*

$$\frac{na_n^d}{\log(1/a_n^d)} \ge c_5 \quad \text{and} \quad b_n \le c_6,$$

*we have, with $\mathcal{H}_n = [a_n, b_n]$, as $n \to \infty$,*

$$\mathbb{P}\left(\sup_{a\in\mathcal{H}_n} \frac{\omega_{n;\mathbb{F}}(a)}{\sqrt{2a^d\log_+(1/a^d)}} \ge C_6\right) \le C_7 b_n^{2d}. \tag{12.69}$$

*Proof* We infer from (12.44) and (12.45) that $0 < 1/C_2 \le 1$. Thus, by (12.55), we have, for all $0 \le a \le \{1/C_2\} \wedge 1 = 1/C_2$,

$$\frac{\omega_{n;\mathbb{F}}(a)}{\sqrt{2a^d \log_+(1/a^d)}} \le \frac{\omega_{n;\mathbb{C}}(C_2 a)}{\sqrt{2(C_2 a)^d \log_+(1/(C_2 a)^d)}} \times C_2^{d/2} \left\{ \frac{\log(1/(C_2 a)^d)}{\log(1/a^d)} \right\}^{1/2}$$

$$= \frac{\omega_{n;\mathbb{C}}(C_2 a)}{\sqrt{2(C_2 a)^d \log_+(1/(C_2 a)^d)}} \times C_2^{d/2} \left\{ 1 + \frac{\log(1/C_2)}{\log(1/a)} \right\}^{1/2}$$

$$\le \frac{\omega_{n;\mathbb{C}}(C_2 a)}{\sqrt{2(C_2 a)^d \log_+(1/(C_2 a)^d)}} \times C_2^{d/2}.$$

□

### 12.3.2 Basic Arguments

For convenience, in the proof of Theorem 12.2.1 below, we will set $\mathbf{I} := \mathbf{I}_0 := [0, 1]^d$. The adaptation of our arguments to a general $\mathbf{I}$ is readily achieved, at the price of heavier notation. Letting $\mathbb{F}(\cdot)$ and $\mathbb{F}_n(\cdot)$ be as in Sect. 12.2.1, we denote by $d\mathbb{F}_n(\cdot)$ (resp. $d\mathbb{F}(\cdot)$) the empirical (resp. underlying) measure pertaining to $\{\mathbf{X}_i : 1 \le i \le n\}$, and write $da_n(\cdot) = n^{1/2}(d\mathbb{F}_n(\cdot) - d\mathbb{F}(\cdot))$, where $a_n(\cdot)$ is as in (12.18). For $N \ge 1$, we denote by $\mathbf{B}_N := \{\mathbf{z} \in \mathbb{R}^N : \|\mathbf{z}\| \le 1\}$ the unit ball of the Euclidian norm $\|\mathbf{z}\| := \left(\mathbf{z}'\mathbf{z}\right)^{1/2}$ in $\mathbb{R}^N$. For each $\mathbf{z} \in \mathbb{R}^N$ and $\varepsilon > 0$, we set $\mathcal{N}_\varepsilon(\mathbf{z}) := \{\mathbf{y} \in \mathbb{R}^N : \|\mathbf{y} - \mathbf{z}\| < \varepsilon\}$, and for each $E \subseteq \mathbb{R}^N$, $E^\varepsilon := \bigcup_{\mathbf{z} \in E} \mathcal{N}_\varepsilon(\mathbf{z})$. For any $E, F \subseteq \mathbb{R}^N$, we write

$$\Delta(E, F) := \inf\{\theta > 0 : E \subseteq F^\theta \text{ and } F \subseteq E^\theta\},$$
$$\text{whenever such a } \theta \text{ exists, and}$$
$$\Delta(E, F) := \infty \quad \text{otherwise.}$$

Fix an integer $M \ge 1$, and select an $0 < a_0 < 1$ such that, for all $0 < a \le a_0$ and $\mathbf{x} \in \mathbf{I}_0 = \left[0, 1\right]^d$, we have $\mathbf{x} + a^{1/d}\mathbf{I}_0 \subseteq \mathbf{J}$. Let $\mathbf{i} := (i_1, \ldots, i_d) \in \mathbb{N}^d$ be such that $\mathbf{0} \le \mathbf{i} \le (M-1) \times \mathbf{1}$, where $\mathbf{0} := (0, \ldots, 0) \in \mathbb{R}^d$ and $\mathbf{1} := (1, \ldots, 1) \in \mathbb{R}^d$. Consider the array of $N := M^d$ random variables, defined, for $\mathbf{0} \le \mathbf{i} \le (M-1) \times \mathbf{1}$, by

$$Z_{n;\mathbf{x};\mathbf{i}}(a) := \frac{\sqrt{N}}{\sqrt{2af(\mathbf{x}) \log_+(1/a)}} \int_{\mathbf{x} + (a/M)^{1/d}(\mathbf{i} + \mathbf{I}_0)} da_n(\mathbf{t}). \tag{12.70}$$

For each $\mathbf{x} \in \mathbf{I}_0$ and $0 < a \leq a_0$, denote by $Z_{n;\mathbf{x}}(a) \in \mathbb{R}^N$ the random vector of $\mathbb{R}^N$ obtained by sorting the array $\{Z_{n;\mathbf{x};\mathbf{i}}(a) : \mathbf{0} \leq \mathbf{i} \leq (M-1) \times \mathbf{1}\}$ in lexicographic order. For each $0 < a \leq a_0$ and $0 < \lambda < 1$ set

$$\mathbf{I}(a;\lambda) = \left\{ \mathbf{x} \in \mathbf{I}_0 : \mathbf{x} = \lambda \mathbf{j}\, a^{1/d} \quad \text{for some} \quad \mathbf{j} \in \mathbb{N}^d \right\}.$$

Consider the set defined by

$$\mathcal{E}_{n;a;N}(\lambda) := \left\{ Z_{n;\mathbf{x}}(a) : \mathbf{x} \in \mathbf{I}(a;\lambda) \right\}.$$

We note for further use that, for $0 < a \leq a_0$ and $0 < \lambda < 1$,

$$\# I(a;\lambda) = \#\{\mathbf{j} \in \mathbb{N}^d : \mathbf{0} \leq \mathbf{j} \leq \lfloor (1/(\lambda a^{1/d}) \rfloor \times \mathbf{1}\} \leq 2^d \lambda^{-d} a.$$

Observe that, for each $\mathbf{x} \in \mathbf{I}_0$, there exists a $\widetilde{\mathbf{x}} = \widetilde{\mathbf{x}}(\mathbf{x}) \in \mathbf{I}(a;\lambda)$ such that

$$\widetilde{\mathbf{x}} \leq \mathbf{x} \leq \widetilde{\mathbf{x}} + \lambda a^{1/d} \mathbf{1}.$$

We will show that Theorem 12.1.1 is equivalent to the following statement.

**Theorem 12.3.1** *Set $\mathcal{H}_n = [a_n, b_n]$, where $0 < a_n \leq b_n$ fulfill, as $n \to \infty$,*

$$b_n \to 0 \quad \text{and} \quad na_n / \log n \to \infty. \tag{12.71}$$

*Then, for each $N = M^d \geq 1$ and $0 < \lambda \leq 1$, we have, as $n \to \infty$,*

$$\sup_{a \in \mathcal{H}_n} \Delta \left( \mathcal{E}_{n;a;N}(\lambda), \mathbf{B}_N \right) = o_{\mathbb{P}}(1). \tag{12.72}$$

*Proof of Theorem 12.3.1* To prove Theorem 12.1.1, we use of a discretization argument due to Deheuvels and Ouadah [10]. For each $0 < \rho < 1$ and $\mathcal{H}_n = [a_n, b_n]$, set

$$\mathcal{H}_n(\rho) = \left\{ \rho^m b_n \in [a_n, b_n] : m \in \mathbb{N} \right\}.$$

We note that $\mathcal{H}_n(\rho)$ is never void, as long as $0 < a_n \leq b_n$. Given this notation, the proof of Theorem 12.1.1 reduces to show that, under (12.21), we have, for each $0 < \rho < 1$,

$$\sup_{a \in \mathcal{H}_n(\rho)} \Delta \left( \mathcal{F}_{n;a;\mathbf{I}}, \mathbb{S}_d \right) = o_{\mathbb{P}}(1). \tag{12.73}$$

The details of this argument are given in [10] for $d = 1$. However, it is easy to see that the same methods apply to an arbitrary $d \geq 1$, so that we omit details.

In a second step, we show that Theorem 12.1.1 is equivalent to Theorem 12.3.1. In view of the above preliminaries, this amounts to show that, under (12.71), the property that the assertion (12.73) holds for each $0 < \rho < 1$, is equivalent to the property that, for each $0 < \rho < 1$ and $N = M^d \geq 1$, as $n \to \infty$,

$$\sup_{a \in \mathcal{H}_n(\rho)} \Delta\left(\mathcal{E}_{n;a;N}, \mathbf{B}_N\right) = o_{\mathbb{P}}(1). \tag{12.74}$$

To show the equivalence between (12.73) and (12.74), we follow the discretization method used by Strassen [33] to establish his law of the iterated logarithm. The corresponding details are given in the forthcoming Sect. 12.3.4 for $d = 1$. Their extension to an arbitrary $d \geq 1$ is mostly a matter of book-keeping, with tedious notation for higher dimensions. We will therefore limit ourselves to the essential part of the argument. Consider the modulus of continuity of $a_n(\cdot)$, defined, for $0 < h \leq 1$, by

$$\omega_n(h) := \sup_{R \in \mathcal{R}} \left| \int_{h^{1/d} R} da_n(\mathbf{x}) \right|, \tag{12.75}$$

where $\mathcal{R}$ denotes the set of all *rectangles* in $\mathbf{I} = [0, 1]^d$. Given these preliminaries, the proof of the equivalence between (12.73) and (12.74) boils down to show that, under (12.72), for each $\varepsilon > 0$, there exists an $N = M^d$ such that

$$\mathbb{P}\left( \sup_{a \in \mathcal{H}_n} \frac{\omega_n(a/M)}{\sqrt{2a \log_+(1/a)}} \geq \varepsilon \right) \to 0. \tag{12.76}$$

This, in turn, will follow directly from Proposition 12.3.1 in the sequel. Given the above arguments, the proof of the equivalence between Theorems 12.1.1 and 12.3.1 is now complete.

It remains to show that (12.74) holds for each choice of $0 < \rho < 1$ and $N = M^d \geq 1$. This property turns out to be a consequence of the limiting results (12.77) and (12.78) below, which must hold, for each choice of $\varepsilon > 0$, $0 < \rho < 1$ and $N = M^d$. In the first place, we have, under (12.72),

$$\sum_{k:\rho^k b_n \in \mathcal{H}_n} \mathbb{P}\left(\mathcal{E}_{n;\rho^k b_n;N} \not\subseteq \mathbf{B}_N^{\varepsilon}\right) \to 0. \tag{12.77}$$

In the second place, we have, for each $0 \leq \|\mathbf{z}\| < 1$,

$$\sum_{k:\rho^k b_n \in \mathcal{H}_n} \mathbb{P}\left(\exists \mathbf{y} \in \mathcal{E}_{n;\rho^k b_n;N} : \mathbf{y} \in \mathcal{N}_{\varepsilon}(\mathbf{z})\right) \to 0. \tag{12.78}$$

The only remaining part of our proof is to obtain the appropriate probabilistic bounds allowing us to establish (12.77) and (12.78). Here, we use a simple trick.

Since the probabilities in (12.77) and (12.78) evaluate deviations of centered and rescaled *multinomial* random vectors in $\mathbb{R}^N$, for a *specified* $N \geq 1$, we may construct these multinomial laws in a space of arbitrary dimension $d$. This allows us to make use of the probabilistic inequalities obtained by Deheuvels and Ouadah [10] for $d = 1$. We note that the latter inequalities rely on strong invariance principles whose extension in higher dimensions is not presently available. Fortunately, the use of multinomial distributions allows us to avoid this technical difficulty. The proof of (12.77) and (12.78), follows directly from the forthcoming Propositions 12.3.2 and 12.3.3. In view of these arguments, the proofs of Theorems 12.1.1 and 12.3.1 is now completed. □

In the remainder of our paper, we outline the proofs of the key properties (12.76)–(12.78), on which rely the above-given proofs of Theorems 12.1.1 and 12.3.1.

### *12.3.3 Multinomial Inequalities*

Let $N \geq 1$ be an integer which will be specified later on. Let $\mathbf{p} := (p_1, \ldots, p_N) \in \mathbb{R}_+^N$ fulfill $p_j > 0$ for $j = 1, \ldots, N$ and $p_{N+1} := 1 - |\mathbf{p}| := 1 - \sum_{j=1}^N p_j > 0$. For each $n \geq 1$, we denote the fact that the random vector $\mathbf{Z}_{n;\mathbf{p};N} := (Z_{n;\mathbf{p};1}, \ldots, Z_{n;\mathbf{p};N}) \in \mathbb{R}^N$ follows a multinomial distribution with parameters $n$ and $\mathbf{p}$, by $\mathbf{Z}_{n;\mathbf{p};N} \stackrel{d}{=} \text{Mult}(n; \mathbf{p})$. This holds whenever, for any $N$-uple of nonnegative integers $\mathbf{k} := (k_1, \ldots, k_N)$, such that $k_{N+1} := n - |\mathbf{k}| := n - \sum_{j=1}^N k_j \geq 0$, we have

$$\mathbb{P}\left(\mathbf{Z}_{n;\mathbf{p};N} = \mathbf{k}\right) = \frac{n!}{k_1! \ldots k_{N+1}!}\, p_1^{k_1} \ldots p_{N+1}^{k_{N+1}}\,.$$

For each $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_N) \in \mathbb{R}_+^N$, set $|\boldsymbol{\delta}| := \sum_{j=1}^N \delta_j$, and consider

$$\mathcal{D}_N = \left\{\boldsymbol{\delta} := (\delta_1, \ldots, \delta_N) \in \mathbb{R}^N : \delta_j > 0,\, j = 1, \ldots, N;\, |\boldsymbol{\delta}| = N\right\}. \tag{12.79}$$

Whenever $\boldsymbol{\delta} \in \mathcal{D}_N$, set

$$0 < \boldsymbol{\delta}_{\min} := \min_{1\leq j\leq N} \delta_j \leq 1 \leq \boldsymbol{\delta}_{\max} := \max_{1\leq j\leq N} \delta_j. \tag{12.80}$$

We will set $\mathbf{p} = a\boldsymbol{\delta}/N$ for some $0 < a \leq 1$, so that $|\mathbf{p}| = aN^{-1}|\boldsymbol{\delta}| = a \leq 1$, and consider the random vector

$$\boldsymbol{\zeta}_{n;a;\boldsymbol{\delta}} := \frac{\sqrt{N}}{\sqrt{2na\log_+(1/a)}} \begin{bmatrix} Z_{n;a\boldsymbol{\delta}/N;1} - na\delta_1/N \\ \vdots \\ Z_{n;a\boldsymbol{\delta}/N;N} - na\delta_N/N \end{bmatrix} \in \mathbb{R}^N. \tag{12.81}$$

Denote by $\mathbf{B}_N := \{\mathbf{z} \in \mathbb{R}^N : \|\mathbf{z}\| \leq 1\}$, the unit ball of the Euclidian norm $\|\mathbf{z}\| := (\mathbf{z}'\mathbf{z})^{1/2}$ in $\mathbb{R}^N$. Let, for each $\mathbf{z} \in \mathbb{R}^N$ and $\varepsilon > 0$, $\mathcal{N}_\varepsilon(\mathbf{z}) := \{\mathbf{y} \in \mathbb{R}^N : \|\mathbf{y} - \mathbf{z}\| < \varepsilon\}$, and set, for each $A \subseteq \mathbb{R}^N$, $A^\varepsilon := \bigcup_{\mathbf{z} \in A} \mathcal{N}_\varepsilon(\mathbf{z})$. We will need the following two propositions.

**Proposition 12.3.2** *There exists a constant $C_0$ such that the following holds. For each $0 < \varepsilon \leq 1$, there exist constants $0 < a_0(\varepsilon) \leq 1/e$ and $0 < c_0(\varepsilon) < \infty$, together with an $n_0(\varepsilon) < \infty$, such that, for all $n \geq n_0(\varepsilon)$ and $a > 0$ fulfilling*

$$na/\log n \geq c_0(\varepsilon) \quad \textit{and} \quad a \leq a_0(\varepsilon), \tag{12.82}$$

*and for all $N \geq 1$ and $\boldsymbol{\delta} \in \mathcal{D}_N$ fulfilling*

$$\sqrt{\boldsymbol{\delta}_{min}} \geq \frac{1 + \frac{1}{2}\varepsilon}{1 + \varepsilon}, \tag{12.83}$$

*we have*

$$\mathbb{P}\left(\boldsymbol{\zeta}_{n;a;\boldsymbol{\delta}} \notin \mathbf{B}_N^\varepsilon\right) \leq C_0 a^{1+\varepsilon/(8N)}. \tag{12.84}$$

The proof of Proposition 12.3.2 is captured in Sects. 12.3.4 and 12.3.5 below.

For the next proposition, we will need the following additional notation. We consider a sequence $\boldsymbol{\delta}(k) = (\delta_1(k), \ldots, \delta_N(k)) \in \mathcal{D}_N$, $k = 1, \ldots, K$, and set $\mathbf{p}(k) = (p_1(k), \ldots, p_N(k)) := a\boldsymbol{\delta}(k)/N$, for $k = 1, \ldots, K$ and $0 < a \leq 1/K$, so that $\sum_{k=1}^K |\mathbf{p}(k)| = aN^{-1} \sum_{k=1}^K |\boldsymbol{\delta}_k| = Ka \leq 1$. Given $\{\boldsymbol{\delta}(k) : k = 1, \ldots, K\}$, we consider a sequence of random vectors

$$\mathbf{Z}_{n;\mathbf{p}(k);N}^{(k)} := (Z_{n;p_1(k);1}^{(k)}, \ldots, Z_{n;p_N(k);N}^{(k)}) \in \mathbb{R}^N, \quad k = 1, \ldots, K,$$

such that, with obvious notation,

$$(\mathbf{Z}_{n;\mathbf{p}(1);N}^{(1)}, \ldots, \mathbf{Z}_{n;\mathbf{p}(K);N}^{(K)}) \stackrel{d}{=} \text{Mult}(n; \mathbf{p}(1), \ldots, \mathbf{p}(K)).$$

In view of (12.81), we consider the random vectors, for $k = 1, \ldots, K$,

$$\boldsymbol{\zeta}_{n;a;\boldsymbol{\delta}(k)}^{(k)} := \frac{\sqrt{N}}{\sqrt{2na \log_+(1/a)}} \begin{bmatrix} Z_{n;p_1(k);1}^{(k)} - np_1(k) \\ \vdots \\ Z_{n;p_N(k);N}^{(k)} - np_N(k) \end{bmatrix} \in \mathbb{R}^N. \tag{12.85}$$

**Proposition 12.3.3** *Fix any $\mathbf{z} \in \mathbf{B}_N$ such that $0 < \|\mathbf{z}\| < 1$. For each $\varepsilon$ such that*

$$0 < \varepsilon < \left\{\frac{1}{2}\|\mathbf{z}\|\right\} \wedge \frac{1}{2N},$$

*there exist an* $a_2(\varepsilon, \mathbf{z})$*, together with* $n_2(\varepsilon) < \infty$ *and* $c_2(\varepsilon)$ *depending upon* $\varepsilon$ *only, such that the following holds. For each* $\boldsymbol{\delta}(1), \ldots, \boldsymbol{\delta}(K) \in \mathcal{D}_N$*, and* $a_1, \ldots, a_k$*, whenever*

$$n \geq n_2(\varepsilon), \quad c_2(\varepsilon) n^{-1} \log n \leq a_1, \ldots, a_k \leq a_2(\varepsilon, g), \quad \sum_{k=1}^{K} a_k \leq \frac{1}{2}, \tag{12.86}$$

*we have, for all* $\boldsymbol{\delta}_1, \ldots, \boldsymbol{\delta}_K$*, fulfilling*

$$\frac{1}{\sqrt{\boldsymbol{\delta}_{max}}} \geq 1 - N\varepsilon \quad \textit{and} \quad \frac{1}{\sqrt{\boldsymbol{\delta}_{min}}} \leq 1 + N\varepsilon, \tag{12.87}$$

$$\mathbb{P}\left( \bigcap_{k=1}^{K} \left\{ \boldsymbol{\zeta}^{(k)}_{n;a_k;\boldsymbol{\delta}(k)} \notin \mathcal{N}_{9N\varepsilon}(\mathbf{z}) \right\} \right) \leq 2 \exp\left( -\tfrac{1}{4} \sum_{k=1}^{K} a_k^{1-\varepsilon/2} \right). \tag{12.88}$$

The proof of Proposition 12.3.3 is postponed until Sect. 12.3.6.

### 12.3.4 Outer Bounds

Let $U_1, U_2, \ldots$ be iid rv's with a uniform (0, 1) distribution. For $n \geq 1$ and $t \in \mathbb{R}$, denote by $\mathbb{U}_n(t) := n^{-1}\#\{U_i \leq t : 1 \leq i \leq n\}$ the empirical df based upon $U_1, \ldots, U_n$, and by $\alpha_n(t) := n^{1/2}(\alpha_n(t) - t)$, the uniform empirical process. For $n \geq 1$, $a > 0$, $t \in [0, 1]$ and $u \in \mathbb{R}$, set

$$\xi_n(a; t; u) = \alpha_n(t + au) - \alpha_n(t). \tag{12.89}$$

The following fact is Proposition 2 of Deheuvels and Ouadah [10].

**Fact 12.3.2** *There exists a constant* $C_2$ *such that the following holds. For each* $0 < \varepsilon \leq 1$*, there exist constants* $0 < a_1(\varepsilon) \leq 1/e$ *and* $0 < c_1(\varepsilon) < \infty$*, together with an* $n_1(\varepsilon) < \infty$*, such that, for all* $n \geq n_1(\varepsilon)$ *and* $a > 0$ *fulfilling*

$$na/\log n \geq c_1(\varepsilon) \quad \textit{and} \quad a \leq a_1(\varepsilon), \tag{12.90}$$

*we have, for all* $t \in [0, 1 - a]$*,*

$$\mathbb{P}\left( \frac{\xi_n(a; t; \cdot)}{\sqrt{2a \log_+(1/a)}} \notin \mathbb{S}^{\varepsilon} \right) \leq C_2 a^{1+\varepsilon}. \tag{12.91}$$

The following lemmas are oriented towards the proof of Proposition 12.3.2.

**Lemma 12.3.3** *For any* $g \in B([0,1])$ *and* $0 \le s, t \le 1$, *we have*

$$|g(t) - g(s)| \le |g|_{\mathbb{H}} \sqrt{|t-s|}\,, \tag{12.92}$$

*and, for any* $0 \le t \le t+h \le 1$, *we have*

$$\sup_{0 \le u \le 1} |g(t+hu) - g(t) - u(g(t+h) - g(t))| \le |g|_{\mathbb{H}} \sqrt{\tfrac{1}{2}h}\,, \tag{12.93}$$

*Proof* When $g \notin AC_0([0,1])$, $|g|_{\mathbb{H}} = \infty$ and (12.92)–(12.93) are trivial. Therefore, we limit ourselves to $g \in AC_0[0,1]$. The Schwarz inequality enables us to write the relations

$$|g(t) - g(s)| = \left| \int_s^t \dot{g}(u)du \right| \le \left| \int_s^t du \right|^{1/2} \left| \int_s^t \dot{g}(u)^2 du \right|^{1/2} \le |g|_{\mathbb{H}} \sqrt{|t-s|}\,,$$

which yield (12.92).

For $g \in AC_0([0,1])$, the function $\phi(u) := g(t+hu) - g(t) - u(g(t+h) - g(t))$, for $0 \le u \le 1$, is such that

$$\phi(0) = \phi(1) = \int_0^1 \dot{\phi}(u)du = 0.$$

Moreover, setting $\psi(u) := h\dot{g}(t+hu)$, for $0 \le u \le 1$, we get

$$\dot{\phi}(u) = h\dot{g}(t+hu) - (g(t+h) - g(t)) = \psi(u) - \int_0^1 \psi(t)dt.$$

Observe that

$$\int_0^1 \dot{\phi}(u)^2 du = \int_0^1 \psi(u)^2 du - \left\{ \int_0^1 \psi(t)dt \right\}^2$$
$$\le \int_0^1 \psi(u)^2 du = h \int_t^{t+h} \dot{g}(s)^2 ds \le h|g|^2_{\mathbb{H}}.$$

An easy argument shows that the supremum of $|\varphi(c)| = |\int_0^c \dot{\varphi}(u)du|$ subject to the constraints $0 \le c \le 1$, $\varphi(0) = 0$, $\int_0^1 \dot{\varphi}(u)du = 0$ and $\int_0^1 \dot{\varphi}(u)^2 du \le \lambda$, is equal to $\frac{1}{2}\sqrt{\lambda}$, and reached when $c = \frac{1}{2}$ and $\dot{\varphi}(u) = \sqrt{\lambda}$, $0 < u < \frac{1}{2}$, $\dot{\varphi}(u) = -\sqrt{\lambda}$, $\frac{1}{2} < u < 1$. Since $\varphi = \phi$ fulfills these conditions with $\lambda := h|g|^2_{\mathbb{H}}$, it follows that the maximal possible value of $\phi$ on $[0,1]$ is less than or equal to $|g|_{\mathbb{H}}\sqrt{\frac{1}{2}h}$. We so obtain (12.93). □

Fix $N \geq 1$, and let $\mathcal{D}_N$ be as in (12.79). For any $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_N) \in \mathcal{D}_N$, set $t_j(\boldsymbol{\delta}) = N^{-1} \sum_{k=1}^{j} \delta_j$, for $j = 0, \ldots, N$, with the convention that $\sum_{\emptyset}(\cdot) := 0$. As in (12.80), set $\boldsymbol{\delta}_{\min} = \min_{1 \leq j \leq N} \delta_j$, and $\boldsymbol{\delta}_{\max} = \max_{1 \leq j \leq N} \delta_j$. Consider the linear maps $\mathcal{P}_{N;\boldsymbol{\delta}}(\cdot)$ and $\mathcal{Q}_{N;\boldsymbol{\delta}}(\cdot)$, defined by

$$g \in B[0,1] \tag{12.94}$$

$$\to \mathcal{P}_{N;\boldsymbol{\delta}}(g) := \begin{bmatrix} \sqrt{\frac{N}{\delta_1}}\,(g(t_1(\boldsymbol{\delta})) - g(t_0(\boldsymbol{\delta}))) \\ \vdots \\ \sqrt{\frac{N}{\delta_N}}\,(g(t_N(\boldsymbol{\delta})) - g(t_{N-1}(\boldsymbol{\delta}))) \end{bmatrix} \in \mathbb{R}^N,$$

$$\mathbf{z} = \begin{bmatrix} z_1 \\ \vdots \\ z_N \end{bmatrix} \in \mathbb{R}^N \to \mathcal{Q}_N(\mathbf{z}) \in AC[0,1], \tag{12.95}$$

where we define $\mathcal{Q}_{N,\boldsymbol{\delta}}(\mathbf{z})$ for $\mathbf{z} = (z_1, \ldots, z_N) \in \mathbb{R}^N$, by setting $z_0 = 0$, $\sum_{\emptyset}(\cdot) = 0$, and, for $k = 1, \ldots, N$,

$$\mathcal{Q}_{N,\boldsymbol{\delta}}(\mathbf{z})(t) = \sum_{j=1}^{k-1} \sqrt{\frac{\delta_j}{N}}\, z_j + \sqrt{\frac{N}{\delta_k}}\, z_k\,(t - t_{k-1}(\boldsymbol{\delta})) \tag{12.96}$$

$$\text{when } t_{k-1}(\boldsymbol{\delta}) \leq t \leq t_k(\boldsymbol{\delta})\,.$$

**Lemma 12.3.4** *For $N \geq 1$, $\boldsymbol{\delta} \in \mathcal{D}_N$, $\mathbf{z} \in \mathbb{R}^N$ and $g \in B([0,1])$, we have*

$$\mathcal{P}_{N,\boldsymbol{\delta}}(\mathcal{Q}_{N,\boldsymbol{\delta}}(\mathbf{z})) = \mathbf{z}; \tag{12.97}$$

$$\left\| \mathcal{Q}_{N,\boldsymbol{\delta}}(\mathcal{P}_{N,\boldsymbol{\delta}}(g)) - g \right\| \leq (2N)^{-1/2} |g|_{\mathbb{H}} \sqrt{\boldsymbol{\delta}_{max}}\,; \tag{12.98}$$

$$\|\mathcal{P}_{N,\boldsymbol{\delta}}(g)\| \leq |g|_{\mathbb{H}} \quad \textit{and} \quad |\mathcal{Q}_{N,\boldsymbol{\delta}}(\mathbf{z})|_{\mathbb{H}} = \|\mathbf{z}\|; \tag{12.99}$$

$$\|\mathcal{P}_{N,\boldsymbol{\delta}}(g)\| \leq 2N\|g\|/\sqrt{\boldsymbol{\delta}_{\,min}}\,; \tag{12.100}$$

$$\mathcal{P}_{N,\boldsymbol{\delta}}(\mathbb{S}) = \mathbf{B}_N := \{\mathbf{t} \in \mathbb{R}^N : \mathbf{t}'\mathbf{t} \leq 1\}; \tag{12.101}$$

$$\mathcal{Q}_{N,\boldsymbol{\delta}}(\mathbf{B}_N) \subseteq \mathbb{S} \subseteq \mathcal{Q}_{N,\boldsymbol{\delta}}(\mathbf{B}_N)^{\sqrt{\boldsymbol{\delta}_{max}/(2N)}}\,. \tag{12.102}$$

*Proof* By (12.96), $\mathcal{Q}_{N,\boldsymbol{\delta}}(\mathbf{z})(t_j(\boldsymbol{\delta})) - \mathcal{Q}_{N,\boldsymbol{\delta}}(\mathbf{z})(t_{j-1}(\boldsymbol{\delta})) = z_j\sqrt{\delta_j/N}$ for $j = 1, \ldots, N$. Thus, by (12.94), we have $\mathcal{P}_{N,\boldsymbol{\delta}}(\mathcal{Q}_{N,\boldsymbol{\delta}}(\mathbf{z})) = \mathbf{z}$, which is (12.97). Since $|g|_{\mathbb{H}} = \infty$ when $g \notin AC_0([0,1])$, there is no loss of generality to assume in our proofs of (12.98)–(12.99) that $g \in AC_0([0,1])$. To establish (12.98) we observe that, for $j = 0, \ldots, N$, $\mathcal{Q}_{N,\boldsymbol{\delta}}(\mathcal{P}_N(g))(t_j(\boldsymbol{\delta})) = g(t_j(\boldsymbol{\delta}))$, so that, by applying (12.93), for $j = 1, \ldots, N$, with $h = \delta_j/N$, we get

$$\|\mathcal{Q}_{N,\boldsymbol{\delta}}(\mathcal{P}_{N,\boldsymbol{\delta}}(g)) - g\| \leq \max_{1\leq j\leq N}\left(\sup_{0\leq u\leq 1}\left|g\left(t_{j-1}(\boldsymbol{\delta}) + u\frac{\delta_j}{N}\right) - g(t_{j-1}(\boldsymbol{\delta}))\right.\right.$$
$$\left.\left. -u\left\{g\left(t_{j-1}(\boldsymbol{\delta}) + \frac{\delta_j}{N}\right) - g\left(\frac{\delta_j}{N}\right)\right\}\right|\right) \leq |g|_{\mathbb{H}} \max_{1\leq j\leq N}\sqrt{\frac{\delta_j}{2N}}\,,$$

which yields (12.98). To establish the first half of (12.99), we select a $g \in AC_0[0,1]$ and set $\mathbf{z} = (z_1, \ldots, z_d) = \mathcal{P}_{N,\boldsymbol{\delta}}(g)$. It follows from (12.94) that $z_j = \sqrt{\frac{N}{\delta_j}}\,(g(t_j(\boldsymbol{\delta})) - g(t_{j-1}(\boldsymbol{\delta})))$, for $j = 1, \ldots, d$. Making use of the Schwarz inequality, we get, in turn,

$$\|\mathcal{P}_{N,\boldsymbol{\delta}}(g)\|^2 = \mathbf{z}'\mathbf{z} = \sum_{j=1}^N z_j^2 = N\sum_{j=1}^N \frac{1}{\delta_j}\left(\int_{t_{j-1}(\boldsymbol{\delta})}^{t_j(\boldsymbol{\delta})} \dot{g}(u)du\right)^2$$
$$\leq \sum_{j=1}^N \frac{N}{\delta_j}\left(\int_{t_{j-1}(\boldsymbol{\delta})}^{t_j(\boldsymbol{\delta})} du\right)\left(\int_{t_{j-1}(\boldsymbol{\delta})}^{t_j(\boldsymbol{\delta})} \dot{g}(u)^2du\right) = \int_0^1 \dot{g}(u)^2du = |g|_{\mathbb{H}}^2,$$

as sought. Next, we choose a $\mathbf{z} \in \mathbb{R}^N$, and set $g = \mathcal{Q}_{N,\boldsymbol{\delta}}(\mathbf{z})$. We infer from (12.96) that, for $j = 1, \ldots, N$,

$$\dot{g}(t) = z_j\sqrt{\frac{N}{\delta_j}} \quad \text{for} \quad t_{j-1}(\boldsymbol{\delta}) \leq t \leq t_j(\boldsymbol{\delta})\,,$$

whence

$$|\mathcal{Q}_{N,\boldsymbol{\delta}}(\mathbf{z})|_{\mathbb{H}}^2 = \sum_{j=1}^N \int_{t_{j-1}(\boldsymbol{\delta})}^{t_j(\boldsymbol{\delta})} \frac{Nz_j^2}{\delta_j}\,du = \sum_{j=1}^N z_j^2 = \|\mathbf{z}\|^2,$$

which yields the second half of (12.99). To establish (12.100), we infer from (12.94) that, for an arbitrary $g \in B([0,1])$,

$$\|\mathcal{P}_{N,\boldsymbol{\delta}}(g)\|^2 = \sum_{j=1}^d \frac{N}{\delta_j}(g(t_j(\boldsymbol{\delta})) - g(t_{j-1}(\boldsymbol{\delta})))^2$$
$$\leq 4N\|g\|^2\sum_{j=1}^n \frac{1}{\delta_j} \leq \frac{(2N\|g\|)^2}{\boldsymbol{\delta}_{\min}}\,.$$

To establish (12.101), we first infer from (12.99) that $\mathcal{P}_{N,\boldsymbol{\delta}}(g) \in \mathbf{B}_N$ for each $g \in \mathbb{S}$, so that $\mathcal{P}_{N,\boldsymbol{\delta}}(\mathbb{S}) \subseteq \mathbf{B}_N$. Conversely, by (12.99), for any $\mathbf{z} \in \mathbf{B}_N$, we have $g := \mathcal{Q}_{N,\boldsymbol{\delta}}(\mathbf{z}) \in \mathbb{S}$. This, in turn, implies, via (12.97), that $\mathcal{P}_{N,\boldsymbol{\delta}}(g) = \mathbf{z}$, whence $\mathbf{B}_N \subseteq \mathcal{P}_{N,\boldsymbol{\delta}}(\mathbb{S})$. We so obtain (12.101). Next, we infer from (12.99) that, for each $\mathbf{z} \in \mathbf{B}_N$, $\mathcal{Q}_{N,\boldsymbol{\delta}}(\mathbf{z}) \in \mathbb{S}$. This, in turn, implies that $\mathcal{Q}_{N,\boldsymbol{\delta}}(\mathbf{B}_N) \subseteq \mathbb{S}$. Finally, we infer from (12.98) and (12.99) that, for each $g \in \mathbb{S}$, we have $\mathbf{y} := \mathcal{P}_{N,\boldsymbol{\delta}}(g) \in \mathbf{B}_N$ and $\|\mathcal{Q}_{N,\boldsymbol{\delta}}(\mathbf{y}) - g\| \leq (2N)^{-1/2}|g|_{\mathbb{H}}\sqrt{\boldsymbol{\delta}_{\max}} \leq (2N)^{-1/2}\sqrt{\boldsymbol{\delta}_{\max}}$. This completes the proof of (12.102). □

Armed with Fact 12.3.1 and Lemmas 12.3.3–12.3.4, we recall (12.79), (12.89), (12.91), and fix an $N \geq 1$. For $n \geq 1$, $0 < a < 1$, $t \in [0, 1-a]$ and $\boldsymbol{\delta} \in \mathcal{D}_N$, we set

$$\mathbf{z}_{n,\boldsymbol{\delta}}(a;t) = \mathcal{P}_{N,\boldsymbol{\delta}}\left(\frac{\xi_n(a;t;\cdot)}{\sqrt{2a\log_+(1/a)}}\right) \in \mathbb{R}^N. \tag{12.103}$$

By combining (12.89) with (12.94) and (12.103), we observe that

$$\mathbf{z}_{n,\boldsymbol{\delta}}(a;t) = \frac{\sqrt{N}}{\sqrt{2na\log_+(1/a)}} \times \begin{bmatrix} \{\alpha_n(t+at_1(\boldsymbol{\delta})) - \alpha_n(t+at_0(\boldsymbol{\delta}))\}/\sqrt{\delta_1} \\ \vdots \\ \{\alpha_n(t+at_N(\boldsymbol{\delta})) - \alpha_n(t+at_{N-1}(\boldsymbol{\delta}))\}/\sqrt{\delta_N} \end{bmatrix}. \tag{12.104}$$

Set, for convenience,

$$\mathbf{z}^*_{n,\boldsymbol{\delta}}(a;t) = \frac{\sqrt{N}}{\sqrt{2a\log_+(1/a)}} \times \begin{bmatrix} \alpha_n(t+at_1(\boldsymbol{\delta})) - \alpha_n(t+at_0(\boldsymbol{\delta})) \\ \vdots \\ \alpha_n(t+at_N(\boldsymbol{\delta})) - \alpha_n(t+at_{N-1}(\boldsymbol{\delta})) \end{bmatrix}. \tag{12.105}$$

Recall the definition (12.81) of $\zeta_{n;a;\boldsymbol{\delta}}$. In view of (12.105), we may write, for each $0 < a < 1$ and $t \in [0, 1-a]$, the distributional equality

$$\zeta_{n;a;\boldsymbol{\delta}} \overset{d}{=} \mathbf{z}^*_{n,\boldsymbol{\delta}}(a;t). \tag{12.106}$$

We infer from (12.104) and (12.105) the inequality

$$\|\mathbf{z}^*_{n,\boldsymbol{\delta}}(a;t)\| \leq \|\mathbf{z}_{n,\boldsymbol{\delta}}(a;t)\|/\sqrt{\boldsymbol{\delta}_{\min}}\,. \tag{12.107}$$

Below, we let $C_2$, $n_1(\cdot)$, $c_1(\cdot)$ and $a_1(\cdot)$ be as in Fact 12.3.2.

**Lemma 12.3.5** *For each $0 < \varepsilon \leq 1$, and for all $n \geq n_1(\varepsilon)$ and $a > 0$ fulfilling*

$$na/\log n \geq c_1(\varepsilon) \quad \textit{and} \quad a \leq a_1(\varepsilon), \tag{12.108}$$

*we have, for all $t \in [0, 1-a]$,*

$$\mathbb{P}\left(\mathbf{z}_{n,\delta}(a;t) \notin \mathbf{B}_N^{\varepsilon}\right) \leq C_2 a^{1+\left(\varepsilon\sqrt{\delta_{min}}\right)/(2N)}. \tag{12.109}$$

*Proof* By (12.100), for any $\phi \in B([0,1])$, $g \in \mathbb{S}$ and $\epsilon > 0$, we have the implication

$$\|\phi - g\| \leq \epsilon \;\Rightarrow\; \|\mathcal{P}_{N,\delta}(\phi) - \mathcal{P}_{N,\delta}(g)\| = \|\mathcal{P}_{N,\delta}(\phi - g)\| \leq 2N\epsilon/\sqrt{\delta_{\min}},$$

which is equivalent to the implication

$$\|\mathcal{P}_{N,\delta}(\phi) - \mathcal{P}_N(g)\| > 2N\epsilon/\sqrt{\delta_{\min}} \;\Rightarrow\; \|\phi - g\| > \epsilon. \tag{12.110}$$

We recall from (12.101) that $\mathcal{P}_{N,\delta}(\mathbb{S}) = \mathbf{B}_N$. Thus, by setting $\mathbf{z} = \mathcal{P}_{N,\delta}(g)$ in (12.110), and letting $g$ vary in $\mathbb{S}$ we obtain the implication

$$\left\{\|\mathcal{P}_{N,\delta}(\phi) - \mathbf{z}\| > 2N\epsilon/\sqrt{\delta_{\min}} \;:\forall\, \mathbf{z} \in \mathbf{B}_N\right\} \;\Rightarrow\; \left\{\|\phi - g\| > \epsilon \;:\forall\, g \in \mathbb{S}\right\},$$

which may be rewritten into

$$\left\{\mathcal{P}_{N,\delta}(\phi) \notin \mathbf{B}_N^{2N\epsilon/\sqrt{\delta_{\min}}}\right\} \;\Rightarrow\; \left\{\phi \notin \mathbb{S}^{\epsilon}\right\}. \tag{12.111}$$

Recalling the definition (12.103) of $\mathbf{z}_{n,\delta}(a;t)$, by setting $\varepsilon = 2N\epsilon/\sqrt{\delta_{\min}}$ and $\phi = \xi_n(a;t;\cdot)/\sqrt{2a\log_+(1/a)}$ in (12.111), we conclude our proof by an application of Fact 12.3.2. □

## *12.3.5 Proof of Proposition 12.3.2*

Fix an $0 < \varepsilon \leq 1$. In view of (12.106) and (12.33), whenever

$$\sqrt{\delta_{\min}} \geq \frac{1 + \frac{1}{2}\varepsilon}{1+\varepsilon}, \tag{12.112}$$

we have, for $0 < a < 1$ and $0 \leq t \leq 1 - a$,

$$\begin{aligned}
&\mathbb{P}\left(\boldsymbol{\zeta}_{n;a;\delta} \notin \mathbf{B}_N^{\varepsilon}\right) = \mathbb{P}\left(\|\boldsymbol{\zeta}_{n;a;\delta}\| > 1+\varepsilon\right) \qquad (12.113)\\
&= \mathbb{P}\left(\|\mathbf{z}^*_{n;\delta}(a;t)\| > 1+\varepsilon\right) \leq \mathbb{P}\left(\|\mathbf{z}_{n;\delta}(a;t)\| > (1+\varepsilon)\sqrt{\delta_{\min}}\right)\\
&\leq \mathbb{P}\left(\|\mathbf{z}_{n;\delta}(a;t)\| > 1 + \tfrac{1}{2}\varepsilon\right) = \mathbb{P}\left(\mathbf{z}_{n;\delta}(a;t) \notin \mathbf{B}_N^{\varepsilon/2}\right).
\end{aligned}$$

The assumption that $0 < \varepsilon \leq 1$, when combined with (12.112) implies that

$$\sqrt{\boldsymbol{\delta}_{\min}} \geq \frac{3}{4} > \frac{1}{2}.$$

By an application of Lemma 12.3.5 with the formal replacement of $\varepsilon$ by $\varepsilon/2$, we see that, for all $n \geq n_0(\varepsilon) := n_1(\varepsilon/2)$ and $a > 0$ fulfilling

$$na/\log n \geq c_0(\varepsilon) := c_1(\varepsilon/2) \quad \text{and} \quad a \leq a_0(\varepsilon) := a_1(\varepsilon/2), \tag{12.114}$$

we have, for all $t \in [0, 1-a]$,

$$\mathbb{P}\left(\mathbf{z}_{n,\boldsymbol{\delta}}(a;t) \notin \mathbf{B}_N^{\varepsilon/2}\right) \leq C_0 a^{1+(\varepsilon\sqrt{\boldsymbol{\delta}_{\min}})/(4N)} \leq C_0 a^{1+\varepsilon/(8N)}. \tag{12.115}$$

By (12.113), this yields (12.84), with $C_0 := C_2$, and completes the proof of Proposition 12.3.2. □

### *12.3.6 Inner Bounds*

The following fact is a version of Proposition 3 of Deheuvels and Ouadah [10], taken with $|\mathcal{I}| = \sum_{k=1}^{K} a_k$.

**Fact 12.3.3** *For each $g \in \mathbb{S}$ such that $0 < |g|_{\mathbb{H}} < 1$, and $0 < \varepsilon < \frac{1}{2}|g|_{\mathbb{H}}$, there exist an $a_2(\varepsilon, g)$, together with $n_2(\varepsilon) < \infty$ and $c_2(\varepsilon)$, depending upon $\varepsilon$ only, such that the following holds. Let, for $K \geq 1$, $t_1, \ldots, t_K \in [0,1]$, and $0 < a_1, \ldots, a_k < 1$, be such that the intervals $(t_k, t_k + a)$, $k = 1, \ldots, K$, are disjoint and in $[0,1]$, with $\sum_{k=1}^{K} a_k \leq \frac{1}{2}$. Then, whenever*

$$n \geq n_2(\varepsilon), \quad c_2(\varepsilon) n^{-1} \log n \leq a_1 \ldots, a_K \leq a_2(\varepsilon, g), \tag{12.116}$$

*we have*

$$\mathbb{P}\left(\bigcap_{k=1}^{K}\left\{\frac{\xi_n(a_k; t_k; \cdot)}{\sqrt{2a_k \log_+(1/a_k)}} \notin \mathcal{N}_\varepsilon(g)\right\}\right) \leq 2\exp\left(-\tfrac{1}{4}\sum_{k=1}^{K} a_k^{1-\varepsilon/2}\right). \tag{12.117}$$

Fix any $\mathbf{z} \in \mathbf{B}_N$, such that $0 < \|\mathbf{z}\| < 1$, and set $g := \mathcal{Q}_{N;\boldsymbol{\delta}}(\mathbf{z})$. Fix $a > 0$ and $t \in [0, 1-a]$, and set, as in (12.103),

$$\phi := \frac{\xi_n(a;t;\cdot)}{\sqrt{2a \log_+(1/a)}} \quad \text{and} \quad \mathbf{z}_{n,\boldsymbol{\delta}}(a;t) = \mathcal{P}_{N,\boldsymbol{\delta}}(\phi) \in \mathbb{R}^N. \tag{12.118}$$

As follows from (12.99) and (12.99), we have $\mathcal{P}_{N;\boldsymbol{\delta}}(g) = \mathbf{z}$ and

$$0 < |g|_{\mathbb{H}} = \|\mathbf{z}\| < 1.$$

Therefore, we infer from the linearity of $\mathcal{P}_{N;\boldsymbol{\delta}}$ and (12.100) that

$$\begin{aligned}\|\mathbf{z}_{n,\boldsymbol{\delta}}(a;t) - \mathbf{z}\| &= \left\|\mathcal{P}_{N,\boldsymbol{\delta}}(\phi) - \mathcal{P}_{N,\boldsymbol{\delta}}(g)\right\| = \left\|\mathcal{P}_{N,\boldsymbol{\delta}}(\phi - g)\right\| \\ &\leq \frac{2N}{\sqrt{\boldsymbol{\delta}_{\min}}}\|\phi - g\| = \frac{2N}{\sqrt{\boldsymbol{\delta}_{\min}}}\left\|\frac{\xi_n(a;t;\cdot)}{\sqrt{2a\log_+(1/a)}} - g\right\|.\end{aligned}$$

We have therefore the implication, for an arbitrary $\varepsilon > 0$,

$$\left\|\frac{\xi_n(a;t;\cdot)}{\sqrt{2a\log_+(1/a)}} - g\right\| \leq \varepsilon \;\Rightarrow\; \|\mathbf{z}_{n,\boldsymbol{\delta}}(a;t) - \mathbf{z}\| \leq \frac{2N\varepsilon}{\sqrt{\boldsymbol{\delta}_{\min}}}$$

which is readily shown to be equivalent to

$$\left\{\|\mathbf{z}_{n,\boldsymbol{\delta}}(a;t) - \mathbf{z}\| > \frac{2N\varepsilon}{\sqrt{\boldsymbol{\delta}_{\min}}}\right\} \subseteq \left\{\frac{\xi_n(a;t;\cdot)}{\sqrt{2a\log_+(1/a)}} \notin \mathcal{N}_\varepsilon(g)\right\}. \tag{12.119}$$

Recalling (12.104), and the definition (12.105) of $\mathbf{z}^*_{n,\boldsymbol{\delta}}(a;t)$, set, for $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_N)$,

$$\mathbf{z} = \begin{bmatrix} z_1 \\ \vdots \\ z_N \end{bmatrix}, \quad \mathbf{z}^*_{n,\boldsymbol{\delta}}(a;t) = \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix} \quad \text{and} \quad \mathbf{z}_{n,\boldsymbol{\delta}}(a;t) = \begin{bmatrix} y_1/\sqrt{\delta_1} \\ \vdots \\ y_N/\sqrt{\delta_N} \end{bmatrix}.$$

By combining the triangle inequality with $\|\mathbf{z}\| < 1$, we see that

$$\begin{aligned}\|\mathbf{z}_{n,\boldsymbol{\delta}}(a;t) - \mathbf{z}\| &= \left\{\sum_{j=1}^N (y_j/\sqrt{\delta_j} - z_j)^2\right\}^{1/2} \qquad (12.120)\\ &\geq \left\{\sum_{j=1}^N (y_j/\sqrt{\delta_j} - z_j/\sqrt{\delta_j})^2\right\}^{1/2} - \left\{\sum_{j=1}^N (z_j\sqrt{\delta_j} - z_j)^2\right\}^{1/2} \\ &\geq \frac{1}{\sqrt{\boldsymbol{\delta}_{\max}}}\|\mathbf{z}^*_{n,\boldsymbol{\delta}}(a;t) - \mathbf{z}\| - \|\mathbf{z}\|\left\{\left(1 - \frac{1}{\sqrt{\boldsymbol{\delta}_{\max}}}\right) \vee \left(\frac{1}{\sqrt{\boldsymbol{\delta}_{\min}}} - 1\right)\right\} \\ &\geq \frac{1}{\sqrt{\boldsymbol{\delta}_{\max}}}\|\mathbf{z}^*_{n,\boldsymbol{\delta}}(a;t) - \mathbf{z}\| - \left\{\left(1 - \frac{1}{\sqrt{\boldsymbol{\delta}_{\max}}}\right) \vee \left(\frac{1}{\sqrt{\boldsymbol{\delta}_{\min}}} - 1\right)\right\}.\end{aligned}$$

Thus, if we assume that

$$\frac{1}{\sqrt{\delta_{\max}}} \geq 1 - N\varepsilon \quad \text{and} \quad \frac{1}{\sqrt{\delta_{\min}}} \leq 1 + N\varepsilon, \tag{12.121}$$

we infer from (12.120) that

$$\|\mathbf{z}_{n,\delta}(a;t) - \mathbf{z}\| \geq \frac{1}{\sqrt{\delta_{\max}}} \|\mathbf{z}^*_{n,\delta}(a;t) - \mathbf{z}\| + N\varepsilon.$$

This, when combined with (12.119), shows that

$$\left\{ \|\mathbf{z}^*_{n,\delta}(a;t) - \mathbf{z}\| > 3N\varepsilon\sqrt{\frac{\delta_{\max}}{\delta_{\min}}} \right\} \subseteq \left\{ \frac{\xi_n(a;t;\cdot)}{\sqrt{2a\log_+(1/a)}} \notin \mathcal{N}_\varepsilon(g) \right\}. \tag{12.122}$$

In view of (12.106), we infer from (12.122) the relation

$$\begin{aligned} &\bigcap_{k=1}^{K} \left\{ \|\boldsymbol{\zeta}^{(k)}_{n;a_k;\delta_k} - \mathbf{z}\| > 3N\varepsilon\sqrt{\frac{\delta_{\max}}{\delta_{\min}}} \right\} \\ &\subseteq \bigcap_{k=1}^{K} \left\{ \frac{\xi_n(a_k;t_k;\cdot)}{\sqrt{2a_k\log_+(1/a_k)}} \notin \mathcal{N}_\varepsilon(g) \right\} \end{aligned} \tag{12.123}$$

Now, we infer from (12.121) that, whenever $N\varepsilon \leq \frac{1}{2}$,

$$\sqrt{\frac{\delta_{\max}}{\delta_{\min}}} \leq \frac{1+N\varepsilon}{1-N\varepsilon} \leq 3.$$

Thus, by (12.123), we have

$$\begin{aligned} &\mathbb{P}\left( \bigcap_{k=1}^{K} \left\{ \|\boldsymbol{\zeta}^{(k)}_{n;a_k;\delta_k} - \mathbf{z}\| > 9N\varepsilon \right\} \right) \\ &\leq \mathbb{P}\left( \bigcap_{k=1}^{K} \left\{ \frac{\xi_n(a_k;t_k;\cdot)}{\sqrt{2a_k\log_+(1/a_k)}} \notin \mathcal{N}_\varepsilon(g) \right\} \right) \leq 2\exp\left( -\frac{1}{4}\sum_{k=1}^{K} a_k^{1-\varepsilon/2} \right). \end{aligned} \tag{12.124}$$

The remainder of the proof is given by routine arguments which we omit. □

# References


1. C.R. Adams, J.A. Clarkson, On definitions of bounded variation for functions of two variables. Trans. Am. Math. Soc. **35**, 824–854 (1933)
2. N. Bouleau, On effective computation of expectations in large or infinite dimensions. J. Comput. Appl. Math. **31**, 23–34 (1990)
3. J.A. Clarkson, C. Raymond Adams, Properties of functions $f(x, y)$ of bounded variation. Trans. Am. Math. Soc. **36**, 711–730 (1933)
4. P. Deheuvels, Laws of the iterated logarithm for density estimators, in *Nonparametric Functional Estimation and Related Topics* (Kluwer, Dordrecht, 1991), pp. 19–29
5. P. Deheuvels, Functional laws of the iterated logarithm for large increments of empirical and quantile processes. Stoch. Process. Appl. **43**, 133–163 (1992)
6. P. Deheuvels, One bootstrap suffices to generate sharp uniform bounds in functional estimation. Kybernetika **47**, 855–865 (2011)
7. P. Deheuvels, J.H.J. Einmahl, Functional limit laws for the increments of Kaplan-Meier product-limit processes and applications. Ann. Probab. **28**, 1301–1335 (2000)
8. P. Deheuvels, D.M. Mason, Functional laws of the iterated logarithm for the increments of empirical and quantile processes. Ann. Probab. **20**, 1248–1287 (1992)
9. P. Deheuvels, D.M. Mason, General asymptotic confidence bands based on kernel-type function estimators. Stat. Infer. Stoch. Process **7**, 225–277 (2004)
10. P. Deheuvels, S. Ouadah, Uniform-in-bandwidth functional limit laws. J. Theor. Probab. **26**(3), 697–721 (2013)
11. J. Dony, Nonparametric regression estimation-An empirical process approach to uniform in bandwidth consistency of kernel-type estimators and conditional $U$-statistics. Doctoral Dissertation. Vrije Universiteit Brussel, Brussels, 2008
12. J. Dony, U. Einmahl, Weighted uniform consistency of kernel density estimators with general bandwidth sequences. Electron. J. Probab. **11**, 844–859 (2006)
13. J. Dony, U. Einmahl, Uniform in bandwidth consistency of kernel-type estimators at a fixed point. Inst. Math. Stat. Collect. **5**, 308–325 (2009)
14. J. Dony, D.M. Mason, Uniform in bandwidth consistency of conditional $U$-statistics. Bernoulli **14**(4), 1108–1133 (2008)
15. J. Dony, U. Einmahl, D.M. Mason, Uniform in bandwidth consistency of local polynomial regression function estimators. Aust. J. Stat. **35**, 105–120 (2006)
16. U. Einmahl, D.M. Mason, An empirical process approach to the uniform consistency of kernel-type function estimators. J. Theor. Probab. **13**, 1–37 (2000)
17. U. Einmahl, D.M. Mason, Uniform in bandwidth consistency of kernel-type function estimators. Ann. Stat. **33**, 1380–1403 (2005)
18. G.H. Hardy, On double Fourier series, and especially those which represent the double zeta-function with real and incommensurable parameters. Q. J. Math. **37**, 53–89 (1905)
19. M. Krause, Über Mittelwertsätze in Gebiete der Doppelsummen und Doppelintegrale. Leipziger Ber. **55**, 239–263 (1903)
20. D.M. Mason, A strong limit theorem for the oscillation modulus of the uniform empirical process. Stoch. Process. Appl. **17**, 127–136 (1984)
21. D.M. Mason, A uniform functional law of the logarithm for the local empirical process. Ann. Probab. **32**, 1391–1418 (2004)
22. D.M. Mason, Proving consistency of non-standard kernel estimators. Stat. Infer. Stoch. Process. **20**(2), 151–176 (2012)
23. D.M. Mason, J. Swanepoel, A general result on the uniform in bandwidth consistency of kernel-type function estimators. Test **20**, 72–94 (2011)
24. D.M. Mason, G.R. Shorack, J.A. Wellner, Strong limit theorems for oscillation moduli of the empirical process. Z. Wahrscheinlichkeitstheorie und Verwandte Geb. **65**, 93–97 (1983)
25. E.A. Nadaraya, On estimating regression. Theor. Probab. Appl. **9**, 141–142 (1964)

26. H. Niederreiter, *Random Number Generation and Quasi-Monte Carlo Methods*. CBMS-NSF Regional Conference Series in Applied Mathematics, vol. 63 (SIAM, Philadelphia, 1992)
27. D. Nolan, J.S. Marron, Uniform consistency and location adaptive delta-sequence estimators. Probab. Theory Relat. Fields **80**, 619–632 (1989)
28. G. Pagès, Y.-J. Xiao, Sequences with low discrepancy and pseudo-random numbers: theoretical results and numerical tests. J. Stat. Comput. Simul. **56**, 163–188 (1997)
29. E. Parzen, On the estimation of a probability density function and mode. Ann. Math. Stat. **33**, 1065–1076 (1962)
30. M. Rosenblatt, Remarks on some nonparametric estimates of a density function. Ann. Math. Stat. **27**, 832–837 (1956)
31. B. Schweizer, E.F. Wolff, On nonparametric measure of dependence for random variables. Ann. Stat. **6**, 177–184 (1981)
32. B. Silverman, Weak and strong consistency of the kernel estimate of a density and its derivatives. Ann. Stat. **6**, 177–184 (1978) (Addendum: (1980). **8**, 1175–1176)
33. V. Strassen, An invariance principle for the law of the iterated logarithm. Z. Wahrscheinlichkeitstheorie und Verwandte Geb. **3**, 211–226 (1964)
34. W. Stute, The oscillation behavior of empirical processes: the multivariate case. Ann. Probab. **12**, 361–379 (1984)
35. I. van Keilegom, D. Varron, Uniform in bandwidth exact rates for a class of kernel estimators. Ann. Inst. Stat. Math. **63**(6), 1077–1102 (2011)
36. D. Varron, Lois fonctionnelles uniforme du logarithme itéré pour les accroissements du processus empirique généralisé, in *Lois limites de type Chung-Mogulskii pour le processus empirique uniforme local*. Doctoral Dissertation, Université Pierre et Marie Curie, Paris, Dec. 17, 2004
37. D. Varron, A limited in bandwidth uniformity for the functional limit law of the increments of the empirical process. Electron. J. Stat. **2**, 1043–1064 (2008)
38. V. Viallon, Functional limit laws for the increments of the quantile process with applications. Electron. J. Stat. **1**, 496–518 (2007)
39. G. Vitali, Sui gruppi di punti e sulle funzioni di variabili reali. Atti Accad. Sci. Torino. **43**, 229–246 (1908)
40. G.S. Watson, Smooth regression analysis. Sankhyā Indian J. Stat. A **26**, 359–372 (1964)

# Chapter 13
# Universality of Limiting Spectral Distribution Under Projective Criteria

**Florence Merlevède and Magda Peligrad**

**Abstract** This paper has double scope. In the first part we study the limiting empirical spectral distribution of a $n \times n$ symmetric matrix with dependent entries. For a class of generalized martingales we show that the asymptotic behavior of the empirical spectral distribution depends only on the covariance structure. Applications are given to strongly mixing random fields. The technique is based on a blend of blocking procedure, martingale techniques and multivariate Lindeberg's method. This means that, for this class, the study of the limiting spectral distribution is reduced to the Gaussian case. The second part of the paper contains a survey of several old and new asymptotic results for the empirical spectral distribution for Gaussian processes, which can be combined with our universality results.

## 13.1 Introduction

The distribution of the eigenvalues of random matrices is useful in many fields of science such as statistics, physics and engineering. The celebrated paper by Wigner [45] deals with symmetric matrices having i.i.d. entries below the diagonal. Wigner proved a global universality result, showing that, asymptotically and with probability one, the empirical distribution of eigenvalues is distributed according to the semicircle law (see Chapter 2 in Bai and Silverstein [1] for more details). The only parameter of this law is the variance of an entry. This result was expanded in various directions. The first generalization was to decrease the degree of stationarity by replacing the condition of equal variance by weaker assumptions of the Lindeberg's type. Another direction of generalization deals with weakening the hypotheses of independence by considering various notions of weak dependence.

F. Merlevède (✉)
Université Paris-Est, LAMA and CNRS UMR 8050, Marne-la-Vallée, France
e-mail: florence.merlevede@u-pem.fr

M. Peligrad
Department of Mathematical Sciences, University of Cincinnati, Cincinnati, OH, USA
e-mail: peligrm@ucmail.uc.edu



N. Gozlan et al. (eds.), *High Dimensional Probability VIII*,
Progress in Probability 74, https://doi.org/10.1007/978-3-030-26391-1_13

For symmetric Gaussian matrices with correlated entries, works of Khorunzhy and Pastur [23], Boutet de Monvel et al. [8], Boutet de Monvel and Khorunzhy [7], Chakrabarty et al. [10], Peligrad and Peligrad [34] showed that the limiting spectral distribution of the symmetric matrix depends only on the covariance structure of the underlying Gaussian process. The limiting spectral distribution is rather complicated and the best way to describe it is by specifying an equation satisfied by its Stieltjes transform.

A way to symmetrize a matrix is to multiply it with its transpose. These matrices, known under the name of Gram matrices or sample covariance matrices, play an important role in statistical studies of large data sets. The spectral analysis of large-dimensional sample covariance matrices has been actively studied starting with the seminal work of Marchenko and Pastur [25] who considered independent random samples from an independent multidimensional vector. A big step forward was the study of the dependent case represented in numerous papers. Basically, the entries of the matrix were allowed to be linear combinations of an independent sequence. The first paper where such a model was considered is by Yin and Krishnaiah [50] followed by important contributions by Yin [49], Silverstein [41], Silverstein and Bai [42], Hachem et al. [22], Pfaffel and Schlemm [35], Yao [46], Pan et al. [31], Davis et al. [12], among many others. Another type of models was considered by Bai and Zhou [2] based on independent columns. The dependence type-condition imposed to the columns is in particular satisfied for isotropic vectors with log-concave distribution (see Pajor and Pastur [30]) but may be hard to verify for non linear time series (such that ARCH models) or requires rate of convergence of mixing coefficients. Let us also mention the recent papers by Yaskov [47, 48] where a weaker version of the Bai-Zhou's dependence type condition has been introduced.

In two recent papers Banna et al. [3] and Merlevède-Peligrad [27], have shown that, for two situations, namely for symmetric matrices whose entries are functions of independent and identically distributed random fields or for large sample covariance matrices generated by random matrices with independent rows, the limiting spectral distribution of eigenvalues counting measure always exists and can be described via an equation satisfied by its Stieltjes transform.

Even if many models encountered in time series analysis can be rewritten as functions of an i.i.d. sequence, this assumption is not completely satisfactory since many stationary processes, even with trivial left sigma field, cannot be in general represented as a function of an i.i.d. sequence, as shown for instance in Rosenblatt [39]. Moreover, the assumption of independence of the rows or of the columns generating the large sample covariance matrices may be too restrictive.

The main goal of our paper is then to continue the study of the asymptotic behavior of the empirical eigenvalues distribution of symmetric matrices and large sample covariance matrices associated with random fields when the variables are not necessarily functions of an i.i.d. sequence or when the rows (or columns) are not necessarily independent. In the first part of the paper we shall show that the universality results hold for both symmetric and symmetrized random matrices when the dependence is controlled by projective type coefficients. These coefficients are easy to estimate in terms of strong mixing coefficients. By "universality" we mean that the limiting distribution of the eigenvalues counting measure depends only of the process' covariance structure. Therefore our result reduces the study of the limiting spectral distribution (LSD) to the case where the entries of the underlying matrix are observations of a Gaussian random field with the same covariance structure. In the second part of the paper we survey old and new results for the Gaussian case, which one can combine with the universality theorems, for obtaining the existence and the characterization of LSD.

Our paper is organized as follows. Section 13.2 contains the notations and the universality results. In Sect. 13.3 we apply our results to classes of strongly mixing random fields. Then, Sect. 13.4 is dedicated to LSD results for symmetrized matrices when the entries are observations of a Gaussian random field. All the proofs are postponed to Sect. 13.5. Several auxiliary results needed in the proofs are given in Sect. 13.6.

Here are some notations used all along the paper. The notation $[x]$ is used to denote the integer part of a real $x$. The notation $\mathbf{0}_{p,q}$ means a matrix of size $p \times q$, $(p, q) \in \mathbb{N}^2$ with entries 0. For a matrix $A$, we denote by $A^T$ its transpose matrix, by $\mathrm{Tr}(A)$ its trace. We shall also use the notation $\|X\|_r$ for the $\mathbb{L}^r$-norm ($r \geq 1$) of a real valued random variable $X$. For two sequences of positive numbers $(a_n)$ and $(b_n)$ the notation $a_n \ll b_n$ means that there is a constant $C$ such that $a_n \leq Cb_n$ for all $n \in \mathbb{N}$. We use bold small letters to denote an element of $\mathbb{Z}^2$, hence $\mathbf{u} = (u_1, u_2) \in \mathbb{Z}^2$. For $\mathbf{u} = (u_1, u_2)$ and $\mathbf{v} = (v_1, v_2)$ in $\mathbb{Z}^2$, the following notations will be used: $|\mathbf{u} - \mathbf{v}| = \max(|u_1 - v_1|, |u_2 - v_2|)$ and $\mathbf{u} \wedge \mathbf{v} = (u_1 \wedge u_2, v_1 \wedge v_2)$ (where $u_1 \wedge u_2 = \min(u_1, u_2)$).

## 13.2 Results

Let $(X_{\mathbf{u}})_{\mathbf{u} \in \mathbb{N}^2}$ be a real-valued random field defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We consider the symmetric $n \times n$ random matrix $\mathbf{X}_n$ such that, for any $i$ and $j$ in $\{1, \ldots, n\}$

$$(\mathbf{X}_n)_{ij} = X_{ij} \ \text{ for } i \geq j \ \text{ and} \qquad (13.1)$$
$$(\mathbf{X}_n)_{ij} = X_{ji} \ \text{ for } i < j \,.$$

Denote by $\lambda_1^n \leq \cdots \leq \lambda_n^n$ the eigenvalues of

$$\mathbb{X}_n := \frac{1}{n^{1/2}} \mathbf{X}_n \qquad (13.2)$$

and define its spectral distribution function by

$$\mathbf{F}^{\mathbb{X}_n}(t) = \frac{1}{n} \sum_{1 \leq k \leq n} I(\lambda_k \leq t) \,,$$

where $I(A)$ denotes the indicator of an event $A$. The Stieltjes transform of $\mathbb{X}_n$ is given by

$$S^{\mathbb{X}_n}(z) = \int \frac{1}{x-z} dF^{\mathbb{X}_n}(x) = \frac{1}{n}\mathrm{Tr}(\mathbb{X}_n - z\mathbf{I}_n)^{-1}\,, \tag{13.3}$$

where $z = u + iv \in \mathbb{C}^+$ (the set of complex numbers with positive imaginary part), and $\mathbf{I}_n$ is the identity matrix of order $n$. In particular, if the random field is an array of i.i.d. random variables with variance $\sigma^2 > 0$, then Wigner [45] proved that, with probability one, and for any $z \in \mathbb{C}^+$, $S^{\mathbb{X}_n}(z)$ converges to $S(z)$, which satisfies the equation $\sigma^2 S^2 + S + z^{-1} = 0$. Its solution

$$S(z) = -(z - \sqrt{z^2 - 4\sigma^2})(2\sigma^2)^{-1} \tag{13.4}$$

is the well-known Stieltjes transform of the semicircle law, which has the density

$$g(x) = \frac{1}{2\pi\sigma^2}\sqrt{4\sigma^2 - x^2}\, I(|x| \le 2\sigma)\,.$$

(See for instance Theorem 2.5 in Bai-Silverstein [1].) Note that it is not necessary for the random variables to have the same law for this result to hold. Indeed, if the random field $(X_{\mathbf{u}})_{\mathbf{u}\in\mathbb{Z}^2}$ is an array of independent centered random variables with common positive variance $\sigma^2$, which satisfies the Lindeberg's condition given in Condition 13.1 below, then for all $z \in \mathbb{C}^+$, $S^{\mathbb{X}_n}(z)$ converges almost surely to the Stieltjes transform of the semicircle law with parameter $\sigma^2$ (see for instance Theorem 2.9 in Bai and Silverstein [1]). Note that the necessity of the Lindeberg's condition has been stated in Girko's book [19].

Another way to state the Wigner's result is to say that the Lévy distance between the distribution function $\mathbf{F}^{\mathbb{X}_n}$ and $G$, defined by $G(x) = \int_{-\infty}^{x} g(u)du$, converges to zero almost surely. Recall that the Lévy metric $d$ between two distribution functions $F$ and $G$, defined by

$$d(F, G) = \inf\{\varepsilon > 0 \;:\; F(x-\varepsilon) - \varepsilon \le G(x) \le F(x+\varepsilon) + \varepsilon\,,\; \forall x \in \mathbb{R}\}\,.$$

The aim of this paper is to specify a class of random fields for which the limiting behavior of $\mathbf{F}^{\mathbb{X}_n}(t)$ depends only on the covariances of the random variables $(X_{\mathbf{u}})_{\mathbf{u}\in\mathbb{N}^2}$ and not on the structural dependence structure. In other words, we shall show that the limiting spectral distribution of $\mathbf{F}^{\mathbb{X}_n}(t)$ can be deduced from that one of $\mathbf{F}^{\mathbb{Y}_n}(t)$ where $\mathbb{Y}_n$ is a Gaussian matrix with the same covariance structure as $\mathbb{X}_n$. Since the estimate of the Lévy distance between $\mathbf{F}^{\mathbb{X}_n}$ and $\mathbf{F}^{\mathbb{Y}_n}$ can be given in terms of their Stieltjes transforms (see, for instance, Theorem B.12 and Lemma B.18 in Bai and Silverstein [1] or Proposition 2.1 in Bobkov et al. [6]), we shall compare their Stieltjes transforms.

Our first result compares the Stieltjes transform of a matrix satisfying martingale-like projective conditions with the Stieltjes transform of an independent matrix whose entries are observations of a Gaussian random field with the same covariance structure. We shall assume that $\mathbb{X}_n$ is defined by (13.2), and satisfies the Lindeberg's condition below:

**Condition 13.1**

*(i) The variables $(X_{ij})_{i,j}$ are centered at expectations.*
*(ii) There exists a positive constant $C$ such that, for any positive integer $n$,*

$$\frac{1}{n^2}\sum_{n\geq i\geq j\geq 1}\mathbb{E}(X_{ij}^2) < C.$$

*(iii) For every $\varepsilon > 0$,*

$$L_n(\varepsilon) = \frac{1}{n^2}\sum_{n\geq i\geq j\geq 1}\mathbb{E}(X_{ij}^2 I(|X_{ij}| > \varepsilon n^{1/2})) \to 0.$$

Clearly the items (ii) and (iii) of this condition are satisfied as soon as the family $(X_{ij}^2)$ is uniformly integrable or the random field is stationary and in $\mathbb{L}^2$ (recall that a random field $(X_{\mathbf{u}})_{\mathbf{u}\in\mathbb{Z}^2}$ is said to be (strictly) stationary if the law of $(X_{\mathbf{u}+\mathbf{v}})_{\mathbf{u}\in\mathbb{Z}^2}$ does not depend on $\mathbf{v}\in\mathbb{Z}^2$).

To introduce our martingale-like projective conditions (13.6) and (13.7) below as well as our regularity-type condition (13.8), we need to introduce the filtrations we shall consider:

For any non-negative integer $a$, let us introduce the following filtrations:

$$\mathcal{F}^a_{i,\infty} = \sigma(X_{uv} : 1\leq u\leq i-a\,,\ v\geq 1) \text{ if } i > a \text{ and } \mathcal{F}^a_{i,\infty} = \{\Omega,\emptyset\} \text{ otherwise} \tag{13.5}$$

$$\mathcal{F}^a_{\infty,j} = \sigma(X_{uv} : u\geq 1\,,\ 1\leq v\leq j-a) \text{ if } j > a \text{ and } \mathcal{F}^a_{\infty,j} = \{\Omega,\emptyset\} \text{ otherwise}$$

$$\mathcal{F}^a_{ij} = \mathcal{F}^a_{i,\infty}\cup\mathcal{F}^a_{\infty,j}\,.$$

Note that $X_{ij}$ is adapted to $\mathcal{F}^0_{ij}$. We are now in position to state our first result.

**Theorem 13.2** *Assume that $\mathbb{X}_n$ satisfies Condition 13.1 and, as $n\to\infty$,*

$$\sup_{i\geq j}\|\mathbb{E}(X_{ij}|\mathcal{F}^n_{ij})\|_2 \to 0 \tag{13.6}$$

*and*

$$n^2\sup\|\mathbb{E}(X_{ij}X_{ab}|\mathcal{F}^n_{i\wedge a,j\wedge b}) - \mathbb{E}(X_{ij}X_{ab})\|_1 \to 0, \tag{13.7}$$

*where the supremum is taken over all pairs* $(i, j) \neq (a, b)$ *with* $i \geq j$ *and* $a \geq b$*. In addition assume that*

$$\sup_{i \geq j} \|\mathbb{E}(X_{ij}^2 | \mathcal{F}_{ij}^n) - \mathbb{E}(X_{ij}^2)\|_1 \to 0 \text{ as } n \to \infty. \tag{13.8}$$

*Then for all* $z \in \mathbb{C}^+$

$$S^{\mathbb{X}_n}(z) - S^{\mathbb{Y}_n}(z) \to 0 \text{ in probability as } n \to \infty, \tag{13.9}$$

*where* $\mathbf{Y}_n$ *is a Gaussian matrix of centered random variables with the same covariance structure as* $\mathbf{X}_n$ *and independent of* $\mathbf{X}_n$ *and* $\mathbb{Y}_n = \mathbf{Y}_n / \sqrt{n}$*.*

**Comment 13.3**

(i) Conditions (13.6) and (13.7) can be viewed as a generalization of the martingale condition given in Basu and Dorea [4] which is $\mathbb{E}(X_{ij} | \mathcal{F}_{i,j}^1) = 0$ a.s. for any $i \geq j \geq 1$. Both conditions (13.6) and (13.7) are obviously satisfied for this type of martingale random field, and then, the conditions of Theorem 13.2 are reduced just to Condition 13.1 and (13.8). Results for other type of martingale random fields based on the lexicographic order can be found in Merlevède et al. [28].

(ii) Note also that Condition (13.8) is a regularity condition. For instance, in case where $\mathcal{F}_{ij}^{\infty} = \bigcap_{n \geq 0} \mathcal{F}_{ij}^n$ is the trivial $\sigma$-field, then this condition is automatically satisfied. Let us also mention that the conditions (13.6)–(13.8) are natural extensions of projective criteria used for obtaining various limit theorems for sequences of random variables. As in the case of random sequences, the conditions (13.6)–(13.8) can be handled either with the help of "physical measure of dependence" as developed in El Machkouri et al. [18] for functions of i.i.d. random fields or by using mixing coefficients (see Sect. 13.3.1).

(iii) We should also mention that we can allow for dependence of $n$ of the variables in $\mathbb{X}_n$. The conditions in the theorem below have to be then generalized in a natural way. For instance, conditions (13.6) and (13.7) should become

$$\lim_{m \to \infty} \sup_{n \geq 1} \sup_{i \geq j} \|\mathbb{E}(X_{ij,n} | \mathcal{F}_{ij,n}^m)\|_2 = 0$$

and

$$\lim_{m \to \infty} m^2 \sup_{n \geq 1} \sup_{(i,j) \neq (a,b)} \|\mathbb{E}(X_{ij,n} X_{ab,n} | \mathcal{F}_{i \wedge a, j \wedge b, n}^m) - \mathbb{E}(X_{ij,n} X_{ab,n})\|_1 = 0\,.$$

Based on the above theorem we shall treat two special cases of symmetric random matrices, namely $\mathbf{X} + \mathbf{X}^T$ and the covariance matrix given in definition (13.17).

We consider first the symmetric $n \times n$ matrix $\mathbf{Z}_n = [Z_{ij}]_{i,j=1}^n$ with $Z_{ij} = X_{ij} + X_{ji}$ and we set

$$\mathbb{Z}_n := \frac{1}{\sqrt{2n}} \mathbf{Z}_n \,. \tag{13.10}$$

This type of symmetrization is important since it leads to a symmetric covariance structure. Indeed, if $(X_{ij})_{(i,j)\in\mathbb{Z}^2}$ is $\mathbb{L}^2$-stationary meaning that, for any $(i, j) \in \mathbb{Z}^2$, $\mathbb{E}(Y_{ij}) = m$ and

$$\mathrm{cov}(X_{u,v}, X_{k+u,\ell+v}) = \mathrm{cov}(X_{0,0}, X_{k,\ell}) = c_{k,\ell} \,,$$

for any integers $u, v, k, \ell$, we get that $(Z_{ij})_{(i,j)\in\mathbb{Z}^2}$ is also a $\mathbb{L}^2$-stationary random field satisfying

$$\mathrm{cov}(Z_{i,j}, Z_{k,\ell}) = b(k-i, \ell-j) + b(k-j, \ell-i) \text{ with } b(u, v) = \gamma_{u,v} + \gamma_{v,u} \,.$$

Notice then that $b(u, v) = b(v, u)$. This symmetry condition on the covariances is used for instance in Khorunzhy and Pastur [23, Theorem 2] to derive the limiting spectral distribution of symmetric matrices associated with a stationary Gaussian random field when its associated series of covariances is absolutely summable.

Our next Theorem 13.4 shows that a similar conclusion as in Theorem 13.2 holds for $\mathbb{Z}_n$ defined above. However, due to the structure of each of the entries, the sequence $(X_{ij})$ has to satisfy the conditions of Theorem 13.2 but with the conditional expectations taken with respect to a larger filtration. Roughly speaking the filtrations in Theorem 13.2 are the union of two half planes, whereas in Theorem 13.4 they are defined as the sigma-algebras generated by all the variables outside the union of two squares. More precisely these latter filtrations are defined as follows: for any non-negative integer $a$,

$$\widetilde{\mathcal{F}}_{ij}^a = \sigma\big(X_{uv} \ : \ (u, v) \in \mathbb{Z}^2 \text{ such that } \max(|i-u|, |j-v|) \geq a\big) \,. \tag{13.11}$$

Note that $X_{ij}$ is adapted to $\widetilde{\mathcal{F}}_{ij}^0$.

**Theorem 13.4** *Assume that $\mathbb{Z}_n$ is defined by (13.10) where the variables $X_{ij}$ satisfy Condition 13.1. In addition assume that*

$$\sup_{i \geq j} \|\mathbb{E}(X_{ij}|\widetilde{\mathcal{F}}_{ij}^n)\|_2 \to 0 \text{ as } n \to \infty, \tag{13.12}$$

$$n^2 \sup \|\mathbb{E}(X_{ij}X_{ab}|\widetilde{\mathcal{F}}_{ij}^n \cap \widetilde{\mathcal{F}}_{ab}^n) - \mathbb{E}(X_{ij}X_{ab})\|_1 \to 0 \text{ as } n \to \infty, \tag{13.13}$$

*where the supremum is taken over all pairs $(i, j) \neq (a, b)$. In addition assume that*

$$\sup_{(i,j)} \|\mathbb{E}(X_{ij}^2|\widetilde{\mathcal{F}}_{ij}^n) - \mathbb{E}(X_{ij}^2)\|_1 \to 0 \text{ as } n \to \infty. \tag{13.14}$$

*Then, for all* $z \in \mathbb{C}^+$,

$$S^{\mathbb{Z}_n}(z) - S^{\mathbb{W}_n}(z) \to 0 \text{ in probability as } n \to \infty, \tag{13.15}$$

*where* $\mathbf{W}_n = [W_{ij}]_{i,j=1}^n$ *with* $W_{ij} = Y_{ij} + Y_{ji}$*,* $(Y_{ij})$ *being a real-valued Gaussian centered random field with the same covariance structure as* $\mathbf{X}_n$ *and independent of* $\mathbf{X}_n$*, and* $\mathbb{W}_n = \mathbf{W}_n/\sqrt{2n}$*.*

Let $(X_{\mathbf{u}})_{\mathbf{u}\in\mathbb{Z}^2}$ be a random field of real-valued square integrable variables and $(Y_{\mathbf{u}})_{\mathbf{u}\in\mathbb{Z}^2}$ be a real-valued Gaussian random field with the same covariances, and independent of $(X_{\mathbf{u}})_{\mathbf{u}\in\mathbb{Z}^2}$. Let $N$ and $p$ be two positive integers and consider the $N \times p$ matrices

$$\mathcal{X}_{N,p} = \left(X_{ij}\right)_{1\le i\le N, 1\le j\le p}, \quad \Gamma_{N,p} = \left(Y_{ij}\right)_{1\le i\le N, 1\le j\le p}. \tag{13.16}$$

Define now the symmetric matrices $\mathbb{B}_N$ and $\mathbb{G}_N$ of order $N$ by

$$\mathbb{B}_N = \frac{1}{N}\mathcal{X}_{N,p}\mathcal{X}_{N,p}^T, \quad \mathbb{G}_N = \frac{1}{N}\Gamma_{N,p}\Gamma_{N,p}^T. \tag{13.17}$$

The matrix $\mathbb{B}_N$ is usually referred to as the sample covariance matrix associated with the process $(X_{\mathbf{u}})_{\mathbf{u}\in\mathbb{Z}^2}$. It is also known under the name of Gram random matrix. In particular, if the random field $(X_{\mathbf{u}})_{\mathbf{u}\in\mathbb{Z}^2}$ is an array of i.i.d. random variables with zero mean and variance $\sigma^2$, then the famous Marchenko and Pastur [25] theorem states that, if $p/N \to c \in (0, \infty)$, then, for all $z \in \mathbb{C}^+$, $S^{\mathbb{B}_N}(z)$ converges almost surely to $S(z) = S$, which is the unique solution with $\text{Im}S(z) \ge 0$ of the quadratic equation: for any $z \in \mathbb{C}^+$,

$$z\sigma^2 S^2 + (z - c\sigma^2 + \sigma^2)S + 1 = 0. \tag{13.18}$$

This means that $\mathbb{P}(d(F^{\mathbb{G}_N}, F_c) \to 0) = 1$, where $F_c$ is a probability distribution function of the so-called Marchenko-Pastur distribution with parameter $c > 0$. That is $F_c$ has density

$$g_c(x) = \frac{1}{2\pi x\sigma^2}\sqrt{(x-a)(b-x)}\,I(a \le x \le b)$$

and a point mass $1 - c$ at the origin if $c < 1$, where $a = \sigma^2(1 - \sqrt{c})^2$ and $b = \sigma^2(1 + \sqrt{c})^2$. Note that this result still holds if the random field $(X_{\mathbf{u}})_{\mathbf{u}\in\mathbb{Z}^2}$ is an array of independent centered random variables with common positive variance $\sigma^2$, which satisfies the Lindeberg's Condition 13.1 (see Pastur [32]). Moreover, in this situation, the Lindeberg's condition is necessary as shown in Girko [20, Theorem 4.1, Chapter 3] (see also Corollary 2.3 in Yaskov [48]).

When we relax the independence assumption of the entries, the following result holds.

**Theorem 13.5** *Assume* $(X_{\mathbf{u}})_{\mathbf{u}\in\mathbb{Z}^2}$ *is as in Theorem 13.2. Then, if* $p/N \to c \in (0,\infty)$*, for all* $z \in \mathbb{C}^+$*,*

$$S^{\mathbb{B}_N}(z) - S^{\mathbb{G}_N}(z) \to 0 \textit{ in probability, as } N \to \infty.$$

All our results can be easily reformulated for random matrices with entries from a stationary random field. For some applications it is interesting to formulate sufficient conditions in terms of the conditional expectation of a single random variable. For this case it is natural to work with the extended filtrations.

Let now $(X_{ij})_{(i,j)\in\mathbb{Z}^2}$ be a stationary real-valued random field. For any non-negative integer $a$, let us introduce the following filtrations:

$$\mathcal{F}^a_{i,\infty} = \sigma(X_{uv} : u \le i-a,\ v \in \mathbb{Z});$$
$$\mathcal{F}^a_{\infty,j} = \sigma(X_{uv} : v \le j-a,\ u \in \mathbb{Z});\ \mathcal{F}^a_{ij} = \mathcal{F}^a_{i,\infty} \cup \mathcal{F}^a_{\infty,j}\,.$$

We call the random field regular if for any $\mathbf{u} \in \mathbb{Z}^2$, $\mathbb{E}(X_{\mathbf{0}}X_{\mathbf{u}}|\mathcal{F}^{\infty}_{\mathbf{u}\wedge\mathbf{0}}) = \mathbb{E}(X_{\mathbf{0}}X_{\mathbf{u}})$ a.s.

**Theorem 13.6** *Assume that* $\mathbb{X}_n$ *is defined by (13.2) where* $(X_{ij})$ *is a stationary, centered and regular random field. Assume the couple of conditions*

$$\sum_{\mathbf{u}\in V_0} \mathbb{E}|X_{\mathbf{u}}\mathbb{E}(X_{\mathbf{0}}|\mathcal{F}^{|u|}_{\mathbf{0}})| < \infty \tag{13.19}$$

*and*

$$p^2 \sup_{\mathbf{u}\in V_0:|\mathbf{u}|>p} \mathbb{E}|X_{\mathbf{u}}\mathbb{E}(X_{\mathbf{0}}|\mathcal{F}^{p}_{\mathbf{0}})| \to 0,\ \textit{as } p \to \infty,$$

*where* $V_0 = \{\mathbf{u} = (u_1,u_2) \in \mathbb{Z}^2 : u_1 \le 0 \textit{ or } u_2 \le 0\}$*. Then the conclusions of Theorems 13.2 and 13.5 hold.*

Condition (13.19) implies that $\sum_{\mathbf{u}\in\mathbb{Z}^2} |\mathrm{cov}(X_{\mathbf{0}}, X_{\mathbf{u}})| < \infty$ and is in the spirit of condition (2.3) given in Dedecker [13] to derive a central limit theorem for stationary random fields. As we shall see in Sect. 13.3.1, when applied to stationary strongly mixing random fields, the conditions of Theorem 13.6 require a rate of convergence of the strong mixing coefficients with only one point in the future whereas the conditions of Theorems 13.2 and 13.5 require a rate of convergence of the strong mixing coefficients with two points in the future.

Combining Theorem 13.6 with Theorem 13.11 concerning Gaussian covariance matrices, the following corollary holds:

**Corollary 13.7** *Let* $\mathbb{B}_N$ *be defined by (13.17). Under the assumptions of Theorem 13.6 and if* $p/N \to c \in (0,\infty)$*,* $d(F^{\mathbb{B}_N}, F) \to 0$ *in probability where* $F$ *is a nonrandom distribution function whose Stieltjes transform* $S(z)$*,* $z \in \mathbb{C}^+$*, is uniquely defined by the spectral density of* $(X_{ij})$ *and satisfies the equation stated in Theorem 13.11.*

## 13.3 Examples

### *13.3.1 Strongly Mixing Random Fields*

Let us first recall the definition of the strong mixing coefficient of Rosenblatt [38]: For any two $\sigma$-algebras $\mathcal{A}$ and $\mathcal{B}$, the strong mixing coefficient $\alpha(\mathcal{A}, \mathcal{B})$ is defined by:

$$\alpha(\mathcal{A}, \mathcal{B}) = \sup\{|\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|; A \in \mathcal{A} \text{ and } B \in \mathcal{B}\}.$$

An equivalent definition is:

$$2\alpha(\mathcal{A}, \mathcal{B}) = \sup\{|\mathbb{E}(|\mathbb{P}(B|\mathcal{A}) - \mathbb{P}(B)|) : B \in \mathcal{B}\},$$

and, according to Bradley [9], Theorem 4.4, item (a2), one also has

$$4\alpha(\mathcal{A}, \mathcal{B}) = \sup\{\|\mathbb{E}(Y|\mathcal{A})\|_1 : Y \ \mathcal{B}\text{-measurable}, \|Y\|_\infty = 1 \text{ and } \mathbb{E}(Y) = 0\}. \tag{13.20}$$

For a random field $\mathbf{X} = (X_\mathbf{u})_{\mathbf{u} \in \mathbb{Z}^2}$, let

$$\alpha_{1,\mathbf{X}}(n) = \sup_{i,j} \alpha(\mathcal{F}_{ij}^n, \sigma(X_{ij})) \text{ and } \alpha_{2,\mathbf{X}}(n) = \sup_{(i,j) \neq (a,b)} \alpha(\mathcal{F}_{i\wedge a, j\wedge b}^n, \sigma(X_{ij}, X_{ab})).$$

Note that $\alpha_{1,\mathbf{X}}(n) \le \alpha_{2,\mathbf{X}}(n)$. For a bounded centered random field, the mixing condition required by Theorem 13.2 (or by Theorem 13.5) is

$$n^2\alpha_{2,\mathbf{X}}(n) \to 0,$$

while for Theorem 13.6, provided the random field is stationary, we need the couple of conditions:

$$\alpha_{2,\mathbf{X}}(n) \to 0 \text{ and } \sum_{n\ge 1} n\alpha_{1,\mathbf{X}}(n) < \infty.$$

If for some $\delta > 0$ we have $\sup_\mathbf{u} \|X_\mathbf{u}\|_{2+\delta} < \infty$ and the random field is centered then, by the properties of the mixing coefficients, applying, for instance, Lemma 4 in Merlevède and Peligrad [26] (see also Bradley [9] and Annex C in Rio [37]), we infer that the conclusions of Theorems 13.2 and 13.5 are implied by

$$n^2(\alpha_{2,\mathbf{X}}(n))^{\delta/(2+\delta)} \to 0.$$

Moreover, if we assume stationarity of the random field, Theorem 13.6 requires the couple of conditions:

$$\alpha_{2,\mathbf{X}}(n) \to 0 \text{ and } \sum_{n \geq 1} n^{1+4/\delta} \alpha_{1,\mathbf{X}}(n) < \infty .$$

Slightly more general results can be given in terms of the quantile function of $|X_{\mathbf{u}}|$ (see Rio [37] or the computations in the proof of Theorem 6.40 in Merlevède et al. [29] where similar projective quantities as those involved in Theorems 13.2 or 13.6 are handled).

We refer to the monograph by Doukhan [17] for examples of strong mixing random fields. Let us also mention the paper by Dombry and Eyi-Minko [16] where, for max-infinitely divisible random fields on $\mathbb{Z}^d$, upper bounds of the strong mixing coefficients are given with the help of the extremal coefficient function (examples such as the Brown-Resnick process and the moving maxima process are considered). Strong mixing coefficients can also be controlled in the case of bounded spin systems. For instance, in case where the family of Gibbs specifications satisfies the weak mixing condition introduced by Dobrushin and Shlosman [15], the coefficient $\alpha_2(n)$ decreases exponentially fast. This is then the case for Ising models with external fields in the regions where the temperature is strictly larger than the critical one (we refer to Dedecker [14, Section 2.3] and to Laroche [24] for more details).

Below, is another example of a random field for which the strong mixing coefficients can be handled.

*Example: Functions of Two Independent Strong Mixing Random Fields* Let us consider two real-valued independent processes $\mathbf{U} = (U_{ij}, i, j \in \mathbb{Z})$ and $\mathbf{V} = (V_{ij}, i, j \in \mathbb{Z})$ such that, setting $\mathbf{U}^{(j)} = (U_{ij}, i \in \mathbb{Z})$, the processes $\mathbf{U}^{(j)}$, $j \in \mathbb{Z}$, are mutually independent and have the same law as $(U_i, i \in \mathbb{Z})$ and, setting $\mathbf{V}_{(i)} = (V_{ij}, j \in \mathbb{Z})$, the processes $\mathbf{V}_{(i)}$, $i \in \mathbb{Z}$ are also mutually independent and have the same law as $(V_j, j \in \mathbb{Z})$. For any measurable function $h$ from $\mathbb{R}^2$ to $\mathbb{R}$, let

$$X_{ij} = h(U_{ij}, V_{ij}) - \mathbb{E}(h(U_{ij}, V_{ij})), \tag{13.21}$$

provided the expectation exists. Note that the random field $\mathbf{X} = (X_{ij}, i, j \in \mathbb{Z})$ does not have independent entries across the rows nor the columns (except if we have that for any $j$ fixed, the r.v.'s $U_{ij}$, $i \in \mathbb{Z}$ are mutually independent as well as the r.v.'s $V_{ij}$, $j \in \mathbb{Z}$, for any $i$ fixed). Hence, the results in Merlevède and Peligrad [27] do not apply. Let $\mathcal{F}_k^{\mathbf{U}} = \sigma(U_\ell, \ell \leq k)$ and $\mathcal{F}_k^{\mathbf{V}} = \sigma(V_\ell, \ell \leq k)$, and define

$$\alpha_{1,\mathbf{U}}(n) = \sup_i \alpha(\mathcal{F}_{i-n}^{\mathbf{U}}, \sigma(U_i)) \, , \alpha_{2,\mathbf{U}}(n) = \sup_{i,j\,:\,j>i} \alpha(\mathcal{F}_{i-n}^{\mathbf{U}}, \sigma(U_i, U_j))$$

and

$$\alpha_{1,\mathbf{V}}(n) = \sup_i \alpha(\mathcal{F}_{i-n}^{\mathbf{V}}, \sigma(V_i)) \, , \alpha_{2,\mathbf{V}}(n) = \sup_{i,j\,:\,j>i} \alpha(\mathcal{F}_{i-n}^{\mathbf{V}}, \sigma(V_i, V_j)) \, .$$

Due to the definition of the strong mixing coefficients, it follows that

$$\alpha_{1,\mathbf{X}}(n) \le \alpha_{1,\mathbf{U}}(n) + \alpha_{1,\mathbf{V}}(n) \text{ and } \alpha_{2,\mathbf{X}}(n) \le \alpha_{2,\mathbf{U}}(n) + \alpha_{2,\mathbf{V}}(n) \, .$$

(See for instance Theorem 6.2 in Bradley [9].) So, if we assume for instance that the function $h$ is bounded and that $n^2(\alpha_{2,\mathbf{U}}(n) + \alpha_{2,\mathbf{V}}(n)) \to 0$, then Theorem 13.2 applies. Moreover if we assume in addition that the sequences $(U_{ij}, i \in \mathbb{Z})$ and $(V_{ij}, j \in \mathbb{Z})$ are stationary and that

$$\sum_{n\ge 1} n\big(\alpha_{2,\mathbf{U}}(n) + \alpha_{2,\mathbf{V}}(n)\big) < \infty \, ,$$

then, according to Corollary 13.7, we derive that, if $p/N \to c \in (0,\infty)$, for all $z \in \mathbb{C}^+$,

$$S^{\mathbb{B}_N}(z) \to S(z) \text{ in probability as } N \to \infty \, ,$$

where $\mathbb{B}_N$ is the Gram random matrix defined by (13.17) and $S$ is defined in Theorem 13.11.

### 13.3.2 A Convolution Example

Let $\mathbf{U} = (U_{ij}, i, j \in \mathbb{Z})$ be a stationary centered regular martingale difference random field in $\mathbb{L}^2$, meaning that $\sup_{i,j} \|U_{ij}\|_2 < \infty$ and that, setting $\mathcal{G}_{ij}^a = \sigma(V_{k\ell}, k \le i-a \text{ or } \ell \le j-a)$,

$$\mathbb{E}(U_{ij}|\mathcal{G}_{ij}^1) = 0 \text{ a.s. and } \|\mathbb{E}(U_{\mathbf{0}}^2|\mathcal{G}_{\mathbf{0}}^n) - \mathbb{E}(U_{\mathbf{0}}^2)\|_1 \to 0 \text{ as } n \to \infty \, .$$

Let $\varepsilon = (\varepsilon_{ij}, i, j \in \mathbb{Z})$ be an iid centered random field in $\mathbb{L}^\infty$, independent of $\mathbf{U}$ and $(a_{k\ell}, k, \ell \in \mathbb{N})$ be a double indexed sequence of real numbers such that $\sum_{k,\ell\in\mathbb{N}}(k^2+\ell^2)|a_{k,\ell}| < \infty$. Set $V_{ij} = \sum_{k,\ell\in\mathbb{Z}} a_{k,\ell}\varepsilon_{i-k,j-\ell}$ and define the stationary centered random field $\mathbf{X} = (X_{\mathbf{u}})_{\mathbf{u}\in\mathbb{Z}^2}$ in $\mathbb{L}^2$ by setting $X_{ij} = U_{ij} + V_{ij}$. It is easy to see that $\mathbf{X}$ satisfies the conditions of Theorem 13.6.

## 13.4 LSD for Stationary Gaussian Random Fields

In this section we survey several old and new results for stationary Gaussian random fields that could be combined with our universality results in order to decide that the LSD exists and to characterize it. Relevant to this part is the notion of spectral density. We consider a centered stationary Gaussian random field $(Y_{ij})_{(i,j)\in\mathbb{Z}^2}$, meaning that for any $(i, j) \in \mathbb{Z}^2$, $\mathbb{E}(Y_{ij}) = 0$ and

$$\mathrm{cov}(Y_{u,v}, Y_{k+u,\ell+v}) = \mathrm{cov}(Y_{0,0}, Y_{k,\ell}) = \gamma_{k,\ell}\,,$$

for any integers $u, v, k, \ell$. According to the Bochner-Herglotz representation (see for instance Theorem 1.7.4 in Sasvári [40]), since the covariance function is positive definite, there exists a unique spectral measure such that

$$\mathrm{cov}(Y_{0,0}, Y_{k,\ell}) = \int_{[0,1]^2} \mathrm{e}^{2\pi \mathrm{i}(ku+\ell v)} F(\mathrm{d}u, \mathrm{d}v), \quad \text{for all } k, \ell \in \mathbb{Z}\,.$$

If $F$ is absolutely continuous with respect to the Lebesgue measure $\lambda \otimes \lambda$, we have

$$\gamma_{k,\ell} := \mathrm{cov}(Y_{0,0}, Y_{k,\ell}) = \int_{[0,1]^2} \mathrm{e}^{2\pi \mathrm{i}(ku+\ell v)} f(u, v)\mathrm{d}u\mathrm{d}v, \quad \text{for all } k, \ell \in \mathbb{Z}\,. \tag{13.22}$$

Khorunzhy and Pastur [23] and Boutet de Monvel and Khorunzhy [7] treated a class of Gaussian fields with absolutely summable covariances,

$$\sum_{k,\ell\in\mathbb{Z}} |\gamma_{k,\ell}| < \infty\,, \tag{13.23}$$

and a certain symmetry condition. They described the limiting spectral distribution via an equation satisfied by the Stieltjes transform of the limiting distribution. Since the covariance structure is determined by the spectral density, this limiting spectral distribution can be expressed in terms of spectral density which generates the covariance structure. More precisely, if we consider the $n \times n$ random matrix $\mathbf{Y}_n$ with entries $Y_{k,\ell}$ and the symmetric matrix

$$\mathbb{W}_n = \frac{1}{\sqrt{2n}}(\mathbf{Y}_n + \mathbf{Y}_n^T)\,, \tag{13.24}$$

Theorem 2 in Khorunzhy and Pastur [23] (see also in Theorem 17.2.1. in Pastur and Shcherbina [33]) gives the following:

**Theorem 13.8** *Let $(Y_{k,\ell})_{(k,\ell)\in\mathbb{Z}^2}$ be a centered stationary Gaussian random field with spectral density $f(x, y)$. Denote $b(x, y) = 2^{-1}(f(x, y) + f(y, x))$. Assume* *that* (13.23) *holds. Let* $\mathbb{W}_n$ *be defined by* (13.24). *Then* $\mathbb{P}(d(F^{\mathbb{W}_n}, F) \to 0) = 1$, *where* $F$ *is a nonrandom distribution function whose Stieltjes transform* $S(z)$ *is uniquely determined by the relations:*

$$S(z) = \int_0^1 g(x, z)dx\,,\ z \in \mathbb{C}^+\,, \tag{13.25}$$

$$g(x, z) = -\Big(z + \int_0^1 b(x, y)g(y, z)dy\Big)^{-1}\,, \tag{13.26}$$

*where for any* $z \in \mathbb{C}^+$ *and any* $x \in [0, 1)$, $g(x, z)$ *is analytic in* $z$ *and*

$$\operatorname{Im} g(x, z) \cdot \operatorname{Im} z > 0\,,\ |g(x, z)| \le (\operatorname{Im} z)^{-1}\,,$$

*and is periodic and continuous in* $x$.

For the symmetric matrix $\mathbb{W}_n$ defined by (13.24) and constructed from a stationary Gaussian random field, Chakrabarty et al. [10] proved the existence of its limiting spectral distribution provided that the spectral density of the Gaussian process exists. Their result goes then beyond the condition (13.23) requiring that the covariances are absolutely summable. It was completed recently by Peligrad and Peligrad [34] who obtained a characterization of the limiting empirical spectral distribution for symmetric matrices with entries selected from a stationary Gaussian field under the sole condition that its spectral density exists. Their Theorem 2 is the following:

**Theorem 13.9** *Let* $(Y_{k,\ell})_{(k,\ell)\in\mathbb{Z}^2}$ *be a centered stationary Gaussian random field with spectral density* $f(x, y)$. *Let* $\mathbb{W}_n$ *be defined by* (13.24). *Then,* $\mathbb{P}(d(F^{\mathbb{W}_n}, F) \to 0) = 1$, *where the Stieltjes transform* $S(z)$ *of* $F$ *is uniquely defined by the relation (13.25) where for almost all* $x$, $g(x, z)$ *is a solution of Eq. (13.26).*

If the spectral density has the structure $f(x, y) = u(x)u(y)$, Eq. (13.25) simplifies as

$$S(z) = -\frac{1}{z}(1 + v^2(z))\,,\ z \in \mathbb{C}^+\,,$$

where $v(z)$ is solution of the equation

$$v(z) = -\int_0^1 \frac{u(y)\mathrm{d}y}{z + u(y)v(z)}, z \in \mathbb{C}^+\,,$$

with $v(z)$ analytic, $\operatorname{Im} v(z) > 0$ and $|v(z)| \le (\operatorname{Im} z)^{-1}\|Y_{0,0}\|_2$ (see the proof of Remark 3 in Peligrad–Peligrad [34]).

In particular, if the random field is an array of i.i.d. random variables with zero mean and variance $\sigma^2$, then $u(x)$ is constant and $S(z)$ satisfies Eq. (13.4).

The following new result gives the existence of LSD for large covariance matrices associated with a stationary Gaussian random field. Its proof is based on the method of proof in Chakrabarty et al. [10].

**Proposition 13.10** *Let* $(Y_{ij})_{(i,j)\in\mathbb{Z}^2}$ *be a stationary real-valued Gaussian process with mean zero. Assume that this process has a spectral density on* $[0,1]^2$ *denoted by* $f$*. Let* $N$ *and* $p$ *be two positive integers and consider* $\Gamma_{N,p}$ *the* $N \times p$ *matrix defined by* $\Gamma_{N,p} = \big(Y_{ij}\big)_{1\le i\le N, 1\le j\le p}$*. Let also* $\mathbb{G}_N = \frac{1}{N}\Gamma_{N,p}\Gamma_{N,p}^T$*. Then, when* $p/N \to c \in (0,\infty)$*, there exists a deterministic probability measure* $\mu_f$ *determined solely by* $c$ *and the spectral density* $f$*, and such that the spectral empirical measure* $\mu_{\mathbb{G}_N}$ *converges weakly in probability to* $\mu_f$*.*

For the case where the covariances are absolutely summable, we cite the following result which is Theorem 2.1 in Boutet de Monvel et al. [8]. It allows to characterize the LSD $\mu_f$ of $\mathbb{G}_N$ via an equation satisfied by its Stieltjes transform.

**Theorem 13.11** *Assume that the assumptions of Proposition 13.10 and that condition (13.23) hold. Then, when* $p/N \to c \in (0,\infty)$*,* $\mathbb{P}(d(F^{\mathbb{G}_N}, F) \to 0) = 1$ *where* $F$ *is a nonrandom distribution function whose Stieltjes transform* $S(z)$*,* $z \in \mathbb{C}^+$ *is uniquely defined by the relations:*

$$S(z) = \int_0^1 h(x,z)dx\,,$$

*where* $h(x,z)$ *is a solution of the equation*

$$h(x,z) = \Big(-z + c\int_0^1 \frac{f(x,s)}{1+\int_0^1 f(u,s)h(u,z)du}ds\Big)^{-1},$$

*with* $f(x,y)$ *the spectral density given in (13.22).*

When we assume that the entries of $\Gamma_{N,p}$ is a sequence of i.i.d. random variables with mean zero and variance $\sigma^2$, then $S(z)$ satisfies Eq. (13.18) of the Marchenko-Pastur distribution. In view of Proposition 13.10 and of Theorem 13.11, it is still an open question if, without imposing the summability condition (13.23) on the covariances, one could still characterize the LSD of $\mathbb{G}_N$.

## 13.5 Proofs

The notation $V_n^1 = \{(i,j); i \ge j \text{ with } i \text{ and } j \text{ in } \{1,\dots,n\}\}$ will be often used along the proofs.

### 13.5.1 *Proof of Theorem 13.2*

The proof is based on a Bernstein-type blocking procedure for random fields and the Lindeberg's method. The blocking argument, originally introduced by Bernstein [5] in order to prove an extension of the central limit theorem to r.v.'s satisfying dependent conditions, consists of making "big blocks" interlaced by "small blocks" which have a negligible behavior compared to the one of the "big blocks". In the context of random fields, this blocking argument can also be used (see for instance Tone [44], where the asymptotic normality of the normalized partial sum of a Hilbert-space valued stationary and mixing random field is proved with the help of a blocking procedure). In our context, the "big" blocks, called $\mathbf{B}_{i,j}$ in the figure (13.28) below, are of size $p$ (with $p$ such that $p/n \to 0$) and the "small" blocks will consist of bands of width $K$ with entries which are zero and with $K$ negligible with respect to $p$. As we shall see below, this blocking procedure can be efficiently done because, roughly speaking, the limiting spectral density distribution is not affected by changing a number of $o(n^2)$ variables.

Now the Lindeberg's method will consist of replacing one by one each of the "big" blocks with blocks of the same size but whose entries are those of a Gaussian random field having the same covariance structure as the initial process.

The blocking procedure combined with the Lindeberg's method does not seem very classical in the context of random matrices. It has been however recently used in Banna et al. [3] and in Merlevède and Peligrad [27], but in the context where the entries of the matrices are functions of an i.i.d. random field in the first mentioned paper, or in the context where the rows or the columns of the matrix are independent, in the second one. These conditions are not assumed in the context of the present paper. This makes the situation more delicate. Indeed, concentration inequalities for the Stieltjes transform around its mean are not available, hence we cannot restrict the study to the difference between the expectations of the two Stieltjes transforms. However, as we shall see, this issue can be bypassed by approximating the random matrix with "big" blocks $\mathbb{B}(\mathbf{X}_n)$ defined in (13.28), by another one where the "big" blocks will have a certain martingale difference property. Hence, in particular, they are uncorrelated. This new uncorrelated block matrix will be called $\mathbb{B}(\mathbf{X}_n')$ in the proof below. A similar treatment will be done to the matrices with the Gaussian field entries, having a suitable covariance structure.

We turn now to the details of the proof of Theorem 13.2, and first, to our blocking procedure, which involves several steps. We then start by some preliminary considerations.

Let $(K)$, $(c_K)$ and $(p_K)$ be sequences of integers converging to $\infty$ such that $p_K = c_K K$. Assume that

$$c_K^2 K^2 \sup_{(i,j)\neq(a,b); i\geq j, a\geq b} \|\mathbb{E}(X_{ij}X_{ab}|\mathcal{F}^K_{i\wedge a, j\wedge b}) - \mathbb{E}(X_{ij}X_{ab})\|_1 \to 0 \text{ as } K \to \infty. \tag{13.27}$$

This selection of $(c_K)$ is possible by (13.7). In what follows, to soothe the notation, we suppress the subscript $K$ to denote the sequences $(c_K)$ and $(p_K)$.

We describe now the blocking procedure which is based on the known fact that the limiting spectral density distribution is not affected by changing a number of $o(n^2)$ variables. We first notice that, without restricting the generality we may and do assume that $n = q(p+K)+p$ where $q := q_n$ is a sequence of positive integers depending on $n$, $K$ and $p$. Indeed, if $(n-p)/(p+K)$ is not an integer, we set $n' = q(p+K)+p$ where $q = [(n-p)/(p+K)]$ and we notice that, by the Cauchy's interlacing law (see for instance Relation (2.96) in Tao's monograph [43] for more details),

$$|S^{\mathbb{X}_n} - S^{\mathbb{X}_{n'}}| \ll \frac{n-n'}{n} \ll \frac{p+K}{n} \to 0 \text{ as } n \to \infty.$$

Therefore, we shall assume, from now on, that $n = q(p+K)+p$.

To introduce the big blocks, for a given symmetric matrix $\mathbf{Z}_n = \{Z_{ij}\}_{i,j=1}^n$ we shall associate the following checkerboard structure

$$\mathbb{B}(\mathbf{Z}_n) = \frac{1}{\sqrt{n}} \begin{pmatrix} 0_{p,p} & & & & & & & \\ 0_{K,p} & 0_{K,K} & & & & & & \\ \mathbf{B}_{1,1} & 0_{p,K} & 0_{p,p} & & & & & \\ 0_{K,p} & 0_{K,K} & 0_{K,p} & 0_{K,K} & & & & \\ \mathbf{B}_{2,1} & 0_{p,K} & \mathbf{B}_{2,2} & 0_{p,K} & 0_{p,p} & & & \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \mathbf{B}_{q-1,1} & 0_{p,K} & \mathbf{B}_{q-1,2} & 0_{p,K} & \mathbf{B}_{q-1,3} & \cdots & 0_{p,p} & \\ 0_{K,p} & 0_{K,K} & 0_{K,p} & 0_{K,K} & 0_{K,p} & \cdots & 0_{K,p} & 0_{K,K} \\ \mathbf{B}_{q,1} & 0_{p,K} & \mathbf{B}_{q,2} & 0_{p,K} & \mathbf{B}_{q,3} & \cdots & \mathbf{B}_{q,q} & 0_{p,K} \; 0_{p,p} \end{pmatrix}. \tag{13.28}$$

The rest of the matrix is completed by symmetry. Here each $\mathbf{B}_{k,\ell}$ denotes a block matrix $p \times p$ indexed by a set of indexes in $\mathcal{E}_{k,\ell}$ defined below, and whose entries are identical to the matrix $\mathbf{Z}_n$. To be more precise, we define

$$\mathcal{E}_{k,\ell} = \{(u,v) \in E_k \times E_{\ell-1}\} \text{ where } E_\ell = \big[\ell(p+K)+1\,,\ \ell(p+K)+p\big] \cap \mathbb{N}\,. \tag{13.29}$$

We shall order the blocks in the lexicographic order starting with the top of the matrix. To soothe the notations all along the paper, we shall use the following convention: for any $k = 1, \ldots, q_n$ and any $\ell = 1, \ldots, k$,

$$\mathbf{B}_u = \mathbf{B}_{k,\ell} \text{ and } \mathbf{I}_u = \mathcal{E}_{k,\ell} \text{ where } u = k(k-1)/2 + \ell. \tag{13.30}$$

To avoid confusion, when block matrices are constructed from different symmetric matrices, we shall also use the notation $\mathbf{B}_i = \mathbf{B}_i(\mathbf{Z}_n)$ to identify the variables in the block matrix. Note that the "big" blocks $\mathbf{B}_i$ are separated by bands of $K$ rows and $K$ columns. Variables in two different blocks are separated by at least either $K$ rows or $K$ columns.

Using Lemma 13.14, Theorem A.43 in Bai and Silverstein [1] and taking into account Condition 13.1, straightforward computations lead to

$$\mathbb{E}|S^{\mathbb{X}_n} - S^{\mathbb{B}(\mathbf{X}_n)}|^2 \to 0 \text{ as } n \to \infty.$$

Similarly, we define $\mathbb{B}(\mathbf{Y}_n)$, and one can prove that

$$\mathbb{E}|S^{\mathbb{Y}_n} - S^{\mathbb{B}(\mathbf{Y}_n)}|^2 \to 0 \text{ as } n \to \infty.$$

We introduce a filtration

$$\mathcal{B}_u = \sigma(\mathbf{B}_1(\mathbf{X}_n), \mathbf{B}_2(\mathbf{X}_n), \ldots, \mathbf{B}_u(\mathbf{X}_n)) \text{ for } u \geq 1 \text{ and } \mathcal{B}_0 = \{\emptyset, \Omega\}. \tag{13.31}$$

To introduce martingale structure, for $1 \leq u \leq q(q+1)/2$ and $j \in \mathbf{I}_u$ define the variables

$$X'_{\mathbf{j}} = X_{\mathbf{j}} - \mathbb{E}(X_{\mathbf{j}}|\mathcal{B}_{u-1}).$$

Then we define a new block matrix, say $\mathbb{B}(\mathbf{X}'_n)$, with blocks $\mathbf{B}'_i = \mathbf{B}_i(\mathbf{X}'_n)$ having a similar structure as $\mathbb{B}(\mathbf{X}_n)$ where the entries in these big blocks are $X'_{\mathbf{j}}$, $j \in \mathbf{I}_u$. Note that by Lemma 13.14

$$\mathbb{E}|S^{\mathbb{B}(\mathbf{X}_n)} - S^{\mathbb{B}(\mathbf{X}'_n)}|^2 \ll \frac{1}{n^2} \sum_{u \geq 1} \sum_{j \in \mathbf{I}_u} \mathbb{E}|\mathbb{E}(X_{\mathbf{j}}|\mathcal{B}_{u-1})|^2 \leq \sup_{i \geq j} \mathbb{E}|\mathbb{E}(X_{ij}|\mathcal{F}^K_{ij})|^2,$$

which converges to 0 uniformly in $n$ when $K \to \infty$ by (13.6). Here and in the sequel we shall keep in mind that the range for the index $u$ is from 1 to $q(q+1)/2$. For simplicity, we shall denote the sum from $u = 1$ to $u = q(q+1)/2$ by a sum over $u \geq 1$.

We proceed similarly for the matrix $\mathbb{B}(\mathbf{Y}_n)$. We introduce the filtration

$$\mathcal{H}_u = \sigma(\mathbf{B}_1(\mathbf{Y}_n), \mathbf{B}_2(\mathbf{Y}_n), \ldots, \mathbf{B}_u(\mathbf{Y}_n)) \text{ for } u \geq 1 \text{ and } \mathcal{H}_0 = \{\emptyset, \Omega\}, \tag{13.32}$$

and for any $j \in \mathbf{I}_u$ define the variables

$$Y'_{\mathbf{j}} = Y_{\mathbf{j}} - \mathbb{E}(Y_{\mathbf{j}}|\mathcal{H}_{u-1}).$$

Notice that $(Y'_{\mathbf{j}}, 1 \leq u \leq q(q+1)/2, \mathbf{j} \in I_u)$ is also a Gaussian vector. In addition, by using the properties of conditional expectation we can easily notice that the random vectors $(Y'_{\mathbf{j}}, \mathbf{j} \in I_u)_u$ are orthogonal. Therefore $(Y'_{\mathbf{j}}, \mathbf{j} \in I_u)_u$ are mutually independent. We shall also prove that for $\mathbf{j} \in I_u$

$$\|\mathbb{E}(Y_{\mathbf{j}}|\mathcal{H}_{u-1})\|_2 \leq \|\mathbb{E}(X_{\mathbf{j}}|\mathcal{B}_{u-1})\|_2. \tag{13.33}$$

To prove the inequality above, it suffices to notice the following facts. Let

$$\mathcal{V}_u = \overline{\text{span}}(1, (Y_{\mathbf{j}}\,,\ 1 \leq v \leq u\,,\ \mathbf{j} \in \mathbf{I}_v))$$

and

$$\mathcal{V}_u^* = \overline{\text{span}}(1, (X_{\mathbf{j}}\,,\ 1 \leq v \leq u\,,\ \mathbf{j} \in \mathbf{I}_v)),$$

where the closure is taken in $\mathbb{L}^2$. Denote by $\Pi_{\mathcal{V}_u}(\cdot)$ the orthogonal projection on $\mathcal{V}_u$ and by $\Pi_{\mathcal{V}_u^*}(\cdot)$ the orthogonal projection on $\mathcal{V}_u^*$. Since $(Y'_{\mathbf{j}}, 1 \leq u \leq q(q+1)/2, \mathbf{j} \in I_u)$ is a Gaussian process,

$$\mathbb{E}(Y_{\mathbf{j}}|\mathcal{H}_{u-1}) = \Pi_{\mathcal{V}_{u-1}}\big(Y_{\mathbf{j}}\big) \text{ a.s. and in } \mathbb{L}^2\,.$$

Since $(Y_{k\ell})_{1\leq \ell\leq k\leq n}$ has the same covariance structure as $(X_{k\ell})_{1\leq \ell\leq k\leq n}$, we observe that

$$\|\Pi_{\mathcal{V}_{u-1}}\big(Y_{\mathbf{j}}\big)\|_2 = \|\Pi_{\mathcal{V}_{u-1}^*}\big(X_{\mathbf{j}}\big)\|_2\,.$$

But,

$$\|\Pi_{\mathcal{V}_{u-1}^*}\big(X_{\mathbf{j}}\big)\|_2 \leq \|\mathbb{E}(X_{\mathbf{j}}|\mathcal{B}_{u-1})\|_2\,,$$

which proves (13.33). Then we define a new block matrix, say $\mathbb{B}(\mathbf{Y}'_n)$, with blocks $\mathbf{\Gamma}'_i = \mathbf{B}_i(\mathbf{Y}'_n)$ having a similar structure as $\mathbb{B}(\mathbf{Y}_n)$ where the entries in these big blocks are $Y'_{\mathbf{j}}$. Therefore, by Lemma 13.14 and (13.33),

$$\begin{aligned}\mathbb{E}|S^{\mathbb{B}(\mathbf{Y}_n)} - S^{\mathbb{B}(\mathbf{Y}'_n)}|^2 &\ll \frac{1}{n^2}\sum_{u\geq 1}\sum_{j\in\mathbf{I}_u}\mathbb{E}|\mathbb{E}(Y_{\mathbf{j}}|\mathcal{H}_{u-1})|^2 \leq \frac{1}{n^2}\sum_{u\geq 1}\sum_{\mathbf{j}\in\mathbf{I}_u}\mathbb{E}|\mathbb{E}(X_{\mathbf{j}}|\mathcal{B}_{u-1})|^2\\ &\leq \sup_{i\geq j}\mathbb{E}|\mathbb{E}(X_{ij}|\mathcal{F}_{ij}^K)|^2,\end{aligned}$$

which converges to 0 as $K \to \infty$ by (13.6), uniformly in $n$. The proof is reduced to showing that

$$\mathbb{E}|S^{\mathbb{B}(\mathbf{X}'_n)} - S^{\mathbb{B}(\mathbf{Y}'_n)}| \to 0 \text{ as } n \to \infty, \tag{13.34}$$

which we shall achieve at the end of several steps.

We write $S^{\mathbb{B}(\mathbf{X}'_n)}$ and $S^{\mathbb{B}(\mathbf{Y}'_n)}$ as function of the entries. So

$$S^{\mathbb{B}(\mathbf{X}'_n)} = s(\mathbf{B}'_1, \ldots, \mathbf{B}'_{q(q+1)/2}) \text{ and } S^{\mathbb{B}(\mathbf{Y}'_n)} = s(\mathbf{\Gamma}'_1, \ldots, \mathbf{\Gamma}'_{q(q+1)/2}) . \tag{13.35}$$

We note that in our proofs the order in which we treat the blocks is critically important for using the power of the martingale structure, but the location of a variable in a block is not going to matter. With our functional notation we use the following decomposition:

$$S^{\mathbb{B}(\mathbf{X}'_n)} - S^{\mathbb{B}(\mathbf{Y}'_n)} = s(\mathbf{B}'_1, \ldots, \mathbf{B}'_{q(q+1)/2}) - s(\mathbf{\Gamma}'_1, \ldots, \mathbf{\Gamma}'_{q(q+1)/2}) \tag{13.36}$$
$$= \sum_{u=1}^{q(q+1)/2} \Big( s(\mathbf{B}'_1, \ldots, \mathbf{B}'_u, \mathbf{\Gamma}'_{u+1}, \ldots, \Gamma'_{q(q+1)/2}) - s(\mathbf{B}'_1, \ldots, \mathbf{B}'_{u-1}, \mathbf{\Gamma}'_u, \ldots, \Gamma'_{q(q+1)/2}) \Big) .$$

We also denote

$$\mathbf{C}_u = (\mathbf{B}'_1, \ldots, \mathbf{B}'_{u-1}, \mathbf{0}_u, \Gamma'_{u+1}, \ldots, \Gamma'_{q(q+1)/2}),$$

where $\mathbf{0}_u$ is a null vector with $p^2$ entries 0. We shall use the Taylor expansion in Lemma 13.12, applied for a fixed index $u$, to the function

$$s(\mathbf{B}'_1, \ldots, \mathbf{B}'_{u-1}, \mathbf{B}_u(\mathbf{X}'_n), \mathbf{\Gamma}'_{u+1}, \ldots, \mathbf{\Gamma}'_{q(q+1)/2}),$$

where $s$ is defined in (13.35). We can view this function simply as a function of a vector $\mathbf{x} = (x_{\mathbf{i}},\ u \in \{1, \ldots, q(q+1)/2\}$ and $\mathbf{i} \in \mathbf{I}_u)$. By using (13.36), Lemma 13.12 with $A = 4\varepsilon n^{1/2}$ and (13.60), we obtain

$$S^{\mathbb{B}(\mathbf{X}'_n)} - S^{\mathbb{B}(\mathbf{Y}'_n)} = R'_1 + R'_2 + R'_3, \tag{13.37}$$

where

$$R'_1 = \sum_{u \geq 1} \sum_{\mathbf{j} \in \mathbf{I}_u} (X'_{\mathbf{j}} - Y'_{\mathbf{j}}) \partial_{\mathbf{j}} s(\mathbf{C}_u),$$

$$R'_2 = \frac{1}{2} \sum_{u \geq 1} \Big( \Big( \sum_{\mathbf{j} \in \mathbf{I}_u} X'_{\mathbf{j}} \partial_{\mathbf{j}} \Big)^2 - \Big( \sum_{\mathbf{j} \in \mathbf{I}_u} Y'_{\mathbf{j}} \partial_{\mathbf{j}} \Big)^2 \Big) s(\mathbf{C}_u)$$

and

$$|R'_3| \leq \sum_{u \geq 1} |R_{u3}| + \sum_{u \geq 1} |R'_{u3}|,$$

with

$$|R_{u3}| \ll \frac{1}{n^2} p^2 \sum_{\mathbf{j} \in \mathbf{I}_u} (X'_{\mathbf{j}})^2 I(|X'_{\mathbf{j}}| > 4\varepsilon n^{1/2}) + \varepsilon n^{1/2} \frac{1}{n^{5/2}} p^4 \sum_{\mathbf{j} \in \mathbf{I}_u} (X'_{\mathbf{j}})^2$$

and

$$|R'_{u3}| \ll \frac{1}{n^2} p^2 \sum_{\mathbf{j}\in \mathbf{I}_u} (Y'_{\mathbf{j}})^2 I(|Y'_{\mathbf{j}}| > 4\varepsilon n^{1/2}) + \varepsilon n^{1/2} \frac{1}{n^{5/2}} p^4 \sum_{\mathbf{j}\in \mathbf{I}_u} (Y'_{\mathbf{j}})^2 .$$

We treat first the term $|R'_3|$. By taking the expected value and considering Condition 13.1, we obtain

$$\sum_{u\geq 1} \mathbb{E}|R_{u3}| \ll p^2 \frac{1}{n^2} \sum_{(i,j)\in V_n^1} \mathbb{E}(|X'_{ij}|^2 I(|X'_{ij}| > 4\varepsilon n^{1/2})) + \varepsilon p^4.$$

Notice now the following fact: If $U$ is a real-valued random variable and $\mathcal{F}$ is a sigma-field, then setting $V = U - \mathbb{E}(U|\mathcal{F})$, the following inequality holds: for any $m \geq 1$ and any $a > 0$,

$$\mathbb{E}\big(|V|^m I(|V| > 4a)\big) \leq 3 \times 2^m \mathbb{E}\big(|U|^m I(|U| > a)\big) . \tag{13.38}$$

This implies that

$$\mathbb{E}(|X'_{ij}|^2 I(|X'_{ij}| > 4\varepsilon n^{1/2})) \ll \mathbb{E}(X_{ij}^2 I(|X_{ij}| > \varepsilon n^{1/2})).$$

Therefore

$$\begin{aligned} \sum_{u\geq 1} \mathbb{E}|R_{u3}| &\ll p^2 \frac{1}{n^2} \sum_{(i,j)\in V_n^1} \mathbb{E}(X_{ij}^2 I(|X_{ij}| > \varepsilon n^{1/2})) + \varepsilon p^4 \\ &= p^2 L_n(\varepsilon) + \varepsilon p^4. \end{aligned}$$

We let first $n \to \infty$ and take into account Condition 13.1 and then we let $\varepsilon \to 0$. It follows that

$$\lim_{n\to\infty} \sum_{u\geq 1} \mathbb{E}|R_{u3}| = 0. \tag{13.39}$$

We handle now the quantity $\sum_{u\geq 1} \mathbb{E}|R'_{u3}|$. Taking into account that $\mathbb{E}(Y_{\mathbf{u}}^2) = \mathbb{E}(X_{\mathbf{u}}^2)$, Condition 13.1 and inequality (13.38), note first that

$$\sum_{u\geq 1} \mathbb{E}|R'_{u3}| \ll p^2 \frac{1}{n^2} \sum_{(i,j)\in V_n^1} \mathbb{E}(Y_{ij}^2 I(|Y_{ij}| > \varepsilon n^{1/2})) + \varepsilon p^4.$$

To treat the first term in the right-hand side, some computations are needed. Note first that if $N$ is a centered Gaussian random variable with variance $\sigma^2$,

$$\mathbb{E}(N^2 I(|N| > \varepsilon n^{1/2})) = \frac{2\sigma}{\sqrt{2\pi}} \varepsilon\sqrt{n} e^{-\varepsilon^2 n/(2\sigma^2)} + \sigma^2 \mathbb{P}(|N| > \varepsilon\sqrt{n})\,. \tag{13.40}$$

Let now $\sigma_{ij}^2 = \mathbb{E}(X_{ij}^2)$. For any $\eta > 0$, we then have

$$\sigma_{ij}^2 \mathbb{P}(|Y_{ij}| > \varepsilon\sqrt{n}) \le \eta^2\varepsilon^2 n \mathbb{P}(|Y_{ij}| > \varepsilon\sqrt{n}) + \mathbb{E}(X_{ij}^2 I(|X_{ij}| > \eta\varepsilon n^{1/2}))$$
$$\le \eta^2 \mathbb{E}(Y_{ij}^2) + \mathbb{E}(X_{ij}^2 I(|X_{ij}| > \eta\varepsilon n^{1/2})) = \eta^2 \mathbb{E}(X_{ij}^2) + \mathbb{E}(X_{ij}^2 I(|X_{ij}| > \eta\varepsilon n^{1/2})).$$

On another hand, let $A$ be a positive real. Observe that, for any $\eta > 0$,

$$\begin{aligned}\sigma_{ij}\varepsilon\sqrt{n} e^{-\varepsilon^2 n/(2\sigma_{ij}^2)} I(\sigma_{ij} > \varepsilon\sqrt{n}/A) &\le A\sigma_{ij}^2 I(\sigma_{ij} > \varepsilon\sqrt{n}/A) \\ &\le A\mathbb{E}(X_{ij}^2 I(|X_{ij}| > \eta n^{1/2})) + An\eta^2 I(\sigma_{ij} > \varepsilon\sqrt{n}/A) \\ &\le A\mathbb{E}(X_{ij}^2 I(|X_{ij}| > \eta n^{1/2})) + \eta^2 A^3 \sigma_{ij}^2/\varepsilon^2.\end{aligned}$$

Moreover

$$\sigma_{ij}\varepsilon\sqrt{n} e^{-\varepsilon^2 n/(2\sigma_{ij}^2)} I(\sigma_{ij} \le \varepsilon\sqrt{n}/A) = \sqrt{2}\sigma_{ij}^2 \frac{\varepsilon\sqrt{n}}{\sqrt{2}\sigma_{ij}} e^{-\varepsilon^2 n/(2\sigma_{ij}^2)} I(\sigma_{ij} \le \varepsilon\sqrt{n}/A) \le 2\sigma_{ij}^2 e^{-A^2/4}.$$

So, overall, taking into account the above considerations and (13.40), it follows that, for any $\varepsilon > 0$, any $\eta > 0$ and any positive real $A$,

$$\begin{aligned}\mathbb{E}(Y_{ij}^2 I(|Y_{ij}| > \varepsilon n^{1/2})) &\ll \eta^2 \mathbb{E}(X_{ij}^2) + \mathbb{E}(X_{ij}^2 I(|X_{ij}| > \eta\varepsilon n^{1/2})) \\ &\quad + A\mathbb{E}(X_{ij}^2 I(|X_{ij}| > \eta n^{1/2})) + \eta^2 A^3 \sigma_{ij}^2/\varepsilon^2 + \mathbb{E}(X_{ij}^2) e^{-A^2/4}.\end{aligned}$$

Hence, taking into account Condition 13.1, it follows that for any $\varepsilon > 0$, any $\eta > 0$ and any $A > 0$,

$$\sum_{u\ge 1} \mathbb{E}|R'_{u3}| \ll \varepsilon p^4 + p^2\eta^2 + p^2 L_n(\eta\varepsilon) + Ap^2 L_n(\eta) + p^2\eta^2 A^3/\varepsilon^2 + p^2 e^{-A^2/4}.$$

Letting $n \to \infty$, then $\eta \to 0$ and finally $A \to \infty$, it follows that

$$\limsup_{n\to\infty} \sum_{u\ge 1} \mathbb{E}|R'_{u3}| \ll \varepsilon p^4.$$

Letting then $\varepsilon \to 0$ and taking into account (13.39), it follows that $\mathbb{E}|R'_3| \to 0$, as $n \to \infty$.

We treat now the term $R'_1$, and we shall compute $\mathbb{E}|R'_1|^2$. Let

$$D_u = \sum_{\mathbf{j}\in\mathbf{I}_u} X'_{\mathbf{j}}\partial_{\mathbf{j}} s(\mathbf{C}_u) \text{ and } \widetilde{D}_u = \sum_{\mathbf{j}\in\mathbf{I}_u} Y'_{\mathbf{j}}\partial_{\mathbf{j}} s(\mathbf{C}_u),$$

and note that

$$\mathbb{E}|R'_1|^2 \le 2\mathbb{E}\Big|\sum_{u\ge1} D_u\Big|^2 + 2\mathbb{E}\Big|\sum_{u\ge1} \widetilde{D}_u\Big|^2.$$

By definition of the $X'_{\mathbf{j}}$ for $\mathbf{j}$ in $\mathbf{I}_u$, the random variables $(D_u)_{u\ge1}$ are orthogonal. Moreover, since the random vectors $(Y'_{\mathbf{j}}, \mathbf{j} \in I_u)_u$ are mutually independent, the variables $(\widetilde{D}_u)_{u\ge1}$ are also orthogonal. Hence we get

$$\mathbb{E}|R'_1|^2 \le 2\sum_{u\ge1}\mathbb{E}|D_u|^2 + 2\sum_{u\ge1}\mathbb{E}|\widetilde{D}_u|^2.$$

Therefore, by using Cauchy–Schwarz's inequality, taking into account (13.60), the fact that $\mathbb{E}(Y_{\mathbf{u}}^2) = \mathbb{E}(X_{\mathbf{u}}^2)$ and Condition 13.1, it follows that

$$\mathbb{E}|R'_1|^2 \ll \frac{p^2}{n^3}\sum_{(i,j)\in V_n^1}\mathbb{E}(X_{ij}^2) \ll \frac{p^2}{n},$$

which converges to 0 when $n \to \infty$.

Now we treat the term $R'_2 = 2^{-1}\sum_{u\ge1} R_{u2}$ where

$$R_{u2} = \Big(\Big(\sum_{\mathbf{j}\in\mathbf{I}_u} X'_{\mathbf{j}}\partial_{\mathbf{j}}\Big)^2 - \Big(\sum_{\mathbf{j}\in\mathbf{I}_u} Y'_{\mathbf{j}}\partial_{\mathbf{j}}\Big)^2\Big)s(\mathbf{C}_u).$$

We write $R_{u2}$ as a sum of differences of the type $\big(X'_{\mathbf{j}}X'_{\mathbf{i}} - Y'_{\mathbf{j}}Y'_{\mathbf{i}}\big)\partial_{\mathbf{j}}\partial_{\mathbf{i}}s(\mathbf{C}_u)$ were $\mathbf{i}, \mathbf{j} \in \mathbf{I}_u$. To introduce martingale structure we add and substract some terms. Hence we write

$$\begin{aligned}\big(X'_{\mathbf{j}}X'_{\mathbf{i}} - Y'_{\mathbf{j}}Y'_{\mathbf{i}}\big)\partial_{\mathbf{j}}\partial_{\mathbf{i}}s(\mathbf{C}_u) &= \big(X'_{\mathbf{j}}X'_{\mathbf{i}} - \mathbb{E}(X'_{\mathbf{j}}X'_{\mathbf{i}}|\mathcal{B}_{u-1})\big)\partial_{\mathbf{j}}\partial_{\mathbf{i}}s(\mathbf{C}_u)\\ &\quad+\big(\mathbb{E}(X'_{\mathbf{j}}X'_{\mathbf{i}}|\mathcal{B}_{u-1}) - \mathbb{E}(X_{\mathbf{j}}X_{\mathbf{i}})\big)\partial_{\mathbf{j}}\partial_{\mathbf{i}}s(\mathbf{C}_u) + \big(\mathbb{E}(X_{\mathbf{j}}X_{\mathbf{i}}) - Y'_{\mathbf{j}}Y'_{\mathbf{i}}\big)\partial_{\mathbf{j}}\partial_{\mathbf{i}}s(\mathbf{C}_u)\\ &:= I^{(1)}_{u\mathbf{ij}} + I^{(2)}_{u\mathbf{ij}} + I^{(3)}_{u\mathbf{ij}}\,. \end{aligned} \tag{13.41}$$

Taking into account the properties of the conditional expectation, we obtain

$$\mathbb{E}(X'_{\mathbf{j}}X'_{\mathbf{i}}|\mathcal{B}_{u-1}) = \mathbb{E}(X_{\mathbf{j}}X_{\mathbf{i}}|\mathcal{B}_{u-1}) - \mathbb{E}(X_{\mathbf{j}}|\mathcal{B}_{u-1})\mathbb{E}(X_{\mathbf{i}}|\mathcal{B}_{u-1}).$$

Therefore

$$\Big|\sum_{\mathbf{i},\mathbf{j}\in\mathbf{I}_u} I^{(2)}_{u\mathbf{i}\mathbf{j}}\Big| \le \Big|\sum_{\mathbf{i},\mathbf{j}\in\mathbf{I}_u}\big(\mathbb{E}(X_{\mathbf{j}}X_{\mathbf{i}}|\mathcal{B}_{u-1})-\mathbb{E}(X_{\mathbf{j}}X_{\mathbf{i}})\big)\partial_{\mathbf{j}}\partial_{\mathbf{i}}s(\mathbf{C}_u)\Big|$$
$$+\Big|\sum_{\mathbf{i},\mathbf{j}\in\mathbf{I}_u}\mathbb{E}(X_{\mathbf{j}}|\mathcal{B}_{u-1})\mathbb{E}(X_{\mathbf{i}}|\mathcal{B}_{u-1})\partial_{\mathbf{j}}\partial_{\mathbf{i}}s(\mathbf{C}_u)\Big|$$
$$:= I_{1,u}+I_{2,u}\,. \tag{13.42}$$

Let us handle the term $I_{2,u}$. By Lemma 13.13,

$$I_{2,u}\le\frac{c_4}{n^2}\sum_{\mathbf{i}\in\mathbf{I}_u}|\mathbb{E}(X_{\mathbf{i}}|\mathcal{B}_{u-1})|^2\,,$$

and then

$$\sum_{u\ge1}\mathbb{E}(I_{2,u})\le\frac{c_4}{n^2}\sum_{u\ge1}\sum_{\mathbf{i}\in\mathbf{I}_u}\|\mathbb{E}(X_{\mathbf{i}}|\mathcal{B}_{u-1})\|_2^2\,. \tag{13.43}$$

Therefore, using the contractivity of conditional expectation, we get

$$\sum_{u=1}^{q(q+1)/2}\mathbb{E}(I_{2,u})\ll\frac{1}{n^2}\Big(\frac{n}{p}\Big)^2p^2\sup_{i\ge j}\mathbb{E}|\mathbb{E}(X_{ij}|\mathcal{F}^K_{ij})|^2\ll\sup_{i\ge j}\mathbb{E}|\mathbb{E}(X_{ij}|\mathcal{F}^K_{ij})|^2\,.$$

Hence, by condition (13.6),

$$\lim_{K\to\infty}\limsup_{n\to\infty}\sum_{u=1}^{q(q+1)/2}\mathbb{E}(I_{2,u})=0.$$

We handle now the term $I_{1,u}$ in (13.42). Using (13.60), we first write

$$\mathbb{E}(I_{1,u})\ll\frac{1}{n^2}\sum_{\mathbf{i},\mathbf{j}\in\mathbf{I}_u}\|\mathbb{E}(X_{\mathbf{j}}X_{\mathbf{i}}|\mathcal{B}_{u-1})-\mathbb{E}(X_{\mathbf{j}}X_{\mathbf{i}})\|_1.$$

By using the contractivity of conditional expectation, we then get

$$\sum_{u\ge1}\mathbb{E}(I_{1,u})\ll\frac{1}{n^2}\sum_{u\ge1}\sum_{\mathbf{i},\mathbf{j}\in\mathbf{I}_u}\mathbb{E}|\mathbb{E}(X_{\mathbf{j}}X_{\mathbf{i}}|\mathcal{B}_{u-1})-\mathbb{E}(X_{\mathbf{j}}X_{\mathbf{i}})| \tag{13.44}$$
$$\ll\frac{1}{n^2}\Big(\frac{n}{p}\Big)^2p^4\sup_{(i,j)\ne(a,b);i\ge j,a\ge b}\mathbb{E}|\mathbb{E}(X_{ij}X_{ab}|\mathcal{F}^K_{i\wedge a,j\wedge b})-\mathbb{E}(X_{ij}X_{ab})|$$

$$+\frac{1}{n^2}\Big(\frac{n}{p}\Big)^2 p^2 \sup_{i\geq j}\mathbb{E}|\mathbb{E}(X_{ij}^2|\mathcal{F}_{ij}^K)-\mathbb{E}(X_{ij}^2)|$$

$$\ll p^2 \sup_{(i,j)\neq(a,b);i\geq j,a\geq b}\mathbb{E}|\mathbb{E}(X_{ij}X_{ab}|\mathcal{F}^K_{i\wedge a,j\wedge b})-\mathbb{E}(X_{ij}X_{ab})|$$

$$+\sup_{i\geq j}\mathbb{E}|\mathbb{E}(X_{ij}^2|\mathcal{F}_{ij}^K)-\mathbb{E}(X_{ij}^2)|.$$

The first term converges to 0 by (13.27) and the second term is convergent to 0 by (13.8). Hence,

$$\lim_{K\to\infty}\limsup_{n\to\infty}\sum_{u\geq 1}\mathbb{E}(I_{1,u})=0.$$

Overall, starting from (13.42) and taking into account the above considerations, we get

$$\sum_{u\geq 1}\mathbb{E}\Big|\sum_{\mathbf{i}\in\mathbf{I}_u}\sum_{\mathbf{j}\in\mathbf{I}_u}I^{(2)}_{u\mathbf{ij}}\Big|=0.$$

We treat now the negligibility of the term $\sum_{u\geq 1}\sum_{\mathbf{i},\mathbf{j}\in\mathbf{I}_u}I^{(1)}_{u\mathbf{ij}}$ in the following way. First we truncate

$$\bar{X}'_{\mathbf{i}}=X'_{\mathbf{i}}I(|X'_{\mathbf{i}}|\leq 4n^{1/2})\text{ and }\tilde{X}'_{\mathbf{i}}=X'_{\mathbf{i}}I(|X'_{\mathbf{i}}|>4n^{1/2})$$

and write

$$X'_{\mathbf{j}}X'_{\mathbf{i}}-\mathbb{E}(X'_{\mathbf{j}}X'_{\mathbf{i}}|\mathcal{B}_{u-1})=X'_{\mathbf{j}}\bar{X}'_{\mathbf{i}}-\mathbb{E}(X'_{\mathbf{j}}\bar{X}'_{\mathbf{i}}|\mathcal{B}_{u-1})+X'_{\mathbf{j}}\tilde{X}'_{\mathbf{i}}-\mathbb{E}(X'_{\mathbf{j}}\tilde{X}'_{\mathbf{i}}|\mathcal{B}_{u-1}).$$

Therefore, by the triangle inequality, the Cauchy–Schwarz inequality, the Minkowski's inequalities and the properties of conditional expectation, we easily obtain

$$\mathbb{E}\Big|\sum_{u\geq 1}\sum_{\mathbf{i},\mathbf{j}\in\mathbf{I}_u}I^{(1)}_{u\mathbf{ij}}\Big|\ll\frac{1}{n^2}\sum_{u\geq 1}\sum_{\mathbf{i},\mathbf{j}\in\mathbf{I}_u}\|X'_{\mathbf{j}}\|_2\|\tilde{X}'_{\mathbf{i}}\|_2$$

$$+\Big\|\sum_{u\geq 1}\sum_{\mathbf{i},\mathbf{j}\in\mathbf{I}_u}[X'_{\mathbf{j}}\bar{X}'_{\mathbf{i}}-\mathbb{E}(X'_{\mathbf{j}}\bar{X}'_{\mathbf{i}}|\mathcal{B}_{u-1})]\partial_{\mathbf{j}}\partial_{\mathbf{i}}s(\mathbf{C}_u)\Big\|_2$$

$$:=A_n+\Big\|\sum_{u\geq 1}D'_u\Big\|_2.$$

By the fact that that the terms $D'_u$ are orthogonal, by (13.60), the level of truncation and Condition 13.1, we have

$$\mathbb{E}\Big|\sum_{u\geq 1} D'_u\Big|^2 \ll \frac{1}{n^4}\sum_{u\geq 1}\mathbb{E}\Big(\sum_{i,j\in\mathbf{I}_u} |X'_{\mathbf{j}}\bar{X}'_{\mathbf{i}} - \mathbb{E}(X'_{\mathbf{j}}\bar{X}'_{\mathbf{i}}|\mathcal{B}_{u-1})|\Big)^2$$

$$\ll \frac{1}{n^4}\sum_{u\geq 1}\Big(\sum_{i,j\in\mathbf{I}_u} \|X'_{\mathbf{j}}\bar{X}'_{\mathbf{i}}\|_2\Big)^2 \ll \frac{p^4}{n^4}\sum_{u\geq 1}\sum_{i,j\in\mathbf{I}_u} \|X'_{\mathbf{j}}\bar{X}'_{\mathbf{i}}\|_2^2$$

$$\ll \frac{p^6}{n^3}\sum_{u\geq 1}\sum_{j\in\mathbf{I}_u}\mathbb{E}(X_{\mathbf{j}}^2) \ll \frac{p^6}{n} \to 0 \text{ as } n\to\infty.$$

Also, by the Cauchy–Schwarz's inequality and Condition 13.1,

$$A_n = \frac{1}{n^2}\sum_{u\geq 1}\sum_{\mathbf{i},\mathbf{j}\in\mathbf{I}_u} \|X'_{\mathbf{j}}\|_2\|\tilde{X}'_{\mathbf{i}}\|_2 \leq p^2\Big(\frac{1}{n^2}\sum_{u\geq 1}\sum_{\mathbf{j}\in\mathbf{I}_u}\|X'_{\mathbf{j}}\|_2^2\Big)^{1/2}\Big(\frac{1}{n^2}\sum_{u\geq 1}\sum_{\mathbf{i}\in\mathbf{I}_u}\|\tilde{X}'_{\mathbf{i}}\|_2^2\Big)^{1/2}$$

$$\leq p^2\Big(\frac{1}{n^2}\sum_{\mathbf{i}\in V_n^1}\mathbb{E}(X'^2_{\mathbf{i}}I(|X'_{\mathbf{i}}| > 4n^{1/2}))\Big)^{1/2}.$$

Using (13.38), we derive that $A_n \ll p^2 L_n^{1/2}(1)$, which converges to 0 for any $p$ fixed as $n\to\infty$. Overall, it follows that

$$\mathbb{E}\Big|\sum_{u\geq 1}\sum_{\mathbf{i},\mathbf{j}\in\mathbf{I}_u} I^{(1)}_{u\mathbf{i}\mathbf{j}}\Big| \to 0, \text{ as } n\to\infty.$$

To end the proof, it remains only to treat the term containing the Gaussian random variables. With this aim, we write $I^{(3)}_{u\mathbf{i}\mathbf{j}} = A_{u\mathbf{i}\mathbf{j}} + B_{u\mathbf{i}\mathbf{j}}$, where

$$A_{u\mathbf{i}\mathbf{j}} := (\mathbb{E}(X_{\mathbf{j}}X_{\mathbf{i}}) - \mathbb{E}(Y'_{\mathbf{j}}Y'_{\mathbf{i}}))\partial_{\mathbf{j}}\partial_{\mathbf{i}}s(\mathbf{C}_u)$$

and

$$B_{u\mathbf{i}\mathbf{j}} := (\mathbb{E}(Y'_{\mathbf{j}}Y'_{\mathbf{i}}) - Y'_{\mathbf{j}}Y'_{\mathbf{i}})\partial_{\mathbf{j}}\partial_{\mathbf{i}}s(\mathbf{C}_u).$$

We use the orthogonality of $\sum_{\mathbf{i},\mathbf{j}\in\mathbf{I}_u} B_{u\mathbf{i}\mathbf{j}}$ and (13.60). This leads to

$$\mathbb{E}\Big|\sum_{u\geq 1}\sum_{\mathbf{i},\mathbf{j}\in\mathbf{I}_u} B_{u\mathbf{i}\mathbf{j}}\Big|^2 \ll \frac{1}{n^4}\sum_{u\geq 1}\Big\|\sum_{\mathbf{i},\mathbf{j}\in\mathbf{I}_u}|\mathbb{E}(Y'_{\mathbf{j}}Y'_{\mathbf{i}}) - Y'_{\mathbf{j}}Y'_{\mathbf{i}}|\Big\|_2^2$$

$$\leq \frac{1}{n^4}\sum_{u\geq 1}\Big(\sum_{\mathbf{i},\mathbf{j}\in\mathbf{I}_u}\|\mathbb{E}(Y'_{\mathbf{j}}Y'_{\mathbf{i}}) - Y'_{\mathbf{j}}Y'_{\mathbf{i}}\|_2\Big)^2 \leq \frac{4}{n^4}\sum_{u\geq 1}\Big(\sum_{\mathbf{i},\mathbf{j}\in\mathbf{I}_u}\|Y'_{\mathbf{j}}\|_4\|Y'_{\mathbf{i}}\|_4\Big)^2$$

$$\leq \frac{4^3}{n^4}\sum_{u\geq 1}\Big(\sum_{\mathbf{i}\in\mathbf{I}_u}\|Y_{\mathbf{i}}\|_4\Big)^4 \leq \frac{4^3p^6}{n^4}\sum_{u\geq 1}\sum_{\mathbf{i}\in\mathbf{I}_u}\|Y_{\mathbf{i}}\|_4^4.$$

Since the r.v.'s $Y_{\mathbf{i}}$ are Gaussian, $\|Y_{\mathbf{i}}\|_4^4 = 3\|Y_{\mathbf{i}}\|_2^4 = 3\|X_{\mathbf{i}}\|_2^4$. Therefore, for any $\varepsilon > 0$,

$$\begin{aligned}
\mathbb{E}\Big|\sum_{u\geq 1}\sum_{\mathbf{i},\mathbf{j}\in\mathbf{I}_u} B_{u\mathbf{ij}}\Big|^2 &\ll \frac{p^6}{n^4}\sum_{u\geq 1}\sum_{\mathbf{i}\in\mathbf{I}_u}\|X_{\mathbf{i}}\|_2^4 \\
&\ll \frac{p^6\varepsilon^2}{n^3}\sum_{u\geq 1}\sum_{\mathbf{i}\in\mathbf{I}_u}\|X_{\mathbf{i}}\|_2^2 + \frac{p^6}{n^4}\sum_{u\geq 1}\sum_{\mathbf{i}\in\mathbf{I}_u}\|X_{\mathbf{i}}^2 I(|X_{\mathbf{i}}| > \varepsilon\sqrt{n})\|_1^2 \\
&\ll \frac{p^6\varepsilon^2}{n^3}\sum_{(i,j)\in V_n^1}\|X_{ij}\|_2^2 + p^6\Big(\frac{1}{n^2}\sum_{(i,j)\in V_n^1}\|X_{ij}^2 I(|X_{ij}| > \varepsilon\sqrt{n})\|_1\Big)^2 .
\end{aligned}$$

Letting $n \to \infty$ and after $\varepsilon \to 0$, and taking into account Condition 13.1, it follows that

$$\mathbb{E}\Big|\sum_{u\geq 1}\sum_{\mathbf{i},\mathbf{j}\in\mathbf{I}_u} B_{u\mathbf{ij}}\Big|^2 \to 0 \text{ as } n \to \infty.$$

On the other hand, since

$$\mathbb{E}(Y'_{\mathbf{j}}Y'_{\mathbf{i}}) = \mathbb{E}(Y_{\mathbf{j}}Y_{\mathbf{i}}) - \mathbb{E}\big(\mathbb{E}(Y_{\mathbf{j}}|\mathcal{H}_{u-1})\mathbb{E}(Y_{\mathbf{i}}|\mathcal{H}_{u-1})\big)$$

and $\mathbb{E}(Y_{\mathbf{j}}Y_{\mathbf{i}}) = \mathbb{E}(X_{\mathbf{j}}X_{\mathbf{i}})$, we get, by the same arguments as those leading to (13.43),

$$\mathbb{E}\Big|\sum_{u\geq 1}\sum_{\mathbf{i},\mathbf{j}\in\mathbf{I}_u} A_{u\mathbf{ij}}\Big| \ll \frac{1}{n^2}\sum_{u\geq 1}\sum_{\mathbf{i}\in\mathbf{I}_u}\|\mathbb{E}(Y_{\mathbf{i}}|\mathcal{H}_{u-1})\|_2^2 .$$

Hence, by (13.33) and the contractivity of conditional expectation,

$$\mathbb{E}\Big|\sum_{u\geq 1}\sum_{\mathbf{i},\mathbf{j}\in\mathbf{I}_u} A_{u\mathbf{ij}}\Big| \ll \frac{1}{n^2}\sum_{u\geq 1}\sum_{\mathbf{i}\in\mathbf{I}_u}\|\mathbb{E}(X_{\mathbf{i}}|\mathcal{B}_{u-1})\|_2^2 \ll \sup_{i\geq j}\|\mathbb{E}(X_{ij}|\mathcal{F}_{ij}^K)\|_2^2 ,$$

which converges to 0, as $K \to \infty$, by condition (13.6). This completes the proof of the theorem. ◊

### 13.5.2 Proof of Theorem 13.4

The proof follows the lines of the proof of Theorem 13.2 with $Z_{ij}$ instead of $X_{ij}$ and with $W_{ij}$ instead of $Y_{ij}$. We point here the differences. The filtrations $\mathcal{B}_u$ and $\mathcal{H}_u$ respectively defined in (13.31) and (13.32) have to be defined as follows. If

$u = k(k-1)/2 + \ell$ with $1 \le \ell \le k$ and $1 \le k \le q$, then

$$\mathcal{B}_u = \mathcal{B}'_u \vee \mathcal{B}''_u \text{ for } u \ge 1 \text{ and } \mathcal{B}_0 = \{\emptyset, \Omega\} \tag{13.45}$$

and

$$\mathcal{H}_u = \mathcal{H}'_u \vee \mathcal{H}''_u \text{ for } u \ge 1 \text{ and } \mathcal{H}_0 = \{\emptyset, \Omega\}, \tag{13.46}$$

where

$$\mathcal{B}'_u = \mathcal{B}'_{k,\ell} = \sigma\big(X_{ab}\,, (a,b) \in \cup_{j=1}^{\ell}\mathcal{E}_{kj} \text{ or } (a,b) \in \cup_{i=1}^{k-1}\cup_{j=1}^{i}\mathcal{E}_{ij}\big)\,,$$

$$\mathcal{B}''_u = \mathcal{B}''_{k,\ell} = \sigma\big(X_{ba}\,, (a,b) \in \cup_{j=1}^{\ell}\mathcal{E}_{kj} \text{ or } (a,b) \in \cup_{i=1}^{k-1}\cup_{j=1}^{i}\mathcal{E}_{ij}\big)\,,$$

and $\mathcal{H}'_u$ and $\mathcal{H}''_u$ defined, respectively, as $\mathcal{B}''_u$ and $\mathcal{B}''_u$ with the $X_{ab}$ (resp. $X_{ba}$) replaced by $Y_{ab}$ (resp. $Y_{ba}$). According to the proof of Theorem 13.2, the proof will be achieved if we can show that if $u = k(k-1)/2 + \ell$ with $1 \le \ell \le k$ and $1 \le k \le q$, then for any $(i, j)$ and $(a, b)$ in $\mathcal{E}_{k,\ell}$

$$\|\mathbb{E}(Z_{ij}|\mathcal{B}_{u-1})\|_2 \le \|\mathbb{E}(X_{ij}|\widetilde{\mathcal{F}}_{ij}^K)\|_2 + \|\mathbb{E}(X_{ji}|\widetilde{\mathcal{F}}_{ji}^K)\|_2 \tag{13.47}$$

and

$$\begin{aligned}\|\mathbb{E}(Z_{ij}Z_{ab}|\mathcal{B}_{u-1}) - \mathbb{E}(Z_{ij}Z_{ab})\|_1 &\le \|\mathbb{E}(X_{ij}X_{ab}|\widetilde{\mathcal{F}}_{ij}^K \cap \widetilde{\mathcal{F}}_{ab}^K) - \mathbb{E}(X_{ij}X_{ab})\|_1 \\ + \|\mathbb{E}(X_{ij}X_{ba}|\widetilde{\mathcal{F}}_{ij}^K \cap \widetilde{\mathcal{F}}_{ba}^K) - \mathbb{E}(X_{ij}X_{ba})\|_1 &+ \|\mathbb{E}(X_{ji}X_{ab}|\widetilde{\mathcal{F}}_{ji}^K \cap \widetilde{\mathcal{F}}_{ab}^K) - \mathbb{E}(X_{ji}X_{ab})\|_1 \\ &+ \|\mathbb{E}(X_{ji}X_{ba}|\widetilde{\mathcal{F}}_{ji}^K \cap \widetilde{\mathcal{F}}_{ba}^K) - \mathbb{E}(X_{ji}X_{ba})\|_1 .\end{aligned} \tag{13.48}$$

To prove the inequalities above, we fix, all along the rest of the proof $k$ and $\ell$ such that $1 \le k \le q$ and $1 \le \ell \le k$ and also a $(i, j)$ in $\mathcal{E}_{k\ell}$. We notice that if $(u, v)$ belongs to $\cup_{m=1}^{\ell-1}\mathcal{E}_{km}$ then $j - v \ge K$, and if $(a, b)$ belongs to $\cup_{r=1}^{k-1}\cup_{m=1}^{r}\mathcal{E}_{rm}$ then $i - a \ge K$. So $\mathcal{H}'_{u-1} \subseteq \widetilde{\mathcal{F}}_{ij}^K$. In addition, if $(a, b)$ belongs to $\cup_{m=1}^{\ell-1}\mathcal{E}_{km}$ then $i - b \ge p + 2K$ and if $(a, b)$ belongs to $\cup_{r=1}^{k-1}\cup_{m=1}^{r}\mathcal{E}_{rm}$ then $i - b \ge 2K + p$. Therefore the distance between $(i, j)$ and all the points $(v, u)$ such that $(a, b)$ belongs either to $\cup_{m=1}^{\ell-1}\mathcal{E}_{km}$ or to $\cup_{r=1}^{k-1}\cup_{m=1}^{r}\mathcal{E}_{rm}$ is larger than $K$. This shows that $\mathcal{B}''_{u-1} \subseteq \widetilde{\mathcal{F}}_{ij}^K$. The two latter inclusions prove that $\mathcal{B}_{u-1} \subseteq \widetilde{\mathcal{F}}_{ij}^K$. Let us prove now that $\mathcal{H}_{u-1} \subseteq \widetilde{\mathcal{F}}_{ji}^K$. If $(a, b)$ belongs to $\cup_{m=1}^{\ell-1}\mathcal{E}_{km}$ then $a - j \ge K$, and if $(a, b)$ belongs to $\cup_{r=1}^{k-1}\cup_{m=1}^{r}\mathcal{E}_{rm}$ then $i - b \ge 2K + p$. So $\mathcal{H}'_{u-1} \subseteq \widetilde{\mathcal{F}}_{ji}^K$. In addition, if $(a, b)$ belongs to $\cup_{m=1}^{\ell-1}\mathcal{E}_{km}$ then $j - b \ge K$ and if $(a, b)$ belongs to $\cup_{r=1}^{k-1}\cup_{m=1}^{r}\mathcal{E}_{rm}$ then $i - a \ge K$. Therefore the distance between $(j, i)$ and all the points $(b, a)$ such that $(a, b)$ belongs to $\cup_{m=1}^{\ell-1}\mathcal{E}_{km}$ or to $\cup_{r=1}^{k-1}\cup_{m=1}^{r}\mathcal{E}_{rm}$ is larger than $K$. Hence $\mathcal{B}''_{u-1} \subseteq \widetilde{\mathcal{F}}_{ji}^K$. This ends the proof of $\mathcal{B}_{u-1} \subseteq \widetilde{\mathcal{F}}_{ji}^K$. Since $\mathcal{B}_{u-1} \subseteq \widetilde{\mathcal{F}}_{ji}^K$ and $\mathcal{B}_{u-1} \subseteq \widetilde{\mathcal{F}}_{ji}^K$, (13.47) follows by applying the tower lemma and using contraction. To prove (13.48), we use again the tower lemma together with the contractivity of the norm for the conditional expectation and the fact that the above inclusions imply that for any $(i, j)$ and $(a, b)$ belonging to $\mathcal{E}_{k,\ell}$, $\mathcal{B}_{u-1} \subseteq \widetilde{\mathcal{F}}^K_{ij} \cap \widetilde{\mathcal{F}}^K_{ab}$, $\mathcal{B}_{u-1} \subseteq \widetilde{\mathcal{F}}^K_{ij} \cap \widetilde{\mathcal{F}}^K_{ba}$, $\mathcal{B}_{u-1} \subseteq \widetilde{\mathcal{F}}^K_{ji} \cap \widetilde{\mathcal{F}}^K_{ab}$ and $\mathcal{B}_{u-1} \subseteq \widetilde{\mathcal{F}}^K_{ji} \cap \widetilde{\mathcal{F}}^K_{ba}$. ♢

### 13.5.3 Proof of Theorem 13.5

The proof is very similar to the proof of Theorem 9 from Merlevède and Peligrad [27]. We give it here for completeness.

Let $n = N + p$ and $\mathbb{X}_n$ the symmetric matrix of order $n$ defined by

$$\mathbb{X}_n = \frac{1}{\sqrt{N}} \begin{pmatrix} \mathbf{0}_{N,N} & \mathcal{X}_{N,p} \\ \mathcal{X}^T_{N,p} & \mathbf{0}_{p,p} \end{pmatrix} .$$

Notice that the eigenvalues of $\mathbb{X}_n^2$ are the eigenvalues of $N^{-1}\mathcal{X}_{N,p}\mathcal{X}^T_{N,p}$ together with the eigenvalues of $N^{-1}\mathcal{X}^T_{N,p}\mathcal{X}_{N,p}$. Assuming that $N \le p$ (otherwise exchange the role of $\mathcal{X}_{N,p}$ and $\mathcal{X}^T_{N,p}$ everywhere), the following relation holds: for any $z \in \mathbb{C}^+$

$$S^{\mathbb{B}_N}(z) = z^{-1/2} \frac{n}{2N} S^{\mathbb{X}_n}(z^{1/2}) + \frac{N-p}{2Nz} . \tag{13.49}$$

(See, for instance, page 549 in Rashidi Far et al. [36] for additional arguments leading to the relation above.) Consider now a real-valued centered Gaussian random field $(Y_{k\ell})_{(k,\ell)\in\mathbb{Z}^2}$ independent of $(X_{k\ell})_{(k,\ell)\in\mathbb{Z}^2}$ and with covariance function given by:

$$\mathbb{E}(Y_{k\ell}Y_{ij}) = \mathbb{E}(X_{k\ell}X_{ij}) \text{ for any } (k,\ell) \text{ and } (i,j) \text{ in } \mathbb{Z}^2 , \tag{13.50}$$

and define the $N \times p$ matrix

$$\Gamma_{N,p} = \big(Y_{ij}\big)_{1\le i\le N, 1\le j\le p} .$$

Let $\mathbb{G}_N = \frac{1}{N}\Gamma_{N,p}\Gamma^T_{N,p}$ and $\mathbb{H}_n$ be the symmetric matrix of order $n$ defined by

$$\mathbb{H}_n = \frac{1}{\sqrt{N}} \begin{pmatrix} \mathbf{0}_{N,N} & \Gamma_{N,p} \\ \Gamma^T_{N,p} & \mathbf{0}_{p,p} \end{pmatrix} .$$

Assuming that $N \le p$, the following relation holds: for any $z \in \mathbb{C}^+$

$$S^{\mathbb{G}_N}(z) = z^{-1/2} \frac{n}{2N} S^{\mathbb{H}_n}(z^{1/2}) + \frac{N-p}{2Nz} . \tag{13.51}$$

In view of relations (13.49) and (13.51), to prove that for any $z \in \mathbb{C}^+$,

$$\left|S^{\mathbb{B}_N}(z) - S^{\mathbb{G}_N}(z)\right| \to 0 \text{ in probability} \tag{13.52}$$

it suffices to prove that, for any $z \in \mathbb{C}^+$,

$$\left|S^{\mathbb{X}_n}(z) - S^{\mathbb{H}_n}(z)\right| \to 0 \text{ in probability} \tag{13.53}$$

(since $n/N \to 1+c$). Clearly (13.53) follows from the proof of Theorem 13.2 together with Comment 13.3 (iii), by noticing the following facts. The entries $x_{i,j}$ and $g_{i,j}$ of the matrices $n^{1/2}\mathbb{X}_n$ and $n^{1/2}\mathbb{H}_n$ respectively, satisfy

$$x_{i,j} = \alpha_{i,j}^{(n)} X_{ji}\,,\, g_{i,j} = \alpha_{i,j}^{(n)} Y_{ji} \text{ if } 1 \le j \le i \le n \text{ and } x_{i,j} = x_{j,i}\,,\, g_{i,j} = g_{j,i} \text{ if } 1 \le j \le i \le n\,,$$

where $(\alpha_{i,j}^{(n)})$ is a sequence of positive numbers defined by: $\alpha_{i,j}^{(n)} = \frac{n^{1/2}}{N^{1/2}} \mathbf{1}_{N+1 \le i \le n}$ $\mathbf{1}_{1 \le j \le N}$. Hence $\mathbb{E}(g_{k,\ell} g_{i,j}) = \alpha_{k,\ell}^{(n)} \alpha_{i,j}^{(n)} \mathbb{E}(X_{k\ell} X_{ij})$, $\max_{1 \le j \le i \le n} \alpha_{i,j} = \frac{n^{1/2}}{N^{1/2}} := \alpha^{(n)}$ and $\lim_{n\to\infty} \alpha^{(n)} = \sqrt{1+c}$. $\Diamond$

### 13.5.4 Proof of Theorem 13.6

The proof of this theorem follows all the steps of the proof of Theorem 13.2 (with the same notations) with the exception of the treatment of terms which appear in (13.44). By stationarity

$$\frac{1}{n^2} \sum_{u=1}^{q(q+1)/2} \sum_{\mathbf{i},\mathbf{j} \in \mathbf{I}_u} \mathbb{E}|\mathbb{E}(X_{\mathbf{j}} X_{\mathbf{i}} | \mathcal{B}_{u-1}) - \mathbb{E}(X_{\mathbf{j}} X_{\mathbf{i}})| \le \frac{1}{p^2} \sum_{\mathbf{i},\mathbf{j} \in \mathcal{E}_p} \mathbb{E}|\mathbb{E}(X_{\mathbf{j}} X_{\mathbf{i}} | \mathcal{F}_{\mathbf{0}}^K) - \mathbb{E}(X_{\mathbf{j}} X_{\mathbf{i}})|\,,$$

where above and below $\mathcal{E}_p = [1,p]^2 \cap \mathbb{N}^2$. For $\mathbf{i}$ fixed in $\mathcal{E}_p$, we shall divide the last sum in three parts according to $\mathbf{j} \in \mathcal{E}_p$, with $|\mathbf{i}-\mathbf{j}| \le d$ or $d \le |\mathbf{i}-\mathbf{j}| \le K$ or $|\mathbf{i}-\mathbf{j}| > K$, where $d$ is a positive integer less than $K$. Since for this case $\mathcal{F}_{\mathbf{0}}^K \subset \mathcal{F}_{\mathbf{i}\wedge\mathbf{j}}^K$, by the properties of conditional expectations and stationarity we have

$$\begin{aligned}\sum_{\mathbf{j} \in \mathcal{E}_p, |\mathbf{i}-\mathbf{j}| \le d} \mathbb{E}|\mathbb{E}(X_{\mathbf{j}} X_{\mathbf{i}} | \mathcal{F}_{\mathbf{0}}^K) - \mathbb{E}(X_{\mathbf{j}} X_{\mathbf{i}})| &\le \sum_{\mathbf{j} \in \mathcal{E}_p, |\mathbf{i}-\mathbf{j}| \le d} \mathbb{E}|\mathbb{E}(X_{\mathbf{j}} X_{\mathbf{i}} | \mathcal{F}_{\mathbf{i}\wedge\mathbf{j}}^K) - \mathbb{E}(X_{\mathbf{j}} X_{\mathbf{i}})| \\ &= \sum_{\mathbf{j} \in \mathcal{E}_p, |\mathbf{i}-\mathbf{j}| \le d} \mathbb{E}|\mathbb{E}(X_{\mathbf{0}} X_{\mathbf{j}-\mathbf{i}} | \mathcal{F}_{(\mathbf{j}-\mathbf{i})\wedge\mathbf{0}}^K) - \mathbb{E}(X_{\mathbf{0}} X_{\mathbf{j}-\mathbf{i}})|\,.\end{aligned}$$

Therefore

$$\frac{1}{p^2} \sum_{\mathbf{i},\mathbf{j} \in \mathcal{E}_p, |\mathbf{i}-\mathbf{j}| \le d} \mathbb{E}|\mathbb{E}(X_{\mathbf{j}} X_{\mathbf{i}} | \mathcal{F}_{\mathbf{0}}^K) - \mathbb{E}(X_{\mathbf{j}} X_{\mathbf{i}})| \le \sum_{\mathbf{u}, |\mathbf{u}| \le d} \mathbb{E}|\mathbb{E}(X_{\mathbf{0}} X_{\mathbf{u}} | \mathcal{F}_{\mathbf{u}\wedge\mathbf{0}}^K) - \mathbb{E}(X_{\mathbf{0}} X_{\mathbf{u}})|,$$

which converges to 0 for $d$ fixed as $K \to \infty$ by the regularity condition of the random field.

Now we treat the part of the sum where $\mathbf{j} \in \mathcal{E}_p$, with $d < |\mathbf{i}-\mathbf{j}| \leq K$. For this case we note that $\mathcal{F}_{\mathbf{0}}^K \subset \mathcal{F}_{\mathbf{i}}^{|\mathbf{i}-\mathbf{j}|}$ and $\mathcal{F}_{\mathbf{0}}^K \subset \mathcal{F}_{\mathbf{j}}^{|\mathbf{i}-\mathbf{j}|}$. By the properties of conditional expectation, stationarity and some computations we infer that

$$\sum_{\mathbf{j}\in\mathcal{E}_p, d<|\mathbf{i}-\mathbf{j}|\leq K} \mathbb{E}|\mathbb{E}(X_{\mathbf{i}}X_{\mathbf{j}}|\mathcal{F}_{\mathbf{0}}^K) - \mathbb{E}(X_{\mathbf{j}}X_{\mathbf{i}})| \leq 2 \sum_{\mathbf{u}\in V_0, |\mathbf{u}|>d} \mathbb{E}|X_{\mathbf{u}}\mathbb{E}(X_{\mathbf{0}}|\mathcal{F}_{\mathbf{0}}^{|\mathbf{u}|})|\,,$$

where we recall that $V_0 = \{\mathbf{u} = (u_1, u_2) \in \mathbb{Z}^2 \, : \, u_1 \leq 0 \text{ or } u_2 \leq 0\}$. It follows that

$$\frac{1}{p^2} \sum_{\mathbf{i},\mathbf{j}\in\mathcal{E}_p, d<|\mathbf{i}-\mathbf{j}|\leq K} \mathbb{E}|\mathbb{E}(X_{\mathbf{i}}X_{\mathbf{j}}|\mathcal{F}_{\mathbf{0}}^K) - \mathbb{E}(X_{\mathbf{j}}X_{\mathbf{i}})| \leq 2 \sum_{\mathbf{u}\in V_0, |\mathbf{u}|>d} \mathbb{E}|X_{\mathbf{u}}\mathbb{E}(X_{\mathbf{0}}|\mathcal{F}_{\mathbf{0}}^{|\mathbf{u}|})|\,,$$

which converges to 0 as $d \to \infty$ uniformly in $p$ and $K$ by (13.19).

Finally, for the third sum where $\mathbf{i}, \mathbf{j} \in \mathcal{E}_p$, $|\mathbf{i}-\mathbf{j}| \geq K$ we either have $\sigma(X_{\mathbf{i}}) \subset \mathcal{F}_{\mathbf{j}}^K$ or $\sigma(X_{\mathbf{j}}) \subset \mathcal{F}_{\mathbf{i}}^K$. Moreover $\mathcal{F}_{\mathbf{0}}^K \subset \mathcal{F}_{\mathbf{i}}^K$ and $\mathcal{F}_{\mathbf{0}}^K \subset \mathcal{F}_{\mathbf{j}}^K$. By the properties of conditional expectation, when $\sigma(X_{\mathbf{i}}) \subset \mathcal{F}_{\mathbf{j}}^K$ we have

$$\mathbb{E}|\mathbb{E}(X_{\mathbf{i}}X_{\mathbf{j}}|\mathcal{F}_{\mathbf{0}}^K) - \mathbb{E}(X_{\mathbf{j}}X_{\mathbf{i}})| \leq 2\mathbb{E}(|\mathbb{E}(X_{\mathbf{i}}X_{\mathbf{j}}|\mathcal{F}_{\mathbf{j}}^K)|) = 2\mathbb{E}(|X_{\mathbf{i}-\mathbf{j}}\mathbb{E}(X_{\mathbf{0}}|\mathcal{F}_{\mathbf{0}}^K)|)\,.$$

When $\sigma(X_{\mathbf{j}}) \subset \mathcal{F}_{\mathbf{i}}^K$, similarly, we have

$$\mathbb{E}|\mathbb{E}(X_{\mathbf{i}}X_{\mathbf{j}}|\mathcal{F}_{\mathbf{0}}^K) - \mathbb{E}(X_{\mathbf{j}}X_{\mathbf{i}})| \leq 2\mathbb{E}(|X_{\mathbf{j}-\mathbf{i}}\mathbb{E}(X_{\mathbf{0}}|\mathcal{F}_{\mathbf{0}}^K)|)\,.$$

Therefore

$$\frac{1}{p^2} \sum_{\mathbf{i},\mathbf{j}\in\mathcal{E}_p, |\mathbf{i}-\mathbf{j}|\geq K} \mathbb{E}|\mathbb{E}(X_{\mathbf{i}}X_{\mathbf{j}}|\mathcal{F}_{\mathbf{0}}^K) - \mathbb{E}(X_{\mathbf{j}}X_{\mathbf{i}})| \leq 2p^2 \sup_{\mathbf{u}\in V_0:|\mathbf{u}|>K} \mathbb{E}|X_{\mathbf{u}}\mathbb{E}(X_{\mathbf{0}}|\mathcal{F}_{\mathbf{0}}^K)|\,.$$

Since we can take $K$ close to $p$, the result follows by letting first $p \to \infty$ followed by $d \to \infty$. ♢

### *13.5.5 Proof of Proposition 13.10*

The proof uses similar arguments as those given in the proof of Theorem 2.1 in Chakrabarty et al. [10].

For any integers $k$ and $\ell$ define

$$c_{k\ell} = \int_{[0,1]^2} \mathrm{e}^{-2\pi \mathrm{i}(kx+\ell y)} \sqrt{f(x, y)}dxdy\,.$$

There are real numbers and satisfy $\sum_{k,\ell\in\mathbb{Z}} c_{k\ell}^2 < \infty$. Let now $(U_{ij})_{(i,j)\in\mathbb{Z}^2}$ be i.i.d. real-valued random variables with law $\mathcal{N}(0,1)$. Then, without restricting the generality, we can write

$$Y_{ij} = \sum_{k,\ell\in\mathbb{Z}} c_{k\ell} U_{i-k,j-\ell}\,. \tag{13.54}$$

(See Fact 4.1 in Chakrabarty et al. [10].)

The result will follow if we can prove that when $N$, $p$ tend to infinity such that $p/N \to c$, then there exists a deterministic probability measure $\mu_f$ depending only on $c$ and $f$, and such that for any $\varepsilon > 0$,

$$\mathbb{P}\big(d(\mu_{\mathbb{G}_N}, \mu_f) > \varepsilon\big) \to 0 \text{ as } N \to \infty. \tag{13.55}$$

Clearly the identity (13.54) holds in distribution, hence to prove (13.55), without loss of generality, we may and do assume from now on, that $Y_{ij}$ is given by (13.54). To prove (13.55), we shall use Fact 4.3 in Chakrabarty et al. [10] and first truncate the series (13.54). Hence we fix a positive integer $m$ and we define

$$Y_{ij}^{(m)} = \sum_{k=-m}^{m} \sum_{\ell=-m}^{m} c_{k\ell} U_{i-k,j-\ell}\,.$$

Let $\Gamma_{N,p}^{(m)} = \big(Y_{ij}^{(m)}\big)_{1\le i\le N, 1\le j\le p}$. Define also $\mathbb{G}_N^{(m)} = \frac{1}{N}\Gamma_{N,p}^{(m)}(\Gamma_{N,p}^{(m)})^T$. By Theorem 2.1 in Boutet de Monvel et al. [8], we have that for any positive integer $m$, there exists a deterministic probability measure $\mu_m$ depending only on $c$ and on the complex-valued function $\chi^{(m)}$ defined on $[0,1]^2$ by $\chi^{(m)} = \sum_{k,\ell\in\mathbb{Z}} \mathbb{E}(Y_{00}^{(m)} Y_{k\ell}^{(m)}) \mathrm{e}^{-2\pi \mathrm{i}(kx+\ell y)}$, and such that for any $\varepsilon > 0$,

$$\mathbb{P}\big(d(\mu_{\mathbb{G}_N^{(m)}}, \mu_m) > \varepsilon\big) \to 0 \text{ as } N \to \infty. \tag{13.56}$$

Notice that

$$\mathbb{E}(Y_{00}^{(m)} Y_{k\ell}^{(m)}) = \sum_{u=\max(k-m,m)}^{\min(k+m,m)} \sum_{v=\max(\ell-m,m)}^{\min(\ell+m,m)} c_{uv} c_{k-u,\ell-v}\,.$$

Since the $c_{k\ell}$'s depend only on $f$, it follows that $\chi^{(m)}$ depends only on $m$ and $f$. Therefore $\mu_m$ can be rewritten as $\mu_{m,f}$. Notice now that by Corollary A.42 in Bai and Silverstein [1], for any $\varepsilon > 0$,

$$\begin{aligned}\mathbb{P}\big(d(\mu_{\mathbb{G}_N}, \mu_{\mathbb{G}_N^{(m)}}) > \varepsilon\big) &\le \frac{1}{\varepsilon^2}\mathbb{E}(d^2(\mu_{\mathbb{G}_N}, \mu_{\mathbb{G}_N^{(m)}}) \\ &\le \frac{\sqrt{2}}{p\sqrt{N}\varepsilon^2}\big\|\mathrm{Tr}^{1/2}\big(\mathbb{G}_N^{(m)} + \mathbb{G}_N\big)\mathrm{Tr}^{1/2}\big((\Gamma_{N,p}^{(m)} - \Gamma_{N,p})(\Gamma_{N,p}^{(m)} - \Gamma_{N,p})^T\big)\big\|_1.\end{aligned}$$

Therefore, by the Cauchy–Schwarz's inequality and simple algebra,

$$\mathbb{P}\big(d(\mu_{\mathbb{G}_N},\mu_{\mathbb{G}_N^{(m)}})>\varepsilon\big)\le\frac{\sqrt{2}}{p\sqrt{N}\varepsilon^2}\big\|\mathrm{Tr}\big(\mathbb{G}_N^{(m)}+\mathbb{G}_N\big)\big\|_1^{1/2}\big\|\mathrm{Tr}^{1/2}\big((\Gamma_{N,p}^{(m)}-\Gamma_{N,p})(\Gamma_{N,p}^{(m)}-\Gamma_{N,p})^T\big)\big\|_1^{1/2}$$
$$\ll\Big(\sum_{k,\ell\in\mathbb{Z}\,:\,|k|\vee|\ell|>m}c_{k\ell}^2\Big)^{1/2}.$$

This proves that, for any $\varepsilon > 0$,

$$\lim_{m\to\infty}\limsup_{n\to\infty}\mathbb{P}\big(d(\mu_{\mathbb{G}_N},\mu_{\mathbb{G}_N^{(m)}})>\varepsilon\big)=0\,. \tag{13.57}$$

Taking into account (13.56) and (13.57), Fact 4.3 in Chakrabarty et al. [10] and the fact that the space of probability measures on $\mathbb{R}$ is a complete metric space when equipped with the Lévy distance, (13.55) follows. ♢

## 13.6 Useful Technical Lemmas

Below we give a Taylor expansion of a more convenient type for using Lindeberg's method:

**Lemma 13.12** *Let $f(\cdot)$ be a function from $\mathbb{R}^\ell$ to $\mathbb{C}$, three times differentiable, with continuous third partial derivatives and such that*

$$|\partial_i\partial_j f(\mathbf{x})|\le L_2 \textit{ and } |\partial_i\partial_j\partial_k f(\mathbf{x})|\le L_3 \textit{ for all } i,j,k\in\{1,\dots,\ell\} \textit{ and } \mathbf{x}\in\mathbb{R}^\ell\,.$$

*Then, for any $\mathbf{a}=(a_1,\dots,a_\ell)$ and $\mathbf{b}=(b_1,\dots,b_\ell)$ in $\mathbb{R}^\ell$,*

$$f(\mathbf{a})-f(\mathbf{b})=\sum_{k=1}^{2}\frac{1}{k!}[(\sum_{j=1}^{\ell}a_j\partial_j)^k-(\sum_{j=1}^{\ell}b_j\partial_j)^k]f(0,\dots,0)+R_3$$

*where $|R_3|\le R(\mathbf{a})+R(\mathbf{b})$, with*

$$R(\mathbf{c})\le 4\ell L_2\sum_{j=1}^{\ell}c_j^2 I(|a_j|>A)+2AL_3\ell^2\Big(\sum_{j=1}^{\ell}c_j^2\Big),$$

*where $\mathbf{c}$ equals $\mathbf{a}$ or $\mathbf{b}$.*

This Lemma can be applied in conjunction with Stieltjes transform. Let $A(\mathbf{x})$ be the matrix defined by

$$(A(\mathbf{x}))_{ij}=\begin{cases}\frac{1}{\sqrt{n}}x_{ij} & i\ge j\\ \frac{1}{\sqrt{n}}x_{ji} & i<j\end{cases} \tag{13.58}$$

Let $z \in \mathbb{C}^+$ and $s := s_z$ be the function defined from $\mathbb{R}^N$ to $\mathbb{C}$ by

$$s(\mathbf{x}) = \frac{1}{n}\mathrm{Tr}(A(\mathbf{x}) - z\mathbf{I}_n)^{-1}, \tag{13.59}$$

where $\mathbf{I}_n$ is the identity matrix of order $n$.

The function $s$, as defined above, admits partial derivatives of all orders. Next we give a lemma concerning the derivatives of $s(\mathbf{x})$ which is easily obtained by using the computations in Chatterjee [11] (see the proof of Lemma 12 in Merlevède and Peligrad [27] for a complete proof of its last inequality).

**Lemma 13.13** *Let $z = u + iv \in \mathbb{C}^+$ and let $(a_{ij})_{1\le j\le i\le n}$ and $(b_{ij})_{1\le j\le i\le n}$ be real numbers. There exist universal positive constants $c_1, c_2$ and $c_3$ depending only on the imaginary part of $z$ such that*

$$|\partial_{\mathbf{u}} s| \le \frac{c_1}{n^{3/2}}, \ |\partial_{\mathbf{u}}\partial_{\mathbf{v}} s| \le \frac{c_2}{n^2} \ \text{and} \ |\partial_{\mathbf{u}}\partial_{\mathbf{v}}\partial_{\mathbf{w}} s| \le \frac{c_3}{n^{5/2}}. \tag{13.60}$$

*Furthermore there exists an universal positive constant $c_4$ depending only on the imaginary part of $z$ such that for any subset $\mathcal{I}_n$ of $\{(i, j)\}_{1\le j\le i\le n}$ and any $\mathbf{x}$,*

$$\Big|\sum_{\mathbf{u}\in\mathcal{I}_n}\sum_{\mathbf{v}\in\mathcal{I}_n} a_{\mathbf{u}} b_{\mathbf{v}} \partial_{\mathbf{u}}\partial_{\mathbf{v}} s_n(\mathbf{x})\Big| \le \frac{c_4}{n^2}\Big(\sum_{\mathbf{u}\in\mathcal{I}_n} a_{\mathbf{u}}^2 \sum_{\mathbf{v}\in\mathcal{I}_n} b_{\mathbf{v}}^2\Big)^{1/2}.$$

The following lemma is Lemma 2.1 in Götze et al. [21].

**Lemma 13.14** *Let $\mathbf{A}_n$ and $\mathbf{B}_n$ be two symmetric $n \times n$ matrices. Then, for any $z \in \mathbb{C}\backslash\mathbb{R}$,*

$$|S_{\mathbb{A}_n}(z) - S_{\mathbb{B}_n}(z)|^2 \le \frac{1}{n^2|\operatorname{Im}(z)|^4}\mathrm{Tr}\Big[(\mathbf{A}_n - \mathbf{B}_n)^2\Big],$$

*where $\mathbb{A}_n = n^{-1/2}\mathbf{A}_n$ and $\mathbb{B}_n = n^{-1/2}\mathbf{B}_n$.*

**Acknowledgements** This paper was supported in part by a Charles Phelps Taft Memorial Fund grant, and the NSF grant DMS-1811373. We are grateful to the referee for carefully reading our manuscript and for helpful comments that improved its presentation.

## References

1. Z. Bai, J.W. Silverstein, *Spectral Analysis of Large Dimensional Random Matrices*, 2nd edn. (Springer, New York, 2010)
2. Z. Bai, W. Zhou, Large sample covariance matrices without independence structures in columns. Stat. Sin. **18**, 425–442 (2008)

3. M. Banna, F. Merlevède, M. Peligrad, On the limiting spectral distribution for a large class of symmetric random matrices with correlated entries. Stoch. Process. Appl. **125**(7), 2700–2726 (2015)
4. A.K. Basu, C.C.Y. Dorea, On functional central limit theorem for stationary martingale random fields. Acta Math. Acad. Sci. Hungar. **33**, 307–316 (1979)
5. S. Bernstein, Sur l'extension du théorème limite du calcul des probabilités aux sommes de quantités dépendantes. Math. Ann. **97**(1), 1–59 (1927)
6. S.G. Bobkov, F. Götze, A.N. Tikhomirov, On concentration of empirical measures and convergence to the semi-circle law. J. Theor. Probab. **23**, 792–823 (2010)
7. A. Boutet de Monvel, A. Khorunzhy, On the norm and eigenvalue distribution of large random matrices. Ann. Probab. **27**, 913–944 (1999)
8. A. Boutet de Monvel, A. Khorunzhy, V. Vasilchuk, Limiting eigenvalue distribution of random matrices with correlated entries. Markov Process. Related Fields **2**, 607–636 (1996)
9. R. Bradley, *Introduction to Strong Mixing Conditions*, vols. 1, 2, 3 (Kendrick Press, Heber City, 2007)
10. A. Chakrabarty, R.S. Hazra, D. Sarkar, From random matrices to long range dependence. Random Matrices Theory Appl. **5**(2), 52 (2016)
11. S. Chatterjee, A generalization of the Lindeberg principle. Ann. Probab. **34**, 2061–2076 (2006)
12. R. Davis, O. Pfaffel, R. Stelzer, Limit theory for the largest eigenvalues of sample covariance matrices with heavy-tails. Stoch. Process. Appl. **124**(1), 18–50 (2014)
13. J. Dedecker, A central limit theorem for stationary random fields. Probab. Theory Relat. Fields **110**, 397–426 (1998)
14. J. Dedecker, Exponential inequalities and functional central limit theorem for random fields. ESAIM Probab. Stat. **5**, 77–104 (2001)
15. R.L. Dobrushin, S.B. Shlosman, Constructive criterion for the uniqueness of Gibbs field, in *Statistical Physics and Dynamical Systems, (Köszeg, 1984)*. Progress in Physics, vol. 10 (Birkhäuser Boston, Boston, 1985), pp. 347–370
16. C. Dombry, F. Eyi-Minko, Strong mixing properties of max-infinitely divisible random fields. Stochastic Process. Appl. **122**(11), 3790–3811 (2012)
17. P. Doukhan, *Mixing: Properties and Examples* (Springer, Berlin, 1984)
18. M. El Machkouri, D. Volný, W.B. Wu, A central limit theorem for stationary random fields. Stoch. Process. Appl. **123**(1), 1–14 (2013)
19. V.L. Girko, *Theory of Random Determinants*.Translated from the Russian. Mathematics and Its Applications (Soviet Series), vol. 45 (Kluwer Academic Publishers Group, Dordrecht, 1990)
20. V.L. Girko, *Statistical Analysis of Observations of Increasing Dimension*, vol. 28 (Springer Science & Business Media, Berlin, 1995)
21. F. Götze, A. Naumov, A. Tikhomirov, Semicircle law for a class of random matrixes with dependent entries (2012). arXiv:math/0702386v1
22. W. Hachem, P. Loubaton, J. Najim, The empirical eigenvalue distribution of a Gram matrix: from independence to stationarity. Markov Process. Related Fields **11**, 629–648 (2005)
23. A. Khorunzhy, L. Pastur, On the eigenvalue distribution of the deformed Wigner ensemble of random matrices, in *Spectral Operator Theory and Related Topics*, ed. by V.A. Marchenko. Advances in Soviet Mathematics, vol. 19 (American Mathematical Society, Providence, 1994), pp. 97–127
24. E. Laroche, Hypercontractivité pour des systèmes de spins de portée infinie. Probab. Theory Relat. Fields. **101**, 89–132 (1995)
25. V. Marchenko, L. Pastur, Distribution of eigenvalues for some sets of random matrices. Mat. Sb. **72**, 507–536 (1967)
26. F. Merlevède, M. Peligrad, On the weak invariance principle for stationary sequences under projective criteria. J. Theor. Probab. **19**(3), 647–689 (2006)
27. F. Merlevède, M. Peligrad, On the empirical spectral distribution for matrices with long memory and independent rows. Stoch. Process. Appl. **126**(9), 2734–2760 (2016)
28. F. Merlevède, C. Peligrad, M. Peligrad, On the universality of spectral limit for random matrices with martingale differences entries. Random Matrices Theor. Appl. **4**(1), 33 (2015)

29. F. Merlevède, M. Peligrad, S. Utev, *Functional Gaussian Approximation for Dependent Structures*. Oxford Studies in Probability, vol. 6 (Oxford University Press, Oxford, 2019)
30. A. Pajor, L. Pastur, On the limiting empirical measure of eigenvalues of the sum of rank one matrices with log-concave distribution. Studia Math. **195**(1), 11–29 (2009)
31. G. Pan, J. Gao, Y. Yang, Testing independence among a large number of high-dimensional random vectors. J. Am. Stat. Assoc. **109**, 600–612 (2014)
32. L. Pastur, On the spectrum of random matrices. Teor. Mat. Fiz. **10**, 102–112 (1972)
33. L. Pastur, M. Shcherbina, *Eigenvalue Distribution of Large Random Matrices*. Mathematical Surveys and Monographs, vol. 171 (American Mathematical Society, Providence, 2011)
34. C. Peligrad, M. Peligrad, The limiting spectral distribution in terms of spectral density. Random Matrices Theory Appl. **5**(1), 19 (2016)
35. O. Pfaffel, E. Schlemm, Eigenvalue distribution of large sample covariance matrices of linear processes. Probab. Math. Stat. **31**, 313–329 (2011)
36. R. Rashidi Far, T. Oraby, W. Bryc, R. Speicher, On slow-fading MIMO systems with nonseparable correlation. IEEE Trans. Inform. Theory **54**, 544–553 (2008)
37. E. Rio, *Asymptotic Theory of Weakly Dependent Random Processes*. Probability Theory and Stochastic Modelling, vol. 80 (Springer, Berlin, 2017)
38. M. Rosenblatt, A central limit theorem and a strong mixing condition. Proc. Nat. Acad. Sci. U. S. A. **42**, 43–47 (1956)
39. M. Rosenblatt, A comment on a conjecture of N. Wiener. Statist. Probab. Lett. **79**, 347–348 (2009)
40. Z. Sasvári, *Multivariate Characteristic and Correlation Functions*. De Gruyter Studies in Mathematics, vol. 50 (Walter de Gruyter & Co., Berlin, 2013), pp. x+366
41. J.W. Silverstein, Strong convergence of the empirical distribution of eigenvalues of large dimensional random matrices. J. Multivar. Anal. **55**, 331–339 (1995)
42. J.W. Silverstein, Z.D. Bai, On the empirical distribution of eigenvalues of a class of large dimensional random matrices. J. Multivar. Anal. **54**, 175–192 (1995)
43. T. Tao, *Topics in Random Matrix Theory*. Graduate Studies in Mathematics, vol. 132 (American Mathematical Society, Providence 2012)
44. C. Tone, Central limit theorems for Hilbert-space valued random fields satisfying a strong mixing condition. ALEA Lat. Am. J. Probab. Math. Stat. **8**, 77–94 (2011)
45. E.P. Wigner, On the distribution of the roots of certain symmetric matrices. Ann. Math. **67**, 325–327 (1958)
46. J. Yao, A note on a Mar Ì†cenko-Pastur type theorem for time series. Statist. Probab. Lett. **82**, 22–28 (2012)
47. P. Yaskov, A short proof of the Marchenko-Pastur theorem. C. R. Math. Acad. Sci. Paris **354**(3), 319–322 (2016)
48. P. Yaskov, Necessary and sufficient conditions for the Marchenko-Pastur theorem. Electron. Commun. Probab. **21**(73), 1–8 (2016)
49. Y.Q. Yin, Limiting spectral distribution for a class of random matrices. J. Multivar. Anal. **20**, 50–68 (1986)
50. Y.Q. Yin, P.R. Krishnaiah, A limit theorem for the eigenvalues of product of two random matrices. J. Multivar. Anal. **13**(4), 489–507 (1983)

# Chapter 14
# Exchangeable Pairs on Wiener Chaos

**Ivan Nourdin and Guangqu Zheng**

*Dedicated to the memory of Charles Stein, in remembrance of his beautiful mind and of his inspiring, creative, very original and deep mathematical ideas, which will, for sure, survive him for a long time.*

**Abstract** Nourdin and Peccati (Probab Theory Relat Fields 145(1):75–118, 2009) combined the Malliavin calculus and Stein's method of normal approximation to associate a rate of convergence to the celebrated fourth moment theorem of Nualart and Peccati (Ann Probab 33(1):177–193, 2005). Their analysis, known as the Malliavin-Stein method nowadays, has found many applications towards stochastic geometry, statistical physics and zeros of random polynomials, to name a few. In this article, we further explore the relation between these two fields of mathematics. In particular, we construct exchangeable pairs of Brownian motions and we discover a natural link between Malliavin operators and these exchangeable pairs. By combining our findings with E. Meckes' infinitesimal version of exchangeable pairs, we can give another proof of the quantitative fourth moment theorem. Finally, we extend our result to the multidimensional case.

**Keywords** Stein's method · Exchangeable pairs · Brownian motion · Malliavin calculus

## 14.1 Introduction

At the beginning of the 1970s, Charles Stein, one of the most famous statisticians of the time, introduced in [24] a new revolutionary method for establishing probabilistic approximations (now known as *Stein's method*), which is based on

I. Nourdin (✉) · G. Zheng
Université du Luxembourg, Esch-sur-Alzette, Luxembourg
e-mail: ivan.nourdin@uni.lu



N. Gozlan et al. (eds.), *High Dimensional Probability VIII*,
Progress in Probability 74, https://doi.org/10.1007/978-3-030-26391-1_14

the breakthrough application of characterizing differential operators. The impact of Stein's method and its ramifications during the last 40 years is immense (see for instance the monograph [3]), and touches fields as diverse as combinatorics, statistics, concentration and functional inequalities, as well as mathematical physics and random matrix theory.

Introduced by Malliavin [7], *Malliavin calculus* can be roughly described as an infinite-dimensional differential calculus whose operators act on sets of random objects associated with Gaussian or more general noises. In 2009, Nourdin and Peccati [14] combined the Malliavin calculus and Stein's method for the first time, thus virtually creating a new domain of research, which is now commonly known as the *Malliavin-Stein method*. The success of their method relies crucially on the existence of integration-by-parts formulae on both sides: on one side, the Stein's lemma is built on the Gaussian integration-by-parts formula and it is one of the cornerstones of the Stein's method; on the other side, the integration-by-parts formula on Gaussian space is one of the main tools in Malliavin calculus. Interested readers can refer to the constantly updated website [13] and the monograph [15] for a detailed overview of this active field of research.

A prominent example of applying Malliavin-Stein method is the obtention (see also (14.1) below) of a Berry-Esseen's type rate of convergence associated to the celebrated fourth moment theorem [19] of Nualart and Peccati, according to which a standardized sequence of multiple Wiener-Itô integrals converges in law to a standard Gaussian random variable if and only if its fourth moment converges to 3.

**Theorem 14.1.1**

(i) (Nualart, Peccati [19]) *Let $(F_n)$ be a sequence of multiple Wiener-Itô integrals of order $p$, for some fixed $p \geqslant 1$. Assume that $E[F_n^2] \to \sigma^2 > 0$ as $n \to \infty$. Then, as $n \to \infty$, we have the following equivalence:*

$$F_n \overset{\text{law}}{\to} N(0, \sigma^2) \quad \Longleftrightarrow \quad E[F_n^4] \to 3\sigma^4.$$

(ii) (Nourdin, Peccati [14, 15]) *Let $F$ be any multiple Wiener-Itô integral of order $p \geqslant 1$, such that $E[F^2] = \sigma^2 > 0$. Then, with $N \sim N(0, \sigma^2)$ and $d_{TV}$ standing for the total variation distance,*

$$d_{TV}(F, N) \leqslant \frac{2}{\sigma^2}\sqrt{\frac{p-1}{3p}}\,\sqrt{E[F^4] - 3\sigma^4}.$$

Of course, (ii) was obtained several years after (i), and (ii) implies '$\Leftarrow$' in (i). Nualart and Peccati's fourth moment theorem has been the starting point of a number of applications and generalizations by dozens of authors. These collective efforts have allowed one to break several long-standing deadlocks in several domains, ranging from stochastic geometry (see e.g. [6, 21, 23]) to statistical physics (see e.g. [8–10]), and zeros of random polynomials (see e.g. [1, 2, 4]), to name a few.

At the time of writing, more than two hundred papers have been written, which use in one way or the other the Malliavin-Stein method (see again the webpage [13]).

Malliavin-Stein method has become a popular tool, especially within the Malliavin calculus community. Nevertheless, and despite its success, it is less used by researchers who are not specialists of the Malliavin calculus. A possible explanation is that it requires a certain investment before one is in a position to be able to use it, and doing this investment may refrain people who are not originally trained in the Gaussian analysis. This paper takes its root from this observation.

During our attempt to make the proof of Theorem 14.1.1(ii) more accessible to readers having no background on Malliavin calculus, we discover the following interesting fact for exchangeable pairs of multiple Wiener-Itô integrals. When $p \geqslant 1$ is an integer and $f$ belongs to $L^2([0,1]^p)$, we write $I_p^B(f)$ to indicate the multiple Wiener-Itô integral of $f$ with respect to Brownian motion $B$, see Sect. 14.2 for the precise meaning.

**Proposition 14.1.2** *Let $(B, B^t)_{t\geqslant 0}$ be a family of exchangeable pairs of Brownian motions (that is, $B$ is a Brownian motion on $[0,1]$ and, for each $t$, one has $(B, B^t) \overset{\text{law}}{=} (B^t, B)$). Assume moreover that*

(a) *for any integer $p \geqslant 1$ and any $f \in L^2([0,1]^p)$,*

$$\lim_{t\downarrow 0} \frac{1}{t} E\Big[I_p^{B^t}(f) - I_p^B(f)\big|\sigma\{B\}\Big] = -p\, I_p^B(f) \quad \text{in } L^2(\Omega).$$

*Then, for any integer $p \geqslant 1$ and any $f \in L^2([0,1]^p)$,*

(b) $\displaystyle\lim_{t\downarrow 0} \frac{1}{t} E\Big[\big(I_p^{B^t}(f) - I_p^B(f)\big)^2|\sigma\{B\}\Big] = 2p^2 \int_0^1 I_{p-1}^B(f(x,\cdot))^2 dx \quad \text{in } L^2(\Omega);$

(c) $\displaystyle\lim_{t\downarrow 0} \frac{1}{t} E\Big[\big(I_p^{B^t}(f) - I_p^B(f)\big)^4\Big] = 0.$

Why is this proposition interesting? Because, as it turns out, it combines perfectly well with the following result, which represents the main ingredient from Stein's method we will rely on and which corresponds to a slight modification of a theorem originally due to Elizabeth Meckes (see [11, Theorem 2.1]).

**Theorem 14.1.3 (Meckes [11])** *Let $F$ and a family of random variables $(F_t)_{t\geqslant 0}$ be defined on a common probability space $(\Omega, \mathcal{F}, P)$ such that $F_t \overset{law}{=} F$ for every $t \geqslant 0$. Assume that $F \in L^3(\Omega, \mathcal{G}, P)$ for some $\sigma$-algebra $\mathcal{G} \subset \mathcal{F}$ and that in $L^1(\Omega)$,*

(a) $\displaystyle\lim_{t\downarrow 0} \frac{1}{t} E[F_t - F|\mathcal{G}] = -\lambda\, F$ *for some* $\lambda > 0$,

(b) $\displaystyle\lim_{t\downarrow 0} \frac{1}{t} E[(F_t - F)^2|\mathcal{G}] = (2\lambda + S)\text{Var}(F)$ *for some random variable* $S$,

(c) $\displaystyle\lim_{t\downarrow 0} \frac{1}{t} (F_t - F)^3 = 0.$

*Then, with* $N \sim N(0, \mathrm{Var}(F))$*,*

$$d_{TV}(F, N) \leqslant \frac{E|S|}{\lambda}.$$

To see how to combine Proposition 14.1.2 with Theorem 14.1.3 (see also point(ii) in Remark 14.5.1), consider indeed a multiple Wiener-Itô integral of the form $F = I_p^B(f)$, with $\sigma^2 = E[F^2] > 0$. Assume moreover that we have at our disposal a family $\{(B, B^t)\}_{t\geqslant 0}$ of exchangeable pairs of Brownian motions, satisfying the assumption (a) in Proposition 14.1.2. Then, putting Proposition 14.1.2 and Theorem 14.1.3 together immediately yields that

$$d_{TV}(F, N) \leqslant \frac{2}{\sigma^2} E\left[\left| p \int_0^1 I_{p-1}^B(f(x, \cdot))^2 dx - \sigma^2 \right|\right]. \tag{14.1}$$

Finally, to obtain the inequality stated Theorem 14.1.1(ii) from (14.1), it remains to 'play' cleverly with the (elementary) product formula (14.7), see Proposition 14.7.1 for the details.

To conclude our elementary proof of Theorem 14.1.1(ii), we are thus left to construct the family $\{(B, B^t)\}_{t>0}$. Actually, we will offer two constructions with different motivations: the first one is inspired by Mehler's formula from Gaussian analysis, whereas the second one is more in the spirit of the so-called *Gibbs sampling procedure* within Stein's method (see e.g. [5, A.2]).

For the first construction, we consider two independent Brownian motions on $[0, 1]$ defined on the same probability space $(\Omega, \mathcal{F}, P)$, namely $B$ and $\widehat{B}$. We interpolate between them by considering, for any $t \geqslant 0$,

$$B^t = e^{-t} B + \sqrt{1 - e^{-2t}}\,\widehat{B}.$$

It is then easy and straightforward to check that, for any $t \geqslant 0$, this new Brownian motion $B^t$, together with $B$, forms an exchangeable pair (see Lemma 14.3.1). Moreover, we will compute below (see (14.10)) that $E\big[I_p^{B^t}(f)\big|\sigma\{B\}\big] = e^{-pt} I_p^B(f)$ for any $p \geqslant 1$ and any $f \in L^2([0, 1]^p)$, from which (a) in Proposition 14.1.2 immediately follows.

For the second construction, we consider two independent Gaussian white noise $W$ and $W'$ on $[0, 1]$ with Lebesgue intensity measure. For each $n \in \mathbb{N}$, we introduce a uniform partition $\{\Delta_1, \ldots, \Delta_n\}$ and a uniformly distributed index $I_n \sim \mathscr{U}_{\{1,\ldots,n\}}$, independent of $W$ and $W'$. For every Borel set $A \subset [0, 1]$, we define $W^n(A) = W'(A \cap \Delta_{I_n}) + W(A \setminus \Delta_{I_n})$. This will give us a new Gaussian white noise $W^n$, which will form an exchangeable pair with $W$. This construction is a particular Gibbs sampling procedure. The analogue of (a) in Proposition 14.1.2 is satisfied, namely, if $f \in L^2([0, 1]^p)$, $F = I_p^W(f)$ is the $p$th multiple integral with respect to $W$ and $F^{(n)} = I_p^{W^n}(f)$, we have

$$nE\big[F^{(n)} - F\big|\sigma\{W\}\big] \to -pF \quad \text{in } L^2(\Omega) \text{ as } n \to \infty.$$

To apply Theorem 14.1.3 in this setting, we only need to replace $\frac{1}{t}$ by $n$ and replace $F_t$ by $F^{(n)}$. To get the exchangeable pairs $(B, B^n)$ of Brownian motions in this setting, it suffices to consider $B(t) = W([0, t])$ and $B^n(t) = W^n([0, t])$, $t \in [0, 1]$. See Sect. 14.4 for more precise statements.

Finally, we discuss the extension of our exchangeable pair approach on Wiener chaos to the multidimensional case. Here again, it works perfectly well, and it allows us to recover the (known) rate of convergence associated with the remarkable Peccati-Tudor theorem [20]. This latter represents a multidimensional counterpart of the fourth moment theorem Theorem 14.1.1(i), exhibiting conditions involving only the second and fourth moments that ensure a central limit theorem for random *vectors* with chaotic components.

**Theorem 14.1.4 (Peccati, Tudor [20])** *Fix $d \geqslant 2$ and $p_1, \dots, p_d \geqslant 1$. For each $k \in \{1, \dots, d\}$, let $(F_n^k)_{n\geqslant 1}$ be a sequence of multiple Wiener-Itô integrals of order $p_k$. Assume that $E[F_n^k F_n^l] \to \sigma_{kl}$ as $n \to \infty$ for each pair $(k, l) \in \{1, \dots, d\}^2$, with $\Sigma = (\sigma_{kl})_{1\leqslant k,l\leqslant d}$ non-negative definite. Then, as $n \to \infty$,*

$$F_n = (F_n^1, \dots, F_n^d) \overset{\text{law}}{\to} N \sim N(0, \Sigma) \quad \Longleftrightarrow \quad E[(F_n^k)^4] \to 3\sigma_{kk}^2 \text{ for all } k \in \{1, \dots, d\}. \tag{14.2}$$

In [16], it is shown that the right-hand side of (14.2) is also equivalent to

$$E[\|F_n\|^4] \to E[\|N\|^4] \quad \text{as } n \to \infty, \tag{14.3}$$

where $\|\cdot\|$ stands for the usual Euclidean $\ell^2$-norm of $\mathbb{R}^d$. Combining the main findings of [17] and [16] yields the following quantitative version associated to Theorem 14.1.4, which we are able to recover by means of our elementary exchangeable approach.

**Theorem 14.1.5 (Nourdin, Peccati, Réveillac, Rosiński [16, 17])** *Let $F = (F^1, \dots, F^d)$ be a vector composed of multiple Wiener-Itô integrals $F^k$, $1 \leqslant k \leqslant d$. Assume that the covariance matrix $\Sigma$ of $F$ is invertible. Then, with $N \sim N(0, \Sigma)$,*

$$d_W(F, N) \leqslant \|\Sigma\|_{op}^{\frac{1}{2}} \|\Sigma^{-1}\|_{op} \sqrt{E[\|F\|^4] - E[\|N\|^4]}, \tag{14.4}$$

*where $d_W$ denotes the Wasserstein distance and $\|\cdot\|_{op}$ the operator norm of a matrix.*

The currently available proof of (14.4) relies on two main ingredients: (1) simple manipulations involving the product formula (14.7) and implying that

$$\sum_{i,j=1}^{d} \mathrm{Var}\Big(p_j \int_0^1 I_{p_i-1}(f_i(x,\cdot)) I_{p_j-1}(f_j(x,\cdot)) dx\Big) \leqslant E[\|F\|^4] - E[\|N\|^4],$$

(see [16, Theorem 4.3] for the details) and (2) the following inequality shown in [17, Corollary 3.6] by means of the Malliavin operators $D$, $\delta$ and $L$:

$$d_W(F,N) \leqslant \|\Sigma\|_{op}^{\frac{1}{2}} \|\Sigma^{-1}\|_{op} \sqrt{\sum_{i,j=1}^{d} \mathrm{Var}\Big(p_j \int_0^1 I_{p_i-1}(f_i(x,\cdot)) I_{p_j-1}(f_j(x,\cdot)) dx\Big)}. \tag{14.5}$$

Here, in the spirit of what we have done in dimension one, we also apply our elementary exchangeable pairs approach to prove (14.5), with slightly different constants.

The rest of the paper is organized as follows. Section 14.2 contains preliminary knowledge on multiple Wiener-Itô integrals. In Sect. 14.3 (resp. 14.4), we present our first (resp. second) construction of exchangeable pairs of Brownian motions, and we give the main associated properties. Section 14.5 is devoted to the proof of Proposition 14.1.2, whereas in Sect. 14.6 we offer a simple proof of Meckes' Theorem 14.1.3. Our new, elementary proof of Theorem 14.1.1(ii) is given in Sect. 14.7. In Sect. 14.8, we further investigate the connections between our exchangeable pairs and the Malliavin operators. Finally, we discuss the extension of our approach to the multidimensional case in Sect. 14.9.

## 14.2 Multiple Wiener-Itô Integrals: Definition and Elementary Properties

In this subsection, we recall the definition of multiple Wiener-Itô integrals, and then we give a few *soft* properties that will be needed for our new proof of Theorem 14.1.1(ii). We refer to the classical monograph [18] for the details and missing proofs.

Let $f : [0,1]^p \to \mathbb{R}$ be a square-integrable function, with $p \geqslant 1$ a given integer. The $p$th multiple Wiener-Itô integral of $f$ with respect to the Brownian motion $B = \big(B(x)\big)_{x\in[0,1]}$ is *formally* written as

$$\int_{[0,1]^p} f(x_1,\ldots,x_p) dB(x_1)\ldots dB(x_p). \tag{14.6}$$

To give a precise meaning to (14.6), Itô's crucial idea from the fifties was to first define (14.6) for elementary functions that vanish on diagonals, and then to approximate any $f$ in $L^2([0,1]^p)$ by such elementary functions.

Consider the diagonal set of $[0,1]^p$, that is, $D = \{(t_1,\ldots,t_p) \in [0,1]^p : \exists i \neq j,\ t_i = t_j\}$. Let $\mathcal{E}_p$ be the vector space formed by the set of elementary functions on

$[0,1]^p$ that vanish over $D$, that is, the set of those functions $f$ of the form

$$f(x_1,\ldots,x_p)=\sum_{i_1,\ldots,i_p=1}^{k}\beta_{i_1\ldots i_p}\mathbf{1}_{[\tau_{i_1-1},\tau_{i_1})\times\ldots\times[\tau_{i_p-1},\tau_{i_p})}(x_1,\ldots,x_p),$$

where $k\geqslant 1$ and $0=\tau_0<\tau_1<\ldots<\tau_k$, and the coefficients $\beta_{i_1\ldots i_p}$ are zero if any two of the indices $i_1,\ldots,i_p$ are equal. For $f\in\mathcal{E}_p$, we define (without ambiguity with respect to the choice of the representation of $f$)

$$I_p^B(f)=\sum_{i_1,\ldots,i_p=1}^{k}\beta_{i_1\ldots i_p}(B(\tau_{i_1})-B(\tau_{i_1-1}))\ldots(B(\tau_{i_p})-B(\tau_{i_p-1})).$$

We also define the symmetrization $\widetilde{f}$ of $f$ by

$$\widetilde{f}(x_1,\ldots,x_p)=\frac{1}{p!}\sum_{\sigma\in\mathfrak{S}_p}f(x_{\sigma(1)},\ldots,x_{\sigma(p)}),$$

where $\mathfrak{S}_p$ stands for the set of all permutations of $\{1,\ldots,p\}$. The following elementary properties are immediate and easy to prove.

1. If $f\in\mathcal{E}_p$, then $I_p^B(f)=I_p^B(\widetilde{f})$.
2. If $f\in\mathcal{E}_p$ and $g\in\mathcal{E}_q$, then $E[I_p^B(f)]=0$ and

$$E[I_p^B(f)I_q^B(g)]=\begin{cases}0 & \text{if } p\neq q\\ p!\langle\widetilde{f},\widetilde{g}\rangle_{L^2([0,1]^p)} & \text{if } p=q\end{cases}.$$

3. The space $\mathcal{E}_p$ is dense in $L^2([0,1]^p)$. In other words, to each $f\in L^2([0,1]^p)$ one can associate a sequence $(f_n)_{n\geqslant 1}\subset\mathcal{E}_p$ such that $\|f-f_n\|_{L^2([0,1]^p)}\to 0$ as $n\to\infty$.
4. Since

$$\begin{aligned}E\big[(I_p^B(f_n)-I_p^B(f_m))^2\big]&=p!\|\widetilde{f_n}-\widetilde{f_m}\|^2_{L^2([0,1]^p)}\\&\leq p!\|f_n-f_m\|^2_{L^2([0,1]^p)}\to 0\end{aligned}$$

as $n,m\to\infty$ for $f$ and $(f_n)_{n\geqslant 1}$ as in the previous point 3, we deduce that the sequence $(I_p(f_n))_{n\geqslant 1}$ is Cauchy in $L^2(\Omega)$ and, as such, it admits a limit denoted by $I_p^B(f)$. It is easy to check that $I_p^B(f)$ only depends on $f$, not on the particular choice of the approximating sequence $(f_n)_{n\geqslant 1}$, and that points 1 to 3 continue to hold for general $f\in L^2([0,1]^p)$ and $g\in L^2([0,1]^q)$.

We will also crucially rely on the following *product formula*, whose proof is elementary and can be made by induction. See, e.g., [18, Proposition 1.1.3].

5. For any $p, q \geqslant 1$, and if $f \in L^2([0,1]^p)$ and $g \in L^2([0,1]^q)$ are symmetric, then

$$I_p^B(f) I_q^B(g) = \sum_{r=0}^{p \wedge q} r! \binom{p}{r} \binom{q}{r} I_{p+q-2r}^B(f \otimes_r g), \tag{14.7}$$

where $f \otimes_r g$ stands for the $r$th-contraction of $f$ and $g$, defined as an element of $L^2([0,1]^{p+q-2r})$ by

$$(f \otimes_r g)(x_1, \ldots, x_{p+q-2r})$$
$$= \int_{[0,1]^r} f(x_1, \ldots, x_{p-r}, u_1, \ldots, u_r) g(x_{p-r+1}, \ldots, x_{p+q-2r}, u_1, \ldots, u_r) du_1 \ldots du_r.$$

Product formula (14.7) has a nice consequence, the inequality (14.8) below. It is a very particular case of a more general phenomenon satisfied by multiple Wiener-Itô integrals, the *hypercontractivity* property.

6. For any $p \geqslant 1$, there exists a constant $c_{4,p} > 0$ such that, for any (symmetric) $f \in L^2([0,1]^p)$,

$$E\left[I_p^B(f)^4\right] \leqslant c_{4,p}\, E\left[I_p^B(f)^2\right]^2. \tag{14.8}$$

Indeed, thanks to (14.7) one can write $I_p^B(f)^2 = \sum_{r=0}^{p} r! \binom{p}{r}^2 I_{2p-2r}^B(f \otimes_r f)$ so that

$$E[I_p^B(f)^4] = \sum_{r=0}^{p} r!^2 \binom{p}{r}^4 (2p-2r)! \| f \widetilde{\otimes}_r f \|^2_{L^2([0,1]^{2p-2r})}.$$

The conclusion (14.8) follows by observing that

$$p!^2 \| f \widetilde{\otimes}_r f \|^2_{L^2([0,1]^{2p-2r})} \leqslant p!^2 \| f \otimes_r f \|^2_{L^2([0,1]^{2p-2r})} \leqslant p!^2 \| f \|^4_{L^2([0,1]^p)} = E[I_p^B(f)^2]^2.$$

Furthermore, for each $n \geqslant 2$, using (14.7) and induction, one can show that, with $c_{2^n,p}$ a constant depending only on $p$ but not on $f$,

$$E\left[I_p^B(f)^{2^n}\right] \leqslant c_{2^n,p}\, E\left[I_p^B(f)^2\right]^{2^{n-1}}.$$

So for any $r > 2$, there exists an absolute constant $c_{r,p}$ depending only on $p, r$ (but not on $f$) such that

$$E\left[|I_p^B(f)|^r\right] \leqslant c_{r,p}\, E\left[I_p^B(f)^2\right]^{r/2}. \tag{14.9}$$

## 14.3 Exchangeable Pair of Brownian Motions: A First Construction

As anticipated in the introduction, for this construction we consider two independent Brownian motions on $[0, 1]$ defined on the same probability space $(\Omega, \mathcal{F}, P)$, namely $B$ and $\widehat{B}$, and we interpolate between them by considering, for any $t \geqslant 0$, $B^t = e^{-t}B + \sqrt{1-e^{-2t}}\widehat{B}$.

**Lemma 14.3.1** *For each $t \geqslant 0$, the pair $(B, B^t)$ is exchangeable, that is, $(B, B^t) \overset{\text{law}}{=} (B^t, B)$. In particular, $B^t$ is a Brownian motion.*

*Proof* Clearly, the bi-dimensional process $(B, B^t)$ is Gaussian and centered. Moreover, for any $x, y \in [0, 1]$,

$$E[B^t(x)B^t(y)] = e^{-2t}E[B(x)B(y)] + (1-e^{-2t})E[\widehat{B}(x)\widehat{B}(y)] = E[B(x)B(y)]$$
$$E[B(x)B^t(y)] = e^{-t}E[B(x)B(y)] = E[B^t(x)B(y)].$$

The desired conclusion follows. □

We can now state that, as written in the introduction, our exchangeable pair indeed satisfies the crucial property (a) of Proposition 14.1.2.

**Theorem 14.3.2** *Let $p \geqslant 1$ be an integer, and consider a kernel $f \in L^2([0, 1]^p)$. Set $F = I_p^B(f)$ and $F_t = I_p^{B^t}(f)$, $t \geqslant 0$. Then,*

$$E\big[F_t\big|\sigma\{B\}\big] = e^{-pt}F. \tag{14.10}$$

*In particular, convergence (a) in Proposition 14.1.2 takes place:*

$$\lim_{t\downarrow 0}\frac{1}{t}E\Big[I_p^{B^t}(f) - I_p^B(f)\big|\sigma\{B\}\Big] = -p\,I_p^B(f) \quad \text{in } L^2(\Omega). \tag{14.11}$$

*Proof* Consider first the case where $f \in \mathcal{E}_p$, that is, $f$ has the form

$$f(x_1, \ldots, x_p) = \sum_{i_1,\ldots,i_p=1}^{k} \beta_{i_1\ldots i_p}\mathbf{1}_{[\tau_{i_1-1},\tau_{i_1})\times\ldots\times[\tau_{i_p-1},\tau_{i_p})}(x_1, \ldots, x_p),$$

with $k \geqslant 1$ and $0 = \tau_0 < \tau_1 < \ldots < \tau_k$, and the coefficients $\beta_{i_1\ldots i_p}$ are zero if any two of the indices $i_1, \ldots, i_p$ are equal. We then have

$$\begin{aligned}F_t &= \sum_{i_1,\ldots,i_p=1}^{k} \beta_{i_1\ldots i_p}(B^t(\tau_{i_1}) - B^t(\tau_{i_1-1}))\ldots(B^t(\tau_{i_p}) - B^t(\tau_{i_p-1}))\\ &= \sum_{i_1,\ldots,i_p=1}^{k} \beta_{i_1\ldots i_p}\big[e^{-t}(B(\tau_{i_1}) - B(\tau_{i_1-1})) + \sqrt{1-e^{-2t}}(\widehat{B}(\tau_{i_1}) - \widehat{B}(\tau_{i_1-1}))\big]\\ &\qquad\times\ldots\times\big[e^{-t}(B(\tau_{i_1}) - B(\tau_{i_1-1})) + \sqrt{1-e^{-2t}}(\widehat{B}(\tau_{i_p}) - \widehat{B}(\tau_{i_p-1}))\big].\end{aligned}$$

Expanding and integrating with respect to $\widehat{B}$ yields (14.10) for elementary $f$. Thanks to point 4 in Sect. 14.2, we can extend it to any $f \in L^2([0,1]^p)$. Indeed, given a general kernel $f \in L^2([0,1]^p)$, there exists a sequence $\{g_m, m \geqslant 1\}$ of simple functions such that $\|g_m - f\|_{L^2([0,1]^p)} \to 0$, as $m \to +\infty$; and this implies $E\big\{[I_p^{B^t}(f) - I_p^{B^t}(g_m)]^2\big\} = p!\|g_m - f\|^2_{L^2([0,1]^p)} \to 0$, as $m \to +\infty$. Since the conditional expectation $E\big[\,\cdot\,|\sigma\{B\}\big]$ is a bounded linear operator in $L^2(\Omega)$, we have

$$E\big[I_p^{B^t}(f)|\sigma\{B\}\big] = L^2\text{-}\lim_{m\to+\infty} E\big[I_p^{B^t}(g_m)|\sigma\{B\}\big] = L^2\text{-}\lim_{m\to+\infty} e^{-pt} I_p^B(g_m) = I_p^B(f)\,.$$

This concludes the proof of (14.10). We then deduce that

$$\frac{1}{t}\,E\big[F_t - F\big|\sigma\{B\}\big] = \frac{e^{-pt}-1}{t}\,F,$$

from which (14.11) now follows immediately, as $F \in L^2(\Omega)$ and $\dfrac{e^{-pt}-1}{t} \to -p$ when $t \downarrow 0$. □

## 14.4 Exchangeable Pair of Brownian Motions: A Second Construction

In this section, we present yet another construction of exchangeable pairs via Gaussian white noise. We believe it is of independent interest, as such a construction can be similarly carried out for other additive noises. This part may be skipped in a first reading, as it is not used in other sections. And we assume that the readers are familiar with the multiple Wiener-Itô integrals with respect to the Gaussian white noise, and refer to [18, pp. 8–13] for all missing details.

Let $W$ be a Gaussian white noise on $[0,1]$ with Lebesgue intensity measure $\nu$, that is, $W$ is a centred Gaussian process indexed by Borel subsets of $[0,1]$ such that for any Borel sets $A, B \subset [0,1]$, $W(A) \sim N\big(0, \nu(A)\big)$ and $E\big[W(A)W(B)\big] = \nu(A\cap B)$. We denote by $\mathscr{G} := \sigma\{W\}$ the $\sigma$-algebra generated by $\big\{W(A)\colon A \text{ Borel subset of } [0,1]\big\}$. Now let $W'$ be an independent copy of $W$ (denote by $\mathscr{G} = \sigma\{W'\}$ the $\sigma$-algebra generated by $W'$) and $I_n$ be a uniform random variable over $\{1,\ldots,n\}$ for each $n \in \mathbb{N}$ such that $I_n$, $W$, $W'$ are independent. For each fixed $n \in \mathbb{N}$, we consider the partition $[0,1] = \bigcup_{j=1}^n \Delta_j$ with $\Delta_1 = [0, \frac{1}{n}]$, $\Delta_2 = (\frac{1}{n}, \frac{2}{n}]$, $\ldots$, $\Delta_n = (1-\frac{1}{n}, 1]$.

**Definition 14.4.1** Set $W^n(A) := W'\big(A\cap\Delta_{I_n}\big) + W\big(A\setminus\Delta_{I_n}\big)$ for any Borel set $A \subset [0,1]$.

*Remark 14.4.2* One can first treat $W$ as the superposition of $\big\{W|_{\Delta_j}, j = 1,\ldots,n\big\}$, where $W|_{\Delta_j}$ denotes the Gaussian white noise on $\Delta_j$. Then according to $I_n = j$, we (only) replace $W|_{\Delta_j}$ by an independent copy $W'|_{\Delta_j}$ so that we get $W^n$. This is nothing else but a particular Gibbs sampling procedure (see [5, A.2]), hence heuristically speaking, the new process $W^n$ shall form an exchangeable pair with $W$.

**Lemma 14.4.3** *$W$ and $W^n$ form an exchangeable pair with $W$, that is,* $(W, W^n) \stackrel{law}{=} (W^n, W)$. *In particular, $W^n$ is a Gaussian white noise on* $[0, 1]$ *with Lebesgue intensity measure.*

*Proof* Let us first consider $m$ mutually disjoint Borel sets $A_1, \ldots, A_m \subset [0, 1]$. Given $D_1, D_2$ Borel subsets of $\mathbb{R}^m$, we have

$$P\Big(\big(W(A_1), \ldots, W(A_m)\big) \in D_1\,,\ \big(W^n(A_1), \ldots, W^n(A_m)\big) \in D_2\Big)$$
$$= \sum_{v=1}^{n} P\Big(\big(W(A_1), \ldots, W(A_m)\big) \in D_1\,,\ \big(W^n(A_1), \ldots, W^n(A_m)\big) \in D_2\,,\ I_n = v\Big)$$
$$= \frac{1}{n} \sum_{v=1}^{n} P\Big(g(X_v, Y_v) \in D_1, g(X'_v, Y_v) \in D_2\Big)\ ,$$

where for each $v \in \{1, \ldots, n\}$,

- $X_v := \big(W(A_1 \cap \Delta_v), \ldots, W(A_m \cap \Delta_v)\big)$, $X'_v := \big(W'(A_1 \cap \Delta_v), \ldots, W'(A_m \cap \Delta_v)\big)$,
- $Y_v := \big(W(A_1 \setminus \Delta_v), \ldots, W(A_m \setminus \Delta_v)\big)$, and $g$ is a function from $\mathbb{R}^{2m}$ to $\mathbb{R}^m$ given by $(x_1, \ldots, x_m, y_1, \ldots, y_m) \mapsto g\big(x_1, \ldots, x_m, y_1, \ldots, y_m\big) = \big(x_1 + y_1, \ldots, x_m + y_m\big)$

It is clear that for each $v \in \{1, \ldots, n\}$, $X_v, X'_v$ and $Y_v$ are independent, therefore $g(X_v, Y_v)$ and $g(X'_v, Y_v)$ form an exchangeable pair. It follows from the above equalities that

$$P\Big(\big(W(A_1), \ldots, W(A_m)\big) \in D_1\,,\ \big(W^n(A_1), \ldots, W^n(A_m)\big) \in D_2\Big)$$
$$= \frac{1}{n} \sum_{v=1}^{n} P\Big(g(X'_v, Y_v) \in D_1, g(X_v, Y_v) \in D_2\Big)$$
$$= P\Big(\big(W^n(A_1), \ldots, W^n(A_m)\big) \in D_1\,,\ \big(W(A_1), \ldots, W(A_m)\big) \in D_2\Big)\ .$$

This proves the exchangeability of $\big(W(A_1), \ldots, W(A_m)\big)$ and $\big(W^n(A_1), \ldots, W^n(A_m)\big)$.

Now let $B_1, \ldots, B_m$ be Borel subsets of $[0, 1]$, then one can find mutually disjoint Borel sets $A_1, \ldots, A_p$ (for some $p \in \mathbb{N}$) such that each $B_j$ is a union of some of $A_i$'s. Therefore we can find some measurable $\phi : \mathbb{R}^p \to \mathbb{R}^m$ such that $\big(W(B_1), \ldots, W(B_m)\big) = \phi\big(W(A_1), \ldots, W(A_p)\big)$. Accordingly, $\big(W^n(B_1), \ldots, W^n(B_m)\big) = \phi\big(W^n(A_1), \ldots, W^n(A_p)\big)$, hence

$\big(W(B_1), \ldots, W(B_m)\big)$ and $\big(W^n(B_1), \ldots, W^n(B_m)\big)$ are exchangeable. Now our proof is complete. □

*Remark 14.4.4* For each $t \in [0,1]$, we set $B(t) := W([0,t])$ and $B^n(t) := W^n([0,t])$. Modulo continuous modifications, one can see from Lemma 14.4.3 that $B$, $B^n$ are two Brownian motions that form an exchangeable pair. An important difference between this construction and the previous one is that $(B, B^t)$ is bi-dimensional Gaussian process whereas $B$, $B^n$ are not *jointly* Gaussian.

Before we state the analogous result to Theorem 14.3.2, we briefly recall the construction of multiple Wiener-Itô integrals in white noise setting.

1. For each $p \in \mathbb{N}$, we denote by $\mathscr{E}_p$ the set of simple functions of the form

$$f\big(t_1, \ldots, t_p\big) = \sum_{i_1,\ldots,i_p=1}^{m} \beta_{i_1\ldots i_p} \mathbf{1}_{A_{i_1}\times\ldots\times A_{i_p}}\big(t_1, \ldots, t_p\big) , \tag{14.12}$$

   where $m \in \mathbb{N}$, $A_1, \ldots, A_m$ are pair-wise disjoint Borel subsets of $[0,1]$, and the coefficients $\beta_{i_1\ldots i_p}$ are zero if any two of the indices $i_1, \ldots i_p$ are equal. It is known that $\mathscr{E}_p$ is dense in $L^2([0,1]^p)$.
2. For $f$ given as in (14.12), the $p$th multiple integral with respect to $W$ is defined as

$$I_p^W(f) := \sum_{i_1,\ldots,i_p=1}^{m} \beta_{i_1\ldots i_p} W(A_{i_1}) \ldots W(A_{i_p}) ,$$

   and one can extend $I_p^W$ to $L^2([0,1]^p)$ via usual approximation argument. Note $I_p^W(f)$ is nothing else but $I_p^B(f)$ with the Brownian motion $B$ constructed in Remark 14.4.4.

**Theorem 14.4.5** *If $F = I_p^W(f)$ for some symmetric $f \in L^2([0,1]^p)$ and we set $F^{(n)} := I_p^{W^n}(f)$, then in $L^2(\Omega, \mathscr{G}, P)$ and as $n \to +\infty$, $n\, E\big[F^{(n)} - F\big|\mathscr{G}\big] \to -pF$.*

*Proof* First we consider the case where $f \in \mathscr{E}_p$, we assume moreover that $F = \prod_{j=1}^p W(A_j)$ with $A_1, \ldots, A_p$ mutually disjoint Borel subsets of $[0,1]$, and accordingly we define $F^{(n)} = \prod_{j=1}^p W^n(A_j)$. Then, (we write $[p] = \{1, \ldots, p\}$, $A^v = A \cap \Delta_v$ for any $A \subset [0,1]$ and $v \in \{1, \ldots, n\}$)

$$\begin{aligned} n\, E\big[F^{(n)}\big|\mathscr{G}\big] &= n\, E\left\{ \sum_{v=1}^{n} \mathbf{1}_{\{I_n=v\}} \prod_{j=1}^{p} \big[W'(A_j^v) + W(A_j \setminus \Delta_v)\big] \,|\mathscr{G} \right\} \\ &= \sum_{v=1}^{n} E\left\{ \prod_{j=1}^{p} \big[W'(A_j^v) + W(A_j \setminus \Delta_v)\big] \,|\mathscr{G} \right\} = \sum_{v=1}^{n} \prod_{j=1}^{p} W\big(A_j \setminus \Delta_v\big) \end{aligned}$$

$$= \sum_{v=1}^{n} \Bigg\{ \left( \prod_{j=1}^{p} W(A_j) \right) - \sum_{k=1}^{p} W(A_k^v) \left( \prod_{j\in[p]\setminus\{k\}} W(A_j) \right) + \sum_{\ell=2}^{p} (-1)^\ell \sum_{\substack{k_1,\dots,k_\ell\in[p]\\ \text{all distinct}}} \left( \prod_{j\in[p]\setminus\{k_1,\dots,k_\ell\}} W(A_j) \right) W\big(A_{k_1}^v\big)\cdots W\big(A_{k_\ell}^v\big) \Bigg\}$$

$$= n\,F - p\,F + R_n(F)\ ,$$

where $$R_n(F) = \sum_{\ell=2}^{p} (-1)^\ell \sum_{\substack{k_1,\dots,k_\ell\in[p]\\ \text{all distinct}}} \left( \prod_{j\in[p]\setminus\{k_1,\dots,k_\ell\}} W(A_j) \right) \sum_{v=1}^{n} W\big(A_{k_1}^v\big)\cdots W\big(A_{k_\ell}^v\big).$$

Then $R_n(F)$ converges in $L^2(\Omega, \mathscr{G}, P)$ to 0, due to the fact that $\sum_{v=1}^{n} \prod_{i=1}^{q} W(A_{k_i}^v)$ converges in $L^2(\Omega)$ to 0, as $n \to +\infty$, if $q \geqslant 2$ and all $k_i$'s are distinct numbers. This proves our theorem when $f \in \mathscr{E}_p$.

By the above computation, we can see that if $F = I_p^W(f)$ with $f$ given in (14.12), then

$$R_n(F) = \sum_{i_1,\dots,i_p=1}^{m} \beta_{i_1 i_2\dots i_p} \sum_{\ell=2}^{p} (-1)^\ell \sum_{\substack{k_1,\dots,k_\ell=1\\ \text{all distinct}}}^{p} \left( \prod_{j\in[p]\setminus\{k_1,\dots,k_\ell\}} W(A_{i_j}) \right) \sum_{v=1}^{n} W\big(A_{i_{k_1}}^v\big)\cdots W\big(A_{i_{k_\ell}}^v\big).$$

Therefore, using Wiener-Itô isometry, we can first write $\big\|R_n(F)\big\|_{L^2(\Omega)}^2$ as

$$p! \sum_{i_1,\dots,i_p=1}^{m} \big(\beta_{i_1 i_2\dots i_p}\big)^2 \sum_{v=1}^{n} \left\| \sum_{\ell=2}^{p} (-1)^\ell \sum_{\substack{k_1,\dots,k_\ell\in[p]\\ \text{all distinct}}} \left( \prod_{j\in[p]\setminus\{k_1,\dots,k_\ell\}} W(A_{i_j}) \right) W\big(A_{i_{k_1}}^v\big)\cdots W\big(A_{i_{k_\ell}}^v\big) \right\|_{L^2(\Omega)}^2 ,$$

and then using the elementary inequality $(a_1 + \ldots + a_m)^\beta \leqslant m^{\beta-1} \sum_{i=1}^{m} |a_i|^\beta$ for $a_i \in \mathbb{R}$, $\beta > 1$, $m \in \mathbb{N}$, we have

$$\left\| \sum_{\ell=2}^{p} (-1)^\ell \sum_{\substack{k_1,\dots,k_\ell\in[p]\\ \text{all distinct}}} \left( \prod_{j\in[p]\setminus\{k_1,\dots,k_\ell\}} W(A_{i_j}) \right) W\big(A_{i_{k_1}}^v\big)\cdots W\big(A_{i_{k_\ell}}^v\big) \right\|_{L^2(\Omega)}^2$$

$$\leqslant \Theta_1 \sum_{\ell=2}^{p} \sum_{\substack{k_1,\dots,k_\ell\in[p]\\ \text{all distinct}}} \left\| \left( \prod_{j\in[p]\setminus\{k_1,\dots,k_\ell\}} W(A_{i_j}) \right) W\big(A_{i_{k_1}}^v\big)\cdots W\big(A_{i_{k_\ell}}^v\big) \right\|_{L^2(\Omega)}^2$$

$$= \Theta_1 \sum_{\ell=2}^{p} \sum_{\substack{k_1,\ldots,k_\ell \in [p] \\ \text{all distinct}}} \left( \prod_{j \in [p] \setminus \{k_1,\ldots,k_\ell\}} \nu(A_{i_j}) \right) \nu\big(A^v_{i_{k_1}}\big) \cdots \nu\big(A^v_{i_{k_\ell}}\big)$$

$$\leqslant \Theta_2 \sum_{\substack{k_1,k_2 \in [p] \\ k_1 \neq k_2}} \left( \prod_{j \in [p] \setminus \{k_1,k_2\}} \nu(A_{i_j}) \right) \nu\big(A^v_{i_{k_1}}\big)\, \nu\big(A^v_{i_{k_2}}\big)$$

where $\Theta_1, \Theta_2$ (and $\Theta_3$ in the following) are some absolute constants that do not depend on $n$ or $F$. Note now for $k_1 \neq k_2$, $\sum_{v=1}^n \nu\big(A^v_{i_{k_1}}\big) \cdot \nu\big(A^v_{i_{k_2}}\big) \leqslant \nu\big(A_{i_{k_1}}\big) \sum_{v=1}^n \nu\big(A^v_{i_{k_2}}\big) = \nu\big(A_{i_{k_1}}\big) \cdot \nu\big(A_{i_{k_2}}\big)$, thus,

$$\|R_n(F)\|^2_{L^2(\Omega)} \leqslant p! \sum_{i_1,\ldots,i_p=1}^{m} \big(\beta_{i_1 i_2 \ldots i_p}\big)^2 \Theta_2 \sum_{\substack{k_1,k_2 \in [p] \\ k_1 \neq k_2}} \left( \prod_{j \in [p] \setminus \{k_1,k_2\}} \nu(A_{i_j}) \right) \nu\big(A_{i_{k_1}}\big) \nu\big(A_{i_{k_2}}\big)$$

$$\leqslant p! \sum_{i_1,\ldots,i_p=1}^{m} \big(\beta_{i_1 i_2 \ldots i_p}\big)^2 \Theta_3 \prod_{j \in [p]} \nu(A_{i_j}) = \Theta_3 \cdot \|F\|^2_{L^2(\Omega)} \,.$$

Since $\big\{I^W_p(f) \,:\, f \in \mathscr{E}_p\big\}$ is dense in the $p$th Wiener chaos $\mathscr{H}_p$, $R_n : \mathscr{H}_p \to L^2(\Omega)$ is a bounded linear operator with operator norm $\|R_n\|_{\text{op}} \leqslant \sqrt{\Theta_3}$ for each $n \in \mathbb{N}$. Note the linearity follows from its definition $R_n(F) := n\, E\big[F^{(n)} - F \big| \mathscr{G}\big] + pF$, $F \in \mathscr{H}_p$.

Now we define

$$\mathscr{C}_p := \left\{ F \in \mathscr{H}_p \,:\, R_\infty(F) := \lim_{n \to +\infty} R_n(F) \quad \text{is well defined in } L^2(\Omega) \right\}.$$

It is easy to see that $\mathscr{C}_p$ is a dense linear subspace of $\mathscr{H}_p$ and for each $f \in \mathscr{E}_p$, $I^W_p(f) \in \mathscr{C}_p$ and $R_\infty(I^W_p(f)) = 0$. As

$$\sup_{n \in \mathbb{N}} \|R_n\|_{\text{op}} \leqslant \sqrt{\Theta_3} < +\infty \ ,$$

$R_\infty$ has a unique extension to $\mathscr{H}_p$ and by density of $\big\{I^W_p(f) \,:\, f \in \mathscr{E}_p\big\}$ in $\mathscr{H}_p$, $R_\infty(F) = 0$ for each $F \in \mathscr{H}_p$. In other words, for any $F \in \mathscr{H}_p$, $n\, E\big[F^{(n)} - F \big| \mathscr{G}\big]$ converges in $L^2(\Omega)$ to $-pF$, as $n \to +\infty$. □

## 14.5 Proof of Proposition 14.1.2

We now give the proof of Proposition 14.1.2, which has been stated in the introduction. We restate it for the convenience of the reader.

**Proposition 14.1.2** *Let* $(B, B^t)_{t\geqslant 0}$ *be a family of exchangeable pairs of Brownian motions (that is,* $B$ *is a Brownian motion on* $[0,1]$ *and, for each* $t$*, one has* $(B, B^t) \overset{\text{law}}{=} (B^t, B)$*). Assume moreover that*

(a) *for any integer* $p \geqslant 1$ *and any* $f \in L^2([0,1]^p)$*,*

$$\lim_{t\downarrow 0}\frac{1}{t}\,E\Big[I_p^{B^t}(f) - I_p^B(f)\big|\sigma\{B\}\Big] = -p\,I_p^B(f) \quad \textit{in } L^2(\Omega).$$

*Then, for any integer* $p \geqslant 1$ *and any* $f \in L^2([0,1]^p)$*,*

(b) $\displaystyle\lim_{t\downarrow 0}\frac{1}{t}\,E\Big[\big(I_p^{B^t}(f) - I_p^B(f)\big)^2|\sigma\{B\}\Big] = 2p^2\int_0^1 I_{p-1}^B(f(x,\cdot))^2dx \quad \textit{in } L^2(\Omega);$

(c) $\displaystyle\lim_{t\downarrow 0}\frac{1}{t}\,E\Big[\big(I_p^{B^t}(f) - I_p^B(f)\big)^4\Big] = 0.$

*Proof* We first concentrate on the proof of (b). Fix $p \geqslant 1$ and $f \in L^2([0,1]^p)$, and set $F = I_p^B(f)$ and $F_t = I_p^{B^t}(f)$. First, we observe that

$$\frac{1}{t}\,E\big[(F_t - F)^2\big|\sigma\{B\}\big] = \frac{1}{t}\,E\big[F_t^2 - F^2\big|\sigma\{B\}\big] - \frac{2}{t}F\,E\big[F_t - F\big|\sigma\{B\}\big].$$

Also, as an immediate consequence of the product formula (14.7) and the definition of $f \otimes_r f$, we have

$$p^2\int_0^1 I_{p-1}^B(f(x,\cdot))^2dx = \sum_{r=1}^p rr!\binom{p}{r}^2 I_{2p-2r}^B(f\otimes_r f).$$

Given (a) and the previous two identities, in order to prove (b) we are thus left to check that

$$\lim_{t\downarrow 0}\frac{1}{t}\,E\big[F_t^2 - F^2\big|\sigma\{B\}\big] = -2p\,F^2 + 2\sum_{r=1}^p rr!\binom{p}{r}^2 I_{2p-2r}^B(f\otimes_r f) \quad \text{in } L^2(\Omega). \tag{14.13}$$

The product formula (14.7) used for multiple integrals with respect to $B^t$ (resp. $B$) yields

$$F_t^2 = \sum_{r=0}^p r!\binom{p}{r}^2 I_{2p-2r}^{B^t}(f\otimes_r f) \quad \Big(\text{resp. } F^2 = \sum_{r=0}^p r!\binom{p}{r}^2 I_{2p-2r}^B(f\otimes_r f)\Big).$$

Hence it follows from (a) that

$$\begin{aligned}\frac{1}{t}E\big[F_t^2-F^2\big|\sigma\{B\}\big] &= \sum_{r=0}^{p-1} r!\binom{p}{r}^2\frac{1}{t}E[I_{2p-2r}^{B^t}(f\otimes_r f)-I_{2p-2r}^{B}(f\otimes_r f)|\sigma\{B\}]\\ &\longrightarrow \sum_{r=0}^{p-1} r!\binom{p}{r}^2(2r-2p)I_{2p-2r}^{B}(f\otimes_r f)\\ &= -2p(F^2-E[F^2])+2\sum_{r=1}^{p-1} rr!\binom{p}{r}^2 I_{2p-2r}^{B}(f\otimes_r f),\end{aligned}$$

which is exactly (14.13). The proof of (b) is complete.

Let us now turn to the proof of (c). Fix $p \geqslant 1$ and $f \in L^2([0,1]^p)$, and set $F = I_p^B(f)$ and $F_t = I_p^{B^t}(f)$, $t \geqslant 0$. We claim that the pair $(F, F_t)$ is exchangeable for each $t$. Indeed, thanks to point 4 in Sect. 14.2, we first observe that it is enough to check this claim when $f$ belongs to $\mathcal{E}_p$, that is, when $f$ has the form

$$f(x_1,\ldots,x_p)=\sum_{i_1,\ldots,i_p=1}^{k}\beta_{i_1\ldots i_p}\mathbf{1}_{[\tau_{i_1-1},\tau_{i_1})\times\ldots\times[\tau_{i_p-1},\tau_{i_p})}(x_1,\ldots,x_p),$$

with $k \geqslant 1$ and $0=\tau_0<\tau_1<\ldots<\tau_k$, and the coefficients $\beta_{i_1\ldots i_p}$ are zero if any two of the indices $i_1,\ldots,i_p$ are equal. But, for such an $f$, one has

$$F=I_p^B(f)=\sum_{i_1,\ldots,i_p=1}^{k}\beta_{i_1\ldots i_p}(B(\tau_{i_1})-B(\tau_{i_1-1}))\ldots(B(\tau_{i_p})-B(\tau_{i_p-1}))$$

$$F_t=I_p^{B^t}(f)=\sum_{i_1,\ldots,i_p=1}^{k}\beta_{i_1\ldots i_p}(B^t(\tau_{i_1})-B^t(\tau_{i_1-1}))\ldots(B^t(\tau_{i_p})-B^t(\tau_{i_p-1})),$$

and the exchangeability of $(F, F_t)$ follows immediately from those of $(B, B^t)$. Since the pair $(F, F_t)$ is exchangeable, we can write

$$\begin{aligned}E\big[(F_t-F)^4\big] &= E\big[F_t^4+F^4-4F_t^3F-4F^3F_t+6F_t^2F^2\big]\\ &= 2E[F^4]-8E\big[F^3F_t\big]+6E\big[F^2F_t^2\big] \quad \text{by exchangeability;}\\ &= 4E\big[F^3(F_t-F)\big]+6E\big[F^2(F_t-F)^2\big] \quad \text{after rearrangement;}\\ &= 4E\big[F^3E[(F_t-F)|\sigma\{B\}]\big]+6E\big[F^2E[(F_t-F)^2|\sigma\{B\}]\big].\end{aligned}$$

Dividing by $t$ and taking the limit $t \downarrow 0$ into the previous identity, we deduce, thanks to (a) and (b) as well, that

$$\lim_{t\downarrow 0} \frac{1}{t}\, E\Big[\big(F_t - F\big)^4\Big] = -4pE[F^4] + 12p^2\, E\left[F^2 \int_0^1 I_{p-1}^B(f(x,\cdot))^2 dx\right]. \tag{14.14}$$

In particular, it appears that the limit of $\frac{1}{t}\, E\big[(F_t - F)^4\big]$ is always the same, irrespective of the choice of our exchangeable pair of Brownian motions $(B, B^t)$ satisfying (a). To compute it, we can then choose the pair $(B, B^t)$ we want, for instance, the pair constructed in Sect. 14.3. This is why, starting from now and for the rest of the proof, $(B, B^t)$ refers to the pair defined in Sect. 14.3 (which satisfies (a), that is, (14.11)). What we gain by considering this particular pair is that it satisfies a hypercontractivity-type inequality. More precisely, there exists $c_p > 0$ (only depending on $p$) such that, for all $t \geqslant 0$,

$$E[(F_t - F)^4] \leqslant c_p\, E[(F_t - F)^2]^2. \tag{14.15}$$

Indeed, going back to the definition of multiple Wiener-Itô integrals as given in Sect. 14.2 (first for elementary functions and then by approximation for the general case), we see that $F_t - F$ is a multiple Wiener-Itô integral of order $p$ with respect to the *two-sided* Brownian motion $B = (\overline{B}(s))_{s\in[-1,1]}$, defined as

$$\overline{B}(s) = B(s)\mathbf{1}_{[0,1]}(s) + \widehat{B}(-s)\mathbf{1}_{[-1,0]}(s).$$

But product formula (14.7) is also true for a two-sided Brownian motion, so the claim (14.15) follows from (14.8) applied to $\overline{B}$. On the other hand, it follows from (b) that $\frac{1}{t}\, E\big[(F_t - F)^2\big]$ converges to a finite number, as $t \downarrow 0$. Hence, combining this fact with (14.15) yields

$$\frac{1}{t}\, E\Big[\big(F_t - F\big)^4\Big] \leqslant c_p\, t\, \left(\frac{1}{t}\, E\Big[\big(F_t - F\big)^2\Big]\right)^2 \to 0\,,$$

as $t \downarrow 0$. □

*Remark 14.5.1*

(i) A byproduct of (14.14) in the previous proof is that

$$\frac{1}{3}\big(E[F^4] - 3\sigma^4\big) = E\left[F^2\left(p\int_0^1 I_{p-1}^B(f(x,\cdot))^2 dx - \sigma^2\right)\right]. \tag{14.16}$$

Note (14.16) was originally obtained by chain rule, see [15, equation (5.2.9)].

(ii) As a consequence of (c) in Proposition 14.1.2, we have $\lim_{t\downarrow 0}\frac{1}{t}\,E\big[|I_p^{B^t}(f) - I_p^B(f)|^3\big] = 0$. Indeed,

$$\frac{1}{t}\,E\Big[|I_p^{B^t}(f) - I_p^B(f)|^3\Big] \leqslant \left(\frac{1}{t}\,E\Big[\big(I_p^{B^t}(f) - I_p^B(f)\big)^2\Big]\right)^{\frac{1}{2}}\left(\frac{1}{t}\,E\Big[\big(I_p^{B^t}(f) - I_p^B(f)\big)^4\Big]\right)^{\frac{1}{2}}$$
$$\to 0\,, \qquad \text{as } t\downarrow 0.$$

(iii) For any $r > 2$, in view of (14.9) and (14.15), there exists an absolute constant $c_{r,p}$ depending only on $p, r$ (but not on $f$) such that

$$E\big[|I_p^B(f) - I_p^{B^t}(f)|^r\big] \leqslant c_{r,p}\,E\big[\big(I_p^B(f) - I_p^{B^t}(f)\big)^2\big]^{r/2}\,.$$

Moreover, if $F \in L^2(\Omega, \sigma\{B\}, P)$ admits a *finite* chaos expansion, say, (for some $p \in \mathbb{N}$) $F = E[F] + \sum_{q=1}^{p} I_q^B(f_q)$, and we set $F_t = E[F] + \sum_{q=1}^{p} I_q^{B^t}(f_q)$, then there exists some absolute constant $C_{r,p}$ that only depends on $p$ and $r$ such that

$$E\big[|F - F_t|^r\big] \leqslant C_{r,p}\,E\big[\big(F - F_t\big)^2\big]^{r/2}\,.$$

## 14.6 Proof of E. Meckes' Theorem 14.1.3

In this section, for sake of completeness and because our version slightly differs from the original one given in [11, Theorem 2.1], we provide a proof of Theorem 14.1.3, which we restate here for convenience.

**Theorem 14.1.3 (Meckes [11])** *Let $F$ and a family of random variables $(F_t)_{t\geqslant 0}$ be defined on a common probability space $(\Omega, \mathcal{F}, P)$ such that $F_t \overset{law}{=} F$ for every $t \geqslant 0$. Assume that $F \in L^3(\Omega, \mathcal{G}, P)$ for some $\sigma$-algebra $\mathcal{G} \subset \mathcal{F}$ and that in $L^1(\Omega)$,*

(a) $\displaystyle\lim_{t\downarrow 0}\frac{1}{t}\,E[F_t - F|\mathcal{G}] = -\lambda\,F$ *for some* $\lambda > 0$,

(b) $\displaystyle\lim_{t\downarrow 0}\frac{1}{t}\,E[(F_t - F)^2|\mathcal{G}] = (2\lambda + S)\mathrm{Var}(F)$ *for some random variable* $S$,

(c) $\displaystyle\lim_{t\downarrow 0}\frac{1}{t}\,(F_t - F)^3 = 0.$

*Then, with $N \sim N(0, \mathrm{Var}(F))$,*

$$d_{TV}(F, N) \leqslant \frac{E|S|}{\lambda}.$$

*Proof* Without loss of generality, we may and will assume that $\mathrm{Var}(F) = 1$. It is known that

$$d_{TV}(F, N) = \frac{1}{2} \sup E\big[\varphi(F) - \varphi(N)\big] , \tag{14.17}$$

where the supremum runs over all smooth functions $\varphi : \mathbb{R} \to \mathbb{R}$ with compact support and such that $\|\varphi\|_\infty \leqslant 1$. For such a $\varphi$, recall (see, e.g. [3, Lemma 2.4]) that

$$g(x) = e^{x^2/2} \int_{-\infty}^{x} \big(\varphi(y) - E[\varphi(N)]\big) e^{-y^2/2}\, dy , \quad x \in \mathbb{R},$$

satisfies

$$g'(x) - xg(x) = \varphi(x) - E[\varphi(N)] \tag{14.18}$$

as well as $\|g\|_\infty \leqslant \sqrt{2\pi}$, $\|g'\|_\infty \leqslant 4$ and $\|g''\|_\infty \leqslant 2\|\varphi'\|_\infty < +\infty$. In what follows, we fix such a pair $(\varphi, g)$ of functions. Let $G$ be a differentiable function such that $G' = g$, then due to $F_t \overset{\text{law}}{=} F$, it follows from the Taylor formula in mean-value form that

$$0 = E\big[G(F_t) - G(F)\big] = E\big[g(F)(F_t - F)\big] + \frac{1}{2} E\big[g'(F)(F_t - F)^2\big] + E[R] ,$$

with remainder $R$ bounded by $\frac{1}{6}\|g''\|_\infty |F_t - F|^3$.

By assumption (c) and as $t \downarrow 0$,

$$\left|\frac{1}{t} E[R]\right| \leqslant \frac{1}{6} \|g''\|_\infty \frac{1}{t} E\big[|F_t - F|^3\big] \to 0.$$

Therefore as $t \downarrow 0$, assumptions (a) and (b) imply that

$$\lambda\, E\big[g'(F) - Fg(F)\big] + \frac{1}{2} E\big[g'(F)S\big] = 0.$$

Plugging this into Stein's equation (14.18) and then using (14.17), we deduce the desired conclusion, namely,

$$d_{TV}(F, N) \leqslant \frac{1}{2} \frac{\|g'\|_\infty}{2\lambda} E|S| \leqslant \frac{E|S|}{\lambda}. \qquad \square$$

*Remark 14.6.1* Unlike the original Meckes' theorem, we do not assume the exchangeability condition $(F_t, F) \overset{law}{=} (F, F_t)$ in our Theorem 14.1.3. Our consideration is motivated by Röllin [22].

## 14.7 Quantitative Fourth Moment Theorem Revisited via Exchangeable Pairs

We give an elementary proof to the quantitative fourth moment theorem, that is, we explain how to prove the inequality of Theorem 14.1.1(ii) by means of our exchangeable pairs approach. For sake of convenience, let us restate this inequality: for any multiple Wiener-Itô integral $F$ of order $p \geqslant 1$ such that $E[F^2] = \sigma^2 > 0$, we have, with $N \sim N(0, \sigma^2)$,

$$d_{TV}(F, N) \leqslant \frac{2}{\sigma^2}\sqrt{\frac{p-1}{3p}}\,\sqrt{E[F^4] - 3\sigma^4}. \tag{14.19}$$

To prove (14.19), we consider, for instance, the exchangeable pairs of Brownian motions $\{(B, B^t)\}_{t>0}$ constructed in Sect. 14.3. We deduce, by combining Proposition 14.1.2 with Theorem 14.1.3 and Remark 14.5.1-(ii), that

$$d_{TV}(F, N) \leqslant \frac{2}{\sigma^2}\, E\left[\left|p\int_0^1 I^B_{p-1}(f(x,\cdot))^2 dx - \sigma^2\right|\right]. \tag{14.20}$$

To deduce (14.19) from (14.20), we are thus left to prove the following result.

**Proposition 14.7.1** *Let $p \geqslant 1$ and consider a symmetric function $f \in L^2([0,1]^p)$. Set $F = I^B_p(f)$ and $\sigma^2 = E[F^2]$. Then*

$$E\left[\left(p\int_0^1 I^B_{p-1}(f(x,\cdot))^2 dx - \sigma^2\right)^2\right] \leqslant \frac{p-1}{3p}\big(E[F^4] - 3\sigma^4\big).$$

*Proof* Using the product formula (14.7), we can write

$$F^2 = \sum_{r=0}^{p} r!\binom{p}{r}^2 I^B_{2p-2r}(f\otimes_r f) = \sigma^2 + \sum_{r=0}^{p-1} r!\binom{p}{r}^2 I^B_{2p-2r}(f\otimes_r f),$$

as well as

$$p\int_0^1 I^B_{p-1}(f(x,\cdot))^2 dx = p\sum_{r=0}^{p-1} r!\binom{p-1}{r}^2 I^B_{2p-2r-2}\left(\int_0^1 f(x,\cdot)\otimes_r f(x,\cdot)dx\right)$$

$$= p\sum_{r=1}^{p}(r-1)!\binom{p-1}{r-1}^2 I^B_{2p-2r}(f\otimes_r f) = \sigma^2 + \sum_{r=1}^{p-1}\frac{r}{p}\, r!\binom{p}{r}^2 I^B_{2p-2r}(f\otimes_r f).$$

Hence, by the isometry property (point 2 in Sect. 14.2),

$$E\left[\left(p\int_0^1 I_{p-1}^B(f(x,\cdot))^2dx-\sigma^2\right)^2\right]=\sum_{r=1}^{p-1}\frac{r^2}{p^2}r!^2\binom{p}{r}^4(2p-2r)!\|f\widetilde{\otimes}_r f\|^2_{L^2([0,1]^{2p-2r})}.$$

On the other hand, one has from (14.16) and the isometry property again that

$$\begin{aligned}\frac{1}{3}\big(E[F^4]-3\sigma^4\big)&=E\left[F^2\left(p\int_0^1 I_{p-1}^B(f(x,\cdot))^2dx-\sigma^2\right)\right]\\&=\frac{1}{3}\big(E[F^4]-3\sigma^4\big)=\sum_{r=1}^{p-1}\frac{r}{p}r!^2\binom{p}{r}^4(2p-2r)!\|f\widetilde{\otimes}_r f\|^2_{L^2([0,1]^{2p-2r})}.\end{aligned}$$

The desired conclusion follows. □

## 14.8 Connections with Malliavin Operators

Our main goal in this paper is to provide an elementary proof of Theorem 14.1.1(ii). Nevertheless, in this section we further investigate the connections we have found between our exchangeable pair approach and the operators of Malliavin calculus. This part may be skipped in a first reading, as it is not used in other sections. It is directed to readers who are already familiar with Malliavin calculus. We use classical notation and so do not introduce them in order to save place. We refer to [18] for any missing detail.

In this section, to stay on the safe side we only consider random variables $F$ belonging to

$$\mathcal{A}:=\bigcup_{p\in\mathbb{N}}\bigoplus_{r\leqslant p}\mathscr{H}_r\ ,\tag{14.21}$$

where $\mathscr{H}_r$ is the $r$th chaos associated to the Brownian motion $B$. In other words, we only consider random variables that are $\sigma\{B\}$-measurable and that admit a *finite* chaotic expansion. Note that $\mathcal{A}$ is an algebra (in view of product formula) that is dense in $L^2\big(\Omega,\sigma\{B\},P\big)$.

As is well-known, any $\sigma\{B\}$-measurable random variable $F$ can be written $F=\psi_F(B)$ for some measurable mapping $\psi_F:\mathbb{R}^{\mathbb{R}_+}\to\mathbb{R}$ determined $P\circ B^{-1}$ almost surely. For such an $F$, we can then define $F_t=\psi_F(B^t)$, with $B^t$ defined in Sect. 14.3. Another equivalent description of $F_t$ is to define it as $F_t=E[F]+\sum_{r=1}^p I_r^{B^t}(f_r)$, if the family $(f_r)_{1\leqslant r\leqslant p}$ is such that $F=E[F]+\sum_{r=1}^p I_r^B(f_r)$.

Our main findings are summarized in the statement below.

**Proposition 14.8.1** *Consider $F, G \in \mathcal{A}$, and define $F_t, G_t$ for each $t \in \mathbb{R}_+$ as is done above. Then, in $L^2(\Omega)$,*

(a) $\lim_{t\downarrow 0} \frac{1}{t}\, E\Big[F_t - F\big|\sigma\{B\}\Big] = LF,$

(b) $\lim_{t\downarrow 0} \frac{1}{t}\, E\Big[\big(F_t - F\big)(G_t - G)|\sigma\{B\}\Big] = L(FG) - FLG - GLF = 2\,\langle DF, DG\rangle.$

*Proof* The proof of (a) is an immediate consequence of (14.11), the linearity of conditional expectation, and the fact that $LI_r^B(f_r) = -r\, I_r^B(f_r)$ by definition of $L$. Let us now turn to the proof of (b). Using elementary algebra and then (a), we deduce that, as $t \downarrow 0$ and in $L^2(\Omega)$,

$$
\begin{aligned}
&\frac{1}{t}\, E\big[(F_t - F)(G_t - G)\big|\sigma\{B\}\big] \\
&= \frac{1}{t}\, E\big[F_t G_t - FG\big|\sigma\{B\}\big] - \frac{1}{t} F\, E\big[G_t - G\big|\sigma\{W\}\big] - \frac{1}{t} G\, E\big[F_t - F\big|\sigma\{B\}\big] \\
&\to L(FG) - FLG - GLF\ .
\end{aligned}
$$

Using $L = -\delta D$, $D(FG) = FDG + GDF$ (Leibniz rule) and $\delta(FDG) = F\delta(DG) - \langle DF, DG\rangle$ (see [18, Proposition 1.3.3]), it is easy to check that $L(FG) - FLG - GLF = 2\langle DF, DG\rangle$, which concludes the proof of Proposition 14.8.1. □

*Remark 14.8.2* The expression appearing in the right-hand side of (b) is nothing else but $2\,\Gamma(F, G)$, the (doubled) carré du champ operator.

To conclude this section, we show how our approach allows to recover the diffusion property of the Ornstein-Uhlenbeck operator.

**Proposition 14.8.3** *Fix $d \in \mathbb{N}$, let $F = (F_1, \ldots, F_d) \in \mathcal{A}^d$ (with $\mathcal{A}$ given in (14.21)), and $\Psi : \mathbb{R}^d \to \mathbb{R}$ be a polynomial function. Then*

$$
L\Psi(F) = \sum_{j=1}^{d} \partial_j \Psi(F) LF_j + \sum_{i,j=1}^{d} \partial_{ij}\Psi(F)\langle DF_i, DF_j\rangle\ . \tag{14.22}
$$

*Proof* We first define $F_t = (F_{1,t}, \ldots, F_{d,t})$ as explained in the beginning of the present section. Using classical multi-index notations, Taylor formula yields that

$$
\begin{aligned}
\Psi(F_t) - \Psi(F) &= \sum_{j=1}^{d} \partial_j \Psi(F)\big(F_{j,t} - F_j\big) + \frac{1}{2}\sum_{i,j=1}^{d} \partial_{i,j}\Psi(F)\big(F_{j,t} - F_j\big)\big(F_{i,t} - F_i\big) \\
&\quad + \sum_{|\beta|=3} \frac{3}{\beta_1! \ldots \beta_d!}(F_t - F)^{\beta} \int_0^1 (1-s)^k \big(\partial_1^{\beta_1} \ldots \partial_d^{\beta_d}\Psi\big)\big(F + s(F_t - F)\big)\, ds\,.
\end{aligned} \tag{14.23}
$$

In view of the previous proposition, the only difficulty in establishing (14.22) is about controlling the last term in (14.23) while passing $t \downarrow 0$. Note first $\big(\partial_1^{\beta_1} \dots \partial_d^{\beta_d} \Psi\big)\big(F + s(F_t - F)\big)$ is polynomial in $F$ and $(F_t - F)$, so our problem reduces to show

$$\lim_{t \downarrow 0} \frac{1}{t} E\big[|F^\alpha (F_t - F)^\beta|\big] = 0\,, \tag{14.24}$$

for $\alpha = (\alpha_1, \dots, \alpha_d)$, $\beta = (\beta_1, \dots, \beta_d) \in \big(\mathbb{N} \cup \{0\}\big)^d$ with $|\beta| \geqslant 3$.

Indeed, (assume $\beta_j > 0$ for each $j$)

$$\begin{aligned}
\frac{1}{t} E\big[|F^\alpha (F_t - F)^\beta|\big] &\leqslant \frac{1}{t} E\big[|F^\alpha|^2\big]^{1/2} E\big[|(F_t - F)^\beta|^2\big]^{1/2} \quad \text{by Cauchy-Schwarz inequality;} \\
&\leqslant E\big[|F^\alpha|^2\big]^{1/2} \frac{1}{t} \left( \prod_{j=1}^d E\Big[(F_{j,t} - F_j)^{2|\beta|}\Big]^{\frac{\beta_j}{|\beta|}} \right)^{1/2} \quad \text{by Hölder inequality;} \\
&\leqslant C\, E\big[|F^\alpha|^2\big]^{1/2} t^{\frac{|\beta|}{2} - 1} \left( \prod_{j=1}^d \frac{1}{t^{\beta_j}} E\Big[(F_{j,t} - F_j)^2\Big]^{\beta_j} \right)^{1/2},
\end{aligned}$$

where the last inequality follows from point-(iii) in Remark 14.5.1 with $C > 0$ independent of $t$. Since $F^\alpha \in \mathcal{A}$ and $|\beta| \geqslant 3$, (14.24) follows immediately from the above inequalities. ◻

## 14.9 Peccati-Tudor Theorem Revisited Too

In this section, we combine a multivariate version of Meckes' abstract exchangeable pairs [12] with our results from Sect. 14.3 to prove (14.5), thus leading to a fully elementary proof of Theorem 14.1.5 as well.

First, we recall the following multivariate version of Meckes' theorem (see [12, Theorem 4]). Unlike in the one-dimensional case, it seems inevitable to impose the exchangeability condition in the following proposition, as we read from its proof in [12].

**Proposition 14.9.1** *For each $t > 0$, let $(F, F_t)$ be an exchangeable pair of centered $d$-dimensional random vectors defined on a common probability space. Let $\mathcal{G}$ be a $\sigma$-algebra that contains $\sigma\{F\}$. Assume that $\Lambda \in \mathbb{R}^{d \times d}$ is an invertible deterministic matrix and $\Sigma$ is a symmetric, non-negative definite deterministic matrix such that*

(a) $\lim_{t \downarrow 0} \frac{1}{t}\, E\big[F_t - F | \mathcal{G}\big] = -\Lambda F$ *in* $L^1(\Omega)$,

(b) $\lim_{t \downarrow 0} \frac{1}{t}\, E\big[(F_t - F)(F_t - F)^T | \mathcal{G}\big] = 2\Lambda\Sigma + S$ *in* $L^1(\Omega, \|\cdot\|_{HS})$ *for some matrix* $S = S(F)$*, and with* $\|\cdot\|_{HS}$ *the Hilbert-Schmidt norm*

(c) $\lim_{t\downarrow 0}\sum_{i=1}^{d}\frac{1}{t}E\big[|F_{i,t}-F_i|^3\big]=0$, *where* $F_{i,t}$ *(resp.* $F_i$*) stands for the* $i$*th coordinate of* $F_t$ *(resp.* $F$*).*

*Then, with* $N\sim N_d(0,\Sigma)$*,*

(1) *for* $g\in C^2(\mathbb{R}^d)$*,*

$$\big|E[g(F)]-E[g(N)]\big|\leqslant\frac{\|\Lambda^{-1}\|_{op}\sqrt{d}\,M_2(g)}{4}E\left[\sqrt{\sum_{i,j=1}^{d}S_{ij}^2}\right],$$

*where* $M_2(g):=\sup_{x\in\mathbb{R}^d}\big\|D^2g(x)\big\|_{op}$ *with* $\|\cdot\|_{op}$ *the operator norm.*

(2) *if, in addition,* $\Sigma$ *is positive definite, then*

$$d_W(F,N)\leqslant\frac{\|\Lambda^{-1}\|_{op}\|\Sigma^{-1/2}\|_{op}}{\sqrt{2\pi}}\,E\left[\sqrt{\sum_{i,j=1}^{d}S_{ij}^2}\right].$$

*Remark 14.9.2* Constant in (2) is different from Meckes' paper [12]. We took this better constant from Christian Döbler's dissertation [5], see page 114 therein.

By combining the previous proposition with our exchangeable pairs, we get the following result, whose point 2 corresponds to (14.5).

**Theorem 14.9.3** *Fix* $d\geqslant 2$ *and* $1\leqslant p_1\leqslant\ldots\leqslant p_d$*. Consider a vector* $F:=\big(I_{p_1}^B(f_1),\ldots,I_{p_d}^B(f_d)\big)$ *with* $f_i\in L^2\big([0,1]^{p_i}\big)$ *symmetric for each* $i\in\{1,\ldots,d\}$*. Let* $\Sigma=(\sigma_{ij})$ *be the covariance matrix of* $F$*, and* $N\sim N_d(0,\Sigma)$*. Then*

(1) *for* $g\in C^2(\mathbb{R}^d)$*,*

$$\big|E[g(F)]-E[g(N)]\big|\leqslant\frac{\sqrt{d}\,M_2(g)}{2p_1}\sqrt{\sum_{i,j=1}^{d}\mathrm{Var}\Big(p_ip_j\int_0^1 I_{p_i-1}(f_i(x,\cdot))I_{p_j-1}(f_j(x,\cdot))dx\Big)},$$

*where* $M_2(g):=\sup_{x\in\mathbb{R}^d}\big\|D^2g(x)\big\|_{op}$.

(2) *if in addition,* $\Sigma$ *is positive definite, then*

$$d_W(F,N)\leqslant\frac{2\|\Sigma^{-1/2}\|_{op}}{q_1\sqrt{2\pi}}\sqrt{\sum_{i,j=1}^{d}\mathrm{Var}\Big(p_ip_j\int_0^1 I_{p_i-1}(f_i(x,\cdot))I_{p_j-1}(f_j(x,\cdot))dx\Big)}.$$

*Proof* We consider $F_t = \big(I_{p_1}^{B^t}(f_1), \ldots, I_{p_d}^{B^t}(f_d)\big)$, where $B^t$ is the Brownian motion constructed in Sect. 14.3. We deduce from (14.10) that

$$\frac{1}{t}\,E\big[F_t - F|\sigma\{B\}\big] = \left(\frac{e^{-p_1 t}-1}{t}\,I_{p_1}^{B^t}(f_1), \ldots, \frac{e^{-p_d t}-1}{t}\,I_{p_d}^{B^t}(f_d)\right)$$

implying in turn that, in $L^2(\Omega)$ and as $t \downarrow 0$,

$$\frac{1}{t}\,E\big[F_t - F|\sigma\{B\}\big] \to -\Lambda F,$$

with $\Lambda = \mathrm{diag}(p_1, \ldots, p_d)$ (in particular, $\|\Lambda^{-1}\|_{\mathrm{op}} = p_1^{-1}$). That is, assumption (a) in Proposition 14.9.1 is satisfied (with $\mathcal{G} = \sigma\{B\}$). That assumption (c) in Proposition 14.9.1 is satisfied as well follows from Proposition 14.1.2(c). Let us finally check that assumption (b) in Proposition 14.9.1 takes place too. First, using the product formula (14.7) for multiple integrals with respect to $B^t$ (resp. $B$) yields

$$F_i F_j = \sum_{r=0}^{p_i \wedge p_j} r!\binom{p_i}{r}\binom{p_j}{r} I_{p_i+p_j-2r}^{B}\big(f_i \otimes_r f_j\big)$$
$$F_{i,t} F_{j,t} = \sum_{r=0}^{p_i \wedge p_j} r!\binom{p_i}{r}\binom{p_j}{r} I_{p_i+p_j-2r}^{B^t}\big(f_i \otimes_r f_j\big).$$

Hence, using (14.11) for passing to the limit,

$$\begin{aligned}
&\frac{1}{t}\,E\big[(F_{i,t}-F_i)(F_{j,t}-F_j)\big|\sigma\{B\}\big] - \frac{1}{t}\,E\big[F_{i,t}F_{j,t} - F_iF_j\big|\sigma\{B\}\big]\\
&= -\frac{1}{t}\,F_i\,E\big[F_{j,t} - F_j|\sigma\{B\}\big] - \frac{1}{t}\,F_j\,E\big[F_{i,t} - F_i\big|\sigma\{B\}\big]\\
&\to (p_i+p_j)F_iF_j = \sum_{r=0}^{p_i \wedge p_j} r!\binom{p_i}{r}\binom{p_j}{r}(p+q) I_{p_i+p_j-2r}^{B}\big(f_i \otimes_r f_j\big) \quad \text{as } t \downarrow 0.
\end{aligned}$$

Now, note in $L^2(\Omega)$,

$$\begin{aligned}
&\frac{1}{t}\,E\big[F_{i,t}F_{j,t} - F_iF_j\big|\sigma\{B\}\big]\\
&= \sum_{r=0}^{p_i \wedge p_j} r!\binom{p_i}{r}\binom{p_j}{r}\frac{1}{t}\,E\Big[I_{p_i+p_j-2r}^{B^t}\big(f_i \otimes_r f_j\big) - I_{p_i+p_j-2r}^{B}\big(f_i \otimes_r f_j\big)\big|\sigma\{B\}\Big]\\
&\to \sum_{r=0}^{p_i \wedge p_j} r!\binom{p_i}{r}\binom{p_j}{r}\,(2r - p_i - p_j) I_{p_i+p_j-2r}^{B}\big(f_i \otimes_r f_j\big)\,, \qquad \text{as } t \downarrow 0, \text{ by (14.11)}\,.
\end{aligned}$$

Thus, as $t \downarrow 0$,

$$\frac{1}{t}E\big[(F_{i,t}-F_i)(F_{j,t}-F_j)\big|\sigma\{B\}\big] \to 2\sum_{r=1}^{p_i\wedge p_j} r!r\binom{p_i}{r}\binom{p_j}{r} I^B_{p_i+p_j-2r}\big(f_i\otimes_r f_j\big)$$
$$= 2p_ip_j\int_0^1 I^B_{p_i-1}(f_i(x,\cdot))I^B_{p_j-1}(f_j(x,\cdot))dx\,,$$

where the last equality follows from a straightforward application of the product formula (14.7). As a result, if we set

$$S_{ij} = 2p_ip_j\int_0^1 I_{p_i-1}(f_i(x,\cdot))I_{p_j-1}(f_j(x,\cdot))dx - 2p_i\sigma_{ij}$$

for each $i, j \in \{1,\ldots,d\}$, then assumption (b) in Proposition 14.9.1 turns out to be satisfied as well. By the isometry property (point 2 in Sect. 14.2), it is straightforward to check that

$$p_j\int_0^1 E\Big[I_{p_i-1}(f_i(x,\cdot))I_{p_j-1}(f_j(x,\cdot))\Big]dx = \sigma_{ij}\ .$$

Therefore,

$$E\left[\sqrt{\sum_{i,j=1}^d S_{ij}^2}\,\right] \leqslant \sqrt{\sum_{i,j=1}^d E\big[S_{ij}^2\big]} = 2\sqrt{\sum_{i,j=1}^d \mathrm{Var}\Big(p_ip_j\int_0^1 I_{p_i-1}(f_i(x,\cdot))I_{p_j-1}(f_j(x,\cdot))dx\Big)}\ .$$

Hence the desired results in (1) and (2) follow from Proposition 14.9.1. □

**Acknowledgements** We would like to warmly thank Christian Döbler and Giovanni Peccati, for very stimulating discussions on exchangeable pairs since the early stage of this work.

## References

1. J.-M. Azaïs, J.R. León, CLT for crossing of random trigonometric polynomials. Electron. J. Probab. **18**(68), 1–17 (2013)
2. J.-M. Azaïs, F. Dalmao, J.R. León, CLT for the zeros of classical random trigonometric polynomials. Ann. Inst. H. Poincaré Probab. Stat. **52**(2), 804–820 (2016)
3. L.H.Y. Chen, L. Goldstein, Q.M. Shao, Normal approximation by Stein's method, in *Probability and Its Applications* (Springer, Berlin, 2011)
4. F. Dalmao, Asymptotic variance and CLT for the number of zeros of Kostlan Shub Smale random polynomials. C. R. Math. **353**(12), 1141–1145 (2015)
5. C. Döbler, New developments in Stein's method with applications. Ph.D. dissertation. Ruhr-Universität, Bochum, 2012. http://www-brs.ub.ruhr-uni-bochum.de/netahtml/HSS/Diss/DoeblerChristian/

6. D. Hug, G. Last, M. Schulte, Second-order properties and central limit theorems for geometric functionals of Boolean models. Ann. Appl. Probab. **26**(1), 73–135 (2016)
7. P. Malliavin, Stochastic calculus of variations and hypoelliptic operators, in *Proceeding International Symposium on Stochastic Differential Equations* (Wiley, Kyoto, 1976/1978), pp. 195–263
8. D. Marinucci, G. Peccati, Ergodicity and Gaussianity for spherical random fields. J. Math. Phys. **51**, 043301 (2010)
9. D. Marinucci, M. Rossi, Stein-Malliavin approximations for nonlinear functionals of random eigenfunctions on $\mathbf{S}^d$. J. Funct. Anal. **268**(8), 2379–2420 (2015)
10. D. Marinucci, G. Peccati, M. Rossi, I. Wigman, Non-universality of nodal length distribution for arithmetic random waves. Geom. Funct. Anal. **26**(3), 926–960 (2016)
11. E. Meckes, An infinitesimal version of Stein's method of exchangeable pairs, Ph.D. dissertation. Stanford University, Stanford, 2006
12. E. Meckes, On Stein's method for multivariate normal approximation, in *IMS Collections, High Dimensional Probability V: The Luminy Volume*, vol. 5 (2009), pp. 153–178
13. I. Nourdin, Malliavin-Stein approach: a webpage maintained by Ivan Nourdin. http://tinyurl.com/kvpdgcy
14. I. Nourdin, G. Peccati, Stein's method on Wiener chaos. Probab. Theory Relat. Fields **145**(1), 75–118 (2009)
15. I. Nourdin, G. Peccati, Normal approximations with Malliavin calculus: from Stein's method to universality, in *Cambridge Tracts in Mathematics*, vol. 192 (Cambridge University Press, Cambridge, 2012)
16. I. Nourdin, J. Rosiński, Asymptotic independence of multiple Wiener-Itô integrals and the resulting limit laws. Ann. Probab. **42**(2), 497–526 (2014)
17. I. Nourdin, G. Peccati, A. Réveillac, Multivariate normal approximation using Stein's method and Malliavin calculus. Ann. Inst. H. Poincaré - Probab. Stat. **46**(1), 45–58 (2010)
18. D. Nualart, *The Malliavin Calculus and Related Topics*, 2nd edn. Probability and Its Applications (Springer, Berlin, 2006)
19. D. Nualart, G. Peccati, Central limit theorems for sequences of multiple stochastic integrals. Ann. Probab. **33**(1), 177–193 (2005)
20. G. Peccati, C.A. Tudor, Gaussian limits for vector-valued multiple stochastic integrals. Séminaire de Probab. **XXXVIII**, 247–262 (2005)
21. M. Reitzner, M. Schulte, Central limit theorems for U-statistics of Poisson point processes. Ann. Probab. **41**(6), 3879–3909 (2013)
22. A. Röllin, A note on the exchangeability condition in Stein's method. Statist. Probab. Lett. **78**, 1800–1806 (2008)
23. M. Schulte, A central limit theorem for the Poisson-Voronoi approximation. Adv. Appl. Math. **49**(3–5), 285–306 (2012)
24. C. Stein, A bound for the error in the normal approximation to the distribution of a sum of dependent random variables, in *Proceeding Sixth Berkeley Symposium on Mathematics Statistical and Probability*, vol. 2 (University of California Press, California, 1972), pp. 583–602

# Chapter 15
# Permanental Processes with Kernels That Are Not Equivalent to a Symmetric Matrix

**Michael B. Marcus and Jay Rosen**

**Abstract** Kernels of $\alpha$-permanental processes of the form

$$\widetilde{u}(x, y) = u(x, y) + f(y), \qquad x, y \in S, \tag{15.1}$$

in which $u(x, y)$ is symmetric, and $f$ is an excessive function for the Borel right process with potential densities $u(x, y)$, are considered. Conditions are given that determine whether $\{\widetilde{u}(x, y); x, y \in S\}$ is symmetrizable or asymptotically symmetrizable.

**Keywords** Permanental processes · Symmetrizable

**AMS 2010 Subject Classification** 60K99, 60J25, 60J27, 60G15, 60G99

## 15.1 Introduction

An $R^n$ valued $\alpha$-permanental random variable $X = (X_1, \ldots, X_n)$ is a random variable with Laplace transform

$$E\left(e^{-\sum_{i=1}^{n} s_i X_i}\right) = \frac{1}{|I + KS|^{\alpha}}, \tag{15.2}$$

where $K$ is an $n \times n$ matrix and $S$ is an $n \times n$ diagonal matrix with diagonal entries $(s_1, \ldots, s_n)$.

M. B. Marcus (✉)
CUNY Graduate Center, New York, NY, USA
e-mail: mbmarcus@optonline.net

J. Rosen
Department of Mathematics, College of Staten Island, CUNY, Staten Island, NY, USA
e-mail: jrosen30@optimum.net



N. Gozlan et al. (eds.), *High Dimensional Probability VIII*,
Progress in Probability 74, https://doi.org/10.1007/978-3-030-26391-1_15

We refer to $K$ as a kernel of $X$. But note that $K$ is not unique. For example, if $K$ satisfies (15.2) so does $\Lambda K \Lambda^{-1}$ for any $\Lambda \in \mathcal{D}_{n,+}$, the set of $n \times n$ diagonal matrices with strictly positive diagonal entries.

Let $\mathcal{K}(X)$ denote the set of all kernels that determine $X$ by (15.2). We are particularly interested in $\alpha$-permanental random variables $X$ for which $\mathcal{K}(X)$ does not contain any symmetric kernels. (We explain at the end of this section why we are interested in such processes and kernels.)

If $\mathcal{K}(X)$ contains a symmetric matrix we say that $X$ is determined by a symmetric matrix or kernel and that any $K \subset \mathcal{K}(X)$ is equivalent to a symmetric matrix, or is symmetrizable. It follows from (15.2) that a kernel $K$ is equivalent to a symmetric matrix if and only if there exists an $n \times n$ symmetric matrix $Q$ such that

$$|I + KS| = |I + QS| \quad \text{for all } S \in \mathcal{D}_{n,+}. \tag{15.3}$$

An $\alpha$-permanental process $\{X_t, t \in T\}$ is a stochastic process that has finite dimensional distributions that are $\alpha$-permanental random variables. An $\alpha$-permanental process is determined by a kernel $\{K(s,t), s, t \in T\}$ with the property that for all distinct $t_1, \ldots, t_n$ in $T$, $\{K(t_i, t_j), i, j \in [1, n]\}$ is the kernel of the $\alpha$-permanental random variable $(X_{t_1}, \ldots, X_{t_n})$.

**Definition** We say that an $\alpha$-permanental process $\{X_t, t \in T\}$ with kernel $\{K(s,t), s, t \in T\}$ is determined by a symmetric kernel if for all $n \geq 1$ and distinct $t_1, \ldots, t_n$ in $T$, $\{K(t_i, t_j), i, j \in [1, n]\}$ is symmetrizable. When this is the case we also say that $\{K(s,t), s, t \in T\}$ is symmetrizable. (In what follows we always take $|T| \geq 3$.)

The next theorem is [6, Theorem 1.9]. It shows that we can modify a very large class of symmetric potentials so that they are no longer symmetric but are still kernels of permanental processes.

**Theorem 15.1.1** *Let $S$ a be locally compact set with a countable base. Let $X = (\Omega, \mathcal{F}_t, X_t, \theta_t, P^x)$ be a transient symmetric Borel right process with state space $S$ and continuous strictly positive potential densities $u(x,y)$ with respect to some $\sigma$-finite measure $m$ on $S$. Then for any finite excessive function $f$ of $X$ and $\alpha > 0$,*

$$\widetilde{u}^f(x,y) = u(x,y) + f(y), \qquad x, y \in S, \tag{15.4}$$

*is the kernel of an $\alpha$-permanental process.*

A function $f$ is said to be excessive for $X$ if $E^x(f(X_t)) \uparrow f(x)$ as $t \to 0$ for all $x \in S$. It is easy to check that for any positive measurable function $h$,

$$f(x) = \int u(x,y)h(y)\,dm(y) = E^x\left(\int_0^\infty h(X_t)\,dt\right) \tag{15.5}$$

is excessive for $X$. Such a function $f$ is called a potential function for $X$.

Unless the function $f$ in (15.4) is constant, $\{\widetilde{u}^f(x,y); x,y \in S\}$ is not symmetric. We now show that, generally, we can choose $f$ so that $\{\widetilde{u}^f(x,y); x,y \in S\}$ is also not equivalent to a symmetric matrix. The next two theorems show how restricted the symmetric matrix $\{u(x,y); x,y \in S\}$ must be for $\{\widetilde{u}^f(x,y); x,y \in S\}$ to be symmetrizable for all potential functions $f$.

We use $\ell_1^+$ to denote strictly positive sequences in $\ell_1$.

**Theorem 15.1.2** *Let $X = (\Omega, \mathcal{F}_t, X_t, \theta_t, P^x)$ be a transient symmetric Borel right process with state space $T \subseteq \mathbb{N}$, and potential $U = \{U_{j,k}\}_{j,k\in T}$. Then*

*(i) Either*

$$U_{j,k} = \Lambda_j \delta_{j,k} + d, \qquad j,k \in T, \tag{15.6}$$

*where $\Lambda_j \geq 0$ and $d \geq 0$,*

*(ii) or we can find a potential function $f = Uh$, with $h \in \ell_1^+$, such that*

$$\widetilde{U}^f_{j,k} := U_{j,k} + f_k, \qquad j,k \in T, \tag{15.7}$$

*is not symmetrizable.*

When we consider limit theorems for infinite sequences of permanental random variables $\{Y(k), k \in \mathbb{N}\}$ with kernel $V = \{v(j,k), j,k \in \mathbb{N}\}$ it is not enough to know that $V$ is not symmetrizable since we are only concerned with the permanental variables generated by $V(n) = \{v(j,k), j,k \geq n\}$ as $n \to \infty$. We would like to know that $V(n)$ is not symmetrizable for large $n$. We say that the kernel $V$ is asymptotically symmetrizable if there exists an $n_0$ such that $V(n)$ is symmetrizable for all $n \geq n_0$. We can modify Theorem 15.1.2 to handle this case also.

**Theorem 15.1.3** *Let $X = (\Omega, \mathcal{F}_t, X_t, \theta_t, P^x)$ be a transient symmetric Borel right process with state space $\mathbb{N}$, and potential $U = \{U_{j,k}\}_{j,k\in\mathbb{N}}$. Then*

*(i) Either there exists an $n_0$ such that*

$$U_{j,k} = \Lambda_j \delta_{j,k} + d, \qquad \forall j,k \geq n_0, \tag{15.8}$$

*where $\Lambda_j \geq 0$ and $d \geq 0$,*

*(ii) or we can find a potential function $f = Uh$, with $h \in \ell_1^+$, such that*

$$\widetilde{U}^f_{j,k} := U_{j,k} + f_k, \qquad j,k \in \mathbb{N}, \tag{15.9}$$

*is not asymptotically symmetrizable.*

The next theorem shows that when the state space of a transient symmetric Borel right process has a limit point, then under reasonable conditions on the potential densities that determine the process, the process is not determined by a kernel that is asymptotically symmetrizable.

**Theorem 15.1.4** *Let $S' = \{x_0, x_1, \ldots\}$ be a countable set with a single limit point $x_0$. Let $\overline{X}$ be a transient symmetric Borel right process with state space $S'$, and continuous strictly positive potential densities $u := \{u(x, y), x, y \in S'\}$ such that $u(y, x_0) < u(x_0, x_0)$ for all $y \neq x_0$. Then we can find a potential function $f = Uh$, with $h \in \ell_1^+$, that is continuous at $x_0$, and is such that,*

$$\widetilde{u}^f(x, y) = u(x, y) + f(y), \qquad x, y \in S', \tag{15.10}$$

*is not asymptotically symmetrizable.*

Theorems 15.1.2–15.1.4 show that generally there exists an excessive function $f$ for $X$ which gives a kernel for an $\alpha$-permanental processes that is not determined by a symmetric matrix. However, in specific examples we deal with specific functions $f$ and want to know that the kernels determined by these functions are not symmetrizable. With some additional structure on the symmetric matrix $u(x, y)$ in (15.4) we can show that $\widetilde{u}^f(x, y)$ in (15.4) is not asymptotically symmetrizable.

**Lemma 15.1.1** *In the notation of (15.4), let $u = \{u(j, k);\ j, k \in \mathbb{N}\}$ be a symmetric Toëplitz matrix, with at least two different off diagonal elements, and set $v(|j-k|) = u(j, k)$. Let*

*(i)*

$$\widetilde{u}^f(j, k) = v(|j - k|) + f(k), \qquad j, k \in \mathbb{N}, \tag{15.11}$$

*where $f$ is a strictly monotone potential for $u$. Then $\{\widetilde{u}^f(j, k);\ j, k \in \mathbb{N}\}$ is not asymptotically symmetrizable.*

*(ii) Let*

$$\widetilde{v}^f(s_j, s_k) = s_j \wedge s_k + f(s_k), \qquad j, k \in \mathbb{N}, \tag{15.12}$$

*where $f$ is a strictly monotone potential for $\{s_j \wedge s_k;\ j, k \in \mathbb{N}\}$. Then for any triple of distinct values $s_j, s_k, s_l$,*

$$\{\widetilde{v}^f(s_p, s_q)\}_{p,q=j,k,l}\,, \tag{15.13}$$

*is not symmetrizable. In particular $\{\widetilde{v}^f(s_j, s_k);\ j, k \in \mathbb{N}\}$ is not asymptotically symmetrizable.*

We can use this lemma to show that certain $\alpha$-permanental processes, studied in [6], are not determined by kernels that are asymptotically symmetrizable. When $S$ is an interval on the real line we say that $\{u(x, y);\ x, y \in S\}$ is not asymptotically symmetrizable at $x_0 \in S$, if we can find a sequence $\{x_k\}$ in $S$ such that $\lim_{k\to\infty} x_k = x_0$, and $\{u(x_j, x_k);\ j, k \in \mathbb{N}\}$ is not asymptotically symmetrizable.

*Example 15.1.1* In [6, Example 1.3] we obtain a limit theorem for the asymptotic behavior of the sample paths at 0 of $\alpha$-permanental processes with the kernel,

$$\widehat{u}^f(s,t) = e^{-\lambda|s-t|} + f(t), \qquad s,t \in [0,1], \tag{15.14}$$

where $f = q + t^\beta$, $\beta > 2$, and $q \geq q_0(\beta)$, a constant depending on $\beta$. We show in Sect. 15.4 that $\widehat{u}^f(s,t)$ is not asymptotically symmetrizable at any $s_0 \in S$.

Similarly

$$\overline{u}^f(j,k) = e^{-\lambda|j-k|} + f(k), \qquad j,k \in \mathbb{N}, \tag{15.15}$$

is not asymptotically symmetrizable.

*Example 15.1.2* In [6, Example 1.4] we obtain limit theorems for the asymptotic behavior of the sample paths at zero and infinity of $\alpha$-permanental processes with the kernel,

$$\widetilde{v}^f(s,t) = s \wedge t + f(t), \qquad s,t \geq 0, \tag{15.16}$$

where $f$ is a concave strictly increasing function. We show in Sect. 15.4 that for any $s_0 \in R^+$ and any sequence of distinct values $\{s_k\}$ such that $\lim_{k\to\infty} s_k = s_0$, $\widetilde{v}^f(s_j, s_k)$ is not asymptotically symmetrizable.

In addition,

$$\overline{v}^f(j,k) = j \wedge k + f(k), \qquad j,k \in \mathbb{N}, \tag{15.17}$$

is not asymptotically symmetrizable.

We explain why we are particularly interested in $\alpha$-permanental processes determined by kernels $K$ that are not equivalent to a symmetric matrix. When $\{u(s,t); s,t \in \mathcal{T}\}$ is symmetric and is a kernel that determines $\alpha$-permanental processes, $Y_\alpha = \{Y_\alpha(t), t \in \mathcal{T}\}$, then

$$Y_{1/2} \overset{law}{=} \{G^2(t)/2, t \in \mathcal{T}\}, \tag{15.18}$$

where $G = \{G(t), t \in \mathcal{T}\}$ is a mean zero Gaussian process with covariance $u(s,t)$.

This is not true when the kernel of $\alpha$-permanental processes is not symmetrizable. In this case we get a new class of processes. These are the processes that we find particularly interesting.

There is another reason why permanental processes with kernels that are not equivalent to a symmetric matrix are interesting. Dynkin's Isomorphism Theorem relates the local times of a symmetric Markov process to the squares of a Gaussian process, with covariance given by the potential densities of the Markov process. This theorem was extended by Eisenbaum and Kaspi [2] to relate the local times of Markov processes that are not symmetric to permanental process, which necessarily, are not determined by symmetric kernels. Results about permanental processes that are not symmetrizable should lead to new results about the local times of Markov processes that are not symmetric.

To study permanental processes with kernels that are not equivalent to a symmetric matrix our first step is to characterize those kernels that are equivalent to a symmetric matrix. This is done in Sect. 15.2. In Sect. 15.3 we give the proofs of Theorems 15.1.2–15.1.4. In Sect. 15.4 we give the proof of Lemma 15.1.1 and details about Examples 15.1.1 and 15.1.2.

## 15.2 Kernels That Are Equivalent to a Symmetric Matrix

Let $M$ be an $n \times n$ matrix. For $\mathcal{I} \subseteq [1, \ldots, n]$ we define $M_{\mathcal{I}}$ to be the $|\mathcal{I}| \times |\mathcal{I}|$ matrix $\{M_{p,q}\}_{p,q\in\mathcal{I}}$. (Recall that $\mathcal{D}_{n,+}$ is the set of all $n \times n$ diagonal matrices with strictly positive diagonal elements.)

**Lemma 15.2.1** *Let $K$ be an $n \times n$ matrix and assume that*

$$|I + KS| = |I + QS| \quad \textit{for all } S \in \mathcal{D}_{n,+}. \tag{15.19}$$

*Then for all $\mathcal{I} \subseteq [1, \ldots, n]$*

$$|K_{\mathcal{I}}| = |Q_{\mathcal{I}}|. \tag{15.20}$$

*In particular*

$$|K| = |Q| \tag{15.21}$$

*and*

$$K_{j,j} = Q_{j,j} \quad \textit{for all} \quad j = 1, \ldots n. \tag{15.22}$$

*Furthermore, if $Q$ is symmetric, then*

$$|Q_{j,k}| = (K_{j,k}K_{k,j})^{1/2} \quad \textit{for all} \quad i, j = 1, \ldots, n \tag{15.23}$$

*and for all distinct $i_1, i_2, i_3 \in [1, \ldots, n]$*

$$K_{i_1,i_2}K_{i_2,i_3}K_{i_3,i_1} = K_{i_1,i_3}K_{i_2,i_1}K_{i_3,i_2}. \tag{15.24}$$

Conditions such as (15.24) appear in [1, 3].

*Proof* Denote the diagonal elements of $S$ by $\{s_i\}_{i=1}^n$. Let $s_i \to 0$ for all $s_i \in \mathcal{I}^c$ in (15.19) to get

$$|I + K_{\mathcal{I}}S| = |I + Q_{\mathcal{I}}S| \quad \text{for all } S \in \mathcal{D}_{|\mathcal{I}|,+}. \tag{15.25}$$

Multiply both sides of (15.25) by $|S^{-1}|$ and let the diagonal components of $S$ go to infinity to get (15.20). The relationships in (15.21) and (15.22) are simply examples of (15.20).

Let $\mathcal{I} = \{j, k\}$. It follows from (15.20) that

$$K_{i,i}K_{j,j} - K_{i,j}K_{j,i} = Q_{i,i}Q_{j,j} - Q_{i,j}^2, \tag{15.26}$$

which by (15.22) implies that $K_{i,j}K_{j,i} = Q_{i,j}^2$. This gives (15.23).

Finally, let $\mathcal{I} = \{i_1, i_2, i_3\}$ and take the determinants $|K(\mathcal{I})|$ and $|Q(\mathcal{I})|$. It follows from (15.20), (15.22) and (15.23) that

$$\begin{aligned} K_{i_1,i_2}K_{i_2,i_3}K_{i_3,i_1} &+ K_{i_1,i_3}K_{i_2,i_1}K_{i_3,i_2} \\ &= Q_{i_1,i_2}Q_{i_2,i_3}Q_{i_3,i_1} + Q_{i_1,i_3}Q_{i_2,i_1}Q_{i_3,i_2} \\ &= 2Q_{i_1,i_2}Q_{i_2,i_3}Q_{i_3,i_1}. \end{aligned} \tag{15.27}$$

By (15.23) this is equal to

$$\pm 2(K_{i_1,i_2}K_{i_2,i_3}K_{i_3,i_1}K_{i_1,i_3}K_{i_2,i_1}K_{i_3,i_2})^{1/2}. \tag{15.28}$$

Set

$$x = K_{i_1,i_2}K_{i_2,i_3}K_{i_3,i_1} \quad \text{and} \quad y = K_{i_1,i_3}K_{i_2,i_1}K_{i_3,i_2}. \tag{15.29}$$

Then we have

$$x + y = \pm 2\sqrt{xy}. \tag{15.30}$$

It is clear from this that $x$ and $y$ have the same sign. If they are both positive, we have

$$x + y = 2\sqrt{xy}, \tag{15.31}$$

That is, $(\sqrt{x} - \sqrt{y})^2 = 0$, which gives (15.24).

On the other hand, if $x$ and $y$ are both negative, (15.30) implies that

$$(-x) + (-y) = 2\sqrt{(-x)(-y)}, \tag{15.32}$$

which also gives (15.24). □

*Remark 15.2.1* Even when $K$ is the kernel of $\alpha$-permanental processes we must have absolute values on the left-hand sides of (15.23). This is because when (15.19) holds it also holds when $|I + QS|$ is replaced by $|I + \mathcal{V}Q\mathcal{V}S|$ for any signature matrix $\mathcal{V}$. (A signature matrix is a diagonal matrix with diagonal entries $\pm 1$.) So the symmetric matrix $Q$ need not be the kernel of $\alpha$-permanental processes On the other hand, by Eisenbaum and Kaspi [2, Lemma 4.2], we can find a symmetric matrix $\widetilde{Q}$ that is the kernel of $\alpha$-permanental processes such that (15.19) holds with $Q$ replaced by $\widetilde{Q}$ and we have $\widetilde{Q}_{j,k} = (K_{j,k}K_{k,j})^{1/2}$.

## 15.3 Proofs of Theorems 15.1.2–15.1.4

We begin with a simple observation that lies at the heart of the proofs of Theorems 15.1.2 and 15.1.3.

For $y \in R^n$ we use $B_\delta(y)$ to denote a Euclidean ball of radius $\delta$ centered at $x$.

**Lemma 15.3.1** *Let $W = \{w_{j,k};\ j,k = 1,2,3\}$ be a positive symmetric matrix such that $w_{j,k} \le w_{j,j} \wedge w_{k,k}$. For any $x = (x_1, x_2, x_3)$ let $\widetilde{W}^x$ be a $3 \times 3$ matrix defined by*

$$\widetilde{W}^x_{j,k} = w_{j,k} + x_k, \qquad j,k = 1,2,3. \tag{15.33}$$

*Suppose that $\widetilde{W}^x$ is symmetrizable for all $x \in B_\delta(x_0)$, for some $x_0 \in R^3$ and $\delta > 0$. Then, necessarily,*

$$w_{j,k} = \Lambda_j \delta_{j,k} + d, \qquad j,k = 1,2,3, \tag{15.34}$$

*where $\Lambda_j \ge 0$ and $d \ge 0$.*

*Proof* It follows from Lemma 15.2.1 that for all $x \in B_\delta(x_0)$

$$\left(w_{1,2} + x_2\right)\left(w_{2,3} + x_3\right)\left(w_{3,1} + x_1\right) = \left(w_{1,3} + x_3\right)\left(w_{2,1} + x_1\right)\left(w_{3,2} + x_2\right). \tag{15.35}$$

We differentiate each side of (15.35) with respect to $x_1$ and $x_2$ in $B_\delta(x_0)$ and see that

$$w_{2,3} + x_3 = w_{1,3} + x_3. \tag{15.36}$$

Therefore, we must have $w_{2,3} = w_{1,3}$. Differentiating twice more with respect to $x_1$ and $x_3$, and $x_2$ and $x_3$, we see that if (15.35) holds for all $x \in B_\delta\,(x_0)$ then

$$w_{2,3} = w_{1,3}, \quad w_{1,2} = w_{3,2}, \quad \text{and} \quad w_{3,1} = w_{2,1}. \tag{15.37}$$

This implies that for some $(d_1, d_2, d_3)$

$$W = \begin{pmatrix} w_{1,1} & d_2 & d_3 \\ d_1 & w_{2,2} & d_3 \\ d_1 & d_2 & w_{3,3} \end{pmatrix}. \tag{15.38}$$

Furthermore, since $W$ is symmetric, we must have $d_1 = d_2 = d_3$.

Set $d = d_i$, $i = 1, 2, 3$. Then, since $w_{i,i} \geq w_{i,j}$, $i, j = 1, 2, 3$, we can write $w_{i,i} = \lambda_i + d$ for some $\lambda_i \geq 0$, $i = 1, 2, 3$. This shows that (15.34) holds. □

In using Lemma 15.3.1 we often consider $3 \times 3$ principle submatrices of a larger matrix. Consider the matrix $\{W(x, y)\}_{x,y \in S}$, for some index set $S$. Let $\{x_1, x_2, x_3\} \subset S$. Consistent with the notation introduced at the beginning of Sect. 15.2 we note that

$$W_{\{x_1,x_2,x_3\}} = \{W_{x_j,x_k}\}_{j,k=1}^{3}. \tag{15.39}$$

We also use $1_n$ to denote an $n \times n$ matrix with all its elements equal to 1.

*Proof of Theorem 15.1.2* If $(i)$ holds then

$$\widetilde{U}^f := \Lambda + 1_{|T|}G, \tag{15.40}$$

where $G$ is a $|T| \times |T|$ diagonal matrix with entries $f_1 + d, f_2 + d, \dots$. Let $\mathcal{I}$ be any finite subset of $T$. Obviously,

$$\left(\widetilde{U}^f\right)_{\mathcal{I}} = \Lambda_{\mathcal{I}} + 1_{|\mathcal{I}|}G_{\mathcal{I}}. \tag{15.41}$$

Since

$$G_{\mathcal{I}}^{1/2}\left(\Lambda_{\mathcal{I}} + 1_{|\mathcal{I}|}G_{\mathcal{I}}\right)G_{\mathcal{I}}^{-1/2} = \Lambda_{\mathcal{I}} + G_{\mathcal{I}}^{1/2}1_{|\mathcal{I}|}G_{\mathcal{I}}^{1/2}, \tag{15.42}$$

and $\Lambda_{\mathcal{I}} + G_{\mathcal{I}}^{1/2}1_{|\mathcal{I}|}G_{\mathcal{I}}^{1/2}$ is symmetric, we see that $\widetilde{U}^f$ is symmetrizable. This shows that if $(i)$ holds then $(ii)$ does not hold.

Suppose that $(i)$ does not hold. We show that in this case we can find a triple $\{t_1, t_2, t_3\}$ such that $U_{\{t_1,t_2,t_3\}}$ does not have all its off diagonal elements equal.

Since $(i)$ does not hold there are two off diagonal elements of $V$ that are not equal, say $u_{l,m} = a$ and $u_{p,q} = b$. Suppose that none of the indices $l, m, p, q$ are equal. The kernel of $(X_l, X_m, X_p)$ has the form.

$$U_{\{l,m,p\}} = \begin{pmatrix} \cdot & a & \cdot \\ a & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{pmatrix}, \tag{15.43}$$

where we use $\cdot$ when we don't know the value of the entry. If any of the off diagonal terms of $U_{\{l,m,p\}}$ are not equal to $a$ we are done.

Assume then that all the off diagonal terms of $U_{\{l,m,p\}}$ are equal. This implies, in particular, that $(U_{\{l,m,p\}})_{m,p} = (U_{\{l,m,p\}})_{p,m} = a$. Therefore, $U_{\{m,p,q\}}$ has the form,

$$U_{\{m,p,q\}} := \begin{pmatrix} \cdot & a & \cdot \\ a & \cdot & b \\ \cdot & b & \cdot \end{pmatrix}. \tag{15.44}$$

Therefore, if none of the indices $l, m, p, q$ are equal we see that there exists a triple $\{t_1, t_2, t_3\}$ such that $U_{\{t_1,t_2,t_3\}}$ does not have all its off diagonal elements equal.

If $l = p$ the argument is simpler, because in this case

$$U_{\{l,m,q\}} = \begin{pmatrix} \cdot & a & b \\ a & \cdot & \cdot \\ b & \cdot & \cdot \end{pmatrix}. \tag{15.45}$$

If $m = q$ the kernel of $(X_l, X_p, X_m)$ is

$$\begin{pmatrix} \cdot & \cdot & a \\ \cdot & \cdot & b \\ a & b & \cdot \end{pmatrix}. \tag{15.46}$$

Using the fact that $U$ is symmetric we see that cases when $l = q$ or $m = p$ are included in the above.

This shows that when $(i)$ does not hold we can find a triple $\{t_1, t_2, t_3\}$ such that $U_{\{t_1,t_2,t_3\}}$ does not have all its off diagonal elements equal. We now show that in this case $(ii)$ holds, that is, we can find a potential $f$ for which (15.7) is not symmetrizable.

For convenience we rearrange the indices so that $\{t_1, t_2, t_3\} = \{1, 2, 3\}$. We take any $h^* \in \ell_1^+$ and consider the potential $f^* = Uh^*$. If $U_{1,2,3} := \{U_{j,k} + f_k^*\}_{j,k=1}^3$ is not symmetrizable, we are done. That is, $(ii)$ holds with $f = f^*$. However, it is possible that $U_{\{1,2,3\}}$ is not of the form of (15.34) but

$$\left(U_{1,2} + f_2^*\right)\left(U_{2,3} + f_3^*\right)\left(U_{3,1} + f_1^*\right) = \left(U_{1,3} + f_3^*\right)\left(U_{2,1} + f_1^*\right)\left(U_{3,2} + f_2^*\right). \tag{15.47}$$

(See (15.35).) Nevertheless, since $U_{\{1,2,3\}}$ is not of the form (15.34), it follows from Lemma 15.3.1 that for all $\delta > 0$ there exists an $(f_1, f_2, f_3) \in B_\delta(f_1^*, f_2^*, f_3^*)$ such that $\{U_{j,k} + f_k\}_{j,k=1}^3$ is not symmetrizable. (Here we use the facts that a symmetric potential density $U_{j,k}$ is always positive and satisfies $U_{j,k} \le U_{j,j} \wedge U_{k,k}$, see [4, (13.2)].)

Note that $U_{\{1,2,3\}}$ is invertible. (See e.g., [5, Lemma A.1].) Therefore, we can find $c_1, c_2, c_3$ such that

$$f_j = f_j^* + \sum_{k=1}^{3} U_{j,k} c_k, \qquad j = 1, 2, 3. \tag{15.48}$$

Now, set $h = h^* + c$, where $c = (c_1, c_2, c_3, 0, 0, \ldots)$, i.e., all the components of $c$ except for the first three are equal to 0 and set $f = Uh$. The components $f_1, f_2, f_3$ are given by (15.48). Furthermore, we can choose $\delta$ sufficiently small so that for $(f_1, f_2, f_3) \in B_\delta(f_1^*, f_2^*, f_3^*)$, $c_1, c_2, c_3$ are small enough so that $h_1, h_2$, and $h_3$ are strictly greater than 0, which, of course, implies that $h \in \ell_1^+$, (defined just prior to Theorem 15.1.2). Therefore, $(ii)$ holds with this potential $f$. □

In Theorem 15.1.2 it is obvious that if $(i)$ does not hold then there are functions $f$ for which (15.7) is not symmetrizable. What was a little difficult was to show that $f = (f_1, f_2, \ldots)$, is a potential for $X$. We have the same problem in the proof of Theorem 15.1.3 but it is much more complicated. If we start with a potential $f^* = Uh^*$, to show that $\widetilde{U}^f$ is not asymptotically symmetrizable, we may need to modify an infinite number of the components of $f^*$ and still end up with a potential $f$. The next lemma is the key to doing this.

**Lemma 15.3.2** *Let $X = (\Omega, \mathcal{F}_t, X_t, \theta_t, P^x)$ be a transient symmetric Borel right process with state space $\mathbb{N}$, and potential $U = \{U_{j,k}\}_{j,k \in \mathbb{N}}$. Then we can find a potential function $f = Uh$, with $h \in \ell_1^+$, such that for all $\alpha > 0$,*

$$\widetilde{U}^f_{j,k} = U_{j,k} + f_k, \qquad j, k \in \mathbb{N}, \tag{15.49}$$

*is the kernel of an $\alpha$-permanental sequence.*

*Moreover, for $I_l = \{3l + 1, 3l + 2, 3l + 3\}$, the following dichotomy holds for each $l \geq 0$:*

*(i) Either $\widetilde{U}^f_{I_l}$ is not symmetrizable,*
*(ii) or*

$$U_{I_l} = \Lambda + d 1_3, \tag{15.50}$$

*where $\Lambda \in D_{3,+}$ and $d \geq 0$.*

*Proof* Let $\{i_{l,j} = 3l + j\}_{l \geq 0, j \in \{1.2.3\}}$. For $f = \{f_k\}_{k=1}^\infty$ define,

$$\begin{aligned} F_l(f) &= F_l(f_{i_{l,1}}, f_{i_{l,2}}, f_{i_{l3}}) \\ &= (U_{i_{l,1},i_{l,2}} + f_{i_{l,2}})(U_{i_{l,2},i_{l,3}} + f_{i_{l,3}})(U_{i_{l,3},i_{l,1}} + f_{i_{l,1}}) \\ &\quad -(U_{i_{l,1},i_{l,3}} + f_{i_{l,3}})(U_{i_{l,3},i_{l,2}} + f_{i_{l,2}})(U_{i_{l,2},i_{l,1}} + f_{i_{l,1}}). \end{aligned} \tag{15.51}$$

We note that when $U_{I_l}$ is given by (15.50), then for any sequence $\{f_{i_1}, f_{i_2}, f_{i_3}\}$, $F_l(f) = 0$ and $\widetilde{U}^f_{I_l}$ is symmetrizable. The first assertion in the previous sentence follows because all the terms $\{U_{i_j, i_k}\}_{j \neq k=1}^3$ are equal $d$. The second is proved in the first paragraph of the proof of Theorem 15.1.2. On the other hand, it follows from Lemma 15.2.1 that if $F_l(f) \neq 0$ then $\widetilde{U}^f_{I_l}$ is not symmetrizable.

Therefore, to prove this theorem it suffices to find an $h \in \ell_1^+$ for which the potential function $f = Uh$ satisfies the following dichotomy for each $l \geq 0$:

$$\text{Either } F_l(f) \neq 0 \quad \text{or} \quad U_{I_l} \text{ has the form (15.50).} \tag{15.52}$$

To find $h$ we take any function $h^* \in \ell_1^+$ and define successively $h^{(n)} \in \ell_1^+$, $n \geq -1$, such that $h^{(-1)} = h^*$ and

$$h_j^{(n+1)} = h_j^{(n)}, \qquad \forall j \notin I_n, \quad \text{and} \quad 0 < \frac{1}{2} h_j^* \leq h_j^{(n)} \leq 2h_j^*, \quad j \geq 1, \tag{15.53}$$

and such that $f^{(n)} := Uh^{(n)}$ satisfies,

$$|F_l(f^{(n+1)}) - F_l(f^{(n)})| \leq \frac{|F_l(f^{(l+1)})|}{2^{n+2}}, \quad n \geq l+1. \tag{15.54}$$

As we point out just below (15.51), if $U_{I_l}$ is of the form (15.50), (15.54) is satisfied trivially since $F_l(f) = 0$ for all $f$. However, when $U_{I_l}$ is not of the form (15.50) we also require that $h^{(l+1)}$ is such that

$$F_l(f^{(l+1)}) \neq 0. \tag{15.55}$$

(The actual construction of $\{h^{(n)}; n \geq -1\}$ is given later in this proof.) □

By (15.53), $\|h^{(n)} - h^{(m)}\|_1 \leq 2 \sum_{j=m}^n h_j^*$ for any $n > m$, hence $h = \lim_{n\to\infty} h^{(n)}$ exists in $\ell_1^+$. We set $f = Uh$ and note that

$$|f_j - f_j^{(n)}| = |(U(h - h^{(n)}))_j| \leq U_{j,j} \|h - h^{(n)}\|_1. \tag{15.56}$$

Here we use the property pointed out in the proof of Theorem 15.1.2 that $U_{i,j} \leq U_{i,i} \wedge U_{j,j}$.

It follows from (15.56) that $f_j = \lim_{n\to\infty} f_j^{(n)}$ for each $j \geq 1$ and consequently, by (15.54),

$$|F_l(f) - F_l(f^{(l+1)})| \leq \sum_{k=l+1}^{\infty} |F_l(f^{(k+1)}) - F_l(f^{(k)})| \leq \frac{|F_l(f^{(l+1)})|}{2}. \tag{15.57}$$

We see from this that when $U_{I_l}$ is not of the form (15.50), it follows from (15.55) and (15.57) that $F_l(f) \neq 0$. This implies that (15.52) holds.

We now describe how the $h^{(j)}$, $j = 0, 1, \ldots$ are chosen. Assume that $h^{(-1)}, \ldots, h^{(n)}$ have been chosen. We choose $h^{(n+1)}$ as follows: If either $F_n(f^{(n)}) \neq 0$ or $U\big|_{I_n \times I_n}$ has the form (15.50), we set $h^{(n+1)} = h^{(n)}$.

Assume then that $F_n(f^{(n)}) = 0$. If $U_{I_n}$ does not have the form of (15.50), it follows from the proof of Lemma 15.3.1 that for all $\epsilon_p \downarrow 0$, there exists a $(g_{1,p}, g_{2,p}, g_{3,p}) \in B_{\epsilon_p}(f_{i_{n,1}}^{(n)}, f_{i_{n,2}}^{(n)}, f_{i_{n,3}}^{(n)})$ such that $F_n(g_{1,p}, g_{2,p}, g_{3,p}) \neq 0$. We choose $f^{(n+1)} = f^{(n)}$ for all indices except $i_{n,1}, i_{n,2}, i_{n,3}$ and $f_{i_{n,1}}^{(n+1)}, f_{i_{n,2}}^{(n+1)}, f_{i_{n,3}}^{(n+1)}$ to be equal to one of these triples $(g_{1,p}, g_{2,p}, g_{3,p})$. This gives (15.55) for $l = n$. Since $\epsilon_p \downarrow 0$ we can take $f^{(n+1)}$ arbitrarily close to $f^{(n)}$ so that it satisfies (15.54).

As in the proof of Theorem 15.1.2 we can solve the equation

$$f^{(n+1)}_{i_{n,j}} = f^{(n)}_{i_{n,j}} + \sum_{k=1}^{3} U_{i_{n,j},i_{n,k}} c_{i_{n,k}}, \qquad j = 1, 2, 3. \tag{15.58}$$

for $c_{i_{n,1}}, c_{i_{n,2}}, c_{i_{n,3}}$. To obtain $h^{(n+1)}$ we set $h^{(n+1)}_q = h^{(n)}_q$ for all $q \notin I_n$ and for $q \in I_n$ we take

$$h^{(n+1)}_q = h^{(n)}_q + c^{(n)}_q. \tag{15.59}$$

where $c^{(n)}_q$ has all its components equal to zero except for the three components $c_{i_{n,1}}, c_{i_{n,2}}, c_{i_{n,3}}$. By taking $\epsilon_p$ sufficiently small we can choose $c_{i_{n,1}}, c_{i_{n,2}}, c_{i_{n,3}}$ so that the third statement in (15.53) holds.

We set $f^{(n+1)} = Uh^{(n+1)}$ and note that this is consistent with (15.58). □

*Proof of Theorem 15.1.3* It is clear from Theorem 15.1.2 that if $(i)$ holds then $U$ is asymptotically symmetrizable, because in this case $\{U_{t_i,t_j}\}_{i,j=1}^{k}$ is symmetrizable for all distinct $t_1, \dots, t_k$ greater than or equal to $n_0$, for all $k$.

Suppose that $(i)$ does not hold. Then, as in the proof of Theorem 15.1.2, we can find a sequence $\{n_k; k \in \mathbb{N}\}$ such that $n_k \to \infty$ and a sequence of triples $3n_k < t_{k,1}, t_{k,2}, t_{k,3} \leq 3n_{k+1}$, such that $U_{\{t_{k,1},t_{k,2},t_{k,3}\}}$ does not have all of its off diagonal elements equal. We interchange the indices $t_{k,1}, t_{k,2}, t_{k,3}$ with the indices in $I_{n_k}$; (see Lemma 15.3.2). We can now use Lemma 15.3.2 to show that $(ii)$ holds. □

*Proof of Theorem 15.1.4* Let $S' = \{x_0, x_1, x_2, \dots\}$ with $\lim_{k\to\infty} x_k = x_0$. Assume that for some integer $n_0$

$$u(x_j, x_k) = \Lambda_j \delta_{x_j,x_k} + d, \qquad \forall j, k \geq n_0. \tag{15.60}$$

Then $u(x_j, x_j) = \Lambda_j + d$, and since, by hypothesis, $u(x, y)$ is continuous,

$$\lim_{j\to\infty} u(x_j, x_j) = u(x_0, x_0), \tag{15.61}$$

which implies that limit $\Lambda_0 := \lim_{j\to\infty} \Lambda_j$ must exist and

$$u(x_0, x_0) = \Lambda_0 + d. \tag{15.62}$$

It also follows from (15.60) that $u(x_j, x_k) = d$ for all $n_0 \leq j < k$. In addition, since $\lim_{k\to\infty} u(x_j, x_k) = u(x_j, x_0)$, we see that for all $j \geq n_0$,

$$u(x_j, x_0) = d. \tag{15.63}$$

Comparing the last two displays we get that for all $j \geq n_0$,

$$u(x_0, x_0) - u(x_j, x_0) = \Lambda_0. \tag{15.64}$$

This contradicts (15.60), because the assumption that $u(x_0, x_0) > u(x_j, x_0)$ implies that $\Lambda_0 > 0$, whereas the assumption that $u$ is continuous and (15.64) implies that $\Lambda_0 = 0$.

Since (15.60) does not hold for any integer $n_0$, (15.10) follows from Theorem 15.1.3. The fact that $f$ is continuous at $x_0$ follows from the Dominated Convergence Theorem since $\lim_{j,k\to\infty} u(x_j, x_k) = u(x_0, x_0)$ implies that $\{u(x, y); x, y \in S'\}$ is uniformly bounded. □

## 15.4 Proof of Lemma 15.1.1 and Examples 15.1.1 and 15.1.2

*Proof of Lemma 15.1.1*

(i) Let $m_1, m_2, m_3$ be increasing integers such that $m_2 - m_1 = m_3 - m_2$ and $u(m_2 - m_1) \neq u(m_3 - m_1)$ and consider the $3 \times 3$ Töeplitz matrix

$$\begin{pmatrix} u(0) + f(m_1) & u(m_2 - m_1) + f(m_2) & u(m_3 - m_1) + f(m_3) \\ u(m_2 - m_1) + f(m_1) & u(0) + f(m_2) & u(m_2 - m_1) + f(m_3) \\ u(m_3 - m_1) + f(m_1) & u(m_2 - m_1) + f(m_2) & u(0) + f(m_3) \end{pmatrix}. \tag{15.65}$$

By Lemma 15.2.1, if $\{\widetilde{u}^f(j, k); j, k \in \mathbb{N}\}$ is symmetrizable we must have

$$\begin{aligned} &(u(m_2 - m_1) + f(m_2))(u(m_2 - m_1) + f(m_3))(u(m_3 - m_1) + f(m_1)) \\ &\quad = (u(m_3 - m_1) + f(m_3))(u(m_2 - m_1) + f(m_1))(u(m_2 - m_1) + f(m_2)). \end{aligned} \tag{15.66}$$

Note that we can cancel the term $u(m_2-m_1)+f(m_2)$ from each side of (15.66) and rearrange it to get

$$(u(m_2 - m_1) - u(m_3 - m_1))(f(m_1) - f(m_3)) = 0. \tag{15.67}$$

This is not possible because $u(m_2 - m_1) \neq u(m_3 - m_1)$ and $f(m_1) \neq f(m_3)$.

Since this holds for all $m_1, m_2, m_3$ satisfying the conditions above we see that Lemma 15.1.1 $(i)$ holds.

(ii) Consider $s_j \wedge s_k$ at the three different values, $s_{j_1}, s_{j_2}, s_{j_3}$, and the matrix

$$\begin{pmatrix} s_{j_1} + f(s_{j_1}) & s_{j_1} + f(s_{j_2}) & s_{j_1} + f(s_{j_3}) \\ s_{j_1} + f(s_{j_1}) & s_{j_2} + f(s_{j_2}) & s_{j_2} + f(s_{j_3}) \\ s_{j_1} + f(s_{j_1}) & s_{j_2} + f(s_{j_2}) & s_{j_3} + f(s_{j_3}) \end{pmatrix}. \tag{15.68}$$

By Lemma 15.2.1, if $\widetilde{v}^{f}_{s_{j_1},s_{j_2},s_{j_3}}$ is symmetrizable we must have

$$(s_{j_1}+f(s_{j_2}))(s_{j_2}+f(s_{j_3}))(s_{j_1}+f(s_{j_1}))=(s_{j_1}+f(s_{j_3}))(s_{j_1}+f(s_{j_1}))(s_{j_2}+f(s_{j_2})) \tag{15.69}$$

or, equivalently,

$$(s_{j_1}-s_{j_2})(f(s_{j_3})-f(s_{j_2}))=0. \tag{15.70}$$

Since $s_{j_1}\neq s_{j_2}$ and $f(s_{j_3})\neq f(s_{j_2})$ this is not possible. Therefore, $\widetilde{v}^{f}_{s_{j_1},s_{j_2},s_{j_3}}$ is not symmetrizable. □

*Proof of Example 15.1.1* Let $s_0\in S$. We choose a sequence $s_j\to s_0$ with the property that it contains a subsequence $\{s_{j_k}\}$, $s_{j_k}\to s_0$, such that

$$s_{j_{3k+1}}-s_{j_{3k}}=s_{j_{3k+2}}-s_{j_{3k+1}}=a_k, \qquad k\geq 1. \tag{15.71}$$

The kernel of the $3\times 3$ matrix

$$\widehat{u}^{f}(s_{j_{3k+p}},s_{j_{3k+q}}), \qquad p,q=0,1,2, \tag{15.72}$$

is

$$\begin{pmatrix} 1+f(s_{j_{3k}}) & e^{-\lambda a_k}+f(s_{j_{3k+1}}) & e^{-\lambda 2a_k}+f(s_{j_{3k+2}}) \\ e^{-\lambda a_k}+f(s_{j_{3k}}) & 1+f(s_{j_{3k+1}}) & e^{-\lambda a_k}+f(s_{j_{3k+2}}) \\ e^{-\lambda 2a_k}+f(s_{j_{3k}}) & e^{-\lambda a_k}+f(s_{j_{3k+1}}) & 1+f(s_{j_{3k+2}}) \end{pmatrix}, \tag{15.73}$$

similar to (15.65). Therefore, following the proof of Lemma 15.1.1, we see that the kernel in (15.72) is not symmetrizable. Since this holds along the subsequence $\{s_{j_k}\}$, $s_{j_k}\to s_0$, we see that $\{\widehat{u}_f(s,t);\, s,t\in S\}$ is not asymptotically symmetrizable at $s_0$.

The result in (15.15) is proved similarly. □

*Proof of Example 15.1.2* The proof of Example 15.1.2 is similar to the proof of Example 15.1.2 but even simpler. This is because for all distinct values, $s_{j_1}, s_{j_2}, s_{j_3}$, the matrix in (15.68) is not symmetrizable. □

**Acknowledgement** Research of Jay Rosen was partially supported by a grant from the Simons Foundation.

## References

1. N. Eisenbaum, Permanental vectors with nonsymmetric kernels. Ann. Probab. **45**, 210–224 (2017)
2. N. Eisenbaum, H. Kaspi, On permanental processes. Stoch. Processes Appl. **119**, 1401–1415 (2009)

3. N. Eisenbaum, F. Maunoury, Existence conditions of permanental and multivariate negative binomial distributions. Ann. Probab. **45**, 4786–4820 (2017)
4. M.B. Marcus, J. Rosen, *Markov Processes, Gaussian Processes and Local Times* (Cambridge University Press, New York, 2006)
5. M.B. Marcus, J. Rosen, Conditions for permanental processes to be unbounded. Ann. Probab. **45**, 2059–2086 (2017)
6. M.B. Marcus, J. Rosen, Sample path properties of permanental processes. Electron. J. Probab. **23**(58), 1–47 (2018)

# Chapter 16
# Pointwise Properties of Martingales with Values in Banach Function Spaces

**Mark Veraar and Ivan Yaroslavtsev**

**Abstract** In this paper we consider local martingales with values in a UMD Banach function space. We prove that such martingales have a version which is a martingale field. Moreover, a new Burkholder–Davis–Gundy type inequality is obtained.

**Keywords** Local martingale · Quadratic variation · UMD Banach function spaces · Burkholder-Davis-Gundy inequalities · Lattice maximal function

**2010 Mathematics Subject Classification** Primary: 60G44; Secondary: 60B11, 60H05, 60G48

## 16.1 Introduction

The discrete Burkholder–Davis–Gundy inequality (see [3, Theorem 3.2]) states that for any $p \in (1, \infty)$ and martingales difference sequence $(d_j)_{j=1}^n$ in $L^p(\Omega)$ one has

$$\Big\| \sum_{j=1}^n d_j \Big\|_{L^p(\Omega)} \eqsim_p \Big\| \Big( \sum_{j=1}^n |d_j|^2 \Big)^{1/2} \Big\|_{L^p(\Omega)}. \tag{16.1}$$

Moreover, there is the extension to continuous-time local martingales $M$ (see [13, Theorem 26.12]) which states that for every $p \in [1, \infty)$,

$$\Big\| \sup_{t \in [0,\infty)} |M_t| \Big\|_{L^p(\Omega)} \eqsim_p \big\| [M]_\infty^{1/2} \big\|_{L^p(\Omega)}. \tag{16.2}$$

Here $t \mapsto [M]_t$ denotes the quadratic variation process of $M$.

M. Veraar (✉) · I. Yaroslavtsev
Delft Institute of Applied Mathematics, Delft University of Technology, Delft, The Netherlands
e-mail: M.C.Veraar@tudelft.nl; yaroslavtsev.i.s@yandex.ru



N. Gozlan et al. (eds.), *High Dimensional Probability VIII*,
Progress in Probability 74, https://doi.org/10.1007/978-3-030-26391-1_16

In the case $X$ is a UMD Banach function space the following variant of (16.1) holds (see [24, Theorem 3]): for any $p \in (1, \infty)$ and martingales difference sequence $(d_j)_{j=1}^n$ in $L^p(\Omega; X)$ one has

$$\Big\| \sum_{j=1}^n d_j \Big\|_{L^p(\Omega;X)} \eqsim_p \Big\| \Big( \sum_{j=1}^n |d_j|^2 \Big)^{1/2} \Big\|_{L^p(\Omega;X)}. \tag{16.3}$$

Moreover, the validity of the estimate also characterizes the UMD property.

It is a natural question whether (16.2) has a vector-valued analogue as well. The main result of this paper states that this is indeed the case:

**Theorem 16.1.1** *Let $X$ be a UMD Banach function space over a $\sigma$-finite measure space $(S, \Sigma, \mu)$. Assume that $N : \mathbb{R}_+ \times \Omega \times S \to \mathbb{R}$ is such that $N|_{[0,t]\times\Omega\times S}$ is $\mathcal{B}([0, t]) \otimes \mathcal{F}_t \otimes \Sigma$-measurable for all $t \geq 0$ and such that for almost all $s \in S$, $N(\cdot, \cdot, s)$ is a martingale with respect to $(\mathcal{F}_t)_{t\geq 0}$ and $N(0, \cdot, s) = 0$. Then for all $p \in (1, \infty)$,*

$$\big\| \sup_{t\geq 0} |N(t, \cdot, \cdot)| \big\|_{L^p(\Omega;X)} \eqsim_{p,X} \sup_{t\geq 0} \big\| N(t, \cdot, \cdot) \big\|_{L^p(\Omega;X)} \eqsim_{p,X} \|[N]_\infty^{1/2}\|_{L^p(\Omega;X)}. \tag{16.4}$$

*where $[N]$ denotes the quadratic variation process of $N$.*

By standard methods we can extend Theorem 16.1.1 to spaces $X$ which are isomorphic to a closed subspace of a Banach function space (e.g. Sobolev and Besov spaces, etc.)

The two-sided estimate (16.4) can for instance be used to obtain two-sided estimates for stochastic integrals for processes with values in infinite dimensions (see [25] and [26]). In particular, applying it with $N(t, \cdot, s) = \int_0^t \Phi(\cdot, s)\, dW$ implies the following maximal estimate for the stochastic integral

$$\begin{aligned} &\Big\| s \mapsto \sup_{t\geq 0} \Big| \int_0^t \Phi(\cdot, s)\, dW \Big| \Big\|_{L^p(\Omega;X)} \\ &\eqsim_{p,X} \sup_{t\geq 0} \Big\| s \mapsto \int_0^t \Phi(\cdot, s)\, dW \Big\|_{L^p(\Omega;X)} \\ &\eqsim_{p,X} \Big\| s \mapsto \Big( \int_0^\infty \Phi^2(t, s)\, dt \Big)^{1/2} \Big\|_{L^p(\Omega;X)}, \end{aligned} \tag{16.5}$$

where $W$ is a Brownian motion and $\Phi : \mathbb{R}_+ \times \Omega \times S \to \mathbb{R}$ is a progressively measurable process such that the right-hand side of (16.5) is finite. The second norm equivalence was obtained in [25]. The norm equivalence with the left-hand side is new in this generality. The case where $X$ is an $L^q$-space was recently obtained in [1] using different methods.

It is worth noticing that the second equivalence of (16.4) in the case of $X = L^q$ was obtained by Marinelli in [18] for some range of $1 < p, q < \infty$ by using an interpolation method.

The UMD property is necessary in Theorem 16.1.1 by necessity of the UMD property in (16.3) and the fact that any discrete martingale can be transformed to a continuous-time one. Also in the case of continuous martingales, the UMD property is necessary in Theorem 16.1.1. Indeed, applying (16.5) with $W$ replaced by an independent Brownian motion $\widetilde{W}$ we obtain

$$\Big\| \int_0^\infty \Phi \, \mathrm{d}W \Big\|_{L^p(\Omega;X)} \eqsim_{p,X} \Big\| \int_0^\infty \Phi \, \mathrm{d}\widetilde{W} \Big\|_{L^p(\Omega;X)},$$

for all predictable step processes $\Phi$. The latter holds implies that $X$ is a UMD Banach space (see [10, Theorem 1]).

In the special case that $X = \mathbb{R}$ the above reduces to (16.2). In the proof of Theorem 16.1.1 the UMD property is applied several times:

- The boundedness of the lattice maximal function (see [2, 9, 24]).
- The $X$-valued Meyer–Yoeurp decomposition of a martingale (see Lemma 16.2.1).
- The square-function estimate (16.3) (see [24]).

It remains open whether there exists a predictable expression for the right-hand side of (16.4). One would expect that one needs simply to replace $[N]$ by its predictable compensator, the *predictable quadratic variation* $\langle N \rangle$. Unfortunately, this does not hold true already in the scalar-valued case: if $M$ is a real-valued martingale, then

$$\mathbb{E}|M|_t^p \lesssim_p \mathbb{E}\langle M \rangle_t^{\frac{p}{2}}, \quad t \geq 0, \quad p < 2,$$

$$\mathbb{E}|M|_t^p \gtrsim_p \mathbb{E}\langle M \rangle_t^{\frac{p}{2}}, \quad t \geq 0, \quad p > 2,$$

where both inequalities are known not to be sharp (see [3, p. 40], [19, p. 297], and [21]). The question of finding such a predictable right-hand side in (16.4) was answered only in the case $X = L^q$ for $1 < q < \infty$ by Dirsken and the second author (see [7]). The key tool exploited there was the so-called *Burkholder-Rosenthal inequalities*, which are of the following form:

$$\mathbb{E}\|M_N\|^p \eqsim_{p,X} \big|\!\big|\!\big|(M_n)_{0 \leq n \leq N}\big|\!\big|\!\big|_{p,X}^p,$$

where $(M_n)_{0 \leq n \leq N}$ is an $X$-valued martingale, $|\!|\!|\cdot|\!|\!|_{p,X}$ is a certain norm defined on the space of $X$-valued $L^p$-martingales which depends only on *predictable moments* of the corresponding martingale. Therefore using approach of [7] one can reduce the problem of continuous-time martingales to discrete-time martingales. However, the Burkholder-Rosenthal inequalities are explored only in the case $X = L^q$.

Thanks to (16.2) the following natural question arises: can one generalize (16.4) to the case $p = 1$, i.e. whether

$$\big\| \sup_{t\geq 0} |N(t,\cdot,\cdot)| \big\|_{L^1(\Omega;X)} \eqsim_{p,X} \|[N]_\infty^{1/2}\|_{L^1(\Omega;X)} \tag{16.6}$$

holds true? Unfortunately the outlined earlier techniques cannot be applied in the case $p = 1$. Moreover, the obtained estimates cannot be simply extrapolated to the case $p = 1$ since those contain the *UMD$_p$ constant*, which is known to have infinite limit as $p \to 1$. Therefore (16.6) remains an open problem. Note that in the case of a continuous martingale $M$ inequalities (16.4) can be extended to the case $p \in (0, 1]$ due to the classical Lenglart approach (see Corollary 16.4.4).

## 16.2 Preliminaries

Throughout the paper any filtration satisfies the *usual conditions* (see [12, Definition 1.1.2 and 1.1.3]), unless the underlying martingale is continuous (then the corresponding filtration can be assumed general).

A Banach space $X$ is called a *UMD space* if for some (or equivalently, for all) $p \in (1,\infty)$ there exists a constant $\beta > 0$ such that for every $n \geq 1$, every martingale difference sequence $(d_j)_{j=1}^n$ in $L^p(\Omega; X)$, and every $\{-1, 1\}$-valued sequence $(\varepsilon_j)_{j=1}^n$ we have

$$\Big(\mathbb{E}\Big\|\sum_{j=1}^n \varepsilon_j d_j\Big\|^p\Big)^{\frac{1}{p}} \leq \beta\Big(\mathbb{E}\Big\|\sum_{j=1}^n d_j\Big\|^p\Big)^{\frac{1}{p}}.$$

The above class of spaces was extensively studied by Burkholder (see [4]). UMD spaces are always reflexive. Examples of UMD space include the reflexive range of $L^q$-spaces, Besov spaces, Sobolev, and Musielak-Orlicz spaces. Example of spaces without the UMD property include all nonreflexive spaces, e.g. $L^1(0, 1)$ and $C([0, 1])$. For details on UMD Banach spaces we refer the reader to [5, 11, 22, 24].

The following lemma follows from [27, Theorem 3.1].

**Lemma 16.2.1 (Meyer-Yoeurp Decomposition)** *Let $X$ be a UMD space and $p \in (1,\infty)$. Let $M : \mathbb{R}_+ \times \Omega \to X$ be an $L^p$-martingale that takes values in some closed subspace $X_0$ of $X$. Then there exists a unique decomposition $M = M^d + M^c$, where $M^c$ is continuous, $M^d$ is purely discontinuous and starts at zero, and $M^d$ and $M^c$ are $L^p$-martingales with values in $X_0 \subseteq X$. Moreover, the following norm estimates hold for every $t \in [0,\infty)$,*

$$\begin{aligned} \|M^d(t)\|_{L^p(\Omega;X)} &\leq \beta_{p,X}\|M(t)\|_{L^p(\Omega;X)}, \\ \|M^c(t)\|_{L^p(\Omega;X)} &\leq \beta_{p,X}\|M(t)\|_{L^p(\Omega;X)}. \end{aligned} \tag{16.7}$$

*Furthermore, if* $A_X^{p,d}$ *and* $A_X^{p,c}$ *are the corresponding linear operators that map* $M$ *to* $M^d$ *and* $M^c$ *respectively, then*

$$A_X^{p,d} = A_{\mathbb{R}}^{p,d} \otimes \mathrm{Id}_X,$$

$$A_X^{c,d} = A_{\mathbb{R}}^{c,d} \otimes \mathrm{Id}_X.$$

Recall that for a given measure space $(S, \Sigma, \mu)$, the linear space of all real-valued measurable functions is denoted by $L^0(S)$.

**Definition 16.2.2** Let $(S, \Sigma, \mu)$ be a measure space. Let $n : L^0(S) \to [0, \infty]$ be a function which satisfies the following properties:

(i) $n(x) = 0$ if and only if $x = 0$,
(ii) for all $x, y \in L^0(S)$ and $\lambda \in \mathbb{R}$, $n(\lambda x) = |\lambda| n(x)$ and $n(x + y) \leq n(x) + n(y)$,
(iii) if $x \in L^0(S)$, $y \in L^0(S)$, and $|x| \leq |y|$, then $n(x) \leq n(y)$,
(iv) if $0 \leq x_n \uparrow x$ with $(x_n)_{n=1}^{\infty}$ a sequence in $L^0(S)$ and $x \in L^0(S)$, then $n(x) = \sup_{n\in\mathbb{N}} n(x_n)$.

Let $X$ denote the space of all $x \in L^0(S)$ for which $\|x\| := n(x) < \infty$. Then $X$ is called the *normed function space associated to* $n$. It is called a *Banach function space* when $(X, \|\cdot\|_X)$ is complete.

We refer the reader to [31, Chapter 15] for details on Banach function spaces.

*Remark 16.2.3* Let $X$ be a Banach function space over a measure space $(S, \Sigma, \mu)$. Then $X$ is continuously embedded into $L^0(S)$ endowed with the topology of convergence in measure on sets of finite measure. Indeed, assume $x_n \to x$ in $X$ and let $A \in \Sigma$ be of finite measure. We claim that $\mathbf{1}_A x_n \to \mathbf{1}_A x$ in measure. For this it suffices to show that every subsequence of $(x_n)_{n\geq 1}$ has a further subsequence which convergences a.e. to $x$. Let $(x_{n_k})_{k\geq 1}$ be a subsequence. Choose a subsubsequence $(\mathbf{1}_A x_{n_{k_\ell}})_{\ell\geq 1} =: (y_\ell)_{\ell\geq 1}$ such that $\sum_{\ell=1}^{\infty} \|y_\ell - x\| < \infty$. Then by [31, Exercise 64.1] $\sum_{\ell=1}^{\infty} |y_\ell - x|$ converges in $X$. In particular, $\sum_{\ell=1}^{\infty} |y_\ell - x| < \infty$ a.e. Therefore, $y_\ell \to x$ a.e. as desired.

Given a Banach function space $X$ over a measure space $S$ and Banach space $E$, let $X(E)$ denote the space of all strongly measurable functions $f : S \to E$ with $\|f\|_{X(E)} := \big\| s \mapsto \|f(s)\|_E \big\|_X \in X$. The space $X(E)$ becomes a Banach space when equipped with the norm $\|f\|_{X(E)}$.

A Banach function space has the UMD property if and only if (16.3) holds for some (or equivalently, for all) $p \in (1, \infty)$ (see [24]). A broad class of Banach function spaces with UMD is given by the reflexive Lorentz–Zygmund spaces (see [6]) and the reflexive Musielak–Orlicz spaces (see [17]).

**Definition 16.2.4** $N : \mathbb{R}_+ \times \Omega \times S \to \mathbb{R}$ is called a (continuous) (local) martingale field if $N|_{[0,t]\times\Omega\times S}$ is $\mathcal{B}([0, t]) \otimes \mathcal{F}_t \otimes \Sigma$-measurable for all $t \geq 0$ and $N(\cdot, \cdot, s)$ is a (continuous) (local) martingale with respect to $(\mathcal{F}_t)_{t\geq 0}$ for almost all $s \in S$.

Let $X$ be a Banach space, $I \subset \mathbb{R}$ be a closed interval (perhaps, infinite). A function $f : I \to X$ is called *càdlàg* (an acronym for the French phrase "continue à droite, limite à gauche") if $f$ is right continuous and has limits from the left-hand side. We define a *Skorohod space* $\mathcal{D}(I; X)$ as a linear space consisting of all càdlàg functions $f : I \to X$. We denote the linear space of all bounded càdlàg functions $f : I \to X$ by $\mathcal{D}_b(I; X)$.

**Lemma 16.2.5** *$\mathcal{D}_b(I; X)$ equipped with the norm $\|\cdot\|_\infty$ is a Banach space.*

*Proof* The proof is analogous to the proof of the same statement for continuous functions. □

Let $X$ be a Banach space, $\tau$ be a stopping time, $V : \mathbb{R}_+ \times \Omega \to X$ be a càdlàg process. Then we define $\Delta V_\tau : \Omega \to X$ as follows

$$\Delta V_\tau := V_\tau - \lim_{\varepsilon \to 0} V_{(\tau-\varepsilon)\vee 0}.$$

## 16.3 Lattice Doob's Maximal Inequality

Doob's maximal $L^p$-inequality immediately implies that for martingale fields

$$\Big\| \sup_{t\geq 0} \|N(t,\cdot)\|_X \Big\|_{L^p(\Omega)} \leq \frac{p}{p-1} \sup_{t\geq 0} \|N(t)\|_{L^p(\Omega;X)}, \quad 1 < p < \infty.$$

In the next lemma we prove a stronger version of Doob's maximal $L^p$-inequality. As a consequence in Theorem 16.3.2 we will obtain the same result in a more general setting.

**Lemma 16.3.1** *Let $X$ be a UMD Banach function space and let $p \in (1, \infty)$. Let $N$ be a càdlàg martingale field with values in a finite dimensional subspace of $X$. Then for all $T > 0$,*

$$\Big\| \sup_{t\in[0,T]} |N(t,\cdot)| \Big\|_{L^p(\Omega;X)} \eqsim_{p,X} \sup_{t\in[0,T]} \|N(t)\|_{L^p(\Omega;X)}$$

*whenever one of the expression is finite.*

*Proof* Clearly, the left-hand side dominates the right-hand side. Therefore, we can assume the right-hand side is finite and in this case we have

$$\|N(T)\|_{L^p(\Omega;X)} = \sup_{t\in[0,T]} \|N(t)\|_{L^p(\Omega;X)} < \infty.$$

Since $N$ takes values in a finite dimensional subspace it follows from Doob's $L^p$-inequality (applied coordinatewise) that the left-hand side is finite.

Since $N$ is a càdlàg martingale field and by Definition 16.2.2(iv) we have that

$$\lim_{n\to\infty} \Big\| \sup_{0\le j\le n} |N(jT/n,\cdot)| \Big\|_{L^p(\Omega;X)} = \Big\| \sup_{t\in[0,T]} |N(t,\cdot)| \Big\|_{L^p(\Omega;X)}.$$

Set $M_j = N_{jT/n}$ for $j \in \{0,\dots,n\}$ and $M_j = M_n$ for $j > n$. It remains to prove

$$\Big\| \sup_{0\le j\le n} |M_j(\cdot)| \Big\|_{L^p(\Omega;X)} \le C_{p,X} \|M_n\|_{L^p(\Omega;X)}.$$

If $(M_j)_{j=0}^n$ is a Paley–Walsh martingale (see [11, Definition 3.1.8 and Proposition 3.1.10]), this estimate follows from the boundedness of the dyadic lattice maximal operator [24, pp. 199–200 and Theorem 3]. In the general case one can replace $\Omega$ by a divisible probability space and approximate $(M_j)$ by Paley-Walsh martingales in a similar way as in [11, Corollary 3.6.7]. □

**Theorem 16.3.2 (Doob's Maximal $L^p$-Inequality)** *Let $X$ be a UMD Banach function space over a $\sigma$-finite measure space and let $p \in (1,\infty)$. Let $M : \mathbb{R}_+ \times \Omega \to X$ be a martingale such that*

1. *for all $t \ge 0$, $M(t) \in L^p(\Omega; X)$;*
2. *for a.a $\omega \in \Omega$, $M(\cdot,\omega)$ is in $\mathcal{D}([0,\infty); X)$.*

*Then there exists a martingale field $N \in L^p(\Omega; X(\mathcal{D}_b([0,\infty))))$ such that for a.a. $\omega \in \Omega$, all $t \ge 0$ and a.a. $s \in S$, $N(t,\omega,s) = M(t,\omega)(s)$ and*

$$\Big\| \sup_{t\ge 0} |N(t,\cdot)| \Big\|_{L^p(\Omega;X)} \eqsim_{p,X} \sup_{t\ge 0} \|M(t,\cdot)\|_{L^p(\Omega;X)}. \tag{16.8}$$

*Moreover, if $M$ is continuous, then $N$ can be chosen to be continuous as well.*

*Proof* We first consider the case where $M$ becomes constant after some time $T > 0$. Then

$$\sup_{t\ge 0} \|M(t,\cdot)\|_{L^p(\Omega;X)} = \|M(T)\|_{L^p(\Omega;X)}.$$

Let $(\xi_n)_{n\ge 1}$ be simple random variables such that $\xi_n \to M(T)$ in $L^p(\Omega; X)$. Let $M_n(t) = \mathbb{E}(\xi_n|\mathcal{F}_t)$ for $t \ge 0$. Then by Lemma 16.3.1

$$\Big\| \sup_{t\ge 0} |N_n(t,\cdot) - N_m(t,\cdot)| \Big\|_{L^p(\Omega;X)} \eqsim_{p,X} \Big\| |M_n(T,\cdot) - M_m(T,\cdot)| \Big\|_{L^p(\Omega;X)} \to 0$$

as $n, m \to \infty$. Therefore, $(N_n)_{n\ge 1}$ is a Cauchy sequence and hence converges to some $N$ from the space $L^p(\Omega; X(\mathcal{D}_b([0,\infty))))$. Clearly, $N(t,\cdot) = M(t)$ and (16.8) holds in the special case that $M$ becomes constant after $T > 0$.

In the case $M$ is general, for each $T > 0$ we can set $M^T(t) = M(t \wedge T)$. Then for each $T > 0$ we obtain a martingale field $N^T$ as required. Since $N^{T_1} = N^{T_2}$ on

$[0, T_1 \wedge T_2]$, we can define a martingale field $N$ by setting $N(t, \cdot) = N^T(t, \cdot)$ on $[0, T]$. Finally, we note that

$$\lim_{T\to\infty} \sup_{t\geq 0} \|M^T(t)\|_{L^p(\Omega;X)} = \sup_{t\geq 0} \|M(t)\|_{L^p(\Omega;X)}.$$

Moreover, by Definition 16.2.2(iv) we have

$$\lim_{T\to\infty} \Big\| \sup_{t\geq 0} |N^T(t, \cdot)|\Big\|_{L^p(\Omega;X)} = \Big\| \sup_{t\geq 0} |N(t, \cdot)|\Big\|_{L^p(\Omega;X)},$$

Therefore the general case of (16.8) follows by taking limits.

Now let $M$ be continuous, and let $(M_n)_{n\geq 1}$ be as before. By the same argument as in the first part of the proof we can assume that there exists $T > 0$ such that $M_t = M_{t\wedge T}$ for all $t \geq 0$. By Lemma 16.2.1 there exists a unique decomposition $M_n = M_n^c + M_n^d$ such that $M_n^d$ is purely discontinuous and starts at zero and $M_n^c$ has continuous paths a.s. Then by (16.7)

$$\|M(T) - M_n^c(T)\|_{L^p(\Omega;X)} \leq \beta_{p,X} \|M(T) - M_n(T)\|_{L^p(\Omega;X)} \to 0.$$

Since $M_n^c$ takes values in a finite dimensional subspace of $X$ we can define a martingale field $N_n$ by $N_n(t, \omega, s) = M_n^c(t, \omega)(s)$. Now by Lemma 16.3.1

$$\Big\| \sup_{0\leq t\leq T} |N_n(t, \cdot) - N_m(t, \cdot)|\Big\|_{L^p(\Omega;X)} \eqsim_{p,X} \Big\||M_n^c(T, \cdot) - M_m^c(T, \cdot)|\Big\|_{L^p(\Omega;X)} \to 0.$$

Therefore, $(N_n)_{n\geq 1}$ is a Cauchy sequence and hence converges to some $N$ from the space $L^p(\Omega; X(C_b([0,\infty))))$. Analogously to the first part of the proof, $N(t, \cdot) = M(t)$ for all $t \geq 0$. □

*Remark 16.3.3* Note that due to the construction of $N$ we have that $\Delta M_\tau(s) = \Delta N(\cdot, s)_\tau$ for any stopping time $\tau$ and almost any $s \in S$. Indeed, let $(M_n)_{n\geq 1}$ and $(N_n)_{n\geq 1}$ be as in the proof of Theorem 16.3.2. Then on the one hand

$$\begin{aligned}\|\Delta M_\tau - \Delta(M_n)_\tau\|_{L^p(\Omega;X)} &\leq \Big\| \sup_{0\leq t\leq T} \|M(t) - M_n(t)\|_X \Big\|_{L^p(\Omega)} \\ &\eqsim_p \|M(T) - M_n(T)\|_{L^p(\Omega;X)} \to 0, \quad n \to \infty.\end{aligned}$$

On the other hand

$$\begin{aligned}\|\Delta N_\tau - \Delta(N_n)_\tau\|_{L^p(\Omega;X)} &\leq \Big\| \sup_{0\leq t\leq T} |N(t) - N_n(t)|\Big\|_{L^p(\Omega;X)} \\ &\eqsim_{p,X} \Big\||N(T) - N_n(T)|\Big\|_{L^p(\Omega;X)} \to 0, \quad n \to \infty.\end{aligned}$$

Since $\|M_n(t) - N_n(t,\cdot)\|_{L^p(\Omega;X)} = 0$ for all $n \geq 0$, we have that by the limiting argument $\|\Delta M_\tau - \Delta N_\tau(\cdot)\|_{L^p(\Omega;X)} = 0$, so the desired follows from Definition 16.2.2(i).

One could hope there is a more elementary approach to derive continuity of $N$ in the case $M$ is continuous: if the filtration $\widetilde{\mathbb{F}} := (\widetilde{\mathcal{F}}_t)_{t\geq 0}$ is generated by $M$, then $M(s)$ is $\widetilde{\mathbb{F}}$-adapted for a.e. $s \in S$, and one might expect that $M$ has a continuous version. Unfortunately, this is not true in general as follows from the next example.

*Example 16.3.4* There exists a continuous martingale $M : \mathbb{R}_+ \times \Omega \to \mathbb{R}$, a filtration $\widetilde{\mathbb{F}} = (\widetilde{\mathcal{F}}_t)_{t\geq 0}$ generated by $M$ and all $\mathbb{P}$-null sets, and a purely discontinuous nonzero $\widetilde{\mathbb{F}}$-martingale $N : \mathbb{R}_+ \times \Omega \to \mathbb{R}$. Let $W : \mathbb{R}_+ \times \Omega \to \mathbb{R}$ be a Brownian motion, $L : \mathbb{R}_+ \times \Omega \to \mathbb{R}$ be a Poisson process such that $W$ and $L$ are independent. Let $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$ be the filtration generated by $W$ and $L$. Let $\sigma$ be an $\mathbb{F}$-stopping time defined as follows

$$\sigma = \inf\{u \geq 0 : \Delta L_u \neq 0\}.$$

Let us define

$$M := \int \mathbf{1}_{[0,\sigma]}\, \mathrm{d}W = W^\sigma.$$

Then $M$ is a martingale. Let $\widetilde{\mathbb{F}} := (\widetilde{\mathcal{F}}_t)_{t\geq 0}$ be generated by $M$. Note that $\widetilde{\mathcal{F}}_t \subset \mathcal{F}_t$ for any $t \geq 0$. Define a random variable

$$\tau = \inf\{t \geq 0 : \exists u \in [0,t) \text{ such that } M \text{ is a constant on } [u,t]\}.$$

Then $\tau = \sigma$ a.s. Moreover, $\tau$ is a $\widetilde{\mathbb{F}}$-stopping time since for each $u \geq 0$

$$\mathbb{P}\{\tau = u\} = \mathbb{P}\{\sigma = u\} = \mathbb{P}\{\Delta L_u^\sigma \neq 1\} \leq \mathbb{P}\{\Delta L_u \neq 1\} = 0,$$

and hence

$$\{\tau \leq u\} = \{\tau < u\} \cup \{\tau = u\} \subset \widetilde{\mathcal{F}}_u.$$

Therefore $N : \mathbb{R}_+ \times \Omega \to \mathbb{R}$ defined by

$$N_t := \mathbf{1}_{[\tau,\infty)}(t) - t \wedge \tau \quad t \geq 0,$$

is an $\widetilde{\mathbb{F}}$-martingale since it is $\widetilde{\mathbb{F}}$-measurable and since $N_t = (L_t - t)^\sigma$ a.s. for each $t \geq 0$, hence for each $u \in [0,t]$

$$\mathbb{E}(N_t|\widetilde{\mathcal{F}}_u) = \mathbb{E}(\mathbb{E}(N_t|\mathcal{F}_u)|\widetilde{\mathcal{F}}_u) = \mathbb{E}(\mathbb{E}((L_t - t)^\sigma|\mathcal{F}_u)|\widetilde{\mathcal{F}}_u) = (L_u - u)^\sigma = N_u$$

due to the fact that $t \mapsto L_t - t$ is an $\widetilde{\mathbb{F}}$-measurable $\mathbb{F}$-martingale (see [15, Problem 1.3.4]). But $(N_t)_{t\geq 0}$ is not continuous since $(L_t)_{t\geq 0}$ is not continuous.

## 16.4 Main Result

Theorem 16.1.1 will be a consequence of the following more general result.

**Theorem 16.4.1** *Let $X$ be a UMD Banach function space over a $\sigma$-finite measure space $(S, \Sigma, \mu)$ and let $p \in (1, \infty)$. Let $M : \mathbb{R}_+ \times \Omega \to X$ be a local $L^p$-martingale with respect to $(\mathcal{F}_t)_{t\geq 0}$ and assume $M(0, \cdot) = 0$. Then there exists a mapping $N : \mathbb{R}_+ \times \Omega \times S \to \mathbb{R}$ such that*

1. *for all $t \geq 0$ and a.a. $\omega \in \Omega$, $N(t, \omega, \cdot) = M(t, \omega)$,*
2. *$N$ is a local martingale field,*
3. *the following estimate holds*

$$\Big\| \sup_{t\geq 0} |N(t, \cdot, \cdot)| \Big\|_{L^p(\Omega;X)} \eqsim_{p,X} \Big\| \sup_{t\geq 0} \|M(t, \cdot)\|_X \Big\|_{L^p(\Omega)} \eqsim_{p,X} \|[N]_\infty^{1/2}\|_{L^p(\Omega;X)}. \tag{16.9}$$

To prove Theorem 16.4.1 we first prove a completeness result.

**Proposition 16.4.2** *Let $X$ be a Banach function space over a $\sigma$-finite measure space $S$, $1 \leq p < \infty$. Let*

$$\begin{aligned} \mathrm{MQ}^p(X) := \{N : \mathbb{R}_+ \times \Omega \times S \to \mathbb{R} : &\ N \textit{ is a martingale field,} \\ &N(0, \cdot, s) = 0\ \forall s \in S, \textit{ and } \|N\|_{\mathrm{MQ}^p(X)} < \infty\}, \end{aligned}$$

*where $\|N\|_{\mathrm{MQ}^p(X)} := \|[N]_\infty^{1/2}\|_{L^p(\Omega;X)}$. Then $(\mathrm{MQ}^p(X), \|\cdot\|_{\mathrm{MQ}^p(X)})$ is a Banach space. Moreover, if $N_n \to N$ in $\mathrm{MQ}^p$, then there exists a subsequence $(N_{n_k})_{k\geq 1}$ such that pointwise a.e. in $S$, we have $N_{n_k} \to N$ in $L^1(\Omega; \mathcal{D}_b([0, \infty)))$.*

*Proof* Let us first check that $\mathrm{MQ}^p(X)$ is a normed vector space. For this only the triangle inequality requires some comments. By the well-known estimate for local martingales $M, N$ (see [13, Theorem 26.6(iii)]) we have that a.s.

$$\begin{aligned} [M + N]_t &= [M]_t + 2[M, N]_t + [N]_t \\ &\leq [M]_t + 2[M]_t^{1/2}[N]_t^{1/2} + [N]_t = \big([M]_t^{1/2} + [N]_t^{1/2}\big)^2, \end{aligned} \tag{16.10}$$

Therefore, $[M + N]_t^{1/2} \leq [M]_t^{1/2} + [N]_t^{1/2}$ a.s. for all $t \in [0, \infty]$.

Let $(N_k)_{k\geq 1}$ be such that $\sum_{k\geq 1} \|N_k\|_{\mathrm{MQ}^p(X)} < \infty$. It suffices to show that $\sum_{k\geq 1} N_k$ converges in $\mathrm{MQ}^p(X)$. Observe that by monotone convergence in $\Omega$ and

Jensen's inequality applied to $\|\cdot\|_X$ for any $n > m \geq 1$ we have

$$\begin{aligned}
&\Big\| \sum_{k=m+1}^{n} \mathbb{E}[N_k]_\infty^{1/2} \Big\|_X \\
&= \Big\| \sum_{k=1}^{n} \mathbb{E}[N_k]_\infty^{1/2} - \sum_{k=1}^{m} \mathbb{E}[N_k]_\infty^{1/2} \Big\|_X \\
&= \Big\| \mathbb{E} \sum_{k=m+1}^{n} [N_k]_\infty^{1/2} \Big\|_X \leq \mathbb{E} \Big\| \sum_{k=m+1}^{n} [N_k]_\infty^{1/2} \Big\|_X \qquad (16.11)\\
&= \Big\| \sum_{k=m+1}^{n} [N_k]_\infty^{1/2} \Big\|_{L^1(\Omega;X)} \leq \Big\| \sum_{k=m+1}^{n} [N_k]_\infty^{1/2} \Big\|_{L^p(\Omega;X)} \\
&\leq \sum_{k=m+1}^{n} \Big\| [N_k]_\infty^{1/2} \Big\|_{L^p(\Omega;X)} \to 0, \quad m, n \to \infty,
\end{aligned}$$

where the latter holds due to the fact that $\sum_{k\geq 1} \Big\| [N_k]_\infty^{1/2} \Big\|_{L^p(\Omega;X)} < \infty$. Thus $\sum_{k=1}^{n} \mathbb{E}[N_k]_\infty^{1/2}$ converges in $X$ as $n \to \infty$, where the corresponding limit coincides with its pointwise limit $\sum_{k\geq 1} \mathbb{E}[N_k]_\infty^{1/2}$ by Remark 16.2.3. Therefore, since any element of $X$ is finite a.s. by Definition 16.2.2, we can find $S_0 \in \Sigma$ such that $\mu(S_0^c) = 0$ and pointwise in $S_0$, we have $\sum_{k\geq 1} \mathbb{E}[N_k]_\infty^{1/2} < \infty$. Fix $s \in S_0$. In particular, we find that $\sum_{k\geq 1} [N_k]_\infty^{1/2}$ converges in $L^1(\Omega)$. Moreover, since by the scalar Burkholder-Davis-Gundy inequalities $\mathbb{E} \sup_{t\geq 0} |N_k(t, \cdot, s)| \eqsim \mathbb{E}[N_k(s)]_\infty^{1/2}$, we also obtain that

$$N(\cdot, s) := \sum_{k\geq 1} N_k(\cdot, s) \ \text{ converges in } L^1(\Omega; \mathcal{D}_b([0, \infty)). \qquad (16.12)$$

Let $N(\cdot, s) = 0$ for $s \notin S_0$. Then $N$ defines a martingale field. Moreover, by the scalar Burkholder-Davis-Gundy inequalities

$$\lim_{m\to\infty} \Big[ \sum_{k=n}^{m} N_k(\cdot, s) \Big]_\infty^{1/2} = \Big[ \sum_{k=n}^{\infty} N_k(\cdot, s) \Big]_\infty^{1/2}$$

in $L^1(\Omega)$. Therefore, by considering an a.s. convergent subsequence and by (16.10) we obtain

$$\Big[ \sum_{k=n}^{\infty} N_k(\cdot, s) \Big]_\infty^{1/2} \leq \sum_{k=n}^{\infty} [N_k(\cdot, s)]_\infty^{1/2}. \qquad (16.13)$$

It remains to prove that $N \in \mathrm{MQ}^p(X)$ and $N = \sum_{k\geq 1} N_k$ with convergence in $\mathrm{MQ}^p(X)$. Let $\varepsilon > 0$. Choose $n \in \mathbb{N}$ such that $\sum_{k\geq n+1} \|N_k\|_{\mathrm{MQ}^p(X)} < \varepsilon$. It follows from (16.11) that $\mathbb{E}\big\| \sum_{k\geq 1}[N_k]_\infty^{1/2}\big\|_X < \infty$, so $\sum_{k\geq 1}[N_k]_\infty^{1/2}$ a.s. converges in $X$. Now by (16.13), the triangle inequality and Fatou's lemma, we obtain

$$\begin{aligned}\Big\|\Big[\sum_{k\geq n+1} N_k\Big]_\infty^{1/2}\Big\|_{L^p(\Omega;X)} &\leq \Big\|\sum_{k=n+1}^{\infty}[N_k]_\infty^{1/2}\Big\|_{L^p(\Omega;X)} \\ &\leq \sum_{k=n+1}^{\infty}\Big\|[N_k]_\infty^{1/2}\Big\|_{L^p(\Omega;X)} \\ &\leq \liminf_{m\to\infty}\sum_{k=n+1}^{m}\Big\|[N_k]_\infty^{1/2}\Big\|_{L^p(\Omega;X)} < \varepsilon^p.\end{aligned}$$

Therefore, $N \in \mathrm{MQ}^p(X)$ and $\|N - \sum_{k=1}^n N_k\|_{\mathrm{MQ}^p(X)} < \varepsilon$.

For the proof of the final assertion assume that $N_n \to N$ in $\mathrm{MQ}^p(X)$. Choose a subsequence $(N_{n_k})_{k\geq 1}$ such that $\|N_{n_k} - N\|_{\mathrm{MQ}^p(X)} \leq 2^{-k}$. Then $\sum_{k\geq 1}\|N_{n_k} - N\|_{\mathrm{MQ}^p(X)} < \infty$ and hence by (16.12) we see that pointwise a.e. in $S$, the series $\sum_{k\geq 1}(N_{n_k} - N)$ converges in $L^1(\Omega; \mathcal{D}_b([0,\infty)))$. Therefore, $N_{n_k} \to N$ in $L^1(\Omega; \mathcal{D}_b([0,\infty); X))$ as required. □

For the proof of Theorem 16.4.1 we will need the following lemma presented in [8, Théorème 2].

**Lemma 16.4.3** *Let $1 < p < \infty$, $M : \mathbb{R}_+ \times \Omega \to \mathbb{R}$ be an $L^p$-martingales. Let $T > 0$. For each $n \geq 1$ define*

$$R_n := \sum_{k=1}^{n}\Big|M_{\frac{Tk}{n}} - M_{\frac{T(k-1)}{n}}\Big|^2.$$

*Then $R_n$ converges to $[M]_T$ in $L^{p/2}$.*

*Proof of Theorem 16.4.1* The existence of the local martingale field $N$ together with the first estimate in (16.9) follows from Theorem 16.3.2. It remains to prove

$$\Big\|\sup_{t\geq 0}\|M(t,\cdot)\|_X\Big\|_{L^p(\Omega)} \eqsim_{p,X} \|[N]_\infty^{1/2}\|_{L^p(\Omega;X)}. \tag{16.14}$$

Due to Definition 16.2.2(iv) it suffices to prove the above norm equivalence in the case $M$ and $N$ becomes constant after some fixed time $T$.

*Step 1: The Finite Dimensional Case* Assume that $M$ takes values in a finite dimensional subspace $Y$ of $X$ and that the right hand side of (16.14) is finite. Then we can write $N(t,s) = M(t)(s) = \sum_{j=1}^n M_j(t)x_j(s)$, where each $M_j$ is a scalar-valued martingale with $M_j(T) \in L^p(\Omega)$ and $x_1,\ldots,x_n \in X$ form a basis of $Y$.

Note that for any $c_1, \dots, c_n \in L^p(\Omega)$ we have that

$$\Big\| \sum_{j=1}^{n} c_j x_j \Big\|_{L^p(\Omega;X)} \eqsim_{p,Y} \sum_{j=1}^{n} \|c_j\|_{L^p(\Omega)}. \tag{16.15}$$

Fix $m \geq 1$. Then by (16.3) and Doob's maximal inequality

$$\begin{aligned} \Big\| \sup_{t\geq 0} \|M(t,\cdot)\|_X \Big\|_{L^p(\Omega)} &\eqsim_p \|M(T,\cdot)\|_{L^p(\Omega;X)} \\ &= \Big\| \sum_{i=1}^{m} M_{\frac{Ti}{m}} - M_{\frac{T(i-1)}{m}} \Big\|_{L^p(\Omega;X)} \\ &\eqsim_{p,X} \Big\| \Big( \sum_{i=1}^{m} \big| M_{\frac{Ti}{m}} - M_{\frac{T(i-1)}{m}} \big|^2 \Big)^{\frac{1}{2}} \Big\|_{L^p(\Omega;X)}, \end{aligned} \tag{16.16}$$

and by (16.15) and Lemma 16.4.3 the right hand side of (16.16) converges to

$$\|[M]_\infty^{1/2}\|_{L^p(\Omega;X)} = \|[N]_\infty^{1/2}\|_{L^p(\Omega;X)}.$$

*Step 2: Reduction to the Case Where $M$ Takes Values in a Finite Dimensional Subspace of $X$* Let $M(T) \in L^p(\Omega; X)$. Then we can find simple functions $(\xi_n)_{n\geq 1}$ in $L^p(\Omega; X)$ such that $\xi_n \to M(T)$. Let $M_n(t) = \mathbb{E}(\xi_n|\mathcal{F}_t)$ for all $t \geq 0$ and $n \geq 1$, $(N_n)_{n\geq 1}$ be the corresponding martingale fields. Then each $M_n$ takes values in a finite dimensional subspace $X_n \subseteq X$, and hence by Step 1

$$\Big\| \sup_{t\geq 0} \|M_n(t,\cdot) - M_m(t,\cdot)\|_X \Big\|_{L^p(\Omega)} \eqsim_{p,X} \|[N_n - N_m]_\infty^{1/2}\|_{L^p(\Omega;X)}$$

for any $m, n \geq 1$. Therefore since $(\xi_n)_{n\geq 1}$ is Cauchy in $L^p(\Omega; X)$, $(N_n)_{n\geq 1}$ converges to some $N$ in $\mathrm{MQ}^p(X)$ by the first part of Proposition 16.4.2.

Let us show that $N$ is the desired local martingale field. Fix $t \geq 0$. We need to show that $N(\cdot,t,\cdot) = M_t$ a.s. on $\Omega$. First notice that by the second part of Proposition 16.4.2 there exists a subsequence of $(N_n)_{n\geq 1}$ which we will denote by $(N_n)_{n\geq 1}$ as well such that $N_n(\cdot,t,\sigma) \to N(\cdot,t,\sigma)$ in $L^1(\Omega)$ for a.e. $\sigma \in S$. On the other hand by Jensen's inequality

$$\big\| \mathbb{E}|N_n(\cdot,t,\cdot) - M_t| \big\|_X = \big\| \mathbb{E}|M_n(t) - M(t)| \big\|_X \leq \mathbb{E}\|M_n(t) - M(t)\|_X \to 0, \quad n \to \infty.$$

Hence $N_n(\cdot,t,\cdot) \to M_t$ in $X(L^1(\Omega))$, and thus by Remark 16.2.3 in $L^0(S; L^1(\Omega))$. Therefore we can find a subsequence of $(N_n)_{n\geq 1}$ (which we will again denote by $(N_n)_{n\geq 1}$) such that $N_n(\cdot,t,\sigma) \to M_t(\sigma)$ in $L^1(\Omega)$ for a.e. $\sigma \in S$ (here we use the fact that $\mu$ is $\sigma$-finite), so $N(\cdot,t,\cdot) = M_t$ a.s. on $\Omega \times S$, and consequently by

Definition 16.2.2(iii), $N(\omega, t, \cdot) = M_t(\omega)$ for a.a. $\omega \in \Omega$. Thus (16.14) follows by letting $n \to \infty$.

*Step 3: Reduction to the Case Where the Left-Hand Side of* (16.14) *is Finite* Assume that the left-hand side of (16.14) is infinite, but the right-hand side is finite. Since $M$ is a local $L^p$-martingale we can find a sequence of stopping times $(\tau_n)_{n\geq 1}$ such that $\tau_n \uparrow \infty$ and $\|M_T^{\tau_n}\|_{L^p(\Omega;X)} < \infty$ for each $n \geq 1$. By the monotone convergence theorem and Definition 16.2.2(iv)

$$\begin{aligned}\|[N]_\infty^{1/2}\|_{L^p(\Omega;X)} &= \lim_{n\to\infty} \|[N^{\tau_n}]_\infty^{1/2}\|_{L^p(\Omega;X)} \eqsim_{p,X} \limsup_{n\to\infty} \|M_T^{\tau_n}\|_{L^p(\Omega;X)} \\ &= \lim_{n\to\infty} \|M_T^{\tau_n}\|_{L^p(\Omega;X)} = \lim_{n\to\infty} \Big\| \sup_{0\leq t\leq T} \|M_t^{\tau_n}\|_X \Big\|_{L^p(\Omega)} \\ &= \Big\| \sup_{0\leq t\leq T} \|M_t\|_X \Big\|_{L^p(\Omega)} = \infty\end{aligned}$$

and hence the right-hand side of (16.14) is infinite as well. □

We use an extrapolation argument to extend part of Theorem 16.4.1 to $p \in (0, 1]$ in the continuous-path case.

**Corollary 16.4.4** *Let $X$ be a UMD Banach function space over a $\sigma$-finite measure space and let $p \in (0, \infty)$. Let $M$ be a continuous local martingale $M : \mathbb{R}_+ \times \Omega \to X$ with $M(0, \cdot) = 0$. Then there exists a continuous local martingale field $N : \mathbb{R}_+ \times \Omega \times S \to \mathbb{R}$ such that for a.a. $\omega \in \Omega$, all $t \geq 0$, and a.a. $s \in S$, $N(t, \omega, \cdot) = M(t, \omega)(s)$ and*

$$\Big\| \sup_{t\geq 0} \|M(t, \cdot)\|_X \Big\|_{L^p(\Omega)} \eqsim_{p,X} \Big\| [N]_\infty^{1/2} \Big\|_{L^p(\Omega;X)}. \tag{16.17}$$

*Proof* By a stopping time argument we can reduce to the case where $\|M(t, \omega)\|_X$ is uniformly bounded in $t \in \mathbb{R}_+$ and $\omega \in \Omega$ and $M$ becomes constant after a fixed time $T$. Now the existence of $N$ follows from Theorem 16.4.1 and it remains to prove (16.17) for $p \in (0, 1]$. For this we can use a classical argument due to Lenglart. Indeed, for both estimates we can apply [16] or [23, Proposition IV.4.7] to the continuous increasing processes $Y, Z : \mathbb{R}_+ \times \Omega \to \mathbb{R}_+$ given by

$$\begin{aligned}Y_u &= \mathbb{E} \sup_{t\in[0,u]} \|M(t, \cdot)\|_X, \\ Z_u &= \|s \mapsto [N(\cdot, \cdot, s)]_u^{1/2}\|_X,\end{aligned}$$

where $q \in (1, \infty)$ is a fixed number. Then by (16.9) for any bounded stopping time $\tau$, we have

$$\mathbb{E}Y_\tau^q = \sup_{t\geq 0} \|M(t \wedge \tau, \cdot)\|_X^q \eqsim_{q,X} \mathbb{E}\|s \mapsto [N(\cdot \wedge \tau, \cdot, s)]_\infty^{1/2}\|_X^q$$
$$\stackrel{(*)}{=} \mathbb{E}\|s \mapsto [N(\cdot, \cdot, s)]_\tau^{1/2}\|_X^q = \mathbb{E}Z_\tau^q,$$

where we used [13, Theorem 17.5] in $(*)$. Now (16.17) for $p \in (0, q)$ follows from [16] or [23, Proposition IV.4.7]. □

As we saw in Theorem 16.3.2, continuity of $M$ implies pointwise continuity of the corresponding martingale field $N$. The following corollaries of Theorem 16.4.1 are devoted to proving the same type of assertions concerning pure discontinuity, quasi-left continuity, and having accessible jumps.

Let $\tau$ be a stopping time. Then $\tau$ is called *predictable* if there exists a sequence of stopping times $(\tau_n)_{n\geq 1}$ such that $\tau_n < \tau$ a.s. on $\{\tau > 0\}$ for each $n \geq 1$ and $\tau_n \nearrow \tau$ a.s. A càdlàg process $V : \mathbb{R}_+ \times \Omega \to X$ is called to have *accessible jumps* if there exists a sequence of predictable stopping times $(\tau_n)_{n\geq 1}$ such that $\{t \in \mathbb{R}_+ : \Delta V \neq 0\} \subset \{\tau_1, \dots, \tau_n, \dots\}$ a.s.

**Corollary 16.4.5** *Let $X$ be a UMD function space over a measure space $(S, \Sigma, \mu)$, $1 < p < \infty$, $M : \mathbb{R}_+ \times \Omega \to X$ be a purely discontinuous $L^p$-martingale with accessible jumps. Let $N$ be the corresponding martingale field. Then $N(\cdot, s)$ is a purely discontinuous martingale with accessible jumps for a.e. $s \in S$.*

For the proof we will need the following lemma taken from [7, Subsection 5.3].

**Lemma 16.4.6** *Let $X$ be a Banach space, $1 \leq p < \infty$, $M : \mathbb{R}_+ \times \Omega \to X$ be an $L^p$-martingale, $\tau$ be a predictable stopping time. Then $(\Delta M_\tau \mathbf{1}_{[0,t]}(\tau))_{t\geq 0}$ is an $L^p$-martingale as well.*

*Proof of Corollary 16.4.5* Without loss of generality we can assume that there exists $T \geq 0$ such that $M_t = M_T$ for all $t \geq T$, and that $M_0 = 0$. Since $M$ has accessible jumps, there exists a sequence of predictable stopping times $(\tau_n)_{n\geq 1}$ such that a.s.

$$\{t \in \mathbb{R}_+ : \Delta M \neq 0\} \subset \{\tau_1, \dots, \tau_n, \dots\}.$$

For each $m \geq 1$ define a process $M^m : \mathbb{R}_+ \times \Omega \to X$ in the following way:

$$M^m(t) := \sum_{n=1}^m \Delta M_{\tau_n} \mathbf{1}_{[0,t]}(\tau_n), \quad t \geq 0.$$

Note that $M^m$ is a purely discontinuous $L^p$-martingale with accessible jumps by Lemma 16.4.6. Let $N^m$ be the corresponding martingale field. Then $N^m(\cdot, s)$ is a purely discontinuous martingale with accessible jumps for almost any $s \in S$ due to Remark 16.3.3. Moreover, for any $m \geq \ell \geq 1$ and any $t \geq 0$ we have that a.s.

$[N^m(\cdot,s)]_t \geq [N^\ell(\cdot,s)]_t$. Define $F : \mathbb{R}_+ \times \Omega \times S \to \mathbb{R}_+ \cup \{+\infty\}$ in the following way:

$$F(t,\cdot,s) := \lim_{m\to\infty} [N^m(\cdot,s)]_t, \quad s \in S, t \geq 0.$$

Note that $F(\cdot,\cdot,s)$ is a.s. finite for almost any $s \in S$. Indeed, by Theorem 16.4.1 and [27, Theorem 4.2] we have that for any $m \geq 1$

$$\big\|[N^m]_\infty^{1/2}\big\|_{L^p(\Omega;X)} \eqsim_{p,X} \|M^m(T,\cdot)\|_{L^p(\Omega;X)} \leq \beta_{p,X}\|M(T,\cdot)\|_{L^p(\Omega;X)},$$

so by Definition 16.2.2(iv), $F(\cdot,\cdot,s)$ is a.s. finite for almost any $s \in S$ and

$$\begin{aligned}\big\|F_\infty^{1/2}\big\|_{L^p(\Omega;X)} = \big\|F_T^{1/2}\big\|_{L^p(\Omega;X)} &= \lim_{m\to\infty}\big\|[N^m]_T^{1/2}\big\|_{L^p(\Omega;X)} \\ &\lesssim_{p,X} \limsup_{m\to\infty}\|M^m(T,\cdot)\|_{L^p(\Omega;X)} \lesssim_{p,X} \|M(T,\cdot)\|_{L^p(\Omega;X)}.\end{aligned}$$

Moreover, for almost any $s \in S$ we have that $F(\cdot,\cdot,s)$ is pure jump and

$$\{t \in \mathbb{R}_+ : \Delta F \neq 0\} \subset \{\tau_1,\ldots,\tau_n,\ldots\}.$$

Therefore to this end it suffices to show that $F(s) = [N(s)]$ a.s. on $\Omega$ for a.e. $s \in S$. Note that by Definition 16.2.2(iv),

$$\big\|(F-[N^m])^{1/2}(\infty)\big\|_{L^p(\Omega;X)} \to 0, \quad m \to \infty \tag{16.18}$$

so by Theorem 16.4.1 $(M^m(T))_{m\geq 1}$ is a Cauchy sequence in $L^p(\Omega;X)$. Let $\xi$ be its limit, $M^0 : \mathbb{R}_+ \times \Omega \to X$ be a martingale such that $M^0(t) = \mathbb{E}(\xi|\mathcal{F}_t)$ for all $t \geq 0$. Then by [27, Proposition 2.14] $M^0$ is purely discontinuous. Moreover, for any stopping time $\tau$ a.s.

$$\Delta M^0_\tau = \lim_{m\to\infty}\Delta M^m_\tau = \lim_{m\to\infty}\Delta M_\tau \mathbf{1}_{\{\tau_1,\ldots,\tau_m\}}(\tau) = \Delta M_\tau,$$

where the latter holds since the set $\{\tau_1,\ldots,\tau_n,\ldots\}$ exhausts the jump times of $M$. Therefore $M = M^0$ since both $M$ and $M^0$ are purely discontinuous with the same jumps, and hence $[N] = F$ (where $F(s) = [M^0(s)]$ by (16.18)). Consequently $N(\cdot,\cdot,s)$ is purely discontinuous with accessible jumps for almost all $s \in S$. □

*Remark 16.4.7* Note that the proof of Corollary 16.4.5 also implies that $M_t^m \to M_t$ in $L^p(\Omega;X)$ for each $t \geq 0$.

A càdlàg process $V : \mathbb{R}_+ \times \Omega \to X$ is called *quasi-left continuous* if $\Delta V_\tau = 0$ a.s. for any predictable stopping time $\tau$.

**Corollary 16.4.8** *Let $X$ be a UMD function space over a measure space $(S,\Sigma,\mu)$, $1 < p < \infty$, $M : \mathbb{R}_+ \times \Omega \to X$ be a purely discontinuous quasi-left continuous*

*$L^p$-martingale. Let $N$ be the corresponding martingale field. Then $N(\cdot, s)$ is a purely discontinuous quasi-left continuous martingale for a.e. $s \in S$.*

The proof will exploit the random measure theory. Let $(J, \mathcal{J})$ be a measurable space. Then a family $\mu = \{\mu(\omega; \mathrm{d}t, \mathrm{d}x), \omega \in \Omega\}$ of nonnegative measures on $(\mathbb{R}_+ \times J; \mathcal{B}(\mathbb{R}_+) \otimes \mathcal{J})$ is called a *random measure*. A random measure $\mu$ is called *integer-valued* if it takes values in $\mathbb{N} \cup \{\infty\}$, i.e. for each $A \in \mathcal{B}(\mathbb{R}_+) \otimes \mathcal{F} \otimes \mathcal{J}$ one has that $\mu(A) \in \mathbb{N} \cup \{\infty\}$ a.s., and if $\mu(\{t\} \times J) \in \{0, 1\}$ a.s. for all $t \geq 0$.

Let $X$ be a Banach space, $\mu$ be a random measure, $F : \mathbb{R}_+ \times \Omega \times J \to X$ be such that $\int_{\mathbb{R}_+ \times J} \|F\| \,\mathrm{d}\mu < \infty$ a.s. Then the integral process $((F \star \mu)_t)_{t \geq 0}$ of the form

$$(F \star \mu)_t := \int_{\mathbb{R}_+ \times J} F(s, \cdot, x)\mathbf{1}_{[0,t]}(s)\mu(\cdot; \mathrm{d}s, \mathrm{d}x), \quad t \geq 0,$$

is a.s. well-defined.

Any integer-valued optional $\mathcal{P} \otimes \mathcal{J}$-$\sigma$-finite random measure $\mu$ has a *compensator*: a unique predictable $\mathcal{P} \otimes \mathcal{J}$-$\sigma$-finite random measure $\nu$ such that $\mathbb{E}(W \star \mu)_\infty = \mathbb{E}(W \star \nu)_\infty$ for each $\mathcal{P} \otimes \mathcal{J}$-measurable real-valued nonnegative $W$ (see [12, Theorem II.1.8]). For any optional $\mathcal{P} \otimes \mathcal{J}$-$\sigma$-finite measure $\mu$ we define the associated compensated random measure by $\bar{\mu} = \mu - \nu$.

Recall that $\mathcal{P}$ denotes the predictable $\sigma$-algebra on $\mathbb{R}_+ \times \Omega$ (see [13] for details). For each $\mathcal{P} \otimes \mathcal{J}$-strongly-measurable $F : \mathbb{R}_+ \times \Omega \times J \to X$ such that $\mathbb{E}(\|F\| \star \mu)_\infty < \infty$ (or, equivalently, $\mathbb{E}(\|F\| \star \nu)_\infty < \infty$, see the definition of a compensator above) we can define a process $F \star \bar{\mu}$ by $F \star \mu - F \star \nu$. Then this process is a purely discontinuous local martingale. We will omit here some technicalities for the convenience of the reader and refer the reader to [12, Chapter II.1], [7, Subsection 5.4–5.5], and [14, 19, 20] for more details on random measures.

*Proof of Corollary 16.4.8* Without loss of generality we can assume that there exists $T \geq 0$ such that $M_t = M_T$ for all $t \geq T$, and that $M_0 = 0$. Let $\mu$ be a random measure defined on $\mathbb{R}_+ \times X$ in the following way

$$\mu(A \times B) = \sum_{t \geq 0} \mathbf{1}_A(t)\mathbf{1}_{B\setminus\{0\}}(\Delta M_t),$$

where $A \subset \mathbb{R}_+$ is a Borel set, and $B \subset X$ is a ball. For each $k, \ell \geq 1$ we define a stopping time $\tau_{k,\ell}$ as follows

$$\tau_{k,\ell} = \inf\{t \in \mathbb{R}_+ : \#\{u \in [0, t] : \|\Delta M_u\|_X \in [1/k, k]\} = \ell\}.$$

Since $M$ has càdlàg trajectories, $\tau_{k,\ell}$ is a.s. well-defined and takes its values in $[0, \infty]$. Moreover, $\tau_{k,\ell} \to \infty$ for each $k \geq 1$ a.s. as $\ell \to \infty$, so we can find a subsequence $(\tau_{k_n,\ell_n})_{n\geq 1}$ such that $k_n \geq n$ for each $n \geq 1$ and $\inf_{m\geq n} \tau_{k_m,\ell_m} \to \infty$ a.s. as $n \to \infty$. Define $\tau_n = \inf_{m\geq n} \tau_{k_m,\ell_m}$ and define $M^n := (\mathbf{1}_{[0,\tau_n]}\mathbf{1}_{B_n}) \star \bar{\mu}$, where $\bar{\mu} = \mu - \nu$ is such that $\nu$ is a compensator of $\mu$ and $B_n = \{x \in X : \|x\| \in$

$[1/n, n]\}$. Then $M^n$ is a purely discontinuous quasi-left continuous martingale by [7]. Moreover, a.s.

$$\Delta M_t^n = \Delta M_t \mathbf{1}_{[0,\tau_n]}(t)\mathbf{1}_{[1/n,n]}(\|\Delta M_t\|), \quad t \geq 0.$$

so by [27] $M^n$ is an $L^p$-martingale (due to the *weak differential subordination* of purely discontinuous martingales).

The rest of the proof is analogous to the proof of Corollary 16.4.5 and uses the fact that $\tau_n \to \infty$ monotonically a.s. □

Let $X$ be a Banach space. A local martingale $M : \mathbb{R}_+ \times \Omega \to X$ is called to have the *canonical decomposition* if there exist local martingales $M^c, M^q, M^a : \mathbb{R}_+ \times \Omega \to X$ such that $M^c$ is continuous, $M^q$ and $M^a$ are purely discontinuous, $M^q$ is quasi-left continuous, $M^a$ has accessible jumps, $M_0^c = M_0^q = 0$, and $M = M^c + M^q + M^a$. Existence of such a decomposition was first shown in the real-valued case by Yoeurp in [30], and recently such an existence was obtained in the UMD space case (see [27, 28]).

*Remark 16.4.9* Note that if a local martingale $M$ has some canonical decomposition, then this decomposition is unique (see [13, 27, 28, 30]).

**Corollary 16.4.10** *Let $X$ be a UMD Banach function space, $1 < p < \infty$, $M : \mathbb{R}_+ \times \Omega \to X$ be an $L^p$-martingale. Let $N$ be the corresponding martingale field. Let $M = M^c + M^q + M^a$ be the canonical decomposition, $N^c$, $N^q$, and $N^a$ be the corresponding martingale fields. Then $N(s) = N^c(s) + N^q(s) + N^a(s)$ is the canonical decomposition of $N(s)$ for a.e. $s \in S$. In particular, if $M_0 = 0$ a.s., then $M$ is continuous, purely discontinuous quasi-left continuous, or purely discontinuous with accessible jumps if and only if $N(s)$ is so for a.e. $s \in S$.*

*Proof* The first part follows from Theorem 16.3.2, Corollaries 16.4.5 and 16.4.8 and the fact that $N(s) = N^c(s) + N^q(s) + N^a(s)$ is then a canonical decomposition of a local martingale $N(s)$ which is unique due to Remark 16.4.9. Let us show the second part. One direction follows from Theorem 16.3.2, Corollaries 16.4.5 and 16.4.8. For the other direction assume that $N(s)$ is continuous for a.e. $s \in S$. Let $M = M^c + M^q + M^a$ be the canonical decomposition, $N^c$, $N^q$, and $N^a$ be the corresponding martingale fields of $M^c$, $M^q$, and $M^a$. Then by the first part of the theorem and the uniqueness of the canonical decomposition (see Remark 16.4.9) we have that for a.e. $s \in S$, $N^q(s) = N^a(s) = 0$, so $M^q = M^a = 0$, and hence $M$ is continuous. The proof for the case of pointwise purely discontinuous quasi-left continuous $N$ or pointwise purely discontinuous $N$ with accessible jumps is similar. □

*Remark 16.4.11* It remains open whether the first two-sided estimate in (16.9) can be extended to $p = 1$. Recently, in [29] the second author has extended the second two-sided estimate in (16.9) to arbitrary UMD Banach spaces and to $p \in [1, \infty)$. Here the quadratic variation has to be replaced by a generalized square function.

**Acknowledgements** The authors would like to thank the referee for helpful comments.

Mark Veraar is supported by the Vidi subsidy 639.032.427 of the Netherlands Organisation for Scientific Research (NWO).

## References

1. M. Antoni, Regular random field solutions for stochastic evolution equations. PhD thesis, 2017
2. J. Bourgain, Extension of a result of Benedek, Calderón and Panzone. Ark. Mat. **22**(1), 91–95 (1984)
3. D.L. Burkholder, Distribution function inequalities for martingales. Ann. Probab. **1**, 19–42 (1973)
4. D.L. Burkholder, A geometrical characterization of Banach spaces in which martingale difference sequences are unconditional. Ann. Probab. **9**(6), 997–1011 (1981)
5. D.L. Burkholder, Martingales and singular integrals in Banach spaces, in *Handbook of the Geometry of Banach Spaces*, vol. I (North-Holland, Amsterdam, 2001), pp. 233–269
6. F. Cobos, Some spaces in which martingale difference sequences are unconditional. Bull. Polish Acad. Sci. Math. **34**(11–12), 695–703 (1986)
7. S. Dirksen, I.S. Yaroslavtsev, $L^q$-valued Burkholder-Rosenthal inequalities and sharp estimates for stochastic integrals. Proc. Lond. Math. Soc. (3) **119**(6), 1633–1693 (2019)
8. C. Doléans, Variation quadratique des martingales continues à droite. Ann. Math. Stat. **40**, 284–289 (1969)
9. J. García-Cuerva, R. Macías, J.L. Torrea, The Hardy-Littlewood property of Banach lattices. Israel J. Math. **83**(1–2), 177–201 (1993)
10. D.J.H. Garling, Brownian motion and UMD-spaces, in *Probability and Banach Spaces (Zaragoza, 1985)*. Lecture Notes in Mathematics, vol. 1221 (Springer, Berlin, 1986), pp. 36–49
11. T. Hytönen, J. van Neerven, M. Veraar, L. Weis, *Analysis in Banach Spaces. Volume I: Martingales and Littlewood-Paley Theory*. Ergebnisse der Mathematik und ihrer Grenzgebiete (3), vol. 63 (Springer, Cham, 2016)
12. J. Jacod, A.N. Shiryaev, *Limit Theorems for Stochastic Processes*. Grundlehren der Mathematischen Wissenschaften, 2nd edn., vol. 288 (Springer, Berlin, 2003)
13. O. Kallenberg, *Foundations of Modern Probability*. Probability and Its Applications (New York), 2nd edn. (Springer, New York, 2002)
14. O. Kallenberg, *Random Measures, Theory and Applications*. Probability Theory and Stochastic Modelling, vol. 77 (Springer, Cham, 2017)
15. I. Karatzas, S.E. Shreve, *Brownian Motion and Stochastic Calculus*. Graduate Texts in Mathematics, 2nd edn., vol. 113 (Springer, New York, 1991)
16. E. Lenglart, Relation de domination entre deux processus. Ann. Inst. H. Poincaré Sect. B (N.S.) **13**(2), 171–179 (1977)
17. N. Lindemulder, M.C. Veraar, I.S. Yaroslavtsev, The UMD property for Musielak–Orlicz spaces, in *Positivity and Noncommutative Analysis. Festschrift in Honour of Ben de Pagter on the Occasion of his 65th Birthday*. Trends in Mathematics (Birkhäuser, Basel, 2019)
18. C. Marinelli, On maximal inequalities for purely discontinuous $L_q$-valued martingales (2013), arXiv:1311.7120
19. C. Marinelli, M. Röckner, On maximal inequalities for purely discontinuous martingales in infinite dimensions, in *Séminaire de Probabilités XLVI*. Lecture Notes in Mathematics, vol. 2123 (Springer, Cham, 2014), pp. 293–315
20. A.A. Novikov, Discontinuous martingales. Teor. Verojatnost. Primemen. **20**, 13–28 (1975)
21. A. Osękowski, A note on the Burkholder-Rosenthal inequality. Bull. Pol. Acad. Sci. Math. **60**(2), 177–185 (2012)

22. A. Osękowski, *Sharp Martingale and Semimartingale Inequalities*. Instytut Matematyczny Polskiej Akademii Nauk. Monografie Matematyczne (New Series), vol. 72 (Birkhäuser/Springer Basel AG, Basel, 2012)
23. D. Revuz, M. Yor, *Continuous Martingales and Brownian Motion*. Grundlehren der Mathematischen Wissenschaften, 3rd edn., vol. 293 (Springer, Berlin, 1999)
24. J.L. Rubio de Francia, Martingale and integral transforms of Banach space valued functions, in *Probability and Banach Spaces (Zaragoza, 1985)*. Lecture Notes in Mathematics, vol. 1221 (Springer, Berlin, 1986), pp. 195–222
25. J.M.A.M. van Neerven, M.C. Veraar, L.W. Weis, Stochastic integration in UMD Banach spaces. Ann. Probab. **35**(4), 1438–1478 (2007)
26. M.C. Veraar, I.S. Yaroslavtsev, Cylindrical continuous martingales and stochastic integration in infinite dimensions. Electron. J. Probab. **21**(59), 53 (2016)
27. I.S. Yaroslavtsev, Martingale decompositions and weak differential subordination in UMD Banach spaces. Bernoulli **25**(3), 1659–1689 (2019)
28. I.S. Yaroslavtsev, On the martingale decompositions of Gundy, Meyer, and Yoeurp in infinite dimensions. Ann. Inst. Henri Poincaré Probab. Stat. (2017, to appear). arXiv:1712.00401
29. I.S. Yaroslavtsev, Burkholder–Davis–Gundy inequalities in UMD Banach spaces (2018). arXiv:1807.05573
30. Ch. Yoeurp, Décompositions des martingales locales et formules exponentielles, in *Séminaire de Probabilités X Université de Strasbourg*. Lecture Notes in Mathematics, vol. 511 (1976), pp. 432–480
31. A.C. Zaanen, *Integration* (North-Holland Publishing/Interscience Publishers Wiley, Amsterdam/New York, 1967). Completely revised edition of An introduction to the theory of integration

# Chapter 17
# Concentration Inequalities for Randomly Permuted Sums

**Mélisande Albert**

**Abstract** Initially motivated by the study of the non-asymptotic properties of non-parametric tests based on permutation methods, concentration inequalities for uniformly permuted sums have been largely studied in the literature. Recently, Delyon et al. proved a new Bernstein-type concentration inequality based on martingale theory. This work presents a new proof of this inequality based on the fundamental inequalities for random permutations of Talagrand. The idea is to first obtain a rough inequality for the square root of the permuted sum, and then, iterate the previous analysis and plug this first inequality to obtain a general concentration of permuted sums around their median. Then, concentration inequalities around the mean are deduced. This method allows us to obtain the Bernstein-type inequality up to constants, and, in particular, to recovers the Gaussian behavior of such permuted sums under classical conditions encountered in the literature. Then, an application to the study of the second kind error rate of permutation tests of independence is presented.

**Keywords** Concentration inequalities · Random permutations

**Mathematics Subject Classification** 60E15, 60C05

## 17.1 Introduction and Motivation

This article presents concentration inequalities for randomly permuted sums defined by

$$Z_n = \sum_{i=1}^{n} a_{i,\Pi(i)},$$

M. Albert (✉)
Institut de Mathématiques de Toulouse; UMR5219, Université de Toulouse; CNRS, INSA, Toulouse, France
e-mail: melisande.albert@insa-toulouse.fr



N. Gozlan et al. (eds.), *High Dimensional Probability VIII*,
Progress in Probability 74, https://doi.org/10.1007/978-3-030-26391-1_17

where $\{a_{i,j}\}_{1\le i,j\le n}$ are real numbers, and $\Pi$ is a uniformly distributed random permutation of the set $\{1,\ldots,n\}$. Initially motivated by hypothesis testing in the non-parametric framework (see [28] for instance), such sums have been largely studied from an asymptotic point of view in the literature. A first combinatorial central limit theorem is proved by Wald and Wolfowitz in [28], in the particular case when the real numbers $a_{i,j}$ are of a product form $b_i \times c_j$, under strong assumptions that have been released for instance by Noether [22]. Then, Hoeffding obtains stronger results in such product case, and generalizes those results to not necessarily product type real terms $a_{i,j}$ in [15]. More precisely, he considers

$$d_{i,j} = a_{i,j} - \frac{1}{n}\sum_{k=1}^{n} a_{k,j} - \frac{1}{n}\sum_{l=1}^{n} a_{i,l} + \frac{1}{n^2}\sum_{k,l=1}^{n} a_{k,l}. \tag{17.1}$$

In particular, $\mathrm{Var}(Z_n) = \frac{1}{n-1}\sum_{i=1}^{n} d_{i,j}^2$. Then he proves (see [15, Theorem 3]) that, if

$$\lim_{n\to+\infty} \frac{\frac{1}{n}\sum_{1\le i,j\le n} d_{i,j}^r}{\left(\frac{1}{n}\sum_{i,j=1}^{n} d_{i,j}^2\right)^{r/2}} = 0, \quad \text{for some } r > 2, \tag{17.2}$$

then the distribution of $Z_n = \sum_{i=1}^{n} a_{i,\Pi(i)}$ is asymptotically normal, that is, for all $x$ in $\mathbb{R}$,

$$\lim_{n\to+\infty} \mathbb{P}\Big(Z_n - \mathbb{E}[Z_n] \le x\sqrt{\mathrm{Var}(Z_n)}\Big) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{y^2}{2}}\,dy.$$

He also considers a stronger (in the sense that it implies (17.2)), but simpler condition in [15, Theorem 3], precisely

$$\frac{\max_{1\le i,j\le n}\left\{\left|d_{i,j}\right|\right\}}{\sqrt{\frac{1}{n}\sum_{i,j=1}^{n} d_{i,j}^2}} \underset{n\to+\infty}{\longrightarrow} 0, \tag{17.3}$$

under which such an asymptotic Gaussian limit holds. Similar results have been obtained later, for instance by Motoo [21], under the following Lindeberg-type condition that is for all $\varepsilon > 0$,

$$\lim_{n\to+\infty} \sum_{1\le i,j\le n} \left(\frac{d_{i,j}}{d}\right)^2 \mathbb{1}_{\left|\frac{d_{i,j}}{d}\right|>\varepsilon} = 0, \tag{17.4}$$

where $d^2 = n^{-1}\sum_{1\le i,j\le n} d_{i,j}^2$. In particular, Motoo proves in [21] that such Lindeberg-type condition is weaker than Hoeffding's ones in the sense that (17.4) is implied by (17.2) (and thus by (17.3)). A few years later, Hájek [13] proves in the product case, that the condition (17.4) is in fact necessary. A simpler proof of the sufficiency of the Lindeberg-type condition is given by Schneller [25] based on Stein's method.

Afterwards, the next step was to study the convergence of the conditional distribution when the terms $a_{i,j}$ in the general case, or $b_i \times c_j$ in the product case, are random. Notably, Dwass studies in [10] the limit of the randomly permuted sum in the product case, where only the $c_j$'s are random, and proves that the conditional distribution given the $c_j$'s converges almost surely (a.s.) to a Gaussian distribution. Then, Shapiro and Hubert [26] generalized this study to weighted $U$-statistics of the form $\sum_{i \neq j} b_{i,j} h(X_i, X_j)$ where the $X_i$'s are independent and identically distributed (i.i.d.) random variables. In a first time, they show some a.s. asymptotic normality of this statistic. In a second time, they complete Jogdeo's [17] work in the deterministic case, proving asymptotic normality of permuted statistics based on the previous weighted $U$-statistic. More precisely, they consider the rank statistic $\sum_{i \neq j} b_{i,j} h(X_{R_i}, X_{R_j})$, where $R_i$ is the rank of $V_i$ in a sample $V_1, \ldots, V_n$ of i.i.d. random variables with a continuous distribution function. In particular, notice that considering such rank statistics is equivalent to considering uniformly permuted statistics. In [2], the previous combinatorial central limit theorems is generalized to permuted sums of non-i.i.d. random variables $\sum_{i=1}^{n} Y_{i,\Pi(i)}$, for particular forms of random variables $Y_{i,j}$. The main difference with the previous results comes from the fact that the random variables $Y_{i,j}$ are not necessarily exchangeable.

Hence, the asymptotic behavior of permuted sums has been vastly investigated in the literature, allowing to deduce good properties for permutation tests based on such statistics, like the asymptotic size, or the power (see for instance [23] or [2]). Yet, such results are purely asymptotic, while, in many application fields, such as neurosciences for instance as described in [2], few exploitable data are available. Hence, such asymptotic results may not be sufficient. This is why a non-asymptotic approach is preferred here, leading to concentration inequalities. In the sequel, unless specified, we will thus drop the index $n$ and denote $Z = Z_n$.

Concentration inequalities have been vastly investigated in the literature, and the interested reader can refer to the books of Ledoux [18], Massart [19], or the more recent one of Boucheron et al. [8] for some overall reviews. Yet in many cases, they provide precise tail bounds for well-behaved functions or sums of independent random variables. For instance, let us recall the classical Bernstein inequality stated for instance in [19, Proposition 2.9 and Corollary 2.10].

**Theorem 17.1.1 (Bernstein's Inequality, Massart [19])** *Let $X_1, \ldots, X_n$ be independent real valued random variables. Assume that there exists some positive numbers $v$ and $c$ such that*

$$\sum_{i=1}^{n} \mathbb{E}\left[X_i^2\right] \leq v,$$

*and for all integers* $k \geq 3$,

$$\sum_{i=1}^{n} \mathbb{E}\left[(X_i)_+^k\right] \leq \frac{k!}{2} v c^{k-2},$$

*where* $(\cdot)_+ = \max\{\cdot, 0\}$ *denotes the positive part.*

*Let* $S = \sum_{i=1}^{n}(X_i - \mathbb{E}[X_i])$*, then for every positive* $x$*,*

$$\mathbb{P}\left(S \geq \sqrt{2vx} + cx\right) \leq e^{-x}. \tag{17.5}$$

*Moreover, for any positive* $t$*,*

$$\mathbb{P}(S \geq t) \leq \exp\left(-\frac{t^2}{2(v + ct)}\right). \tag{17.6}$$

Notice that both forms of Bernstein's inequality appear in the literature. Yet, due to its form, (17.5) is rather preferred in statistics, even though (17.6) is more classical.

The work in this article is based on the pioneering work of Talagrand (see [27] for a review) who investigates the concentration of measure phenomenon for product measures. Of main interest here, he proves the following inequality for random permutations in [27, Theorem 5.1].

**Theorem 17.1.2 (Talagrand [27])** *Denote by* $\mathfrak{S}_n$ *the set of all permutations of* $\{1, \ldots, n\}$*. Define for any subset* $A \subset \mathfrak{S}_n$*, and permutation* $\tau \in \mathfrak{S}_n$*,*

$$U_A(\tau) = \left\{s \in \{0,1\}^n \; ; \; \exists \sigma \in A \text{ such that } \forall 1 \leq i \leq n,\; s_i = 0 \implies \sigma(i) = \tau(i)\right\}.$$

*Then, consider* $V_A(\tau) = \text{ConvexHull}\,(U_A(\tau))$*, and*

$$f(A, \tau) = \min\left\{\sum_{i=1}^{n} v_i^2 \; ; \; v = (v_i)_{1 \leq i \leq n} \in V_A(\tau)\right\}.$$

*Then, if* $P_n$ *denotes the uniform distribution on* $\mathfrak{S}_n$*,*

$$\int_{\mathfrak{S}_n} e^{\frac{1}{16} f(A,\tau)} dP_n(\tau) \leq \frac{1}{P_n(A)}.$$

*Therefore, by Markov's inequality, for all* $t > 0$*,*

$$P_n\left(\tau \; ; \; f(A, \tau) \geq t^2\right) \leq \frac{e^{-t^2/16}}{P_n(A)}. \tag{17.7}$$

This result on random permutations is fundamental, and is a key point to many other non-asymptotic works on random permutations. Among them emerges McDiarmid's article [20] in which he derives from Talagrand's inequality, exponential concentration inequalities around the median for randomly permuted functions of the observation under Lipschitz-type conditions and applied to randomized methods for graph coloring. More recently, Adamczak et al. obtained in [4] some concentration inequality under convex-Lipschitz conditions when studying the empirical spectral distribution of random matrices. In particular, they prove the following theorem (precisely [4, Theorem 3.1]).

**Theorem 17.1.3 (Adamczak et al. [4])** *Consider* $x_1, \ldots, x_n$ *in* $[0, 1]$ *and let* $\varphi : [0, 1]^n \to \mathbb{R}$ *be an L-Lipschitz convex function. Let* $\Pi$ *be a uniform random permutation of the set* $\{1, \ldots, n\}$ *and denote* $Y = \varphi\big(x_{\Pi(1)}, \ldots, x_{\Pi(n)}\big)$*. Then, there exists some positive absolute constant c such that, for all* $t > 0$*,*

$$\mathbb{P}(Y - \mathbb{E}[Y] \geq t) \leq 2 \exp\left(-\frac{ct^2}{L^2}\right).$$

Yet, the Lipschitz assumptions may be very restrictive and may not be satisfied by the functions considered in the application fields (see Sect. 17.3.1 for instance). Hence, the idea is to exploit the attractive form of a sum. Based on Stein's method, initially introduced to study the Gaussian behavior of sums of dependent random variables, Chatterjee studies permuted sums of non-negative numbers in [9]. He obtains in [9, Proposition 1.1] the following first Bernstein-type concentration inequality for non-negative terms around the mean.

**Theorem 17.1.4 (Chatterjee [9])** *Let* $\left\{a_{i,j}\right\}_{1 \leq i,j \leq n}$ *be a collection of numbers from* $[0, 1]$*. Let* $Z = \sum_{i=1}^{n} a_{i,\Pi(i)}$*, where* $\Pi$ *is drawn from the uniform distribution over the set of all permutations of* $\{1, \ldots, n\}$*. Then, for any* $t \geq 0$*,*

$$\mathbb{P}(|Z - \mathbb{E}[Z]| \geq t) \leq 2 \exp\left(-\frac{t^2}{4\mathbb{E}[Z] + 2t}\right). \tag{17.8}$$

Notice that because of the expectation term in the right-hand side of (17.8), the link with Hoeffding's combinatorial central limit theorem (for instance) is not so clear.

In [6, Theorem 4.3], this result is sharpened in the sense that this expectation term is replaced by a variance term, allowing us to provide a non-asymptotic version of such combinatorial central limit theorem. This result is moreover generalized to any real numbers (not necessarily non-negative). More precisely, based on martingale theory, they prove the following result.

**Theorem 17.1.5 (Bercu et al. [6])** *Let* $\left\{a_{i,j}\right\}_{1\leq i,j\leq n}$ *be an array of real numbers from* $[-M_a, M_a]$. *Let* $Z = \sum_{i=1}^{n} a_{i,\Pi(i)}$, *where* $\Pi$ *is drawn from the uniform distribution over the set of all permutations of* $\{1,\ldots,n\}$. *Then, for any* $t > 0$,

$$\mathbb{P}(|Z - \mathbb{E}[Z]| \geq t) \leq 4\exp\left(-\frac{t^2}{16(\theta\frac{1}{n}\sum_{i,j=1}^{n} a_{i,j}^2 + M_a t/3)}\right), \tag{17.9}$$

*where* $\theta = \dfrac{5}{2}\ln(3) - \dfrac{2}{3}$.

In this work, we obtain a similar result (up to constants) but based on a completely different approach. Moreover, this approach provides a direct proof for a concentration inequality of a permuted sum around its median.

The present work is organized as follows. In Sect. 17.2 are formulated the main results. Section 17.2.1 is devoted to the permuted sums of non-negative numbers. Based on Talagrand's result, a first rough concentration inequality for the square root of permuted sum is obtained in Lemma 17.2.1. Then by iterating the previous analysis and plugging this first inequality, a general concentration of permuted sums around their median is obtained in Proposition 17.2.1. Finally, the concentration inequality of Proposition 17.2.2 around the mean is deduced. In Sect. 17.2.2, the previous inequalities are generalized to permuted sums of not necessarily non-negative terms in Theorem 17.2.1. Section 17.3 presents an application to the study of non-asymptotic properties of a permutation independence test in statistics. In particular, a sharp control of the critical value of the test is deduced from the main result. The proofs are detailed in Sect. 17.4. Finally, the Appendix contains technical results for the non-asymptotic control of the second kind error rate of the permutation test introduced in Sect. 17.3.

## 17.2 Bernstein-Type Concentration Inequalities for Permuted Sums

Let us first introduce some general notation. In the sequel, denote by $\mathfrak{S}_n$ the set of permutations of $\{1, 2, \ldots, n\}$. For all collection of real numbers $\left\{a_{i,j}\right\}_{1\leq i,j\leq n}$, and for each $\tau$ in $\mathfrak{S}_n$, consider the permuted sum

$$Z(\tau) = \sum_{i=1}^{n} a_{i,\tau(i)}.$$

Let $\Pi$ be a uniform random permutation in $\mathfrak{S}_n$, and $Z := Z(\Pi)$. Denote $\mathrm{med}\,(Z)$ its median, that is which satisfies

$$\mathbb{P}(Z \geq \mathrm{med}\,(Z)) \geq 1/2 \quad \text{and} \quad \mathbb{P}(Z \leq \mathrm{med}\,(Z)) \geq 1/2.$$

This study is divided in two steps. The first one is restrained to non-negative terms. The second one extends the previous results to general terms, based on a trick involving both non-negative and negative parts.

### *17.2.1 Concentration of Permuted Sums of Non-negative Numbers*

In the present section, consider a non-negative collection of numbers $\{a_{i,j}\}_{1\le i,j\le n}$. The proof of the concentration inequality around the median in Proposition 17.2.1 needs a preliminary step which is presented in Lemma 17.2.1. It provides concentration inequality for the square root of the sum. It allows us then by iterating the same argument, and plugging the obtained inequality to the square root of the sum of the squares, namely $\sqrt{\sum_{i=1}^{n} a_{i,\Pi(i)}^2}$, to be able to sharpen Chatterjee's concentration inequality (17.8).

**Lemma 17.2.1** *Let* $\{a_{i,j}\}_{1\le i,j\le n}$ *be a collection of non-negative numbers, and* $\Pi$ *be a uniform random permutation in* $\mathfrak{S}_n$*. Consider* $Z = \sum_{i=1}^{n} a_{i,\Pi(i)}$*. Then, for all* $t > 0$*,*

$$\mathbb{P}\left(\sqrt{Z} \ge \sqrt{\mathrm{med}\,(Z)} + t\sqrt{\max_{1\le i,j\le n}\left\{a_{i,j}\right\}}\right) \le 2e^{-t^2/16}, \tag{17.10}$$

*and*

$$\mathbb{P}\left(\sqrt{Z} \le \sqrt{\mathrm{med}\,(Z)} - t\sqrt{\max_{1\le i,j\le n}\left\{a_{i,j}\right\}}\right) \le 2e^{-t^2/16}. \tag{17.11}$$

*In particular, one obtains the following two-sided concentration for the square root of a randomly permuted sum of non-negative numbers,*

$$\mathbb{P}\left(\left|\sqrt{Z} - \sqrt{\mathrm{med}\,(Z)}\right| > t\sqrt{\max_{1\le i,j\le n}\left\{a_{i,j}\right\}}\right) \le 4e^{-t^2/16}.$$

The idea of the proof is the same that the one of Adamczak et al. in [4, Theorem 3.1], but with a sum instead of a convex Lipschitz function. In a similar way, it is based on Talagrand's inequality for random permutations recalled in Theorem 17.1.2.

In the following are presented two concentration inequalities in the non-negative case; the first one around the median, and the second one around the mean. It is well known that both are equivalent up to constants, but here, both are detailed in order to give the order of magnitude of the constants. The transition from the median to the mean can be obtained thanks to Ledoux' trick in the proof of [18, Proposition 1.8] allowing to reduce exponential concentration inequalities around any constant $m$ (corresponding in our case to med $(Z)$) to similar inequalities around the mean. This trick consists in using the exponentially fast decrease around $m$ to upper bound the difference between $m$ and the mean. Yet, this approach leads to drastic multiplicative constants (of the order $8e^{16\pi}$ as shown in [1]). Better constants can be deduced from the following lemma.

**Lemma 17.2.2** *For any real valued random variable $X$,*

$$|\mathbb{E}[X] - \mathrm{med}(X)| \leq \sqrt{\mathrm{Var}(X)}.$$

In particular, we obtain the following results.

**Proposition 17.2.1** *Let $\{a_{i,j}\}_{1\leq i,j\leq n}$ be a collection of non-negative numbers and $\Pi$ be a uniform random permutation in $\mathfrak{S}_n$. Consider $Z = \sum_{i=1}^{n} a_{i,\Pi(i)}$. Then, for all $x > 0$,*

$$\mathbb{P}\left(|Z - \mathrm{med}(Z)| > \sqrt{\mathrm{med}\left(\sum_{i=1}^{n} a_{i,\Pi(i)}^2\right) x} + x \max_{1\leq i,j\leq n}\{a_{i,j}\}\right) \leq 8\exp\left(\frac{-x}{16}\right). \tag{17.12}$$

Since in many applications, the concentration around the mean is more adapted, the following proposition shows that one may obtain a similar behavior around the mean, at the cost of higher constants.

**Proposition 17.2.2** *Let $\{a_{i,j}\}_{1\leq i,j\leq n}$ be a collection of non-negative numbers, and $\Pi$ be a uniform random permutation in $\mathfrak{S}_n$. Consider $Z = \sum_{i=1}^{n} a_{i,\Pi(i)}$.*

*Then, for all $x > 0$,*

$$\mathbb{P}\left(|Z - \mathbb{E}[Z]| \geq 2\sqrt{\left(\frac{1}{n}\sum_{i,j=1}^{n} a_{i,j}^2\right) x} + \max_{1\leq i,j\leq n}\{a_{i,j}\} x\right) \leq 8e^{1/16}\exp\left(-\frac{x}{16}\right). \tag{17.13}$$

This concentration inequality is called a Bernstein-type inequality restricted to non-negative sums, due to its resemblance to the standard Bernstein inequality, as recalled in Theorem 17.1.1. The main difference here lies in the fact that the random variables in the sum are not independent. Moreover, this inequality implies a more popular form of Bernstein's inequality stated in Corollary 17.2.1.

**Corollary 17.2.1** *With the same notation and assumptions as in Proposition 17.2.2, for all $t > 0$,*

$$\mathbb{P}(|Z - \mathbb{E}[Z]| \geq t) \leq 8e^{1/16} \exp\left(\frac{-t^2}{16\left(4\frac{1}{n}\sum_{i,j=1}^{n} a_{i,j}^2 + 2\max_{1\leq i,j\leq n}\left\{a_{i,j}\right\} t\right)}\right). \tag{17.14}$$

*Comment* Recall Chatterjee's result in [9, Proposition 2.1], quoted in Theorem 17.1.4, which can easily be rewritten with our notation, and for a collection of non-negative numbers not necessarily in [0, 1], by

$$\forall t > 0, \quad \mathbb{P}(|Z - \mathbb{E}[Z]| \geq t) \leq 2\exp\left(\frac{-t^2}{4M_a\frac{1}{n}\sum_{i,j=1}^{n} a_{i,j} + 2M_a t}\right),$$

where $M_a$ denotes the maximum $\max_{1\leq i,j\leq n}\left\{a_{i,j}\right\}$. As mentioned in [6], the inequality in (17.14) is sharper up to constants, because of the quadratic term, since the inequality $\sum_{i,j=1}^{n} a_{i,j}^2 \leq M_a \sum_{i,j=1}^{n} a_{i,j}$ always holds.

### *17.2.2 Concentration of Permuted Sums in the General Case*

In this section, the collection of numbers $\left\{a_{i,j}\right\}_{1\leq i,j\leq n}$ is no longer assumed to be non-negative. The following general concentration inequality for randomly permuted sums directly derives from Proposition 17.2.2.

**Theorem 17.2.1** *Let $\left\{a_{i,j}\right\}_{1\leq i,j\leq n}$ be a collection of any real numbers, and $\Pi$ be a uniform random permutation in $\mathfrak{S}_n$. Consider $Z = \sum_{i=1}^{n} a_{i,\Pi(i)}$. Then, for all $x > 0$,*

$$\mathbb{P}\left(|Z - \mathbb{E}[Z]| \geq 2\sqrt{2\left(\frac{1}{n}\sum_{i,j=1}^{n} a_{i,j}^2\right)x} + 2\max_{1\leq i,j\leq n}\left\{\left|a_{i,j}\right|\right\}x\right) \leq 16e^{1/16}\exp\left(-\frac{x}{16}\right). \tag{17.15}$$

Once again, the obtained inequality is a Bernstein-type inequality. Moreover, it is also possible to obtain a more popular form of Bernstein-type inequalities applying the same trick based on the non-negative and the negative parts from Corollary 17.2.1.

**Corollary 17.2.2** *With the same notation as in Theorem 17.2.1, for all $t > 0$,*

$$\mathbb{P}(|Z - \mathbb{E}[Z]| \geq t) \leq 16e^{1/16}\exp\left(\frac{-t^2}{256\left(\mathrm{Var}(Z) + \max_{1\leq i,j\leq n}\left\{\left|a_{i,j}\right|\right\} t\right)}\right).$$

*Comments* One recovers a Gaussian behavior of the centered permuted sum obtained by Hoeffding in [15, Theorem 3] under the same assumptions. Indeed, in the proof of Corollary 17.2.2, one obtains the following intermediate result (see (17.41)), that is

$$\mathbb{P}(|Z - \mathbb{E}[Z]| \geq t) \leq 16e^{1/16} \exp\left(\frac{-t^2}{64\left(4\frac{1}{n}\sum_{i,j=1}^n d_{i,j}^2 + \max_{1\leq i,j\leq n}\left\{\left|d_{i,j}\right|\right\} t\right)}\right),$$

where the $d_{i,j}$'s are defined in (17.1). Yet, $\mathrm{Var}(Z) = (n-1)^{-1}\sum_{i,j=1}^n d_{i,j}^2$ (see [15, Theorem 2]). Hence, applying this inequality to

$$t = x\sqrt{\mathrm{Var}(Z)} \geq x\sqrt{\frac{1}{n}\sum_{i,j=1}^n d_{i,j}^2},$$

for $x > 0$ leads to

$$\mathbb{P}\left(|Z - \mathbb{E}[Z]| \geq x\sqrt{\mathrm{Var}(Z)}\right) \leq 16e^{1/16} \exp\left(\frac{-x^2}{256\left(1 + \frac{\max_{1\leq i,j\leq n}\left\{\left|d_{i,j}\right|\right\}}{\sqrt{\frac{1}{n}\sum_{i,j=1}^n d_{i,j}^2}} x\right)}\right),$$

Hence, under Hoeffding's simpler condition (17.3), namely

$$\lim_{n\to+\infty} \frac{\max_{1\leq i,j\leq n} d_{i,j}^2}{\frac{1}{n}\sum_{i,j=1}^n d_{i,j}^2} = 0,$$

one recovers, (denoting $Z = Z_n$ depending on $n$),

$$\lim_{n\to+\infty} \mathbb{P}\left(|Z_n - \mathbb{E}[Z_n]| \geq x\sqrt{\mathrm{Var}(Z_n)}\right) \leq 16e^{1/16}e^{-x^2/256},$$

which is a Gaussian tail that is, up to constants, close in spirit to the one obtained by Hoeffding in [15, Theorem 3].

## 17.3 Application to Independence Testing

### *17.3.1 Statistical Motivation*

Let $\mathcal{X}$ represent a separable set. Given an i.i.d. $n$-sample $\mathbb{X}_n = (X_1, \ldots, X_n)$, where each $X_i$ is a couple $(X_i^1, X_i^2)$ in $\mathcal{X}^2$ with distribution $P$ of marginals $P^1$ and $P^2$, we aim at testing the null hypothesis ($\mathcal{H}_0$) "$P = (P^1 \otimes P^2)$" against the alternative ($\mathcal{H}_1$) "$P \neq (P^1 \otimes P^2)$". The considered test statistic is defined by

$$T(\mathbb{X}_n) = \frac{1}{n-1}\left(\sum_{i=1}^{n}\varphi(X_i^1, X_i^2) - \frac{1}{n}\sum_{i,j=1}^{n}\varphi(X_i^1, X_j^2)\right), \tag{17.16}$$

where $\varphi$ is a measurable real-valued function on $\mathcal{X}^2$. Denoting for any real-valued measurable function $g$ on $\mathcal{X}^2$,

$$\mathbb{E}_P[g] = \int_{\mathcal{X}^2} g\left(x^1, x^2\right) dP\left(x^1, x^2\right) \quad \text{and} \quad \mathbb{E}_{\perp\!\!\!\perp}[g] = \int_{\mathcal{X}^2} g\left(x^1, x^2\right) dP^1\left(x^1\right) dP^2\left(x^2\right), \tag{17.17}$$

one may notice that, $T(\mathbb{X}_n)$ is an unbiased estimator of

$$\mathbb{E}\left[T(\mathbb{X}_n)\right] = \mathbb{E}_P[\varphi] - \mathbb{E}_{\perp\!\!\!\perp}[\varphi],$$

which is equal to 0 under ($\mathcal{H}_0$).

For more details on the choice of the test statistic, the interested reader can refer to [2] (motivated by synchrony detection in neuroscience for instance). The particular case where $\mathcal{X} = [0, 1]$ and $\varphi$ is a two-dimensional isotropic Haar wavelet is studied in [1, Chapter 4] and recalled below. More precisely, consider a resolution scale $j$ in $\mathbb{N}$, a translation $k = (k_1, k_2)$ in $\mathcal{K}_j := \{0, 1, \ldots, 2^j - 1\}^2$. Consider the functions defined for all $(x^1, x^2)$ in $[0, 1]^2$ by

$$\varphi_0(x^1, x^2) = \phi(x^1)\phi(x^2), \quad \text{and} \quad \begin{cases} \varphi_{(1,j,k)}(x^1, x^2) = \phi_{j,k_1}(x^1)\psi_{j,k_2}(x^2), \\ \varphi_{(2,j,k)}(x^1, x^2) = \psi_{j,k_1}(x^1)\phi_{j,k_2}(x^2), \\ \varphi_{(3,j,k)}(x^1, x^2) = \psi_{j,k_1}(x^1)\psi_{j,k_2}(x^2), \end{cases}$$

where $\phi = \mathbb{1}_{[0,1)}$ and $\psi = \mathbb{1}_{[0,1/2)} - \mathbb{1}_{[1/2,1)}$ are respectively the one-dimensional Haar father and Haar mother wavelets and

$$\Phi_{j,k}(\cdot) = 2^{j/2}\Phi(2^j \cdot - k)$$

denotes the dilated/translated wavelet at scale $j$ in $\mathbb{N}$ for $\Phi$ being either $\phi$ or $\psi$. Notice that $\mathcal{K}_j$ corresponds to the set of translations $k$ such that for any $1 \leq i \leq 3$, the intersection between the supports of the wavelets $\varphi_{(i,j,k)}$ and $[0, 1)^2$ is not empty.

Then the function $\varphi$ is taken out of the family $\{\varphi_\delta, \delta \in \Lambda\}$, with

$$\Lambda = \{0\} \cup \left\{(i, j, k) \in \{1, 2, 3\} \times \mathbb{N} \times \mathcal{K}_j\right\},$$

which constitutes an orthonormal basis of $\mathbb{L}_2\left([0, 1]^2\right)$. Notice that in this case, the Lipschitz assumptions of Adamczak et al. (see Theorem 17.1.3) are not satisfied, since the Haar wavelet functions are not even continuous.

The critical value of the test is obtained from the permutation approach, inspired by Hoeffding [16], and Romano [23]. Let $\Pi$ be a uniformly distributed random permutation of $\{1, \ldots, n\}$ independent of $\mathbb{X}_n$ and consider the permuted sample

$$\mathbb{X}_n^\Pi = (X_1^\Pi, \ldots X_n^\Pi), \quad \text{where } \forall 1 \leq i \leq n, \quad X_i^\Pi = (X_i^1, X_{\Pi(i)}^2),$$

obtained from permuting only the second coordinates. Then, under $(\mathcal{H}_0)$, the original sample $\mathbb{X}_n$ and the permuted one $\mathbb{X}_n^\Pi$ have the same distribution. Hence, the critical value of the upper-tailed test, denoted by $q_{1-\alpha}(\mathbb{X}_n)$, is the $(1-\alpha)$-quantile of the conditional distribution of the permuted statistic $T(\mathbb{X}_n^\Pi)$ given the sample $\mathbb{X}_n$, where the permuted test statistic is equal to

$$T(\mathbb{X}_n^\Pi) = \frac{1}{n-1}\left(\sum_{i=1}^n \varphi(X_i^1, X_{\Pi(i)}^2) - \frac{1}{n}\sum_{i,j=1}^n \varphi(X_i^1, X_j^2)\right).$$

More precisely, given $\mathbb{X}_n$, if

$$T^{(1)}(\mathbb{X}_n) \leq T^{(2)}(\mathbb{X}_n) \leq \cdots \leq T^{(n!)}(\mathbb{X}_n)$$

denote the ordered values of all the permuted test statistic $T(\mathbb{X}_n^\tau)$, when $\tau$ describes the set of all permutations of $\{1, \ldots, n\}$, then the critical value is equal to

$$q_{1-\alpha}(\mathbb{X}_n) = T^{(n! - \lfloor n!\alpha \rfloor)}(\mathbb{X}_n). \tag{17.18}$$

The corresponding test rejects the null hypothesis when $T(\mathbb{X}_n) > q_{1-\alpha}(\mathbb{X}_n)$, here denoted by

$$\Delta_\alpha(\mathbb{X}_n) = \mathbb{1}_{T(\mathbb{X}_n) > q_{1-\alpha}(\mathbb{X}_n)}. \tag{17.19}$$

In [2], the asymptotic properties of such test are studied. Based on a combinatorial central limit theorem in a non-i.i.d. case, the test is proved to be, under mild conditions, asymptotically of prescribed size, and power equal to one under any reasonable alternatives. Yet, as explained above, such purely asymptotic properties may be insufficient when applying these tests in neuroscience for instance. Moreover, the delicate choice of $\varphi$, generally out of a parametric family $\{\varphi_\delta\}_\delta$ (which reduces to the choice of the parameter $\delta$), is a real question, especially, in neuroscience, where it has some biological meaning, as mentioned in [2] and [3]. A possible approach to overcome this issue is to aggregate several tests for different parameters $\delta$, and reject independence if at least one of them does. In particular, this approach should give us information on how to choose this parameter. Yet, to do so, non-asymptotic controls are necessary.

From a non-asymptotic point of view, since the test is non-asymptotically of prescribed level by construction, remains the non-asymptotic control of the second kind error rate, that is the probability of wrongly accepting the null hypothesis. In the spirit of [11, 12, 24], the idea is to study the uniform separation rates of testing, in order to study the optimality in the minimax sense (see [5]).

From now on, consider an alternative $P$ satisfying $(\mathcal{H}_1)$, and an i.i.d. sample $\mathbb{X}_n$ from such distribution $P$. Assume moreover that the alternative satisfies $\mathbb{E}_P[\varphi] > \mathbb{E}_{\perp\!\!\perp}[\varphi]$, that is $\mathbb{E}[T(\mathbb{X}_n)] > 0$. The initial step is to find some condition on $P$ guaranteeing the control of the second kind error rate, namely $\mathbb{P}(\Delta_\alpha(\mathbb{X}_n) = 0)$, by a prescribed value $\beta > 0$. Intuitively, since the expectation of the test statistic $\mathbb{E}[T(\mathbb{X}_n)]$ is equal to zero under the null hypothesis, the test should be more efficient in rejecting $(\mathcal{H}_0)$ for large values of this expectation. So, the aim is to find conditions of the form $\mathbb{E}[T(\mathbb{X}_n)] \geq s$ for some threshold $s$ to be determined. Yet, one of the main difficulties here comes from the randomness of the critical value. The idea, as in [11], is thus to introduce $q^\alpha_{1-\beta/2}$ the $(1 - \beta/2)$-quantile of the critical value $q_{1-\alpha}(\mathbb{X}_n)$ and deduce from Chebychev's inequality (see section "A First Condition Ensuing from Chebychev's Inequality" in the Appendix), that the second kind error rate is controlled by $\beta$ as soon as

$$\mathbb{E}[T(\mathbb{X}_n)] \geq q^\alpha_{1-\beta/2} + \sqrt{\frac{2}{\beta}\operatorname{Var}(T(\mathbb{X}_n))}. \tag{17.20}$$

Usually, the goal in general minimax approaches is to express, for well-chosen functions $\varphi$, some distance between the alternative $P$ and the null hypothesis $(\mathcal{H}_0)$ in terms of $\mathbb{E}[T(\mathbb{X}_n)]$ for which minimax lower-bounds are known (see for instance [11, 12]). The objective is then to control, up to a constant, such distance (and in particular each term in the right-hand side of (17.20)) by the minimax rate of independence testing with respect to such distance on well-chosen regularity subspaces of alternatives, in order to prove the optimality of the method from a theoretical point of view. The interested reader could refer to the thesis [1, Chapter 4] for more details about this kind of development in the density case. It is not in the scope of the present article to develop such minimax theory in the general case, but to provide some general tools providing some sharp control of each term in the right-hand side of (17.20) which consists in a very first step of this approach. Some technical computations imply that the variance term can be upper bounded, up to a multiplicative constant, by $n^{-1}(\mathbb{E}_P[\varphi^2] + \mathbb{E}_{\perp\!\!\perp}[\varphi^2])$ (see Lemma 17.3.1). Hence, the challenging part relies in the quantile term. At this point, several ideas have been explored.

### 17.3.2 *Why Concentration Inequalities Are Necessary*

A first idea to control the conditional quantile of the permuted test statistic is based on the non-asymptotic control of the critical value obtained in section

"Control of the Critical Value Based on Hoeffding's Approach" in the Appendix (see Eq. (17.49)), following Hoeffding's idea (see [16, Theorem 2.1]), that leads to the condition

$$\mathbb{E}\left[T\left(\mathbb{X}_n\right)\right] \geq \frac{4}{\sqrt{\alpha}}\sqrt{\frac{2}{\beta}\frac{\mathbb{E}_P\left[\varphi^2\right]+\mathbb{E}_{\perp\!\!\!\perp}\left[\varphi^2\right]}{n}}. \tag{17.21}$$

The proof of this result is detailed in section "A First Condition Ensuing from Hoeffding's Approach" in the Appendix. Yet, this result may not be sharp enough, especially in $\alpha$. Indeed, as explained above, the next step consists in aggregating several tests for different functions $\varphi$ out of a parametric family $\{\varphi_\delta\}_\delta$ in a purpose of adaptivity. Generally, when aggregating tests, as in multiple testing methods, the multiplicity of the tests has to be taken into account. In particular, the single prescribed level of each individual test should be corrected. Several corrections exist, such as the Bonferroni one, which consists in dividing the global desired level $\alpha$ by the number of tests $M$. Yet, for such correction, the lower-bound in (17.21) comes with a cost in $\sqrt{M}$, which is too large to provide optimal rates. Even with more sophisticated corrections than the Bonferroni one (see, e.g., [11, 12, 24]), the control by a term of order $\sqrt{1/\alpha}$ is too large, since classically in the literature, the dependence on $\alpha$ should be of the order of $\sqrt{\ln(1/\alpha)}$. Hence, the bound ensuing from this first track being not sharp enough, the next idea was to investigate other non-asymptotic approaches for permuted sums.

Such approaches have also been studied in the literature. For instance, Ho and Chen [14] obtain non-asymptotic Berry-Esseen type bounds in the $\mathbb{L}^p$-distance between the cumulative distribution function (c.d.f.) of the standardized permuted sum of i.i.d. random variables and the c.d.f. of the normal distribution, based on Stein's method. In particular, they obtain the rate of convergence to a normal distribution in $\mathbb{L}^p$-distance under Lindeberg-type conditions. Then, Bolthausen [7] considers a different approach, also based on Stein's method allowing to extend Ho and Chen's results in the non-identically distributed case. More precisely, he obtains bounds in the $\mathbb{L}^\infty$-distance in the non-random case. In particular, in the deterministic case (which can easily be generalized to random cases), considering the notation introduced above, he obtains the following non-asymptotic bound:

$$\sup_{x\in\mathbb{R}}\left|\mathbb{P}\left(Z-\mathbb{E}\left[Z\right]\leq x\sqrt{\operatorname{Var}(Z)}\right)-\Phi_{0,1}(x)\right| \leq \frac{C}{n\sqrt{\operatorname{Var}(Z)}^3}\sum_{i,j=1}^{n}\left|d_{i,j}\right|^3,$$

where $C$ is an absolute constant, and $\Phi_{0,1}$ denotes the standard normal distribution function. In particular, when applying this result to answer our motivation by considering random variables $\varphi(X_i^1, X_j^2)$ instead of the deterministic terms $a_{i,j}$, and working conditionally on the sample $\mathbb{X}_n$, the permuted statistic $T(\mathbb{X}_n^\Pi)$ corresponds to $(n-1)^{-1}(Z - \mathbb{E}[Z])$. Therefore, the previous inequality implies that, for all $t$ in $\mathbb{R}$,

$$\mathbb{P}\big(T(\mathbb{X}_n^\Pi) > t \big| \mathbb{X}_n\big) \leq \left[1 - \Phi_{0,1}\left(\frac{t}{\sqrt{\operatorname{Var}\big(T(\mathbb{X}_n^\Pi)\big|\mathbb{X}_n\big)}}\right)\right] + \frac{C}{n(n-1)^{2/3}\sqrt{\operatorname{Var}\big(T(\mathbb{X}_n^\Pi)\big|\mathbb{X}_n\big)}^3} \sum_{i,j} \big|D_{i,j}\big|^3, \tag{17.22}$$

where $D_{i,j}$ is defined by

$$D_{i,j} = \varphi(X_i^1, X_j^2) - \frac{1}{n}\sum_{l=1}^n \varphi(X_i^1, X_l^2) - \frac{1}{n}\sum_{k=1}^n \varphi(X_k^1, X_j^2) + \frac{1}{n^2}\sum_{k,l=1}^n \varphi(X_k^1, X_l^2).$$

Yet, by definition of conditional quantiles, the critical value $q_{1-\alpha}(\mathbb{X}_n)$ is the smallest value of $t$ such that $\mathbb{P}(T(\mathbb{X}_n) > t|\mathbb{X}_n) \leq \alpha$. Hence, considering (17.22), one can easily make the first term of the sum in the right-hand side of the inequality as small as one wants by choosing $t$ large enough. However, the second term being fixed, nothing guarantees that the upper-bound in (17.22) can be constrained to be smaller than $\alpha$. Thus, this result cannot be applied in order to control non-asymptotically the critical value. Concentration inequalities seem thus to be adequate here, as they provide sharp non-asymptotic results, with usually exponentially small controls which leads to the desired logarithmic dependency in $\alpha$, as mentioned above.

### *17.3.3 A Sharp Control of the Conditional Quantile and a New Condition Guaranteeing a Control of the Second Kind Error Rate*

Sharp controls of the quantiles are provided in the following proposition.

**Proposition 17.3.1** *Consider the same notation as in Sect. 17.3.1 and let $q_{1-\beta/2}^\alpha$ be the $(1-\beta/2)$-quantile of the conditional quantile $q_{1-\alpha}(\mathbb{X}_n)$. Then, there exists two universal positive constants $C'$ and $c_0$ such that*

$$q_{1-\alpha}(\mathbb{X}_n) \leq \frac{C'}{n-1}\left\{\sqrt{\frac{1}{n}\sum_{i,j=1}^n \varphi^2(X_i^1, X_j^2)}\sqrt{\ln\left(\frac{c_0}{\alpha}\right)} + \|\varphi\|_\infty \ln\left(\frac{c_0}{\alpha}\right)\right\}. \tag{17.23}$$

*As a consequence, there exists a universal positive constants C such that*

$$q^{\alpha}_{1-\beta/2} \leq C \left\{ \sqrt{\frac{2}{\beta} \ln\left(\frac{c_0}{\alpha}\right)} \left( \frac{\sqrt{\mathbb{E}_P[\varphi^2]}}{n} + \frac{\sqrt{\mathbb{E}_{\perp\!\!\!\perp}[\varphi^2]}}{\sqrt{n}} \right) + \frac{\|\varphi\|_\infty}{n} \ln\left(\frac{c_0}{\alpha}\right) \right\}. \tag{17.24}$$

Moreover, a control of the variance term is obtained in the following lemma based on the Cauchy-Schwartz inequality.

**Lemma 17.3.1** *Let $n \geq 4$ and $\mathbb{X}_n$ be a sample of n i.i.d. random variables with distribution P and marginals $P^1$ and $P^2$. Let T be the test statistic defined in (17.16), and $\mathbb{E}_P[\cdot]$ and $\mathbb{E}_{\perp\!\!\!\perp}[\cdot]$ be notation introduced in (17.17). Then, if both $\mathbb{E}_P\left[\varphi^2\right] < +\infty$ and $\mathbb{E}_{\perp\!\!\!\perp}\left[\varphi^2\right] < +\infty$,*

$$\mathrm{Var}(T(\mathbb{X}_n)) \leq \frac{1}{n}\left(\sqrt{\mathbb{E}_P\left[\varphi^2\right]} + 2\sqrt{\mathbb{E}_{\perp\!\!\!\perp}\left[\varphi^2\right]}\right)^2.$$

Proposition 17.3.1 and Lemma 17.3.1 both imply that the right-hand side of (17.20) is upper bounded by

$$C'' \left\{ \sqrt{\frac{2}{\beta}\left[\ln\left(\frac{c_0}{\alpha}\right) + 1\right] \frac{\left(\mathbb{E}_P\left[\varphi^2\right] + \mathbb{E}_{\perp\!\!\!\perp}\left[\varphi^2\right]\right)}{n}} + \frac{\|\varphi\|_\infty}{n} \ln\left(\frac{c_0}{\alpha}\right) \right\}, \tag{17.25}$$

where $C''$ is a universal constant.

Indeed, the control of $q^{\alpha}_{1-\beta/2}$ is implied by (17.24) combined with the concavity property of the square-root function. Lemma 17.3.1 directly implies that the variance term satisfies

$$\mathrm{Var}(T(\mathbb{X}_n)) \leq \frac{8}{n}\left(\mathbb{E}_P\left[\varphi^2\right] + \mathbb{E}_{\perp\!\!\!\perp}\left[\varphi^2\right]\right),$$

Finally, if $\mathbb{E}\left[T(\mathbb{X}_n)\right]$ is larger than the quantity in (17.25), then condition (17.20) is satisfied which directly provides that $\mathbb{P}(\Delta_\alpha(\mathbb{X}_n) = 0) \leq \beta$, that is the second kind error rate of the test $\Delta_\alpha$ is less than or equal to the prescribed value $\beta$. One may notice that this time, the dependence in $\alpha$ is, as expected, of the order of $\sqrt{\ln(1/\alpha)}$.

## 17.4 Proofs

### *17.4.1 Proof of Lemma 17.2.1*

**Sketch of Proof** From now on, fix $t > 0$. Recall the notation introduced by Talagrand in Theorem 17.1.2. The main purpose of these notation is to introduce some notion of distance between a permutation $\tau$ in $\mathfrak{S}_n$ and a subset $A$ of $\mathfrak{S}_n$. To do so, the idea is to reduce the set of interest to a simpler one, that is $[0, 1]^n$, by considering

$$U_A(\tau) = \left\{ s \in \{0, 1\}^n \ ; \ \exists \sigma \in A \text{ such that } \forall 1 \leq i \leq n, \ s_i = 0 \implies \sigma(i) = \tau(i) \right\}.$$

One may notice that the permutation $\tau$ belongs to $A$ if and only if 0 belongs to the set $U_A(\tau)$. Hence, the corresponding distance between the permutation $\tau$ and the set $A$ is coded by the distance between 0 and the set $U_A(\tau)$ and thus defined by

$$f(A, \tau) = \min \left\{ \sum_{i=1}^{n} v_i^2 \ ; \ v = (v_i)_{1 \leq i \leq n} \in V_A(\tau) \right\},$$

where $V_A(\tau) = \text{ConvexHull}\,(U_A(\tau))$. One may notice in particular that $A$ contains $\tau$ if and only if the distance $f(A, \tau) = 0$.

The global frame of the proof of Lemma 17.2.1 (and also Proposition 17.2.1) relies on the following steps. The first step consists in proving that

$$\mathbb{P}\left( \sqrt{Z} \geq \sqrt{C_A} + t \sqrt{\max_{1 \leq i, j \leq n} \left\{ a_{i,j} \right\}} \right) \leq \frac{e^{-t^2/16}}{\mathbb{P}(Z \in A)}, \tag{17.26}$$

for some subset $A$ of $\mathfrak{S}_n$ of the shape $A = \{\sigma \in \mathfrak{S}_n \ ; \ Z(\sigma) \leq C_A\}$ for some constant $C_A$ to be chosen later. For this purpose, since Talagrand's inequality for random permutations (see Theorem 17.1.2) provides that

$$\mathbb{P}\Big( f(A, \Pi) \geq t^2 \Big) \leq \frac{e^{-t^2/16}}{\mathbb{P}(\Pi \in A)},$$

it is sufficient to prove that

$$\mathbb{P}\Big( f(A, \Pi) \geq t^2 \Big) \geq \mathbb{P}\left( \sqrt{Z} \geq \sqrt{C_A} + t \sqrt{\max_{1 \leq i, j \leq n} \left\{ a_{i,j} \right\}} \right),$$

to obtain (17.26). To do so, the idea, as in [4], is to show that the assertion $f(A, \Pi) < t^2$ implies that $\sqrt{Z} < \sqrt{C_A} + t\sqrt{\max_{1 \leq i, j \leq n} \left\{ a_{i,j} \right\}}$, and to conclude by contraposition.

Then, the two following steps consist in choosing appropriate constants $C_A$ in (17.26) depending on the median of $Z$, such that both $\mathbb{P}\Big( \sqrt{Z} \geq \sqrt{C_A} + t\sqrt{\max_{1 \leq i, j \leq n} \left\{ a_{i,j} \right\}} \Big)$ and $\mathbb{P}(Z \in A)$ are greater than 1/2, in order to control both probabilities

$$\mathbb{P}\left( \sqrt{Z} \geq \sqrt{\text{med}\,(Z)} + t \sqrt{\max_{1 \leq i, j \leq n} \left\{ a_{i,j} \right\}} \right) \text{ and } \mathbb{P}\left( \sqrt{Z} \leq \sqrt{\text{med}\,(Z)} - t \sqrt{\max_{1 \leq i, j \leq n} \left\{ a_{i,j} \right\}} \right)$$

respectively in (17.10) and (17.11).

**First Step: Preliminary Study** Assume $f(A, \Pi) < t^2$. Then, by definition of the distance $f$, there exists some $s^1, \dots, s^m$ in $U_A(\Pi)$, and some non-negative weights $p_1, \dots, p_m$ satisfying $\sum_{j=1}^m p_j = 1$ such that

$$\sum_{i=1}^n \left[ \left( \sum_{j=1}^m p_j s_i^j \right)^2 \right] < t^2.$$

For each $1 \leq j \leq m$, since $s^j$ belongs to $U_A(\Pi)$, one may consider a permutation $\sigma_j$ in $A$ associated to $s^j$ (that is satisfying $s_i^j = 0 \implies \sigma_j(i) = \Pi(i)$). Then, since the $a_{i,j}$ are non-negative, and from the Cauchy-Schwartz inequality,

$$\begin{aligned}
Z - \sum_{j=1}^m p_j Z(\sigma_j) &= \sum_{i=1}^n \sum_{j=1}^m p_j \left( a_{i,\Pi(i)} - a_{i,\sigma_j(i)} \right) \\
&= \sum_{i=1}^n \sum_{j=1}^m p_j \left( a_{i,\Pi(i)} - a_{i,\sigma_j(i)} \right) s_i^j \\
&\leq \sum_{i=1}^n \left[ \left( \sum_{j=1}^m p_j s_i^j \right) a_{i,\Pi(i)} \right] \\
&\leq \sqrt{\sum_{i=1}^n \left( \sum_{j=1}^m p_j s_i^j \right)^2} \sqrt{\sum_{i=1}^n a_{i,\Pi(i)}^2} \\
&< t \sqrt{\max_{1 \leq i,j \leq n} \left\{ a_{i,j} \right\}} \sqrt{Z}.
\end{aligned}$$

Thus, as the $\sigma_j$'s all belong to $A = \{\sigma \ ; \ Z(\sigma) \leq C_A\}$,

$$Z < C_A + t \sqrt{\max_{1 \leq i,j \leq n} \left\{ a_{i,j} \right\}} \sqrt{Z}.$$

Therefore, by solving the second-order polynomial in $\sqrt{Z}$ above, one obtains

$$\sqrt{Z} < \frac{t \sqrt{\max_{1 \leq i,j \leq n} \left\{ a_{i,j} \right\}} + \sqrt{t^2 \max_{1 \leq i,j \leq n} \left\{ a_{i,j} \right\} + 4C_A}}{2} \leq t \sqrt{\max_{1 \leq i,j \leq n} \left\{ a_{i,j} \right\}} + \sqrt{C_A}.$$

Finally, by contraposition,

$$\mathbb{P}\left(\sqrt{Z} \geq \sqrt{C_A} + t\sqrt{\max_{1\leq i,j\leq n}\left\{a_{i,j}\right\}}\right) \leq \mathbb{P}\Big(f(A,\Pi) \geq t^2\Big),$$

which, combined with (17.7) of Theorem 17.1.2 provides (17.26).

**Second Step: Proof of** (17.10) Taking $C_A = \operatorname{med}(Z)$ guarantees $\mathbb{P}(Z \in A) \geq 1/2$ and thus, (17.26) provides (17.10).

**Third Step: Proof of** (17.11) Taking $C_A = \left(\sqrt{\operatorname{med}(Z)} - t\sqrt{\max_{1\leq i,j\leq n}\left\{a_{i,j}\right\}}\right)^2$ implies

$$\mathbb{P}\left(\sqrt{Z} \geq \sqrt{C_A} + t\sqrt{\max_{1\leq i,j\leq n}\left\{a_{i,j}\right\}}\right) = \mathbb{P}\Big(\sqrt{Z} \geq \sqrt{\operatorname{med}(Z)}\Big) = \mathbb{P}(Z \geq \operatorname{med}(Z)) \geq \frac{1}{2}.$$

So finally, again by (17.26),

$$\begin{aligned}
\mathbb{P}\left(\sqrt{Z} \leq \sqrt{\operatorname{med}(Z)} - t\sqrt{\max_{1\leq i,j\leq n}\left\{a_{i,j}\right\}}\right) &= \mathbb{P}(Z \in A)\\
&\leq \frac{e^{-t^2/16}}{\mathbb{P}\left(\sqrt{Z} \geq \sqrt{C_A} + t\sqrt{\max_{1\leq i,j\leq n}\left\{a_{i,j}\right\}}\right)}\\
&\leq 2e^{-t^2/16},
\end{aligned}$$

which ends the proof of the Lemma.

### 17.4.2 Proof of Lemma 17.2.2

Let $X$ be any real random variable. Recall that

$$\operatorname{med}(X) \in \operatorname*{argmin}_{m\in\mathbb{R}} \mathbb{E}\left[|X - m|\right].$$

In particular, thanks to Jensen's inequality,

$$\begin{aligned}
\left|\mathbb{E}\left[X\right] - \operatorname{med}(X)\right| &\leq \mathbb{E}\left[|X - \operatorname{med}(X)|\right]\\
&\leq \mathbb{E}\left[|X - \mathbb{E}\left[X\right]|\right]\\
&\leq \sqrt{\mathbb{E}\left[(X - \mathbb{E}\left[X\right])^2\right]}\\
&\leq \sqrt{\operatorname{Var}(X)}. \qquad (17.27)
\end{aligned}$$

### *17.4.3 Proof of Proposition 17.2.1*

From now on, fix $x > 0$, and consider $t = x^2$. This proof is again based on Talagrand's inequality for random permutations, combined with (17.10) in Lemma 17.2.1. It follows exactly the same progression as in the proof of Lemma 17.2.1; the preliminary step consists in working with subsets $A \subset \mathfrak{S}_n$ of the form $A = \{\sigma \in \mathfrak{S}_n \ ; \ Z(\sigma) \leq C_A\}$ for some constant $C_A$, in order to obtain for all $v > 0$,

$$\mathbb{P}\left(Z \geq C_A + t\left(\sqrt{\operatorname{med}\left(\sum_{i=1}^{n} a_{i,\Pi(i)}^2\right)} + v \max_{1\leq i,j\leq n}\{a_{i,j}\}\right)\right) \leq \frac{e^{-t^2/16}}{\mathbb{P}(Z \in A)} + 2e^{-v^2/16}. \tag{17.28}$$

The second and third step consist in picking up a well-chosen constant $C_A$ and a well-chosen $v > 0$ in order to obtain respectively

$$\mathbb{P}\left(Z \geq \operatorname{med}(Z) + t\left(\sqrt{\operatorname{med}\left(\sum_{i=1}^{n} a_{i,\Pi(i)}^2\right)} + (t \vee C_0) \max_{1\leq i,j\leq n}\{a_{i,j}\}\right)\right) \leq 4e^{-t^2/16}, \tag{17.29}$$

and

$$\mathbb{P}\left(Z \leq \operatorname{med}(Z) - t\left(\sqrt{\operatorname{med}\left(\sum_{i=1}^{n} a_{i,\Pi(i)}^2\right)} + (t \vee C_0) \max_{1\leq i,j\leq n}\{a_{i,j}\}\right)\right) \leq 4e^{-t^2/16}, \tag{17.30}$$

where $C_0 = 4\sqrt{\ln(8)}$. The final step combines (17.29) and (17.30) in order to prove (17.12).

**First Step: Preliminary Study** Let $A = \{\sigma \in \mathfrak{S}_n \ ; \ Z(\sigma) \leq C_A\}$ with $C_A$ a general constant, and fix $v > 0$. Assume, this time, that both

$$f(A, \Pi) < t^2 \quad \text{and} \quad \sqrt{\sum_{i=1}^{n} a_{i,\Pi(i)}^2} < \sqrt{\operatorname{med}\left(\sum_{i=1}^{n} a_{i,\Pi(i)}^2\right)} + v \max_{1\leq i,j\leq n}\{a_{i,j}\}. \tag{17.31}$$

Then, as in the preliminary study of the proof of Lemma 17.2.1, from the first assumption in (17.31), there exists some $s^1, \ldots, s^m$ in $U_A(\Pi)$, and some non-negative weights $p_1, \ldots, p_m$ satisfying $\sum_{j=1}^m p_j = 1$ such that

$$\sum_{i=1}^n \left[ \left( \sum_{j=1}^m p_j s_i^j \right)^2 \right] < t^2.$$

For each $1 \leq j \leq m$, consider $\sigma_j$ in $A$ associated to $s^j$, that is a permutation $\sigma_j$ in $A$ satisfying $s_i^j = 0 \implies \sigma_j(i) = \Pi(i)$. Then, combining the Cauchy-Shwartz inequality with the second assumption in (17.31) leads to

$$\begin{aligned}
Z - \sum_{j=1}^m p_j Z(\sigma_j) &= \sum_{i=1}^n \sum_{j=1}^m p_j \left( a_{i,\Pi(i)} - a_{i,\sigma_j(i)} \right) s_i^j \\
&\leq \sum_{i=1}^n \left[ \left( \sum_{j=1}^m p_j s_i^j \right) a_{i,\Pi(i)} \right] \\
&\leq \sqrt{\sum_{i=1}^n \left( \sum_{j=1}^m p_j s_i^j \right)^2} \sqrt{\sum_{i=1}^n a_{i,\Pi(i)}^2} \\
&< t \left( \sqrt{\operatorname{med}\left( \sum_{i=1}^n a_{i,\Pi(i)}^2 \right) + v \max_{1 \leq i,j \leq n} \{a_{i,j}\}} \right).
\end{aligned}$$

Notice that here, the reasoning begins exactly as in the proof of Lemma 17.2.1. Yet, the second assumption in (17.31), which can be controlled using that lemma, allows us to sharpen the inequality. Thus, as the $\sigma_j$'s all belong to $A = \{\sigma\ ;\ Z(\sigma) \leq C_A\}$,

$$Z < C_A + t \left( \sqrt{\operatorname{med}\left( \sum_{i=1}^n a_{i,\Pi(i)}^2 \right) + v \max_{1 \leq i,j \leq n} \{a_{i,j}\}} \right). \tag{17.32}$$

Hence, by contraposition of (17.31) $\implies$ (17.32), one obtains

$$\begin{aligned}
&\mathbb{P}\left( Z \geq C_A + t \left( \sqrt{\operatorname{med}\left( \sum_{i=1}^n a_{i,\Pi(i)}^2 \right) + v \max_{1 \leq i,j \leq n} \{a_{i,j}\}} \right) \right) \\
&\leq \mathbb{P}\Big( f(A, \Pi) \geq t^2 \Big) + \mathbb{P}\left( \sqrt{\sum_{i=1}^n a_{i,\Pi(i)}^2} \geq \sqrt{\operatorname{med}\left( \sum_{i=1}^n a_{i,\Pi(i)}^2 \right) + v \max_{1 \leq i,j \leq n} \{a_{i,j}\}} \right),
\end{aligned}$$

and (17.28) follows from Theorem 17.1.2 and (17.10) in Lemma 17.2.1.

**Second Step: Proof of** (17.29) Consider $C_A = \text{med}(Z)$ so that $\mathbb{P}(Z \in A) \geq 1/2$. Thus, if $v = t$ in (17.28),

$$\mathbb{P}\left(Z \geq \text{med}(Z) + t\left(\sqrt{\text{med}\left(\sum_{i=1}^{n} a_{i,\Pi(i)}^2\right)} + (t \vee C_0) \max_{1 \leq i,j \leq n}\left\{a_{i,j}\right\}\right)\right)$$
$$\leq \mathbb{P}\left(Z \geq \text{med}(Z) + t\left(\sqrt{\text{med}\left(\sum_{i=1}^{n} a_{i,\Pi(i)}^2\right)} + t \max_{1 \leq i,j \leq n}\left\{a_{i,j}\right\}\right)\right)$$
$$\leq 4e^{-t^2/16}.$$

Notice that the maximum with the constant in $(t \vee C_0)$ is not necessary in the case only a control of the right-tail is wanted.

**Third Step: Proof of** (17.30) Consider now

$$C_A = \text{med}(Z) - t\left(\sqrt{\text{med}\left(\sum_{i=1}^{n} a_{i,\Pi(i)}^2\right)} + v \max_{1 \leq i,j \leq n}\left\{a_{i,j}\right\}\right),$$

so that

$$\mathbb{P}\left(Z \geq C_A + t\left(\sqrt{\text{med}\left(\sum_{i=1}^{n} a_{i,\Pi(i)}^2\right)} + v \max_{1 \leq i,j \leq n}\left\{a_{i,j}\right\}\right)\right) = \mathbb{P}(Z \geq \text{med}(Z)) \geq \frac{1}{2}.$$

Hence, on the one hand, from (17.28),

$$\mathbb{P}(Z \in A) \leq \frac{e^{-t^2/16}}{\left(\frac{1}{2} - 2e^{-v^2/16}\right)}.$$

Thus, if $v = C_0 = 4\sqrt{\ln(8)}$, then $\left(1/2 - 2e^{-v^2/16}\right) = 1/4$, and $\mathbb{P}(Z \in A) \leq 4e^{-t^2/16}$.

On the other hand, as $(t \vee C_0) \geq C_0 = v$,

$$\mathbb{P}(Z \in A) \geq \mathbb{P}\left(Z \leq \text{med}(Z) - t\left(\sqrt{\text{med}\left(\sum_{i=1}^{n} a_{i,\Pi(i)}^2\right)} + (t \vee C_0) \max_{1 \leq i,j \leq n}\left\{a_{i,j}\right\}\right)\right),$$

which ends the proof of (17.30).

**Fourth Step: Proof of** (17.12) Both (17.29) and (17.30) lead to

$$\mathbb{P}\left(|Z - \text{med}(Z)| > t\left(\sqrt{\text{med}\left(\sum_{i=1}^{n} a_{i,\Pi(i)}^2\right)} + (t \vee C_0) \max_{1\le i,j\le n}\left\{a_{i,j}\right\}\right)\right) \le 8e^{-t^2/16}.$$

Thus, on the one hand, if $t \ge C_0$, that is $t \vee C_0 = t$, and (17.12) holds. On the other hand, if $t < C_0$,

$$\mathbb{P}\left(|Z - \text{med}(Z)| > t\left(\sqrt{\text{med}\left(\sum_{i=1}^{n} a_{i,\Pi(i)}^2\right)} + t \max_{1\le i,j\le n}\left\{a_{i,j}\right\}\right)\right) \le 1$$

$$\le e^{C_0^2/16 - t^2/16} = 8e^{-t^2/16},$$

which ends the proof of the Proposition by taking $x = \sqrt{t}$.

### *17.4.4 Proof of Proposition 17.2.2*

First, for a better readability, let

$$M = \max_{1\le i,j\le n}\left\{a_{i,j}\right\} \quad \text{and} \quad V = \mathbb{E}\left[\sum_{i=1}^{n} a_{i,\Pi(i)}^2\right] = \frac{1}{n}\sum_{i,j=1}^{n} a_{i,j}^2.$$

Then, $\text{med}\left(\sum_{i=1}^{n} a_{i,\Pi(i)}^2\right) \le 2V$ since by Markov's inequality, for all non-negative random variable X, $\text{med}(X) \le 2\mathbb{E}[X]$. Indeed,

$$\frac{1}{2} \le \mathbb{P}(X \ge \text{med}(X)) \le \frac{\mathbb{E}[X]}{\text{med}(X)}.$$

Thus, by Proposition 17.2.1, one obtains that, for all $x > 0$,

$$\mathbb{P}\left(|Z - \text{med}(Z)| \ge \sqrt{2Vx} + Mx\right) \le 8e^{-x/16}. \tag{17.33}$$

The following is based on Lemma 17.2.2, and provides an upper-bound of the difference between the expectation and the median of $Z$.

**Lemma 17.4.1** *With the notation defined above,*

$$|\mathbb{E}[Z] - \text{med}(Z)| \le \sqrt{2V}.$$

*Proof (Proof of Lemma 17.4.1)* Lemma 17.2.2 implies that

$$|\mathbb{E}[Z] - \text{med}(Z)| \leq \sqrt{\text{Var}(Z)}.$$

Let us prove that

$$\text{Var}(Z) \leq 2V. \tag{17.34}$$

Indeed,

$$\begin{aligned}\text{Var}(Z) &= \mathbb{E}\left[\left(\sum_{i=1}^{n} a_{i,\Pi(i)} - \frac{1}{n}\sum_{i,j=1}^{n} a_{i,j}\right)^2\right] \\ &= \mathbb{E}\left[\left(\sum_{i,j=1}^{n} a_{i,j}\left(\mathbb{1}_{\Pi(i)=j} - \frac{1}{n}\right)\right)^2\right] \\ &= \sum_{i,j=1}^{n}\sum_{k,l=1}^{n} a_{i,j}a_{k,l}E_{i,j,k,l},\end{aligned}$$

where

$$E_{i,j,k,l} = \mathbb{E}\left[\left(\mathbb{1}_{\Pi(i)=j} - \frac{1}{n}\right)\left(\mathbb{1}_{\Pi(k)=l} - \frac{1}{n}\right)\right] = \mathbb{E}\left[\mathbb{1}_{\Pi(i)=j}\mathbb{1}_{\Pi(k)=l}\right] - \frac{1}{n^2}.$$

In particular,

$$E_{i,j,k,l} = \begin{cases} \frac{1}{n} - \frac{1}{n^2} \leq \frac{1}{n} & \text{if } i = k \text{ and } j = l, \\ \frac{-1}{n^2} \leq 0 & \text{if } i = k \text{ and } j \neq l \text{ or } i \neq k \text{ and } j = l, \\ \frac{1}{n(n-1)} - \frac{1}{n^2} = \frac{1}{n^2(n-1)} & \text{if } i \neq k \text{ and } j \neq l. \end{cases}$$

Therefore, from the Cauchy-Schwarz inequality applied to the second sum below (of $n^2(n-1)^2$ terms), one obtains

$$\begin{aligned}\text{Var}(Z) &\leq \frac{1}{n}\sum_{i,j=1}^{n} a_{i,j}^2 + \frac{1}{n^2(n-1)}\sum_{i\neq k}\sum_{j\neq l} a_{i,j}a_{k,l} \\ &\leq V + \frac{\sqrt{n^2(n-1)^2}}{n^2(n-1)}\sqrt{\sum_{i\neq k}\sum_{j\neq l} a_{i,j}^2 a_{k,l}^2}\end{aligned}$$

$$\begin{aligned}\mathrm{Var}(Z) &\le V + \frac{1}{n}\sqrt{\sum_{i,j} a_{i,j}^2 \sum_{k,l} a_{k,l}^2} \\ &= 2V.\end{aligned}$$

Finally, combining (17.27) and (17.34) ends the proof of Lemma 17.4.1. □

Therefore, one deduces from Lemma 17.4.1 and Eq. (17.33) that for all $x > 0$,

$$\mathbb{P}\Big(|Z - \mathbb{E}[Z]| \ge \sqrt{2V} + \sqrt{2Vx} + Mx\Big) \le 8e^{-x/16}. \tag{17.35}$$

Now, as in [8, Corollary 2.11], introduce $h_1 : u \in \mathbb{R}^+ \mapsto 1 + u - \sqrt{1+2u}$. Then, in particular, $h_1$ is non-decreasing, convex, one to one function on $\mathbb{R}^+$ with inverse function $h_1^{-1} : v \in \mathbb{R}^+ \mapsto v + \sqrt{2v}$. Indeed,

$$\begin{aligned}h_1\left(h_1^{-1}(v)\right) &= 1 + v + \sqrt{2v} - \sqrt{1 + 2v + 2\sqrt{2v}} \\ &= 1 + v + \sqrt{2v} - \sqrt{\left(1+\sqrt{2v}\right)^2} = v,\end{aligned}$$

and

$$\begin{aligned}h_1^{-1}\left(h_1(u)\right) &= 1 + u - \sqrt{1+2u} + \sqrt{2 + 2u - 2\sqrt{1+2u}} \\ &= u + 1 - \sqrt{1+2u} + \sqrt{1 - 2\sqrt{1+2u} + 1 + 2u} \\ &= 1 + u - \sqrt{1+2u} + \sqrt{\left(1 - \sqrt{1+2u}\right)^2} = u.\end{aligned}$$

Consider $\mathfrak{a}$ and $\mathfrak{c}$ defined by $\mathfrak{a} = V/M$ and $\mathfrak{c} = M^2/V$, such that $\mathfrak{a}\mathfrak{c} = M$ and $\mathfrak{a}^2\mathfrak{c} = V$ and thus

$$\sqrt{2Vx} + Mx = \mathfrak{a}h_1^{-1}(\mathfrak{c}x).$$

Then, from (17.35),

$$\mathbb{P}\Big(|Z - \mathbb{E}[Z]| \ge \sqrt{2\mathfrak{a}^2\mathfrak{c}} + \mathfrak{a}h_1^{-1}(\mathfrak{c}x)\Big) \le 8e^{-x/16}.$$

Let $t > 0$, and consider the two following cases.

**First Case** If $t \geq \sqrt{2V} = \sqrt{2\mathfrak{a}^2\mathfrak{c}}$, then define $x = \frac{1}{\mathfrak{c}} h_1 \left( \frac{t}{\mathfrak{a}} - \sqrt{2\mathfrak{c}} \right)$ such that $t = \sqrt{2\mathfrak{a}^2\mathfrak{c}} + \mathfrak{a} h_1^{-1}(\mathfrak{c}x)$. Then,

$$\mathbb{P}(|Z - \mathbb{E}[Z]| \geq t) \leq 8 \exp\left( -\frac{1}{16\mathfrak{c}} h_1 \left( \frac{t}{\mathfrak{a}} - \sqrt{2\mathfrak{c}} \right) \right).$$

Yet, by convexity of $h_1$,

$$h_1 \left( \frac{t}{\mathfrak{a}} - \sqrt{2\mathfrak{c}} \right) \geq 2h_1 \left( \frac{t}{2\mathfrak{a}} \right) - h_1 \left( \sqrt{2\mathfrak{c}} \right).$$

Hence,

$$\mathbb{P}(|Z - \mathbb{E}[Z]| \geq t) \leq 8 \exp\left( \frac{1}{16\mathfrak{c}} h_1 \left( \sqrt{2\mathfrak{c}} \right) \right) \exp\left( -\frac{1}{8\mathfrak{c}} h_1 \left( \frac{t}{2\mathfrak{a}} \right) \right).$$

Moreover, $\sqrt{2\mathfrak{c}} \leq \mathfrak{c} + \sqrt{2\mathfrak{c}} = h_1^{-1}(\mathfrak{c})$, hence

$$\frac{1}{16\mathfrak{c}} h_1 \left( \sqrt{2\mathfrak{c}} \right) \leq \frac{1}{16}.$$

So finally in this case,

$$\mathbb{P}(|Z - \mathbb{E}[Z]| \geq t) \leq 8e^{1/16} \exp\left( -\frac{1}{8\mathfrak{c}} h_1 \left( \frac{t}{2\mathfrak{a}} \right) \right). \tag{17.36}$$

**Second Case** If $t < \sqrt{2V} = \sqrt{2\mathfrak{a}^2\mathfrak{c}}$,

$$\mathbb{P}(|Z - \mathbb{E}[Z]| \geq t) \quad \leq \quad 1 \quad = \quad \exp\left( \frac{1}{8\mathfrak{c}} h_1 \left( \frac{t}{2\mathfrak{a}} \right) \right) \exp\left( -\frac{1}{8\mathfrak{c}} h_1 \left( \frac{t}{2\mathfrak{a}} \right) \right)$$

Moreover, in this case, since $\sqrt{2\mathfrak{c}}/2 \leq h_1^{-1}(\mathfrak{c}/4)$, hence

$$\frac{1}{8\mathfrak{c}} h_1 \left( \frac{t}{2\mathfrak{a}} \right) \leq \frac{1}{8\mathfrak{c}} h_1 \left( \frac{\sqrt{2\mathfrak{c}}}{2} \right) \leq \frac{1}{32},$$

and thus

$$\mathbb{P}(|Z - \mathbb{E}[Z]| \geq t) \leq e^{1/32} \exp\left( -\frac{1}{8\mathfrak{c}} h_1 \left( \frac{t}{2\mathfrak{a}} \right) \right). \tag{17.37}$$

Finally, combining (17.36) and (17.37) leads, in all cases, to

$$\mathbb{P}(|Z - \mathbb{E}[Z]| \geq t) \leq 8e^{1/16} \exp\left(-\frac{1}{8\mathfrak{c}} h_1\left(\frac{t}{2\mathfrak{a}}\right)\right). \tag{17.38}$$

Now, in order to obtain the Bernstein-type inequality, let $x = \frac{2}{\mathfrak{c}} h_1\left(\frac{t}{2\mathfrak{a}}\right)$, then

$$t = 2\mathfrak{a} h_1^{-1}\left(\frac{\mathfrak{c}x}{2}\right) = \mathfrak{a}\mathfrak{c}x + 2\sqrt{\mathfrak{a}^2\mathfrak{c}x} = 2\sqrt{Vx} + Mx,$$

and thus for all $x > 0$,

$$\mathbb{P}\left(|Z - \mathbb{E}[Z]| \geq 2\sqrt{Vx} + Mx\right) \leq 8e^{1/16} \exp\left(-\frac{x}{16}\right), \tag{17.39}$$

which ends the proof of the Proposition.

### *17.4.5 Proof of Corollary 17.2.1*

Consider the same notation as in both Proposition 17.2.2 and its proof. This proof follows the one of [19, Corollary 2.10]. Notice that for all $u \geq 0$,

$$h_1(u) \geq \frac{u^2}{2(1+u)}.$$

Hence, from (17.38) in the proof of Proposition 17.2.2, for all $t \geq 0$,

$$\begin{aligned}
\mathbb{P}(|Z - \mathbb{E}[Z]| \geq t) &\leq 8e^{1/16} \exp\left(-\frac{1}{8\mathfrak{c}} h_1\left(\frac{t}{2\mathfrak{a}}\right)\right) \\
&\leq 8e^{1/16} \exp\left(-\frac{t^2}{64\mathfrak{a}^2\mathfrak{c}\,(1 + t/2\mathfrak{a})}\right) \\
&= 8e^{1/16} \exp\left(-\frac{t^2}{32\left(2\mathfrak{a}^2\mathfrak{c} + \mathfrak{a}\mathfrak{c}t\right)}\right) \\
&= 8e^{1/16} \exp\left(-\frac{t^2}{32\,(V + Mt)}\right).
\end{aligned}$$

which ends the proof of the corollary.

### 17.4.6 Proof of Theorem 17.2.1

For a better readability, let us introduce $a^+_{i,j} = a_{i,j}\mathbb{1}_{a_{i,j}\geq 0}$ (respectively $a^-_{i,j} = -a_{i,j}\mathbb{1}_{a_{i,j}<0}$), and denote $Z^+ = \sum_{i=1}^n a^+_{i,\Pi(i)}$ (respectively $Z^- = \sum_{i=1}^n a^-_{i,\Pi(i)}$). Then

$$Z = \sum_{i=1}^n a_{i,\Pi(i)} = Z^+ - Z^-.$$

Moreover, if $v$ (respectively $v^+$ and $v^-$) denotes $\frac{1}{n}\sum_{i,j=1}^n a^2_{i,j}$ (respectively $\frac{1}{n}\sum_{i,j=1}^n (a^+_{i,j})^2$ and $\frac{1}{n}\sum_{i,j=1}^n (a^-_{i,j})^2$), then $v = v^+ + v^-$ and, from the concavity property of the square root function,

$$\sqrt{2v} \geq \sqrt{v^+} + \sqrt{v^-}.$$

Furthermore, if $M^+$ (respectively $M^-$) denotes $\max_{1\leq i,j\leq n}\{a^+_{i,j}\}$ (respectively $\max_{1\leq i,j\leq n}\{a^-_{i,j}\}$), then $2M = 2\max_{1\leq i,j\leq n}\left\{|a_{i,j}|\right\} \geq M^+ + M^-$.

Finally, applying Proposition 17.2.2 to $Z^+$ and $Z^-$ which are both sums of non-negative numbers leads to

$$\begin{aligned}
&\mathbb{P}\Big(|Z-\mathbb{E}[Z]| \geq 2\sqrt{2vx} + 2Mx\Big)\\
&\quad\leq \mathbb{P}\Big(\left|Z^+ - \mathbb{E}\left[Z^+\right]\right| + \left|Z^- - \mathbb{E}\left[Z^-\right]\right| \geq 2\sqrt{v^+x} + M^+x + 2\sqrt{v^-x} + M^-x\Big)\\
&\quad\leq \mathbb{P}\Big(\left|Z^+ - \mathbb{E}\left[Z^+\right]\right| \geq 2\sqrt{v^+x} + M^+x\Big) + \mathbb{P}\Big(\left|Z^- - \mathbb{E}\left[Z^-\right]\right| \geq 2\sqrt{v^-x} + M^-x\Big)\\
&\quad\leq 16e^{1/16}\exp\left(-\frac{x}{16}\right),
\end{aligned}$$

which ends the proof of the Theorem.

### 17.4.7 Proof of Corollary 17.2.2

Consider the same notation as in the proof of Theorem 17.2.1, and let $t > 0$. Let $M$ denote the maximum $\max_{1\leq i,j\leq n}\left\{|a_{i,j}|\right\}$. On the one hand, $M^+ \leq M$ and $M^- \leq M$, and on the other hand, $v^+ \leq v$ and $v^- \leq v$. Therefore, applying Corollary 17.2.1, one obtains

$$\begin{aligned}
\mathbb{P}(|Z - \mathbb{E}[Z]| \geq t) &\leq \mathbb{P}\big(\left|Z^+ - \mathbb{E}\left[Z^+\right]\right| + \left|Z^- - \mathbb{E}\left[Z^-\right]\right| \geq t\big)\\
&\leq \mathbb{P}\big(\left|Z^+ - \mathbb{E}\left[Z^+\right]\right| \geq t/2\big) + \mathbb{P}\big(\left|Z^- - \mathbb{E}\left[Z^-\right]\right| \geq t/2\big)
\end{aligned}$$

$$\mathbb{P}(|Z - \mathbb{E}[Z]| \geq t) \leq 8e^{1/16} \exp\left(\frac{-(t/2)^2}{16\left(4v^+ + 2M^+ t/2\right)}\right) + 8e^{1/16} \exp\left(\frac{-(t/2)^2}{16\left(4v^- + 2M^- t/2\right)}\right)$$
$$\leq 16e^{1/16} \exp\left(\frac{-t^2}{64\,(4v + Mt)}\right),$$

which leads to the following intermediate result

$$\mathbb{P}(|Z - \mathbb{E}[Z]| \geq t) \leq 16e^{1/16} \exp\left(\frac{-t^2}{64\left(4\frac{1}{n}\sum_{i,j=1}^n a_{i,j}^2 + \max_{1\leq i,j\leq n}\left\{\left|a_{i,j}\right|\right\} t\right)}\right). \tag{17.40}$$

In order to make the variance appear, consider Hoeffding's centering trick recalled in (17.1) and introduce

$$d_{i,j} = a_{i,j} - \frac{1}{n}\sum_{k=1}^n a_{k,j} - \frac{1}{n}\sum_{l=1}^n a_{i,l} + \frac{1}{n^2}\sum_{k,l=1}^n a_{k,l} = \frac{1}{n^2}\sum_{k,l=1}^n \left(a_{i,j} - a_{k,j} - a_{i,l} + a_{k,l}\right).$$

One may easily verify that for all $i_0$ and $j_0$, $\sum_{i=1}^n d_{i,j_0} = \sum_{j=1}^n d_{i_0,j} = 0$. Moreover,

$$\sum_{i=1}^n d_{i,\Pi(i)} = \sum_{i=1}^n a_{i,\Pi(i)} - \frac{1}{n}\sum_{i,j=1}^n a_{i,j} = Z - \mathbb{E}[Z] \quad \text{and} \quad \mathbb{E}\left[\sum_{i=1}^n d_{i,\Pi(i)}\right] = \frac{1}{n}\sum_{i,j=1}^n d_{i,j} = 0.$$

In particular, applying Eq. (17.40) to the permuted sum of the $d_{i,j}$'s leads to

$$\mathbb{P}(|Z - \mathbb{E}[Z]| \geq t) \leq 16e^{1/16} \exp\left(\frac{-t^2}{64\left(4\frac{1}{n}\sum_{i,j=1}^n d_{i,j}^2 + \max_{1\leq i,j\leq n}\left\{\left|d_{i,j}\right|\right\} t\right)}\right). \tag{17.41}$$

Then, it is sufficient to notice that, on the one hand, from [15, Theorem 2],

$$\mathrm{Var}(Z) = \frac{1}{n-1}\sum_{i,j=1}^n d_{i,j}^2 \geq \frac{1}{n}\sum_{i,j=1}^n d_{i,j}^2,$$

and on the other hand,

$$\max_{1\leq i,j\leq n}\left\{\left|d_{i,j}\right|\right\} \leq 4 \max_{1\leq i,j\leq n}\left\{\left|a_{i,j}\right|\right\},$$

to end the proof of Corollary 17.2.2.

### 17.4.8 Proof of Proposition 17.3.1

The proof of Proof of Proposition 17.3.1 is divided into two steps. The first step consists in controlling the conditional quantile $q_{1-\alpha}(\mathbb{X}_n)$ and the second step provides an upper-bound for $q^{\alpha}_{1-\beta/2}$.

**First Step** Let us prove (17.23), that is

$$q_{1-\alpha}(\mathbb{X}_n) \leq \frac{C'}{n-1}\left\{\sqrt{\frac{1}{n}\sum_{i,j=1}^{n}\varphi^2(X_i^1,X_j^2)}\sqrt{\ln\left(\frac{c_0}{\alpha}\right)} + \|\varphi\|_\infty \ln\left(\frac{c_0}{\alpha}\right)\right\}.$$

Introduce $\tilde{Z}(\mathbb{X}_n) = \sum_{i=1}^n \varphi(X_i^1, X^2_{\Pi(i)})$. Then, notice that

$$T^{\Pi}(\mathbb{X}_n) = \frac{1}{n-1}\left(\tilde{Z}(\mathbb{X}_n) - \mathbb{E}\left[\tilde{Z}(\mathbb{X}_n)\middle|\mathbb{X}_n\right]\right). \tag{17.42}$$

Therefore, applying Theorem 17.2.1 to the conditional probability given $\mathbb{X}_n$, one obtains that there exist universal positive constants $c_0$ and $c_1$ such that, for all $x > 0$,

$$\mathbb{P}\left(\left|\tilde{Z}(\mathbb{X}_n) - \mathbb{E}\left[\tilde{Z}(\mathbb{X}_n)\middle|\mathbb{X}_n\right]\right| \geq 2\sqrt{2\left(\frac{1}{n}\sum_{i,j=1}^{n}\varphi^2(X_i^1,X_j^2)\right)x} + 2\|\varphi\|_\infty x \,\middle|\, \mathbb{X}_n\right) \leq c_0\exp\left(-c_1 x\right).$$

In particular, from (17.42), one obtains

$$\mathbb{P}\left(\left|T(\mathbb{X}_n^{\Pi})\right| \geq \frac{2}{n-1}\left(\sqrt{2\left(\frac{1}{n}\sum_{i,j=1}^{n}\varphi^2(X_i^1,X_j^2)\right)x} + \|\varphi\|_\infty x\right)\middle|\,\mathbb{X}_n\right) \leq c_0\exp\left(-c_1 x\right).$$

Yet, by definition of the quantile, $q_{1-\alpha}(\mathbb{X}_n)$ is the smallest $u$ such that

$$\mathbb{P}\left(\left|T(\mathbb{X}_n^{\Pi})\right| \geq u \middle| \mathbb{X}_n\right) \leq \alpha.$$

Thus taking $x$ such that $c_0\exp(-c_1 x) = \alpha$, that is $x = c_1^{-1}\ln(c_0/\alpha)$, one obtains (17.23) with $C' = 2\max\left\{\sqrt{2/c_1}, 1/c_1\right\}$ which is a universal positive constant.

**Second Step** Let us now control the quantile $q^{\alpha}_{1-\beta/2}$. Since (17.23) is always true, by definition of $q^{\alpha}_{1-\beta/2}$, one has that $q^{\alpha}_{1-\beta/2}$ is upper bounded by the $(1-\beta/2)$-quantile of the right-hand side of (17.23). Yet, the only randomness left in the right-hand side of (17.23) comes from the randomness of $\frac{1}{n}\sum_{i,j=1}^{n}\varphi^2(X_i^1, X_j^2)$, and thus it is sufficient to control its $(1-\beta/2)$-quantile.

Besides, applying Markov's inequality, one obtains for all $x > 0$,

$$\mathbb{P}\left(\frac{1}{n}\sum_{i,j=1}^{n}\varphi^2(X_i^1, X_j^2) \geq x\right) \leq \frac{\mathbb{E}\left[\frac{1}{n}\sum_{i,j=1}^{n}\varphi^2(X_i^1, X_j^2)\right]}{x},$$

with $\mathbb{E}\left[\frac{1}{n}\sum_{i,j=1}^{n}\varphi^2(X_i^1, X_j^2)\right] = \mathbb{E}_P\left[\varphi^2\right] + (n-1)\mathbb{E}_{\perp\!\!\!\perp}\left[\varphi^2\right]$, and thus, taking

$$x = \frac{2}{\beta}\left(\mathbb{E}_P\left[\varphi^2\right] + (n-1)\mathbb{E}_{\perp\!\!\!\perp}\left[\varphi^2\right]\right),$$

one has that the $(1-\beta/2)$-quantile of $\frac{1}{n}\sum_{i,j=1}^{n}\varphi^2(X_i^1, X_j^2)$ is upper bounded by $x$, and thus, the $(1-\beta/2)$-quantile of $\sqrt{\frac{1}{n}\sum_{i,j=1}^{n}\varphi^2(X_i^1, X_j^2)}$ is itself upper bounded by

$$\sqrt{\frac{2}{\beta}}\left(\sqrt{\mathbb{E}_P[\varphi^2]} + \sqrt{n}\sqrt{\mathbb{E}_{\perp\!\!\!\perp}[\varphi^2]}\right).$$

Finally,

$$q^{\alpha}_{1-\beta/2} \leq \frac{2C'}{n}\left\{\sqrt{\frac{2}{\beta}}\left(\sqrt{\mathbb{E}_P[\varphi^2]} + \sqrt{n}\sqrt{\mathbb{E}_{\perp\!\!\!\perp}[\varphi^2]}\right)\sqrt{\ln\left(\frac{c_0}{\alpha}\right)} + \|\varphi\|_\infty \ln\left(\frac{c_0}{\alpha}\right)\right\}.$$

which is exactly (17.24) for any constant $C \geq 2C'$.

## 17.5 Proof of Lemma 17.3.1

Let us now prove Lemma 17.3.1. Let $n \geq 4$ and $\mathbb{X}_n$ be an i.i.d. sample with distribution $P$. First notice that one can write

$$T(\mathbb{X}_n) = \frac{1}{n(n-1)}\sum_{i\neq j}\left(\varphi(X_i^1, X_i^2) - \varphi(X_i^1, X_j^2)\right).$$

In particular, one recovers that $\mathbb{E}\left[T(\mathbb{X}_n)\right] = \mathbb{E}_P[\varphi] - \mathbb{E}_{\perp\!\!\!\perp}[\varphi]$.

For a better readability, let us introduce for all $i \neq j$ in $\{1, 2, \ldots, n\}$,

$$Y_i = \varphi(X_i^1, X_i^2) - \mathbb{E}_P[\varphi] \quad \text{and} \quad Z_{i,j} = \varphi(X_i^1, X_j^2) - \mathbb{E}_{\perp\!\!\!\perp}[\varphi].$$

Then,

$$\mathbb{E}[Y_i] = \mathbb{E}\left[Z_{i,j}\right] = 0, \quad \text{and} \quad \begin{cases} \mathbb{E}\left[Y_i^2\right] = \mathrm{Var}_P(\varphi) \leq \mathbb{E}_P\left[\varphi^2\right], \\ \mathbb{E}\left[Z_{i,j}^2\right] = \mathrm{Var}_{\perp\!\!\!\perp}(\varphi) \leq \mathbb{E}_{\perp\!\!\!\perp}\left[\varphi^2\right]. \end{cases} \tag{17.43}$$

One can write

$$T(\mathbb{X}_n) - \mathbb{E}[T(\mathbb{X}_n)] = \frac{1}{n(n-1)} \sum_{i \neq j} \left(Y_i - Z_{i,j}\right),$$

and thus,

$$\begin{aligned} \mathrm{Var}(T(\mathbb{X}_n)) &= \mathbb{E}\left[\left(\frac{1}{n(n-1)} \sum_{i \neq j} \left(Y_i - Z_{i,j}\right)\right)^2\right] \\ &= \frac{1}{n^2(n-1)^2} \sum_{i \neq j} \sum_{k \neq l} \mathbb{E}\left[\left(Y_i - Z_{i,j}\right)\left(Y_k - Z_{k,l}\right)\right] \\ &= A_n - 2B_n + C_n, \end{aligned}$$

with

$$A_n = \frac{1}{n^2} \sum_{i,k=1}^{n} \mathbb{E}[Y_i Y_k],$$

$$B_n = \frac{1}{n^2(n-1)} \sum_{i=1}^{n} \sum_{k \neq l} \mathbb{E}\left[Y_i Z_{k,l}\right],$$

$$C_n = \frac{1}{n^2(n-1)^2} \sum_{i \neq j} \sum_{k \neq l} \mathbb{E}\left[Z_{i,j} Z_{k,l}\right],$$

where each sum is taken for indexes contained in $\{1, 2, \ldots, n\}$. In particular, since just an upper-bound of the variance is needed, it is sufficient to write

$$\mathrm{Var}(T(\mathbb{X}_n)) \leq |A_n| + 2|B_n| + |C_n|, \tag{17.44}$$

and to study each term separately.

**Study of $A_n$** Since by construction, the $Y_i$'s are centered, and independent (as the $X_i$'s are),

$$A_n = \frac{1}{n^2}\left(\sum_i \mathbb{E}\left[Y_i^2\right] + \sum_{i\neq k}\mathbb{E}\left[Y_i\right]\mathbb{E}\left[Y_k\right]\right)$$
$$= \frac{1}{n}\mathbb{E}\left[Y_1^2\right],$$

and in particular, from (17.43),

$$|A_n| \leq \frac{1}{n}\mathbb{E}_P\left[\varphi^2\right]. \tag{17.45}$$

**Study of $B_n$** If $i, k$ and $l$ are all different, using once again the independence of the $X_i$'s and a centering argument, then $\mathbb{E}\left[Y_i Z_{k,l}\right] = \mathbb{E}\left[Y_i\right]\mathbb{E}\left[Z_{k,l}\right] = 0$. Thus

$$B_n = \frac{1}{n^2(n-1)}\sum_{i\neq k}\left(\mathbb{E}\left[Y_i Z_{i,k}\right] + \mathbb{E}\left[Y_i Z_{k,i}\right]\right)$$
$$= \frac{1}{n}\left(\mathbb{E}\left[Y_1 Z_{1,2}\right] + \mathbb{E}\left[Y_1 Z_{2,1}\right]\right).$$

In particular, applying the Cauchy-Schwartz inequality, and from (17.43), one obtains

$$|B_n| \leq \frac{2}{n}\sqrt{\mathbb{E}\left[Y_1^2\right]\mathbb{E}\left[Z_{1,2}^2\right]} \leq \frac{2}{n}\sqrt{\mathbb{E}_P\left[\varphi^2\right]\mathbb{E}_{\perp\!\!\!\perp}\left[\varphi^2\right]}. \tag{17.46}$$

**Study of $C_n$** Still by an independence and a centering argument, if $i, j, k$ and $l$ are all different, $\mathbb{E}\left[Z_{i,j}Z_{k,l}\right] = \mathbb{E}\left[Z_{i,j}\right]\mathbb{E}\left[Z_{k,l}\right] = 0$. Thus, if $I_n^{[3]}$ denotes the set of triplets $(i, j, k)$ in $\{1, \ldots, n\}^3$ which are all different, one obtains

$$C_n = \frac{1}{n^2(n-1)^2}\left\{\sum_{(i,j,k)\in I_n^{[3]}}\left(\mathbb{E}\left[Z_{i,j}Z_{i,k}\right] + 2\mathbb{E}\left[Z_{i,j}Z_{k,i}\right] + \mathbb{E}\left[Z_{j,i}Z_{k,i}\right]\right)\right.$$
$$\left. + \sum_{i\neq j}\left(\mathbb{E}\left[Z_{i,j}^2\right] + \mathbb{E}\left[Z_{i,j}Z_{j,i}\right]\right)\right\}$$
$$= \frac{n-2}{n(n-1)}\left(\mathbb{E}\left[Z_{1,2}Z_{1,3}\right] + 2\mathbb{E}\left[Z_{1,2}Z_{3,1}\right] + \mathbb{E}\left[Z_{2,1}Z_{3,1}\right]\right)$$
$$+ \frac{1}{n(n-1)}\left(\mathbb{E}\left[Z_{1,2}^2\right] + \mathbb{E}\left[Z_{1,2}Z_{2,1}\right]\right).$$

In particular, applying the Cauchy-Schwartz inequality, and using (17.43), each expectation in the previous equation satisfies $\mathbb{E}\left[Z_{i,j}Z_{k,l}\right] \le \mathbb{E}\left[Z_{1,2}^2\right] \le \mathbb{E}_{\perp\!\!\perp}[\varphi^2]$, and thus

$$|C_n| \le \left(\frac{4(n-2)}{n(n-1)} + \frac{2}{n(n-1)}\right)\mathbb{E}_{\perp\!\!\perp}\left[\varphi^2\right] \le \frac{4}{n}\mathbb{E}_{\perp\!\!\perp}\left[\varphi^2\right]. \tag{17.47}$$

Finally, combining (17.44), (17.45), (17.46), and (17.47) leads to

$$\mathrm{Var}(T(\mathbb{X}_n)) \le \frac{1}{n}\left(\sqrt{\mathbb{E}_P[\varphi^2]} + 2\sqrt{\mathbb{E}_{\perp\!\!\perp}[\varphi^2]}\right)^2,$$

which ends the proof of the lemma.

**Acknowledgements** The author is grateful to Jean-François Coeurjolly for his insightful discussions. This work was supported in part by the French Agence Nationale de la Recherche (ANR 2011 BS01 010 01 projet Calibration).

## Appendix: A Non-asymptotic Control of the Second Kind Error Rates

Consider the notation from Sect. 17.3. Since this section focuses on the study of the second kind error rate of the test, in all the sequel, the observation is assumed to satisfy the alternative $(\mathcal{H}_1)$. Let thus $P$ be an alternative, that is $P \neq (P^1 \otimes P^2)$, $n \ge 4$ and $\mathbb{X}_n = (X_i, \dots, X_n)$ be an i.i.d. sample from distribution $P$. Fix $\alpha$ and $\beta$ be two fixed values in $(0, 1)$. Consider $T$ the test statistic introduced in (17.16), the (random) critical value $q_{1-\alpha}(\mathbb{X}_n)$ defined in (17.18), and the corresponding permutation test defined in (17.19) by

$$\Delta_\alpha(\mathbb{X}_n) = \mathbb{1}_{T(\mathbb{X}_n) > q_{1-\alpha}(\mathbb{X}_n)},$$

which precisely rejects independence when $T(\mathbb{X}_n) > q_{1-\alpha}(\mathbb{X}_n)$. Notice that this test is exactly the upper-tailed test by permutation introduced in [2].

The aim of this section is to provide different conditions on the alternative $P$ ensuring a control of the second kind error rate by a fixed value $\beta > 0$, that is $\mathbb{P}(\Delta_\alpha(\mathbb{X}_n) = 0) \le \beta$. The following steps constitute the first steps of a general study of the separation rates for the previous independence test, and is worked through in the specific case of continuous real-valued random variables in [1, Chapter 4].

Recall the notation introduced in (17.17) for a better readability. For all real-valued measurable function $g$ on $\mathcal{X}^2$, denote respectively

$$\mathbb{E}_P[g] = \mathbb{E}\left[g(X_1^1, X_1^2)\right] \quad \text{and} \quad \mathbb{E}_{\perp\!\!\perp}[g] = \mathbb{E}\left[g(X_1^1, X_2^2)\right],$$

the expectations of $g(X)$ under the alternative $P$, that is if $X \sim P$, and under the null hypothesis $(\mathcal{H}_0)$, that is if $X \sim (P^1 \otimes P^2)$.

Assume the following moment assumption holds, that is

$$\left(\mathcal{A}_{Mmt,2}\right) \quad \Big| \text{ both } \mathbb{E}_P\left[\varphi^2\right] < +\infty \text{ and } \mathbb{E}_{\perp\!\!\!\perp}\left[\varphi^2\right] < +\infty,$$

so that all variance and second-order moments exist. Then, the following statements hold.

1. By Chebychev's inequality, one has $\mathbb{P}(\Delta_\alpha(\mathbb{X}_n) = 0) \leq \beta$ as soon as Condition (17.20) is satisfied, that is

$$\mathbb{E}\left[T(\mathbb{X}_n)\right] \geq q^\alpha_{1-\beta/2} + \sqrt{\frac{2}{\beta}\operatorname{Var}(T(\mathbb{X}_n))}.$$

2. On the one hand,

$$\operatorname{Var}(T(\mathbb{X}_n)) \leq \frac{8}{n}\left(\mathbb{E}_P\left[\varphi^2\right] + \mathbb{E}_{\perp\!\!\!\perp}\left[\varphi^2\right]\right), \tag{17.48}$$

3. On the other hand, in order to control the quantile $q^\alpha_{1-\beta/2}$, let us first upper bound the conditional quantile, following Hoeffding's approach based on the Cauchy-Schwarz inequality, by

$$q_{1-\alpha}(\mathbb{X}_n) \leq \sqrt{\frac{1-\alpha}{\alpha}\operatorname{Var}\left(T\left(\mathbb{X}_n^\Pi\right) \middle| \mathbb{X}_n\right)}. \tag{17.49}$$

4. Markov's inequality allows us to deduce the following bound for the quantile:

$$q^\alpha_{1-\beta/2} \leq 2\sqrt{\frac{1-\alpha}{\alpha}}\sqrt{\frac{2}{\beta}\frac{\left(\mathbb{E}_{\perp\!\!\!\perp}\left[\varphi^2\right] + \mathbb{E}_P\left[\varphi^2\right]\right)}{n}}. \tag{17.50}$$

5. Finally, combining (17.20), (17.48) and (17.50) ensures that $\mathbb{P}(\Delta_\alpha(\mathbb{X}_n) = 0) \leq \beta$ as soon as Condition (17.21) is satisfied, that is

$$\mathbb{E}\left[T(\mathbb{X}_n)\right] \geq \frac{4}{\sqrt{\alpha}}\sqrt{\frac{2}{\beta}\frac{\mathbb{E}_P\left[\varphi^2\right] + \mathbb{E}_{\perp\!\!\!\perp}\left[\varphi^2\right]}{n}}.$$

This section is divided in five subsections, each one of them respectively proving a point stated above. The first one proves the sufficiency of Condition (17.20) in order to control the second kind error rate. The second, third and fourth ones provide respectively upper-bounds of the variance term, the critical value and the quantile $q^\alpha_{1-\beta/2}$. Finally, the fifth one provides the sufficiency of Condition (17.21).

### *A First Condition Ensuing from Chebychev's Inequality*

In this section, we prove the sufficiency of a first simple condition, derived from Chebychev's inequality in order to control the second error rate. Assume that (17.20) is satisfied, that is

$$\mathbb{E}\left[T(\mathbb{X}_n)\right] \geq q^{\alpha}_{1-\beta/2} + \sqrt{\frac{2}{\beta}\operatorname{Var}(T(\mathbb{X}_n))}.$$

Then,

$$\begin{aligned}
\mathbb{P}(\Delta_\alpha(\mathbb{X}_n) = 0) &= \mathbb{P}(T(\mathbb{X}_n) \leq q_{1-\alpha}(\mathbb{X}_n)) \qquad (17.51)\\
&= \mathbb{P}\Big(\{T(\mathbb{X}_n) \leq q_{1-\alpha}(\mathbb{X}_n)\} \cap \Big\{q_{1-\alpha}(\mathbb{X}_n) \leq q^{\alpha}_{1-\beta/2}\Big\}\Big)\\
&\quad + \mathbb{P}\Big(\{T(\mathbb{X}_n) \leq q_{1-\alpha}(\mathbb{X}_n)\} \cap \Big\{q_{1-\alpha}(\mathbb{X}_n) > q^{\alpha}_{1-\beta/2}\Big\}\Big)\\
&\leq \mathbb{P}\Big(T(\mathbb{X}_n) \leq q^{\alpha}_{1-\beta/2}\Big) + \mathbb{P}\Big(q_{1-\alpha}(\mathbb{X}_n) > q^{\alpha}_{1-\beta/2}\Big)\\
&\leq \mathbb{P}\Big(T(\mathbb{X}_n) \leq q^{\alpha}_{1-\beta/2}\Big) + \frac{\beta}{2}, \qquad (17.52)
\end{aligned}$$

by definition of the quantile $q^{\alpha}_{1-\beta/2}$. Yet, from (17.20) one obtains from Chebychev's inequality that

$$\begin{aligned}
\mathbb{P}\Big(T(\mathbb{X}_n) \leq q^{\alpha}_{1-\beta/2}\Big) &\leq \mathbb{P}\left(T(\mathbb{X}_n) \leq \mathbb{E}\left[T(\mathbb{X}_n)\right] - \sqrt{\frac{2}{\beta}\operatorname{Var}(T(\mathbb{X}_n))}\right)\\
&\leq \mathbb{P}\left(\left|T(\mathbb{X}_n) - \mathbb{E}\left[T(\mathbb{X}_n)\right]\right| \geq \sqrt{\frac{2}{\beta}\operatorname{Var}(T(\mathbb{X}_n))}\right)\\
&\leq \frac{\beta}{2}. \qquad (17.53)
\end{aligned}$$

Finally, both (17.52) and (17.53) lead to the desired control $\mathbb{P}(\Delta_\alpha(\mathbb{X}_n) = 0) \leq \beta$ which ends the proof.

### *Control of the Variance in the General Case*

To upper bound the variance term, we apply Lemma 17.3.1 which directly implies that

$$\operatorname{Var}(T(\mathbb{X}_n)) \leq \frac{2}{n}\left(\mathbb{E}_P\left[\varphi^2\right] + 4\mathbb{E}_{\perp\!\!\!\perp}\left[\varphi^2\right]\right),$$

which directly leads to (17.48).

### *Control of the Critical Value Based on Hoeffding's Approach*

This section is devoted to the proof the inequality (17.49), namely

$$q_{1-\alpha}(\mathbb{X}_n) \leq \sqrt{\frac{1-\alpha}{\alpha}\operatorname{Var}\big(T\left(\mathbb{X}_n^{\Pi}\right)\big|\mathbb{X}_n\big)}.$$

The proof of this upper-bound follows Hoeffding's approach in [16], and relies on a normalizing trick, and the Cauchy-Schwarz inequality. From now on, for a better readability, denote respectively $\mathbb{E}^*[\cdot]$ and $\operatorname{Var}^*(\cdot)$ the conditional expectation and variance given the sample $\mathbb{X}_n$.

As in Hoeffding [16], the first step is to center and normalize the permuted test statistic. Yet, by construction the permuted test statistic is automatically centered, that is $\mathbb{E}^*\big[T\left(\mathbb{X}_n^{\Pi}\right)\big] = 0$, as one can notice that

$$T\left(\mathbb{X}_n^{\Pi}\right) = \frac{1}{n-1}\left(\sum_{i=1}^n \varphi\left(X_i^1, X_{\Pi(i)}^2\right) - \mathbb{E}^*\left[\sum_{i=1}^n \varphi\left(X_i^1, X_{\Pi(i)}^2\right)\right]\right).$$

Therefore, just consider the normalizing term

$$v(\mathbb{X}_n) = \operatorname{Var}^*\big(T\left(\mathbb{X}_n^{\Pi}\right)\big) = \mathbb{E}^*\left[T\left(\mathbb{X}_n^{\Pi}\right)^2\right] = \frac{1}{n!}\sum_{\tau\in\mathfrak{S}_n}\left(T\left(\mathbb{X}_n^{\tau}\right)\right)^2.$$

Two cases appear: either $v(\mathbb{X}_n) = 0$ or not.

In the first case, the nullity of the conditional variance implies that all the permutations of the test statistic are equal. Hence, for all permutation $\tau$ of $\{1,\ldots,n\}$, one has $T(\mathbb{X}_n^{\tau}) = T(\mathbb{X}_n)$. Since the centering term $\mathbb{E}^*\left[\sum_{i=1}^n \varphi\left(X_i^1, X_{\Pi(i)}^2\right)\right] = n^{-1}\sum_{i,j=1}^n \varphi(X_i^1, X_j^2)$ is permutation invariant, one obtains the equality of the permuted sums, that is

$$\sum_{i=1}^n \varphi\left(X_i^1, X_{\tau(i)}^2\right) = \sum_{i=1}^n \varphi\left(X_i^1, X_i^2\right),$$

and this for all permutation $\tau$. In particular, the centering term is also equal to $\sum_{i=1}^n \varphi\left(X_i^1, X_i^2\right)$. Indeed, by invariance of the sum (applied in the third equality below),

$$\begin{aligned}\frac{1}{n}\sum_{i,j=1}^n \varphi\left(X_i^1, X_j^2\right) &= \frac{1}{n}\sum_{i,j=1}^n \varphi\left(X_i^1, X_j^2\right)\left[\frac{1}{(n-1)!}\sum_{\tau\in\mathfrak{S}_n}\mathbb{1}_{\tau(i)=j}\right]\\ &= \frac{1}{n!}\sum_{\tau\in\mathfrak{S}_n}\sum_{i=1}^n \varphi\left(X_i^1, X_{\tau(i)}^2\right)\left[\sum_{j=1}^n \mathbb{1}_{\tau(i)=j}\right]\end{aligned}$$

$$\frac{1}{n}\sum_{i,j=1}^{n}\varphi\left(X_i^1, X_j^2\right) = \frac{1}{n!}\sum_{\tau\in\mathfrak{S}_n}\left(\sum_{i=1}^{n}\varphi\left(X_i^1, X_i^2\right)\right)$$
$$= \sum_{i=1}^{n}\varphi\left(X_i^1, X_i^2\right).$$

Therefore, $T(\mathbb{X}_n)$ is equal to zero, and thus, so is $q_{1-\alpha}(\mathbb{X}_n)$. Finally, inequality (17.55) is satisfied since

$$q_{1-\alpha}(\mathbb{X}_n) = 0 \leq 0 = \sqrt{\frac{1-\alpha}{\alpha}\operatorname{Var}\big(T\left(\mathbb{X}_n^{\Pi}\right)\big|\mathbb{X}_n\big)}.$$

Consider now the second case, and assume $\nu(\mathbb{X}_n) > 0$. Let us introduce the (centered and) normalized statistic

$$T'(\mathbb{X}_n) = \frac{1}{\sqrt{\nu(\mathbb{X}_n)}}\left(T(\mathbb{X}_n)\right).$$

In particular, the new statistic $T'(\mathbb{X}_n)$ satisfies

$$\mathbb{E}^*\big[T'\big(\mathbb{X}_n^{\Pi}\big)\big] = 0 \quad \text{and} \quad \operatorname{Var}^*\big(T'\big(\mathbb{X}_n^{\Pi}\big)\big) \leq 1.$$

One may moreover notice that the normalizing term $\nu(\mathbb{X}_n)$ is permutation invariant, that is, for all permutations $\tau$ and $\tau'$ in $\mathfrak{S}_n$,

$$\nu\big(\mathbb{X}_n^{\tau}\big) = \nu(\mathbb{X}_n) = \nu\left(\mathbb{X}_n^{\tau'}\right).$$

In particular, since $\nu(\mathbb{X}_n) > 0$,

$$T\left(\mathbb{X}_n^{\tau}\right) \leq T\left(\mathbb{X}_n^{\tau'}\right) \quad \Leftrightarrow \quad T'\big(\mathbb{X}_n^{\tau}\big) \leq T'\left(\mathbb{X}_n^{\tau'}\right).$$

Therefore, as the test $\Delta_\alpha$ depends only on the comparison of the $\left\{T\left(\mathbb{X}_n^{\tau}\right)\right\}_{\tau\in\mathfrak{S}_n}$, the test statistic $T$ can be replaced by $T'$, and the new critical value becomes

$$q'_{1-\alpha}(\mathbb{X}_n) = T'^{(n!-\lfloor n!\alpha\rfloor)}(\mathbb{X}_n) = \frac{T^{(n!-\lfloor n!\alpha\rfloor)}(\mathbb{X}_n)}{\nu(\mathbb{X}_n)} = \frac{q_{1-\alpha}(\mathbb{X}_n)}{\nu(\mathbb{X}_n)}. \tag{17.54}$$

Moreover, following the proof of Theorem 2.1. of Hoeffding [16], one can show (as below) that

$$q'_{1-\alpha}(\mathbb{X}_n) \leq \sqrt{\frac{1-\alpha}{\alpha}}. \tag{17.55}$$

Hence, combining (17.55) with (17.54) leads straightforwardly to (17.49).

Finally, remains the proof of (17.55). There are two cases:

**First Case** If $q'_{1-\alpha}(\mathbb{X}_n) \leq 0$, then (17.55) is satisfied.

**Second Case** If $q'_{1-\alpha}(\mathbb{X}_n) > 0$, then introduce $Y = q'_{1-\alpha}(\mathbb{X}_n) - T'\big(\mathbb{X}_n^{\Pi}\big)$.

First, since by construction, $\mathbb{E}^*\big[T'\big(\mathbb{X}_n^{\Pi}\big)\big] = 0$, one directly obtains $\mathbb{E}^*[Y] = q'_{1-\alpha}(\mathbb{X}_n)$. Hence,

$$0 < q'_{1-\alpha}(\mathbb{X}_n) = \mathbb{E}^*[Y] \leq \mathbb{E}^*[Y \mathbb{1}_{Y>0}],$$

and by the Cauchy-Schwarz inequality,

$$\big(q'_{1-\alpha}(\mathbb{X}_n)\big)^2 \quad \leq \quad \big(\mathbb{E}^*[Y\mathbb{1}_{Y>0}]\big)^2 \quad \leq \quad \mathbb{E}^*\Big[Y^2\Big]\,\mathbb{E}^*[\mathbb{1}_{Y>0}],$$

Yet, on one hand,

$$\begin{aligned}
\mathbb{E}^*\Big[Y^2\Big] &= \mathbb{E}^*\Big[\big(q'_{1-\alpha}(\mathbb{X}_n) - T'\big(\mathbb{X}_n^{\Pi}\big)\big)^2\Big] \\
&= \big(q'_{1-\alpha}(\mathbb{X}_n)\big)^2 + \mathbb{E}^*\Big[\big(T'\big(\mathbb{X}_n^{\Pi}\big)\big)^2\Big] - 2q'_{1-\alpha}(\mathbb{X}_n)\mathbb{E}^*\big[T'\big(\mathbb{X}_n^{\Pi}\big)\big] \\
&= \big(q'_{1-\alpha}(\mathbb{X}_n)\big)^2 + \mathrm{Var}^*\big(T'\big(\mathbb{X}_n^{\Pi}\big)\big) \\
&\leq \big(q'_{1-\alpha}(\mathbb{X}_n)\big)^2 + 1,
\end{aligned}$$

since by the normalizing initial step, $\mathrm{Var}^*\big(T'\big(\mathbb{X}_n^{\Pi}\big)\big) \leq 1$.

And, on the other hand,

$$\begin{aligned}
\mathbb{E}^*[\mathbb{1}_{Y>0}] &= \mathbb{E}^*\Big[\mathbb{1}_{T'(')(\mathbb{X}_n^{\Pi}) < q'_{1-\alpha}(\mathbb{X}_n)}\Big] \\
&= \frac{\#\big\{\tau \in \mathfrak{S}_n \;;\; T'\big(\mathbb{X}_n^{\tau}\big) < T'^{(n! - \lfloor n!\alpha \rfloor)}(\mathbb{X}_n)\big\}}{n!} \\
&\leq \frac{(n! - \lfloor n!\alpha \rfloor) - 1}{n!} \;=\; 1 - \frac{\lfloor n!\alpha \rfloor + 1}{n!} \\
&< 1 - \frac{n!\alpha}{n!} \;=\; 1 - \alpha.
\end{aligned}$$

So finally,

$$\big(q'_{1-\alpha}(\mathbb{X}_n)\big)^2 \leq (1-\alpha)\left(\big(q'_{1-\alpha}(\mathbb{X}_n)\big)^2 + 1\right),$$

which is equivalent to $\big(q'_{1-\alpha}(\mathbb{X}_n)\big)^2 \leq (1-\alpha)/\alpha$, and thus ends the proof of (17.55).

### *Control of the Quantile of the Critical Value*

The control of the conditional quantile allows us to upper bound its own quantile $q^{\alpha}_{1-\beta/2}$ as stated in (17.50), that is

$$q^{\alpha}_{1-\beta/2} \le 2\sqrt{\frac{1-\alpha}{\alpha}}\sqrt{\frac{2}{\beta}\frac{\left(\mathbb{E}_{\perp\!\!\!\perp}\left[\varphi^2\right] + \mathbb{E}_P\left[\varphi^2\right]\right)}{n}}.$$

Indeed, (17.49) ensures that

$$q_{1-\alpha}(\mathbb{X}_n) \le \sqrt{\frac{1-\alpha}{\alpha}}\sqrt{\mathbb{E}\left[T\left(\mathbb{X}_n^{\Pi}\right)^2 \middle| \mathbb{X}_n\right]},$$

and in particular, the $(1-\beta/2)$-quantile of $q_{1-\alpha}(\mathbb{X}_n)$ satisfies

$$q^{\alpha}_{1-\beta/2} \le \sqrt{\frac{1-\alpha}{\alpha}}\sqrt{\zeta_{1-\beta/2}}, \tag{17.56}$$

where $\zeta_{1-\beta/2}$ is the $(1-\beta/2)$-quantile of $\mathbb{E}\left[T\left(\mathbb{X}_n^{\Pi}\right)^2 \middle| \mathbb{X}_n\right]$. Yet, from Markov's inequality, for all positive $x$,

$$\mathbb{P}\left(\mathbb{E}\left[T\left(\mathbb{X}_n^{\Pi}\right)^2 \middle| \mathbb{X}_n\right] \ge x\right) \le \frac{\mathbb{E}\left[T\left(\mathbb{X}_n^{\Pi}\right)^2\right]}{x}.$$

In particular, the choice of $x = 2\mathbb{E}\left[T\left(\mathbb{X}_n^{\Pi}\right)^2\right]/\beta$ leads to the control of the quantile

$$\zeta_{1-\beta/2} \le \frac{2\mathbb{E}\left[T\left(\mathbb{X}_n^{\Pi}\right)^2\right]}{\beta}. \tag{17.57}$$

Moreover, noticing that one can write

$$T\left(\mathbb{X}_n^{\Pi}\right) = \frac{1}{n-1}\sum_{i,j=1}^{n}\left(\mathbb{1}_{\Pi(i)=j} - \frac{1}{n}\right)\varphi(X_i^1, X_j^2),$$

the second-order moment in (17.57) can be rewritten

$$\begin{aligned}\mathbb{E}\left[T\left(\mathbb{X}_n^{\Pi}\right)^2\right] &= \frac{1}{(n-1)^2}\mathbb{E}\left[\left(\sum_{i,j=1}^{n}\left(\mathbb{1}_{\Pi(i)=j} - \frac{1}{n}\right)\varphi(X_i^1, X_j^2)\right)^2\right] \\ &= \frac{1}{(n-1)^2}\sum_{i,j=1}^{n}\sum_{k,l=1}^{n} E_{i,j,k,l} \times \mathbb{E}\left[\varphi(X_i^1, X_j^2)\varphi(X_k^1, X_l^2)\right],\end{aligned}$$

by independence between $\Pi$ and $\mathbb{X}_n$, where

$$E_{i,j,k,l} = \mathbb{E}\left[\left(\mathbb{1}_{\Pi(i)=j} - \frac{1}{n}\right)\left(\mathbb{1}_{\Pi(k)=l} - \frac{1}{n}\right)\right] = \mathbb{E}\left[\mathbb{1}_{\Pi(i)=j}\mathbb{1}_{\Pi(k)=l}\right] - \frac{1}{n^2}.$$

On the one hand, for all $1 \leq i, j, k, l \leq n$, the Cauchy-Schwarz inequality always ensures

$$\mathbb{E}\left[\varphi(X_i^1, X_j^2)\varphi(X_k^1, X_l^2)\right] \leq \sqrt{\mathbb{E}\left[\varphi^2(X_i^1, X_j^2)\right]\mathbb{E}\left[\varphi^2(X_k^1, X_l^2)\right]} \leq \mathbb{E}_{\perp\!\!\perp}\left[\varphi^2\right] + \mathbb{E}_P\left[\varphi^2\right], \tag{17.58}$$

since for all $1 \leq i, j \leq n$, $\mathbb{E}\left[\varphi^2(X_i^1, X_j^2)\right] \leq \mathbb{E}_{\perp\!\!\perp}\left[\varphi^2\right] + \mathbb{E}_P\left[\varphi^2\right]$.

On the other hand, remains to control the sum $(n-1)^{-2}\sum_{i,j=1}^n \sum_{k,l=1}^n E_{i,j,k,l}$. Three cases appear.

**First Case** If $i \neq k$ and $j \neq l$ (occurring $[n(n-1)]^2$ times), then

$$E_{i,j,k,l} = \frac{1}{n(n-1)} - \frac{1}{n^2} = \frac{1}{n^2(n-1)}.$$

**Second Case** If $[i \neq k$ and $j = l]$ or $[i = k$ and $j \neq l]$, then

$$E_{i,j,k,l} = 0 - 1/n^2 \leq 0.$$

**Third Case** If $i = k$ and $j = l$ (occurring $n(n-1)$ times), then

$$E_{i,j,k,l} = \frac{1}{n} - \frac{1}{n^2} = \frac{n-1}{n^2} \leq \frac{1}{n}.$$

Therefore,

$$\begin{aligned}\frac{1}{(n-1)^2}\sum_{i,j=1}^n \sum_{k,l=1}^n E_{i,j,k,l} &\leq \frac{1}{(n-1)^2}\left([n(n-1)]^2 \times \frac{1}{n^2(n-1)} + n(n-1) \times \frac{1}{n}\right) \\ &\leq \frac{2}{n-1} \\ &\leq \frac{4}{n}.\end{aligned} \tag{17.59}$$

Finally, both (17.58) and (17.59) imply that

$$\mathbb{E}\left[T\left(\mathbb{X}_n^\Pi\right)^2\right] \leq \frac{4}{n}\left(\mathbb{E}_{\perp\!\!\perp}\left[\varphi^2\right] + \mathbb{E}_P\left[\varphi^2\right]\right), \tag{17.60}$$

Therefore, combining (17.56), (17.57) and (17.60) ends the proof of (17.50).

### A First Condition Ensuing from Hoeffding's Approach

Back to the condition (17.20) derived from Chebychev's inequality, both (17.48) and (17.50) imply that

$$q^{\alpha}_{1-\beta/2} + \sqrt{\frac{2}{\beta}\operatorname{Var}(T(\mathbb{X}_n))} \leq \sqrt{\frac{2}{\beta}\frac{\left(\mathbb{E}_P\left[\varphi^2\right] + \mathbb{E}_{\perp\!\!\!\perp}\left[\varphi^2\right]\right)}{n}}\left(2\sqrt{\frac{1-\alpha}{\alpha}} + \sqrt{8}\right),$$

with $2\sqrt{(1-\alpha)/\alpha} + \sqrt{8} \leq 4/\sqrt{\alpha}$, since $\sqrt{1-\alpha} + \sqrt{\alpha} \leq \sqrt{2}$. Finally, the right-hand side of condition (17.20) being upper bounded by

$$\frac{4}{\sqrt{\alpha}}\sqrt{\frac{2}{\beta}\frac{\left(\mathbb{E}_P\left[\varphi^2\right] + \mathbb{E}_{\perp\!\!\!\perp}\left[\varphi^2\right]\right)}{n}},$$

which is exactly the right-hand side of (17.21), this ensures the sufficiency of condition 17.21 to control the second kind error rate by $\beta$.

## References


1. M. Albert, Tests of independence by bootstrap and permutation: an asymptotic and non-asymptotic study. Application to Neurosciences. PhD thesis, Nice, 2015
2. M. Albert, Y. Bouret, M. Fromont, P. Reynaud-Bouret, Bootstrap and permutation tests of independence for point processes. Ann. Stat. **43**(6), 2537–2564 (2015)
3. M. Albert, Y. Bouret, M. Fromont, P. Reynaud-Bouret, Surrogate data methods based on a shuffling of the trials for synchrony detection: the centering issue. Neural Comput. **28**(11), 2352–2392 (2016)
4. R. Adamczak, D. Chafaï, P. Wolff, Circular law for random matrices with exchangeable entries (2014), arXiv preprint arXiv:1402.3660
5. Y. Baraud, Non-asymptotic minimax rates of testing in signal detection. Bernoulli **8**(5), 577–606 (2002)
6. B. Bercu, B. Delyon, E. Rio, *Concentration Inequalities for Sums and Martingales* (Springer, Cham, 2015)
7. E. Bolthausen, An estimate of the remainder in a combinatorial central limit theorem. Probab. Theory Relat. Fields **66**(3), 379–386 (1984)
8. S. Boucheron, G. Lugosi, P. Massart, *Concentration Inequalities: A Nonasymptotic Theory of Independence* (Oxford University Press, Oxford, 2013)
9. S. Chatterjee, Stein's method for concentration inequalities. Probab. Theory Relat. Fields **138**(1), 305–321 (2007)
10. M. Dwass, On the asymptotic normality of some statistics used in non-parametric tests. Ann. Math. Stat. **26**, 334–339 (1955)
11. M. Fromont, B. Laurent, P. Reynaud-Bouret, Adaptive tests of homogeneity for a Poisson process. Ann. Inst. Henri Poincaré Probab. Stat. **47**(1), 176–213 (2011)
12. M. Fromont, B. Laurent, P. Reynaud-Bouret, The two-sample problem for poisson processes: adaptive tests with a nonasymptotic wild bootstrap approach. Ann. Stat. **41**(3), 1431–1461 (2013)

13. J. Hájek, Some extensions of the Wald-Wolfowitz-Noether theorem. Ann. Math. Stat. **32**(2), 506–523 (1961)
14. S.T. Ho, L.H.Y. Chen, An $L_p$ bound for the remainder in a combinatorial central limit theorem. Ann. Probab. **6**(2), 231–249 (1978)
15. W. Hoeffding, A combinatorial central limit theorem. Ann. Math. Stat. **22**(4), 558–566 (1951)
16. W. Hoeffding, The large-sample power of tests based on permutation of the observations. Ann. Math. Stat. **23**(2), 169–192 (1952)
17. K. Jogdeo, Asymptotic normality in nonparametric methods. Ann. Math. Stat. **39**, 905–922 (1968)
18. M. Ledoux, *The Concentration of Measure Phenomenon*, vol. 89 (American Mathematical Society, Providence, 2005)
19. P. Massart, J. Picard, *Concentration Inequalities and Model Selection*, vol. 1896 (Springer, Berlin, 2007)
20. C. McDiarmid, Concentration for independent permutations. Comb. Probab. Comput. **11**(02), 163–178 (2002)
21. M. Motoo, On the Hoeffding's combinatrial central limit theorem. Ann. Inst. Stat. Math. **8**(1), 145–154 (1956)
22. G.E. Noether, On a theorem by Wald and Wolfowitz. Ann. Math. Stat. **20**, 455–458 (1949)
23. J.P. Romano, Bootstrap and randomization tests of some nonparametric hypotheses. Ann. Stat. **17**(1), 141–159 (1989)
24. L. Sansonnet, C. Tuleau-Malot, A model of Poissonian interactions and detection of dependence. Stat. Comput. **25**(2), 449–470 (2015)
25. W. Schneller, A short proof of Motoo's combinatorial central limit theorem using stein's method. Probab. Theory Relat. Fields **78**(2), 249–252 (1988)
26. C.P. Shapiro, L. Hubert, Asymptotic normality of permutation statistics derived from weighted sums of bivariate functions. Ann. Stat. **7**, 788–794 (1979)
27. M. Talagrand, Concentration of measure and isoperimetric inequalities in product spaces. Publ. Math. Inst. Hautes Etudes Sci. **81**(1), 73–205 (1995)
28. A. Wald, J. Wolfowitz, Statistical tests based on permutations of the observations. Ann. Math. Stat. **15**(4), 358–372 (1944)

# Chapter 18
# Uncertainty Quantification for Matrix Compressed Sensing and Quantum Tomography Problems

**Alexandra Carpentier, Jens Eisert, David Gross, and Richard Nickl**

**Abstract** We construct minimax optimal non-asymptotic confidence sets for low rank matrix recovery algorithms such as the Matrix Lasso or Dantzig selector. These are employed to devise adaptive sequential sampling procedures that guarantee recovery of the true matrix in Frobenius norm after a data-driven stopping time $\hat{n}$ for the number of measurements that have to be taken. With high probability, this stopping time is minimax optimal. We detail applications to quantum tomography problems where measurements arise from Pauli observables. We also give a theoretical construction of a confidence set for the density matrix of a quantum state that has optimal diameter in nuclear norm. The non-asymptotic properties of our confidence sets are further investigated in a simulation study.

**Keywords** Low rank recovery · Quantum information · Confidence sets · Sequential sampling

A. Carpentier
Institut für Mathematische Stochastik, Otto von Guericke Universität Magdeburg, Magdeburg, Germany
e-mail: alexandra.carpentier@ovgu.de

J. Eisert
Department of Physics, Dahlem Center for Complex Quantum Systems, Freie Universität zu Berlin, Berlin, Germany

D. Gross
Institute of Theoretical Physics, Universität Köln, Köln, Germany
e-mail: david.gross@thp.uni-koeln.de

R. Nickl (✉)
Statistical Laboratory, University of Cambridge, Cambridge, UK
e-mail: r.nickl@statslab.cam.ac.uk



N. Gozlan et al. (eds.), *High Dimensional Probability VIII*,
Progress in Probability 74, https://doi.org/10.1007/978-3-030-26391-1_18

## 18.1 Introduction

### *18.1.1 Uncertainty Quantification in Compressed Sensing*

Compressed sensing and related convex relaxation algorithms have had a profound impact on high-dimensional statistical modelling in recent years. They provide efficient recovery of low-dimensional objects that sit within a high-dimensional ill-posed system of linear equations. Prototypical low-dimensional structures are described by *sparsity* and *low rank* hypotheses. The statistical analysis of such algorithms has mostly been concerned with recovery rates, or with the closely related question of how many measurements are sufficient to reach a prescribed recovery level—some key references are Refs. [3, 5–7, 16, 24, 25, 33].

A statistical question of fundamental importance that has escaped a clear answer so far is the question of *uncertainty quantification*: Can we tell from the observations how well the algorithm has worked? In technical terms: can we report confidence sets for the unknown parameter? Or, in the sequential sampling setting, can we give *data-driven* rules that ensure recovery of the true parameter at a given precision? Answers to this question are of great importance in various applications of compressed sensing. For one-dimensional subproblems, such as projection onto a fixed coordinate of the parameter vector, recent advances have provided some useful confidence intervals (see Refs. [8, 9, 22, 36]), but our understanding of valid inference procedures for the *entire* parameter remains limited.

Whereas the 'estimation theory' for compressed sensing is quite similar for sparsity and low rank constraints, this is not so for the theory of confidence sets. On the one hand, if one is interested in inference on the full parameter, *sparsity* conditions induce information theoretic barriers, as shown in Ref. [30]: unless one is willing to make additional signal strength assumptions (inspired by the literature on nonparametric confidence sets, such as Refs. [13, 19]), a uniformly valid confidence set for the unknown parameter vector $\theta$ cannot have a better performance than $1/\sqrt{n}$ in quadratic loss. This significantly falls short of the optimal recovery rates $(k \log p)/n$ for the most interesting sparsity levels $k$. On the other hand, and perhaps surprisingly, we will show in this article that the *low-rank constraint* is naturally compatible with certain risk estimation approaches to confidence sets. This will be seen to be true for general sub-Gaussian sensing matrices, but also for sensing matrices arising from Pauli observables, as is specifically relevant in quantum state tomography problems (see the next section). In the latter case it will be helpful to enforce the additional 'quantum state shape constraint' on the unknown matrix to obtain optimal results. One can conclude that, in contrast to 'sparse models', *no signal strength assumptions* are necessary for the existence of adaptive confidence statements in low rank recovery problems. Our findings are confirmed in a simple simulation study, see Sect. 18.4.

The honest non-asymptotic confidence sets we will derive below can be used for the construction of *adaptive sampling procedures*: An experimenter wants to know—at least with a prescribed probability of $1 - \alpha$—that the matrix recovery algorithm, the 'estimator', has produced an output $\tilde{\theta}$ that is close to the 'true state' $\theta$. The sequential protocol advocated here—which is related to 'active learning algorithms' from machine learning, e.g., Ref. [29]—should tell the experimenter whether a new batch of measurements has to be taken to decrease the recovery error, or whether the collected observations are already sufficient. The data-driven stopping time for this protocol should not exceed the minimax optimal stopping time, again with high probability. We shall show that for Pauli and sub-Gaussian sensing ensembles, such algorithms exist under mild assumptions on the true matrix $\theta$. These assumptions are in particular always satisfied under the 'quantum shape constraint' that naturally arises in quantum tomography problems.

Our results depend on the choice of the Frobenius norm and the Hilbert space geometry induced by it. For other natural matrix norms, such as for instance the trace-(nuclear) norm, the theory is more difficult. We show as a first step that at least theoretically a trace-norm optimal confidence set can be constructed for the unknown quantum state (Theorem 18.4)—this suggests interesting directions for future research.

### *18.1.2 Application to Quantum State Estimation*

This work was partly motivated by a problem arising in present-day physics experiments that aim at estimating quantum states. Conceptually, a quantum mechanical experiment involves two stages (c.f. Fig. 18.1): A *source* (or *preparation procedure*) that emits quantum mechanical systems with unknown properties, and a *measurement device* that interacts with incoming quantum systems and produces real-valued measurement outcomes, e.g. by pointing a dial to a value on a scale. Quantum mechanics stipulates that both stages are completely described by certain matrices.

The properties of the source are represented by a positive semi-definite unit trace matrix $\theta$, the *quantum state*, also referred to as *density matrix*. In turn, the measurement device is modelled by a Hermitian matrix $X$, which is referred to as an *observable* in physics jargon. A key axiom of the quantum mechanical formalism

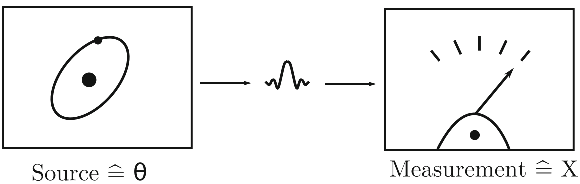


**Fig. 18.1** Caricature of a quantum mechanical experiment. With every source of quantum systems, one associates a *density matrix* $\theta$. Observations systems are performed by measurement devices, which interact with incoming systems and produce real-valued outcomes. Each such devices is modelled mathematically by a Hermitian matrix $X$, referred to as an *observable*

states that if the measurement $X$ is repeatedly performed on systems emitted by the source that is preparing $\theta$, then the real-valued measurement outcomes will fluctuate randomly with expected value

$$\langle X, \theta \rangle_F = \mathrm{tr}(X\theta) \tag{18.1}$$

(referred to as *expectation value* in the quantum physics literature). The precise way in which physical properties are represented by these matrices is immaterial to our discussion (cf. any textbook, e.g. Ref. [32]). We merely note that, while in principle *any* Hermitian $X$ can be measured by some physical apparatus, the required experimental procedures are prohibitively complicated for all but a few highly structured matrices. This motivates the introduction of *Pauli designs* below, which correspond to fairly tractable 'spin parity measurements'.

The *quantum state estimation* or *quantum state tomography*[1] problem is to estimate an unknown density matrix $\theta$ from the measurement of a collection of observables $X^1, \ldots, X^n$. This task is of particular importance to the young field of quantum information science [31]. There, the sources might be a carefully engineered component used for technological applications such as quantum key distribution or quantum computing. In this context, quantum state estimation is the process of characterising the components one has built—clearly an important capability for any technology.

A major challenge lies in the fact that relevant instances are described by $d \times d$-matrices for fairly large dimensions $d$ ranging from 100 to 10,000 in presently performed experiments [18]. Such high-dimensional estimation problems can benefit substantially from structural properties of the objects to be recovered. Fortunately, the density matrices occurring in quantum information experiments are typically well-approximated by matrices of *low rank* $r \ll d$. In fact, in the practically most important applications, one usually even aims at preparing a state of rank one—a so-called *pure quantum state*. While environmental noise will drive the actual state away from the perfect rank-one case, the error will usually be small.

As a result, quantum physicists have early on shown an interest in low-rank matrix recovery methods [12, 15–17, 28]. Initial works [15, 16] focused on the minimal number $n$ of observables $X^1, \ldots, X^n$ required for reconstructing a rank-$r$ density matrix $\theta$ in the noiseless case, i.e. under the idealised assumption that the expectation values $\mathrm{tr}(\theta X^i)$ are known exactly. The practically highly relevant problem of quantifying the uncertainty of an estimate $\hat{\theta}$ arising from noisy observations on low-rank states was addressed only later [12] and remains less well understood.

[1] The term 'tomography' goes back to the use of *Radon transforms* in early schemes for estimating quantum states of electromagnetic fields [1, 27]. It has become synonymous with 'quantum density matrix estimation', even though current methods applied to quantum systems with a finite dimension $d$ have no technical connection to classical tomographic reconstruction algorithms.

More concretely, the basic approach taken in Ref. [12] for uncertainty quantification is similar to the one pursued in the present paper. In a first step, one uses a Matrix Lasso or Dantzig Selector to construct an estimate. Then, a confidence region is obtained by comparing predictions derived from the initial estimate to new samples. However, Ref. [12] suffers from two demerits. First, and most importantly, the performance analysis of the scheme relies on a bound on the rank $r$ of the unknown true $\theta$. Such a bound is not available in practice. Second, the dependence of the rate on $r$ is not tight. Both of these demerits will be addressed here.

We close this section pointing to more broadly related works. Uncertainty quantification in quantum state tomography in general has been treated by numerous authors—a highly incomplete list is Refs. [2, 4, 10, 34, 35]. However, the concept of dimension reduction for low-rank states does not feature explicitly in these papers. This contrasts with Ref. [17], where the authors propose model selection techniques based on information criteria to arrive at low-rank estimates. The use of general-purpose methods—like maximum likelihood estimation and the Akaike Information Criterion—in Ref. [17] means that it is applicable to very general experimental designs. In contrast to this, the present paper relies on compressed sensing ideas to arrive at rigorous a priori guarantees on statistical and computational performance. Also, it remains non-obvious how such model selection steps can be transformed into 'post-model selection' confidence sets—typically such constructions result in sub-optimal signal strength conditions that ensure model selection consistency (see Ref. [26] and also the discussion after Theorem 2 in Ref. [30]). Our confidence procedures never estimate the unknown rank of the quantum state—not even implicitly. Rather, they estimate the performance of a dimension-reduction technique directly based on sample splitting.

## 18.2 Matrix Compressed Sensing

We consider inference on a $d \times d$ matrix $\theta$ that is symmetric, or, if it consists of possibly complex entries, assumed to be Hermitian (that is $\theta = \theta^*$ where $\theta^*$ is the conjugate transpose of $\theta$). Denote by $\mathbb{M}_d(\mathbb{K})$ the space of $d \times d$ matrices with entries in $\mathbb{K} = \mathbb{C}$ or $\mathbb{K} = \mathbb{R}$. We write $\|\cdot\|_F$ for the usual Frobenius norm on $\mathbb{M}_d(\mathbb{K})$ arising from the inner product $tr(AB) = \langle A, B\rangle_F$. Moreover let $\mathbb{H}_d(\mathbb{C})$ be the set of all Hermitian matrices, and $\mathbb{H}_d(\mathbb{R})$ for the set of all symmetric $d \times d$ matrices with real entries. The norm symbol $\|\cdot\|$ without subindex denotes the standard Euclidean norm on $\mathbb{R}^n$ or on $\mathbb{C}^n$ arising from the Euclidean inner product $\langle\cdot,\cdot\rangle$.

We denote the usual operator norm on $\mathbb{M}_d(\mathbb{C})$ by $\|\cdot\|_{op}$. For $M \in \mathbb{M}_d(\mathbb{C})$ let $(\lambda_k^2 : k = 1, \ldots, d)$ be the eigenvalues of $M^T M$ (which are all real-valued and positive). The $l_1$-*Schatten*, *trace*, or *nuclear* norm of $M$ is defined as

$$\|M\|_{S_1} = \sum_{j \le d} |\lambda_j|.$$

Note that for any matrix $M$ of rank $1 \leq r \leq d$ the following inequalities are easily shown,

$$\|M\|_F \leq \|M\|_{S_1} \leq \sqrt{r}\|M\|_F. \tag{18.2}$$

We will consider parameter subspaces of $\mathbb{H}_d(\mathbb{C})$ described by low rank constraints on $\theta$, and denote by $R(k)$ the space of all Hermitian $d \times d$ matrices that have rank at most $k$, $k \leq d$. In quantum tomography applications, we may assume an additional 'shape constraint', namely that $\theta$ is a density matrix of a quantum state, and hence contained in *state space*

$$\Theta_+ = \{\theta \in \mathbb{H}_d(\mathbb{C}) : \mathrm{tr}(\theta) = 1, \theta \succeq 0\},$$

where $\theta \succeq 0$ means that $\theta$ is positive semi-definite. In fact, in most situations, we will only require the bound $\|\theta\|_{S_1} \leq 1$ which trivially holds for any $\theta$ in $\Theta_+$.

We have at hand measurements arising from inner products $\langle X^i, \theta\rangle_F = \mathrm{tr}(X^i\theta)$, $i = 1, \ldots, n$, of $\theta$ with $d \times d$ (random) matrices $X^i$. This measurement process is further subject to independent additive noise $\varepsilon$. Formally, the measurement model is

$$Y_i = tr(X^i\theta) + \varepsilon_i, \quad i = 1, \ldots, n, \tag{18.3}$$

where the $\varepsilon_i$'s and $X^i$'s are independent of each other. We write $Y = (Y_1, \ldots, Y_n)^T$, and for probability statements under the law of $Y, X, \varepsilon$ given fixed $\theta$ we will use the symbol $\mathbb{P}_\theta$. Unless mentioned otherwise we will make the basic assumption of Gaussian noise

$$\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)^T \sim N(0, \sigma^2 I_n),$$

where $\sigma > 0$ is known. See Remark 18.6 for some discussion of the unknown variance case. In the context of quantum mechanics, the inner product $tr(X^i\theta)$ gives the expected value of the observable $X^i$ when measured on a system in state $\theta$ (cf. Sect. 18.1.2). A class of physically realistic measurements (correlations among spin-1/2 particles) is described by $X^i$'s drawn from the *Pauli basis*. Our main results also hold for measurement processes of this type. Before we describe this in Sect. 18.2.2, let us first discuss our assumptions on the matrices $X^i$.

### *18.2.1 Sensing Matrices and the RIP*

When $\theta \in \mathbb{M}_d(\mathbb{R})$, we shall restrict to design matrices $X^i$ that have real-valued entries, too, and when $\theta \in \mathbb{H}_d(\mathbb{C})$ we shall consider designs where $X^i \in \mathbb{H}_d(\mathbb{C})$. This way, in either case, the measurements $tr(X_i\theta)$'s and hence the $Y_i$'s are all real-valued. More concretely, the sensing matrices $X^i$ that we shall consider are described in the following assumption, which encompasses both a prototypical compressed sensing setting—where we can think of the matrices $X^i$ as i.i.d. draws from a Gaussian ensemble $(X_{m,k}) \sim^{iid} N(0, 1)$—as well as the 'random sampling from a basis of $\mathbb{M}_d(\mathbb{C})$' scenario. The systematic study of the latter has been initiated by quantum physicists [15, 28], as it contains, in particular, the case of Pauli basis measurements [12, 16] frequently employed in quantum tomography problems. Note that in Part (a) the design matrices are not Hermitian but our results can easily be generalised to symmetrised sub-Gaussian ensembles (as those considered in Ref. [24]).

**Condition 18.1**

*(a)* $\theta \in \mathbb{H}_d(\mathbb{R})$**, 'isotropic' sub-Gaussian design***: The random variables* $(X^i_{m,k})$*,* $1 \le m, k \le d, i = 1, \ldots, n$*, generating the entries of the random matrix* $X^i$ *are i.i.d. distributed across all indices* $i, m, k$ *with mean zero and unit variance. Moreover, for every* $\theta \in \mathbb{M}_d(\mathbb{R})$ *such that* $\|\theta\|_F \le 1$*, the real random variables* $Z_i = tr(X^i\theta)$ *are sub-Gaussian: for some fixed constants* $\tau_i > 0$ *independent of* $\theta$*,*

$$\mathbb{E}e^{\lambda Z_i} \le \tau_1 e^{\lambda^2 \tau_2^2} \ \forall \lambda \in \mathbb{R}.$$

*(b)* $\theta \in \mathbb{H}_d(\mathbb{C})$**, random sampling from a basis ('Pauli design')***: Let* $\{E_1, \ldots, E_{d^2}\} \subset \mathbb{H}_d(\mathbb{C})$ *be a basis of* $\mathbb{M}_d(\mathbb{C})$ *that is orthonormal for the scalar product* $\langle \cdot, \cdot \rangle_F$ *and such that the operator norms satisfy, for all* $i = 1, \ldots, d^2$*,*

$$\|E_i\|_{op} \le \frac{K}{\sqrt{d}},$$

*for some universal 'coherence' constant* $K$*. [In the Pauli basis case we have* $K = 1$*.] Assume the* $X^i$*,* $i = 1, \ldots, n$*, are draws from the finite family* $\mathcal{E} = \{dE_i : i = 1, \ldots, d^2\}$ *sampled uniformly at random.*

The above examples all obey the *matrix restricted isometry property*, that we describe now. Note first that if $\mathcal{X} : \mathbb{R}^{d\times d} \to \mathbb{R}^n$ is the linear 'sampling' operator

$$\mathcal{X} : \theta \mapsto \mathcal{X}\theta = (tr(X^1\theta), \ldots, \mathrm{tr}(X^n\theta))^T, \tag{18.4}$$

so that we can write the model equation (18.3) as $Y = \mathcal{X}\theta + \varepsilon$, then in the above examples we have the 'expected isometry'

$$\mathbb{E}\frac{1}{n}\|\mathcal{X}\theta\|^2 = \|\theta\|_F^2.$$

Indeed, in the isotropic design case we have

$$\frac{1}{n}\mathbb{E}\|\mathcal{X}\theta\|^2 = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left(\sum_m\sum_k X^i_{m,k}\theta_{m,k}\right)^2 = \sum_m\sum_k \mathbb{E}X^2_{m,k}\theta^2_{m,k} = \|\theta\|_F^2, \tag{18.5}$$

and in the 'basis case' we have, from Parseval's identity and since the $X^i$'s are sampled uniformly at random from the basis,

$$\frac{1}{n}\mathbb{E}\|\mathcal{X}\theta\|^2 = \frac{d^2}{n}\sum_{i=1}^{n}\sum_{j=1}^{d^2}\Pr(X^i = E_j)|\langle E_j, \theta\rangle_F|^2 = \|\theta\|_F^2. \tag{18.6}$$

The restricted isometry property (RIP) then requires that this 'expected isometry' actually holds, up to constants and with probability $\geq 1-\delta$, for a given realisation of the sampling operator, and for all $d \times d$ matrices $\theta$ of rank at most $k$:

$$\sup_{\theta\in R(k)}\left|\frac{\frac{1}{n}\|\mathcal{X}\theta\|^2 - \|\theta\|_F^2}{\|\theta\|_F^2}\right| \leq \tau_n(k), \tag{18.7}$$

where $\tau_n(k)$ are some constants that may depend, among other things, on the rank $k$ and the 'exceptional probability' $\delta$. For the above examples of isotropic and Pauli basis design inequality (18.7) can be shown to hold with

$$\tau_n^2(k) = c^2\frac{kd\cdot\overline{\log}d}{n}, \tag{18.8}$$

where

$$\overline{\log}x := (\log x)^\eta,$$

for some $\eta > 0$ denotes a 'polylog function', and where $c = c(\delta)$ is a constant. See Refs. [6, 28] for these results, where it is also shown that $c(\delta)$ can be taken to be at least $O(1/\delta^2)$ as $\delta \to 0$ (sufficient for our purposes below).

### *18.2.2 Quantum Measurements*

Here, we introduce a paradigmatic set of quantum measurements that is frequently used in both theoretical and practical treatments of quantum state estimation (e.g. [16, 18]). For a more general account, we refer to standard textbooks [20, 31]. The purpose of this section is to motivate the 'Pauli design' case (Condition 18.1(b) of the main theorem, as well as the approximate Gaussian noise model. Beyond this, the technical details presented here will not be used.

#### 18.2.2.1 Pauli Spin Measurements on Multiple Particles

We start by describing 'spin measurements' on a single 'spin-1/2 particle'. Such a measurement corresponds to the situation of having $d = 2$. Without worrying about the physical significance, we accept as fact that on such particles, one may measure one of three properties, referred to as the 'spin along the $x$, $y$, or $z$-axis' of $\mathbb{R}^3$. Each of these measurements may yield one of two outcomes, denoted by $+1$ and $-1$ respectively.

The mathematical description of these measurements is derived from the *Pauli matrices*

$$\sigma^1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \sigma^2 = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \sigma^3 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \tag{18.9}$$

in the following way. Recall that the Pauli matrices have eigenvalues $\pm 1$. For $x \in \{1, 2, 3\}$ and $j \in \{+1, -1\}$, we write $\psi_j^x$ for the normalised eigenvector of $\sigma^x$ with eigenvalue $j$. The spectral decomposition of each Pauli spin matrix can hence be expressed as

$$\sigma^x = \pi_+^x - \pi_-^x, \tag{18.10}$$

with

$$\pi_\pm^x = \psi_\pm^x (\psi_\pm^y)^* \tag{18.11}$$

denoting the projectors onto the eigenspaces. Now, a physical measurement of the 'spin along direction $x$' on a system in state $\theta$ will give rise to a $\{-1, 1\}$-valued random variable $C^x$ with

$$\mathbb{P}(C^x = j) = \operatorname{tr}\left(\pi_j^x \theta\right), \tag{18.12}$$

where $\theta \in \mathbb{H}_2(\mathbb{C})$. Using Eq. (18.10), this is equivalent to stating that the expected value of $C^x$ is given by

$$\mathbb{E}(C^x) = \operatorname{tr}\left(\sigma^x \theta\right). \tag{18.13}$$

Next, we consider the case of joint spin measurements on a collection of $N$ particles. For each, one has to decide on an axis for the spin measurement. Thus, the joint *measurement setting* is now described by a word $x = (x_1, \ldots, x_N) \in \{1, 2, 3\}^N$. The axioms of quantum mechanics posit that the joint state $\theta$ of the $N$ particles acts on the tensor product space $(\mathbb{C}^2)^{\otimes N}$, so that $\theta \in \mathbb{H}_{2^N}(\mathbb{C})$.

Likewise, the *measurement outcome* is a word $j = (j_1, \ldots, j_N) \in \{1, -1\}^N$, with $j_i$ the value of the spin along axis $x_i$ of particle $i = 1, \ldots, N$. As above, this prescription gives rise to a $\{1, -1\}^N$-valued random variable $C^x$. Again, the axioms of quantum mechanics imply that the distribution of $C^x$ is given by

$$\mathbb{P}(C^x = j) = \mathrm{tr}\left((\pi^{x_1}_{j_1} \otimes \cdots \otimes \pi^{x_N}_{j_N})\theta\right). \tag{18.14}$$

Note that the components of the random vector $C^x$ are not necessarily independent, as $\theta$ will generally not factorise

It is often convenient to express the information in Eq. (18.14) in a way that involves tensor products of Pauli matrices, rather than their spectral projections. In other words, we seek a generalisation of Eq. (18.13) to $N$ particles. As a first step toward this goal, let

$$\chi(j) = \begin{cases} -1 & \text{number of } -1 \text{ elements in } j \text{ is odd} \\ 1 & \text{number of } -1 \text{ elements in } j \text{ is even} \end{cases} \tag{18.15}$$

be the *parity function*. Then one easily verifies

$$tr((\sigma^{x_1} \otimes \cdots \otimes \sigma^{x_N})\theta) = \sum_{j \in \{1,-1\}^N} \chi(j)\, tr\left(\theta(\pi^{x_1}_{j_1} \otimes \cdots \otimes \pi^{x_N}_{j_N})\right) = \mathbb{E}\big(\chi(C^x)\big). \tag{18.16}$$

In this sense, the tensor product $\sigma^{x_1} \otimes \cdots \otimes \sigma^{x_N}$ describes a measurement of the parity of the spins along the respective directions given by $x$.

In fact, the entire distribution of $C^x$ can be expressed in terms of tensor products of Pauli matrices and suitable parity functions. To this end, we extend the definitions above. Write

$$\sigma^0 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{18.17}$$

for the identity matrix in $\mathbb{M}_2(\mathbb{C})$. For every subset $S$ of $\{1, \ldots, N\}$, define the 'parity function restricted to $S$' via

$$\chi_S(j) = \begin{cases} -1 & \text{number of } -1 \text{ elements } j_i \text{ for } i \in S \text{ is odd} \\ 1 & \text{number of } -1 \text{ elements } j_i \text{ for } i \in S \text{ is even.} \end{cases} \tag{18.18}$$

Lastly, for $S \subset \{1, \ldots, N\}$ and $x \in \{1, 2, 3\}^N$, the *restriction of $x$ to $S$* is

$$x^S_i = \begin{cases} x_i & i \in S \\ 0 & i \notin S. \end{cases} \tag{18.19}$$

Then for every such $x$, $S$ one verifies the identity

$$tr((\sigma^{x_1^S} \otimes \cdots \otimes \sigma^{x_N^S})\theta) = \mathbb{E}\big(\chi_S(C^x)\big). \tag{18.20}$$

In other words, the distribution of $C^x$ contains enough information to compute the expectation value of all observables $(\sigma^{x_1^S} \otimes \cdots \otimes \sigma^{x_N^S})$ that can be obtained by replacing the Pauli matrices on an arbitrary subset $S$ of particles by the identity $\sigma^0$. The converse is also true: the set of all such expectation values allows one to recover the distribution of $C^x$. The explicit formula reads

$$\mathbb{P}(C^x = j) = \frac{1}{2^N} \sum_{S \subset \{1,\ldots,N\}} \chi_S(j)\, \mathbb{E}\big(\chi_S(C^x)\big) = \frac{1}{2^N} \sum_{S \in \{1,\ldots,N\}} \chi_S(j)\, tr\big(\theta(\sigma^{x_1^S} \otimes \cdots \otimes \sigma^{x_N^S})\big) \tag{18.21}$$

and can be verified by direct computation.[2]

In this sense, the information obtainable from joint spin measurements on $N$ particles can be encoded in the $4^N$ real numbers

$$2^{-N/2}\, tr((\sigma^{y_1} \otimes \cdots \otimes \sigma^{y_N})\theta), \qquad y \in \{0, 1, 2, 3\}^N. \tag{18.22}$$

Indeed, every such $y$ arises as $y = x^S$ for some (generally non-unique) combination of $x$ and $S$. This representation is particularly convenient from a mathematical point of view, as the collection of matrices

$$E^y := 2^{-N/2}\sigma^{y_1} \otimes \cdots \otimes \sigma^{y_N}, \qquad y \in \{0, 1, 2, 3\}^N \tag{18.23}$$

forms an ortho-normal basis with respect to the $\langle\cdot,\cdot\rangle_F$ inner product. Thus the terms in Eq. (18.22) are just the coefficients of a basis expansion of the density matrix $\theta$.[3]

From now on, we will use Eq. (18.22) as our model for quantum tomographic measurements. Note that the $E^y$ satisfy Condition 18.1(b) with coherence constant $K = 1$ and $d = 2^N$.

[2] A more insightful way of proving the first identity is to realise that $\mathbb{E}\big(\chi_S(C^x)\big)$ is effectively a Fourier coefficient (over the group $\mathbb{Z}_2^N$) of the distribution function of the $\{-1, 1\}^N$-valued random variable $C^x$ (e.g., [11]). Equation (18.21) is then nothing but an inverse Fourier transform.

[3] We note that quantum mechanics allows to design measurement devices that directly probe the observable of $\sigma^{y_1} \otimes \cdots \otimes \sigma^{y_N}$, without first measuring the spin of every particle and then computing a parity function. In fact, the ability to perform such correlation measurements is crucial for *quantum error correction protocols* [31]. For practical reasons these setups are used less commonly in tomography experiments, though.

#### 18.2.2.2 Bernoulli Errors and Pauli Observables

In the model (18.3) under Condition 18.1(b) we wish to approximate $d \cdot \mathrm{tr}(E^y\theta)$ for a fixed observable $E^y$ (we fix the random values of the $X^i$'s here) and for $d = 2^N$. If $y = x^S$ for some setting $x$ and subset $S$, then the parity function $B^y := \chi_S(C^x)$ has expected value $2^{N/2} \cdot \mathrm{tr}(E^y\theta) = \sqrt{d} \cdot tr(E^y\theta)$ (see Eqs. (18.20) and (18.23)), and itself is a Bernoulli variable taking values $\{1, -1\}$ with

$$p = \mathbb{P}(B^y = 1) = \frac{1 + \sqrt{d}\mathrm{tr}(E^y\theta)}{2}.$$

Note that

$$\sqrt{d}|tr(E^y\theta)| \leq \sqrt{d}\|E^y\|_{op}\|\theta\|_{S_1} \leq 1,$$

so indeed $p \in [0, 1]$ and the variance satisfies

$$\mathrm{Var}B^y = 1 - d \cdot \mathrm{tr}(E^y\theta)^2 \leq 1.$$

This is the error model considered in Ref. [12].

In order to estimate all $Y_i$, $i = 1, \ldots, n$, for given $E_i := E^y$, a total number $nT$ of identical preparations of the quantum state $\theta$ are being performed, divided into batches of $T$ Bernoulli variables $B_{i,j} := B^y_j$, $j = 1, \ldots, T$. The measurements of the sampling model Eq. (18.3) are thus

$$Y_i = \frac{\sqrt{d}}{T}\sum_{j=1}^{T} B_{i,j} = d \cdot \mathrm{tr}(E_i\theta) + \varepsilon_i$$

where

$$\varepsilon_i = \frac{\sqrt{d}}{T}\sum_{j=1}^{T}(B_{i,j} - \mathbb{E}B_{i,j})$$

is the effective error arising from the measurement procedure making use of $T$ preparations to estimate each quantum mechanical expectation value. Now note that

$$|\varepsilon_i| \leq 2\sqrt{d},\ \mathbb{E}\varepsilon_i^2 \leq \frac{d}{T}\mathrm{Var}(B_{i,1}) \leq \frac{d}{T}. \tag{18.24}$$

We see that since the $\varepsilon_i$'s are themselves sums of independent random variables, an approximate Gaussian error model with variance $\sigma^2$ will be roughly appropriate. If $T \geq n$ then $\sigma^2 = \mathbb{E}\varepsilon_1^2$ is no greater than $d/n$, and if in addition $T \geq d^2$ then all results in Sect. 18.3 below can be proved for this Bernoulli noise model too, see Remarks 18.5 and 18.6 for details.

### 18.2.3 Minimax Estimation Under the RIP

Assuming the matrix RIP to hold and Gaussian noise $\varepsilon$, one can show that the minimax risk for recovering a Hermitian rank $k$ matrix is

$$\inf_{\hat{\theta}} \sup_{\theta \in R(k)} \mathbb{E}_\theta \|\hat{\theta} - \theta\|_F^2 \simeq \sigma^2 \frac{dk}{n}, \tag{18.25}$$

where $\simeq$ denotes two-sided inequality up to universal constants.

For the upper bound one can use the nuclear norm minimisation procedure or matrix Dantzig selector from Candès and Plan [6], and needs $n$ to be large enough so that the matrix RIP holds with $\tau_n(k) < c_0$ where $c_0$ is a small enough numerical constant. Such an estimator $\tilde{\theta}$ then satisfies, for every $\theta \in R(k)$ and those $n \in \mathbb{N}$ for which $\tau_n(k) < c_0$,

$$\|\tilde{\theta} - \theta\|_F^2 \leq D(\delta)\sigma^2 \frac{kd}{n}, \tag{18.26}$$

with probability greater than $1 - 2\delta$, and with the constant $D(\delta)$ depending on $\delta$ and also on $c_0$ (suppressed in the notation). Note that the results in Ref. [6] use a different scaling in sample size in their Theorem 2.4, but eq. (II.7) in that reference explains that this is just a question of renormalisation. The same result holds for randomly sampled 'Pauli bases', see Ref. [28] (and take note of the slightly different normalisation in the notation there, too), and also for the Bernoulli noise model from Sect. 18.2.2.2, see Ref. [12].

A key interpretation for quantum tomography applications is that, instead of having to measure all $n = d^2$ basis coefficients $tr(E_i\theta)$, $i = 1, \ldots, d^2$, a number

$$n \approx kd\overline{\log}d$$

of *randomly* chosen basis measurements is sufficient to reconstruct $\theta$ in Frobenius norm loss (up to a small error). In situations where $d$ is large compared to $k$ such a gain can be crucial.

*Remark 18.1 (Uniqueness)* It is worth noting that in the absence of errors, so when $Y_0 = \mathcal{X}\theta_0$ in terms of the sampling operator of Eq. (18.4), the quantum shape constraint ensures that under a suitable RIP condition, only the single matrix $\theta_0$ is compatible with the data. More specifically, let $Y_0 = \mathcal{X}\theta_0$ for some $\theta_0 \in \Theta_+$ of rank $k$, and assume that $\mathcal{X}$ satisfies RIP with $\tau_n(4k) < \sqrt{2} - 1$. Then

$$\{\theta \in \Theta_+ : \mathcal{X}\theta = Y_0\} = \{\theta_0\}. \tag{18.27}$$

This is a direct consequence of Theorem 3.2 in Ref. [33], which states that if RIP is satisfied with $\tau_n(4k) < \sqrt{2} - 1$ and $Y_0 = \mathcal{X}\theta_0$, the unique solution of

$$\begin{aligned} &\text{argmin} \quad \|\theta\|_{S_1} \\ &\text{subject to } \mathcal{X}\theta = Y_0 \end{aligned} \tag{18.28}$$

is given by $\theta_0$. If $\theta_0 \in \Theta_+$, then the minimisation can be replaced by (compare also Ref. [23]).

$$\begin{aligned} &\text{argmin} \qquad \text{tr}(\theta) \\ &\text{subject to } \mathcal{X}\theta = Y_0, \ \ \theta \geq 0, \end{aligned} \tag{18.29}$$

giving rise to the above remark. This observation further signifies the role played by the quantum shape constraint.

## 18.3 Uncertainty Quantification for Low-Rank Matrix Recovery

We now turn to the problem of quantifying the uncertainty of estimators $\tilde{\theta}$ that satisfy the risk bound (18.26). In fact the confidence sets we construct could be used for any estimator of $\theta$, but the conclusions are most interesting when used for minimax optimal estimators $\tilde{\theta}$. For the main flow of ideas we shall assume $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)^T \sim N(0, \sigma^2 I_n)$ but the results hold for the Bernoulli measurement model from Sect. 18.2.2.2 as well—this is summarised in Remark 18.5.

From a statistical point of view, we phrase the problem at hand as the one of constructing a confidence set for $\theta$: a data-driven subset $C_n$ of $\mathbb{M}_d(\mathbb{C})$ that is 'centred' at $\tilde{\theta}$ and that satisfies

$$\mathbb{P}_\theta(\theta \in C_n) \geq 1 - \alpha, \quad 0 < \alpha < 1,$$

for a chosen 'coverage' or significance level $1 - \alpha$, and such that the Frobenius norm diameter $|C_n|_F$ reflects the accuracy of estimation, that is, it satisfies, with high probability,

$$|C_n|_F^2 \approx \|\tilde{\theta} - \theta\|_F^2.$$

In particular such a confidence set provides, through its diameter $|C_n|_F$, a data-driven estimate of how well the algorithm has recovered the true matrix $\theta$ in Frobenius-norm loss, and in this sense provides a quantification of the uncertainty in the estimate.

In the situation of an experimentalist this can be used to decide sequentially whether more measurements should be taken (to improve the recovery rate), or whether a satisfactory performance has been reached. Concretely, if for some $\epsilon > 0$ a recovery level $\|\tilde{\theta} - \theta\|_F \le \epsilon$ is desired for an estimator $\tilde{\theta}$, then assuming $\tilde{\theta}$ satisfies the minimax optimal risk bound $dk/n$ from (18.26), we expect to need, ignoring constants,

$$\frac{dk}{n} < \epsilon^2 \text{ and hence at least } n > \frac{dk}{\epsilon^2}$$

measurements. Note that we also need the RIP to hold with $\tau_n(k)$ from (18.8) less than a small constant $c_0$, which requires the same number of measurements, increased by a further poly-log factor of $d$ (and independently of $\sigma$).

Since the rank $k$ of $\theta$ remains unknown after estimation we cannot obviously guarantee that the recovery level $\epsilon$ has been reached after a given number of measurements. A confidence set $C_n$ for $\tilde{\theta}$ provides such certificates with high probability, by checking whether $|C_n|_F \le \epsilon$, and by continuing to take further measurements if not. The main goal is then to prove that a sequential procedure based on $C_n$ does *not* require more than approximately

$$n > \frac{dk\overline{\log}d}{\epsilon^2}$$

samples (with high probability). We construct confidence procedures in the following subsections that work with at most as many measurements, for the designs from Condition 18.1.

### *18.3.1 Adaptive Sequential Sampling*

Before we describe our confidence procedures, let us make the following definition, where we recall that $R(k)$ denotes the set of $d \times d$ Hermitian matrices of rank at most $k \le d$.

**Definition 18.1** Let $\epsilon > 0, \delta > 0$ be given constants. An algorithm $\mathcal{A}$ returning a $d \times d$ matrix $\hat{\theta}$ after $\hat{n} \in \mathbb{N}$ measurements in model (18.3) is called an $(\epsilon, \delta)$—adaptive sampling procedure if, with $\mathbb{P}_\theta$-probability greater than $1-\delta$, the following properties hold for every $\theta \in R(k)$ and every $1 \le k \le d$:

$$\|\hat{\theta} - \theta\|_F \le \epsilon, \tag{18.30}$$

and, for positive constants $C(\delta), \gamma$, the stopping time $\hat{n}$ satisfies

$$\hat{n} \le C(\delta)\frac{kd(\log d)^\gamma}{\epsilon^2}. \tag{18.31}$$

Such an algorithm provides recovery at given accuracy level $\epsilon$ with $\hat{n}$ measurements of minimax optimal order of magnitude (up to a poly-log factor), and with probability greater than $1-\delta$. The sampling algorithm is adaptive since it does not require the knowledge of $k$, and since the number of measurements required depends only on $k$ and not on the 'worst case' rank $d$.

The construction of non-asymptotic confidence sets $C_n$ for $\theta$ at any sample size $n$ in the next subsections will imply that such algorithms exist for low rank matrix recovery problems. The main idea is to check sequentially, for a geometrically increasing number $2^m$ of samples, $m = 1, 2, \ldots,$ if the diameter $|C_{2^m}|_F$ of a confidence set exceeds $\epsilon$. If this is not the case, the algorithm terminates. Otherwise one takes $2^{m+1}$ additional measurements and evaluates the diameter $|C_{2^{m+1}}|_F$. A precise description of the algorithm is given in the proof of the following theorem, which we detail for the case of 'Pauli' designs. The isotropic design case is discussed in Remark 18.9.

**Theorem 18.1** *Consider observations in the model (18.3) under Condition 18.1(b) with $\theta \in \Theta_+$. Then an adaptive sampling algorithm in the sense of Definition 18.1 exists for any $\epsilon, \delta > 0$.*

*Remark 18.2 (Dependence in $\sigma$ of Definition 18.1 and Theorem 18.1)* Definition 18.1 and Theorem 18.1 are stated for the case where the standard deviation of the noise $\sigma$ is assumed to be bounded by an absolute constant. It is straight-forward to modify the proofs to obtain a version where the dependency of the constants on the variance is explicit. Indeed, under Condition 1(a), Theorem 18.1 continues to hold if Eq. (18.31) is replaced by

$$\hat{n} \leq C(\delta)\frac{\sigma^2 kd(\log d)^\gamma}{\epsilon^2}.$$

For the 'Pauli design case'—Condition 1(b)—Eq. (18.31) can be modified to

$$\hat{n} \leq C(\delta)\Big(\frac{\sigma^2 kd(\log d)^\gamma}{\epsilon^2} \vee \frac{d(\log d)^\gamma}{\epsilon^2}\Big).$$

*Remark 18.3 (Necessity of the Quantum Shape Constraint)* Note that the assumption $\theta \in \Theta_+$ in the previous theorem is necessary (in the case of Pauli design): Else the example of $\theta = 0$ or $\theta = E_i$—where $E_i$ is an arbitrary element of the Pauli basis—demonstrates that the number of measurements has to be at least of order $d^2$: otherwise with positive probability, $E_i$ is not drawn at a fixed sample size. On this event, both the measurements and $\hat{\theta}$ coincide under the laws $\mathbb{P}_0$ and $\mathbb{P}_{E_i}$, so we cannot have $\|\hat{\theta} - 0\|_F < \epsilon$ *and* $\|\hat{\theta} - E_i\|_F < \epsilon$ simultaneously for every $\epsilon > 0$, disproving existence of an adaptive sampling algorithm. In fact, the crucial condition for Theorem 18.1 to work is that the nuclear norms $\|\theta\|_{S_1}$ are bounded by an absolute constant (here $= 1$), which is violated by $\|E_i\|_{S_1} = \sqrt{d}$.

### 18.3.2 A Non-asymptotic Confidence Set Based on Unbiased Risk Estimation and Sample-Splitting

We suppose that we have two samples at hand, the first being used to construct an estimator $\tilde{\theta}$, such as the one from (18.26). We freeze $\tilde{\theta}$ and the first sample in what follows and all probabilistic statements are under the distribution $\mathbb{P}_\theta$ of the second sample $Y, X$ of size $n \in \mathbb{N}$, conditional on the value of $\tilde{\theta}$. We define the following residual sum of squares statistic (recalling that $\sigma^2$ is known):

$$\hat{r}_n = \frac{1}{n}\|Y - \mathcal{X}\tilde{\theta}\|_F^2 - \sigma^2,$$

which satisfies $\mathbb{E}_\theta \hat{r}_n = \|\theta - \tilde{\theta}\|_F^2$ as is easily seen (see the proof of Theorem 18.2 below). Given $\alpha > 0$, let $\xi_{\alpha,\sigma}$ be quantile constants such that

$$\Pr\left(\sum_{i=1}^{n}(\varepsilon_i^2 - 1) > \xi_{\alpha,\sigma}\sqrt{n}\right) = \alpha \tag{18.32}$$

(these constants converge to the quantiles of a fixed normal distribution as $n \to \infty$), let $z_\alpha = \log(3/\alpha)$ and, for $z \geq 0$ a fixed constant to be chosen, define the confidence set

$$C_n = \left\{v \in \mathbb{H}_d(\mathbb{C}) : \|v - \tilde{\theta}\|_F^2 \leq 2\left(\hat{r}_n + z\frac{d}{n} + \frac{\bar{z} + \xi_{\alpha/3,\sigma}}{\sqrt{n}}\right)\right\}, \tag{18.33}$$

where

$$\bar{z}^2 = \bar{z}^2(\alpha, d, n, \sigma, v) = z_{\alpha/3}\sigma^2 \max(3\|v - \tilde{\theta}\|_F^2, 4zd/n).$$

Note that in the 'quantum shape constraint' case we can always bound $\|v - \tilde{\theta}\|_F \leq 2$ which gives a confidence set that is easier to compute and of only marginally larger overall diameter. In many important situations, however, the quantity $\bar{z}/\sqrt{n}$ is of smaller order than $1/\sqrt{n}$, and the more complicated expression above is preferable.

It is not difficult to see (using that $x^2 \lesssim y + x/\sqrt{n}$ implies $x^2 \lesssim y + 1/n$) that the square Frobenius norm diameter of this confidence set is, with high probability, of order

$$|C_n|_F^2 \lesssim \|\tilde{\theta} - \theta\|_F^2 + \frac{zd + z_{\alpha/3}}{n} + \frac{\xi_{\alpha/3,\sigma}}{\sqrt{n}}. \tag{18.34}$$

Whenever $d \geq \sqrt{n}$—so as long as at most $n \leq d^2$ measurements have been taken—the deviation terms are of smaller order than $kd/n$, and hence $C_n$ has minimax optimal expected squared diameter whenever the estimator $\tilde{\theta}$ is minimax optimal as in (18.26). Improvements for $d < \sqrt{n}$, corresponding to $n > d^2$ measurements, will be discussed in the next subsections.

The following result shows that $C_n$ is an honest confidence set for arbitrary $d \times d$ matrices (without any rank constraint). Note that the result is non-asymptotic—it holds for every $n \in \mathbb{N}$.

**Theorem 18.2** *Let $\theta \in \mathbb{H}_d(\mathbb{C})$ be arbitrary and let $\mathbb{P}_\theta$ be the distribution of $Y, X$ from model (18.3).*

*(a) Assume Condition 18.1(a) and let $C_n$ be given by (18.33) with $z = 0$. We then have for every $n \in \mathbb{N}$ that*

$$\mathbb{P}_\theta(\theta \in C_n) \geq 1 - \frac{2\alpha}{3} - 2e^{-cn}$$

*where $c$ is a numerical constant. In the case of standard Gaussian design, $c = 1/24$ is admissible.*

*(b) Assume Condition 18.1(b), let $C_n$ be given by (18.33) with $z > 0$ and assume also that $\theta \in \Theta_+$ and $\tilde{\theta} \in \Theta_+$ (that is, both satisfy the 'quantum shape constraint'). Then for every $n \in \mathbb{N}$,*

$$\mathbb{P}_\theta(\theta \in C_n) \geq 1 - \frac{2\alpha}{3} - 2e^{-C(K)z}$$

*where, for $K$ the coherence constant of the basis,*

$$C(K) = \frac{1}{(16 + 8/3)K^2}.$$

In Part (a), if we want to control the coverage probability at level $1 - \alpha$, $n$ needs to be large enough so that the third deviation term is controlled at level $\alpha/3$. In the Gaussian design case with $\alpha = 0.05$, $n \geq 100$ is sufficient, for smaller sample sizes one can reduce the coverage level. The bound in (b) is entirely non-asymptotic (using the quantum constraint) for suitable choices of $z$. Also note that the quantile constants $z, z_\alpha, \xi_\alpha$ all scale at least as $O(\log(1/\alpha))$ in the desired coverage level $\alpha \to 0$.

*Remark 18.4 (Dependence of the Confidence Set's Diameter on $K$ (Pauli Design) and $\sigma$)* Note that in the case of the Pauli design from Condition 1(b), the confidence set's diameter depends on $K$ only through the potential dependence of $\|\theta - \tilde{\theta}\|_F^2$ on $K$—the constants involved in the construction of $\tilde{C}_n$ and on the bound on its diameter do not depend on $K$. On the other hand, the coverage probability of the confidence set depends on $K$, see Theorem 18.2, (b).

In this paper we assume that $\sigma$ is a universal constant, and so as such it does not appear in Eqs. (18.33) and (18.34). It can however be interesting to investigate the dependence in $\sigma$. In the case of isotropic design from Condition 1(a), we could set

$$C_n = \left\{ v \in \mathbb{H}_d(\mathbb{C}) : \|v - \tilde{\theta}\|_F^2 \le 2\left(\hat{r}_n + zc\sigma^2 \frac{d}{n} + \sigma^2 \frac{\bar{z} + \xi_{\alpha/3,\sigma}}{\sqrt{n}}\right)\right\},$$

(where $\sigma^2$ could be replaced by twice the plug-in estimator of $\sigma^2$, using $\hat{\theta}$) and one would get

$$\mathbb{E}_\theta |C_n|_F^2 \lesssim \|\tilde{\theta} - \theta\|_F^2 + \sigma^2 \frac{zd + z_{\alpha/3}}{n} + \sigma^2 \frac{\xi_{\alpha/3,\sigma}}{\sqrt{n}},$$

and Theorem 18.2 also holds by introducing minor changes in the proof. In the case of the Pauli design from Condition 1(b), we could set

$$C_n = \left\{ v \in \mathbb{H}_d(\mathbb{C}) : \|v - \tilde{\theta}\|_F^2 \le 2\left(\hat{r}_n + zc \frac{d}{n} + \sigma^2 \frac{\bar{z} + \xi_{\alpha/3,\sigma}}{\sqrt{n}}\right)\right\},$$

(where $\sigma^2$ could be replaced by twice the plug-in estimator of $\sigma^2$, using $\hat{\theta}$) and one would get

$$\mathbb{E}_\theta |C_n|_F^2 \lesssim \|\tilde{\theta} - \theta\|_F^2 + \frac{zd + z_{\alpha/3}}{n} + \sigma^2 \frac{\xi_{\alpha/3,\sigma}}{\sqrt{n}},$$

and Theorem 18.2 also holds by introducing minor changes in the proof. In this case we do not get a full dependence in $\sigma$ as in the isotropic design case from Condition 1(a). However if $k^2 d \lesssim n$, we could also obtain a result similar to the one for the Gaussian design, using part (c) of Lemma 18.1.

*Remark 18.5 (Bernoulli Noise)* Theorem 18.2(b) holds as well for the Bernoulli measurement model from Sect. 18.2.2.2 with $T \ge d^2$, with slightly different constants in the construction of $C_n$ and the coverage probabilities. See Remark 18.10 after the proof of Theorem 18.2(b) below. The modified quantile constants $z, z_\alpha, \xi_\alpha$ still scale as $O(\sqrt{1/\alpha})$ in the desired coverage level $\alpha \to 0$, and hence the adaptive sampling Theorem 18.1 holds for such noise too, if the number $T$ of preparations of the quantum state exceeds $d^2$.

*Remark 18.6 (Unknown Variance)* The above confidence set $C_n$ can be constructed with $\tilde{r}_n = \frac{1}{n}\|Y - \mathcal{X}\tilde{\theta}\|^2$ replacing $\hat{r}_n$—so without requiring knowledge of $\sigma$—if an a priori bound $\sigma^2 \le vd/n$ is available, with $v$ a known constant. An example of such a situation was discussed at the end of Sect. 18.2.2.2 above in quantum tomography problems: when $T \ge n$, the constant $z$ should be increased by $v$ in the construction of $C_n$, and the coverage proof goes through as well by compensating for the centring at $\mathbb{E}\varepsilon_i^2 = \sigma^2$ by the additional deviation constant $v$.

*Remark 18.7 (Anisotropic Design Instead of Condition 1(a))* It is also interesting to consider the case of anisotropic design. This case is not very different, when it comes to confidence sets, than isotropic design, as long as the variance-covariance matrix of the anisotropic sub-Gaussian design is such that the ratio of its largest eigenvalue with the smallest eigenvalue is bounded. Lemma 18.1(a), which quantifies the effect of the design, would change as follows: There exist constants $c_-, c_+, c > 0$ that depend only on the variance-covariance matrix of the anisotropic sub-Gaussian design and that are such that

$$\Pr\left(c_-\|\vartheta\|_F^2 \leq \frac{1}{n}\|\mathcal{X}\vartheta\|^2 \leq c_+\|\vartheta\|_F^2\right) \geq 1 - 2e^{-cn}.$$

Using this instead of the inequality in Lemma 18.1(a) in the proof of Theorem 18.2, part (a) leads to a similar result as Theorem 18.2, part (a).

### 18.3.3 Improvements When $d \leq \sqrt{n}$

The confidence set from Theorem 18.2 is optimal whenever the desired performance of $\|\theta - \tilde{\theta}\|_F^2$ is no better than of order $1/\sqrt{n}$. From a minimax point of view we expect $\|\theta - \tilde{\theta}\|_F^2$ to be of order $kd/n$ for low rank $\theta \in R(k)$. In absence of knowledge about $k \geq 1$ the confidence set from Theorem 18.2 can hence be guaranteed to be optimal whenever $d \geq \sqrt{n}$, corresponding to the important regime $n \leq d^2$ for sequential sampling algorithms. Refinements for measurement scales $n \geq d^2$ are also of interest—we present two optimal approaches in this subsection for the designs from Condition 18.1.

#### 18.3.3.1 Isotropic Design and $U$-Statistics

Consider first isotropic i.i.d design from Condition 18.1(a), and an estimator $\tilde{\theta}$ based on an initial sample of size $n$ (all statements that follow are conditional on that sample). Collect another $n$ samples to perform the uncertainty quantification step. Define the $U$-statistic

$$\hat{R}_n = \frac{2}{n(n-1)} \sum_{i<j} \sum_{m,k} (Y_i X^i_{m,k} - \tilde{\theta}_{m,k})(Y_j X^j_{m,k} - \tilde{\theta}_{m,k}) \tag{18.35}$$

whose $\mathbb{E}_\theta$-expectation, conditional on $\tilde{\theta}$, equals $\|\theta - \tilde{\theta}\|_F^2$ in view of

$$\mathbb{E} Y_i X^i_{m,k} = \mathbb{E} \sum_{m',k'} X^i_{m',k'} X^i_{m,k} \theta_{m',k'} = \theta_{m,k}.$$

Define

$$C_n = \left\{ v \in \mathbb{H}_d(\mathbb{R}) : \|v - \tilde{\theta}\|_F^2 \leq \hat{R}_n + z_{\alpha,n} \right\} \tag{18.36}$$

where

$$z_{\alpha,n} = \frac{C_1 \|\theta - \tilde{\theta}\|_F}{\sqrt{n}} + \frac{C_2 d}{n}$$

and $C_1 \geq \zeta_1 \|\theta\|_F,\ C_2 \geq \zeta_2 \|\theta\|_F^2$ with $\zeta_i$ constants depending on $\alpha, \sigma$. Note that if $\theta \in \Theta_+$ then $\|\theta\|_F \leq 1$ can be used as an upper bound. In practice the constants $\zeta_i$ can be calibrated by Monte Carlo simulations (see the implementation section below), or chosen based on concentration inequalities for $U$-statistics (see Ref. [14], Theorem 4.4.8). This confidence set has expected diameter

$$\mathbb{E}_\theta |C_n|_F^2 \lesssim \|\tilde{\theta} - \theta\|_F^2 + \frac{C_1 + C_2 d}{n},$$

and hence is compatible with any minimax recover rate $\|\tilde{\theta} - \theta\|_F^2 \lesssim kd/n$ from (18.26), where $k \geq 1$ is now arbitrary. For suitable choices of $\zeta_i$ we now show that $C_n$ also has non-asymptotic coverage.

**Theorem 18.3** *Assume Condition 18.1(a), and let $C_n$ be as in (18.36). For every $\alpha > 0$ we can choose $\zeta_i(\alpha) = O(\sqrt{1/\alpha}), i = 1, 2,$ large enough so that for every $n \in \mathbb{N}$ we have*

$$\mathbb{P}_\theta(\theta \in C_n) \geq 1 - \alpha.$$

*Remark 18.8 (Dependence of the Confidence Set's Diameter on $\sigma$)* As what was noted in Remark 18.4, Theorem 18.3 does not make explicit the dependence on $\sigma$, which is assumed to be (bounded by) an universal constant. In order to take the dependence on $\sigma$ into account, we could replace $z_{\alpha,n}$ in Eq. (18.36) by $\frac{C_1\|\theta - \tilde{\theta}\|_F}{\sqrt{n}} + \sigma^2 \frac{C_2 d}{n}$ (where $\sigma^2$ could be replaced by twice the plug-in estimator of $\sigma^2$, using $\hat{\theta}$), and we would get

$$\mathbb{E}_\theta |C_n|_F^2 \lesssim \|\tilde{\theta} - \theta\|_F^2 + \sigma^2 \frac{C_1 + C_2 d}{n},$$

and Theorem 18.3 also holds by introducing minor changes in the proof.

#### 18.3.3.2 Re-averaging Basis Elements When $d \leq \sqrt{n}$

Consider the setting of Condition 18.1(b) where we sample uniformly at random from a (scaled) basis $\{dE_1, \ldots, dE_{d^2}\}$ of $\mathbb{M}_d(\mathbb{C})$. When $d \leq \sqrt{n}$ we are taking

$n \geq d^2$ measurements, and there is no need to sample at *random* from the basis as we can measure each individual coefficient, possibly even multiple times. Repeatedly sampling a basis coefficient $tr(E_k\theta)$ leads to a reduction of the variance of the measurement by averaging. More precisely, when taking $n = md^2$ measurements for some (for simplicity integer) $m \geq 1$, and if $(Y_{k,l} : l = 1, \dots, m)$ are the measurements $Y_i$ corresponding to the basis element $E_k, k \in \{1, \dots, d^2\}$, we can form averaged measurements

$$Z_k = \frac{1}{\sqrt{m}} \sum_{l=1}^{m} Y_{k,l} = \sqrt{m} d \langle E_k, \theta \rangle_F + \epsilon_k, \quad \epsilon_k = \frac{1}{\sqrt{m}} \sum_{l=1}^{m} \varepsilon_l \sim N(0, \sigma^2).$$

We can then define the new measurement vector $\tilde{Z} = (\tilde{Z}_1, \dots, \tilde{Z}_{d^2})^T$ (using also $m = n/d^2$)

$$\tilde{Z}_k = Z_k - \sqrt{n} \langle \tilde{\theta}, E_k \rangle = \sqrt{n} \langle E_k, \theta - \tilde{\theta} \rangle_F + \epsilon_k, \quad k = 1, \dots, d^2$$

and the statistic

$$\hat{R}_n = \frac{1}{n} \|\tilde{Z}\|^2_{\mathbb{R}^{d^2}} - \frac{\sigma^2 d^2}{n}$$

which estimates $\|\theta - \tilde{\theta}\|_F^2$ with precision

$$\begin{aligned} \hat{R}_n - \|\theta - \tilde{\theta}\|_F^2 &= \frac{2}{\sqrt{n}} \sum_{k=1}^{d^2} \epsilon_k \langle E_k, \theta - \tilde{\theta} \rangle_F + \frac{1}{n} \sum_{k=1}^{d^2} (\epsilon_k^2 - \mathbb{E}\epsilon^2) \\ &= O_P \left( \frac{\sigma \|\theta - \tilde{\theta}\|_F}{\sqrt{n}} + \frac{\sigma^2 d}{n} \right). \end{aligned}$$

Hence, for $z_\alpha$ the quantiles of a $N(0, 1)$ distribution and $\xi_{\alpha,\sigma}$ as in (18.32) with $d^2$ replacing $n$ there, we can define a confidence set

$$\bar{C}_n = \left\{ v \in \mathbb{H}_d(\mathbb{C}) : \|v - \tilde{\theta}\|_F^2 \leq \hat{R}_n + \frac{z_{\alpha/2} \sigma \|\theta - \tilde{\theta}\|_F}{\sqrt{n}} + \frac{\xi_{\alpha/2,\sigma} d}{n} \right\} \tag{18.37}$$

which has non-asymptotic coverage

$$\mathbb{P}_\theta(\theta \in \bar{C}_n) \geq 1 - \alpha$$

for every $n \in \mathbb{N}$, by similar (in fact, since Lemma 18.1 is not needed, simpler) arguments as in the proof of Theorem 18.2 below. The expected diameter of $\bar{C}_n$ is by construction

$$\mathbb{E}_\theta |\bar{C}_n|_F^2 \lesssim \|\theta - \tilde{\theta}\|_F^2 + \frac{\sigma^2 d}{n}, \tag{18.38}$$

now compatible with *any* rate of recovery $kd/n$, $1 \leq k \leq d$.

### *18.3.4 A Confidence Set in Trace Norm Under Quantum Shape Constraints*

The confidence sets from the previous subsections are all valid in the sense that they contain information about the recovery of $\theta$ by $\tilde{\theta}$ in Frobenius norm $\|\cdot\|_F$. It is of interest to obtain results in stronger norms, such as for instance the nuclear norm $\|\cdot\|_{S_1}$, which is particularly meaningful for quantum tomography problems since it then corresponds to the total variation distance on the set of 'probability density matrices'. In fact, since

$$\frac{1}{2}\|\theta - \tilde{\theta}\|_{S_1} = \sup_{\|X\|_{op}=1} tr\left(X(\theta - \tilde{\theta})\right), \tag{18.39}$$

the nuclear norm has a clear interpretation in terms of the maximum probability with which two quantum states can be distinguished by arbitrary measurements.

The absence of the 'Hilbert space geometry' induced by the relationship of the Frobenius norm to the inner product $\langle \cdot, \cdot \rangle_F$ makes this problem significantly harder, both technically and from an information-theoretic point of view. In particular it appears that the quantum shape constraint $\theta \in \Theta_+$ is crucial to obtain any results whatsoever, and for the theoretical results presented here it will be more convenient to perform an asymptotic analysis where $\min(n, d) \to \infty$ (with $o$, $O$-notation to be understood accordingly).

Instead of Condition 18.1 we shall now consider any design $(X^1, \ldots, X^n)$ in model (18.3) that satisfies the matrix RIP (18.7) with

$$\tau_n(k) = c\sqrt{kd\frac{\overline{\log}(d)}{n}}. \tag{18.40}$$

As discussed above, this covers in particular the designs from Condition 18.1. We shall still use the convention discussed before Condition 18.1 that $\theta$ and the matrices $X^i$ are such that $tr(X^i\theta)$ is always real-valued.

In contrast to the results from the previous section we shall now assume a minimal low rank constraint on the parameter space:

**Condition 18.2** $\theta \in R^+(k) := R(k) \cap \Theta_+$ *for some k satisfying*

$$k\sqrt{\frac{d\overline{\log}d}{n}} = o(1),$$

This in particular implies that the RIP holds with $\tau_n(k) = o(1)$. Given this minimal rank constraint $\theta \in R^+(k)$, we now show that it is possible to construct a confidence set $C_n$ that adapts to any low rank $1 \leq k_0 < k$. Here we may choose $k = d$ but note that this forces $n \gg d^2$ (for Condition 18.2 to hold with $k = d$).

We assume that there exists an estimator $\tilde{\theta}_{\text{Pilot}}$ that satisfies, uniformly in $R(k_0)$ for any $k_0 \leq k$ and for $n$ large enough,

$$\|\tilde{\theta}_{\text{Pilot}} - \theta\|_F^2 \leq D\sigma^2 \frac{k_0 d}{n} := \frac{r_n^2(k_0)}{4} \tag{18.41}$$

where $D = D(\delta)$ depends on $\delta$, and where so-defined $r_n$ will be used frequently below. Such estimators exist as has already been discussed before (18.26). We shall in fact require a little more, namely the following oracle inequality: for any $k$ and any matrix $S$ of rank $k \leq d$, with high probability and for $n$ large enough,

$$\|\tilde{\theta}_{\text{Pilot}} - \theta\|_F \lesssim \|\theta - S\|_F + r_n(k), \tag{18.42}$$

which in fact implies (18.41). Such inequalities exist assuming the RIP and Condition 18.2, see, e.g., Theorem 2.8 in Ref. [6]. Starting from $\tilde{\theta}_{\text{Pilot}}$ one can construct (see Theorem 18.5 below) an estimator that recovers $\theta \in R(k)$ in nuclear norm at rate $k\sqrt{d/n}$, which is again optimal from a minimax point of view, even under the quantum constraint (as discussed, e.g., in Ref. [24]). We now construct an adaptive confidence set for $\theta$ centred at a suitable projection of $\tilde{\theta}_{\text{Pilot}}$ onto $\Theta_+$.

In the proof of Theorem 18.4 below we will construct estimated eigenvalues $(\hat{\lambda}_j, j = 1, \ldots, d)$ of $\theta$ (see after Lemma 18.3). Given those eigenvalues and $\tilde{\theta}_{\text{Pilot}}$, we choose $\hat{k}$ to equal the smallest integer $\leq d$ such that there exists a rank $\hat{k}$ matrix $\tilde{\theta}'$ for which

$$\|\tilde{\theta}' - \tilde{\theta}_{\text{Pilot}}\|_F \leq r_n(\hat{k}) \text{ and } 1 - \sum_{J \leq \hat{k}} \hat{\lambda}_J \leq 2\hat{k}\sqrt{d/n}$$

is satisfied. Such $\hat{k}$ exists with high probability (since the inequalities are satisfied for the true $\theta$ and $\lambda_j$'s, as our proofs imply). Define next $\hat{\vartheta}$ to be the $\langle \cdot, \cdot \rangle_F$-projection of $\tilde{\theta}_{\text{Pilot}}$ onto

$$R^+(2\hat{k}) := R(2\hat{k}) \cap \Theta_+$$

and note that, since $2\hat{k} \geq \hat{k}$,

$$\|\tilde{\theta}_{\text{Pilot}} - \hat{\vartheta}\|_F = \|\tilde{\theta}_{\text{Pilot}} - R^+(2\hat{k})\|_F \leq \|\tilde{\theta}_{\text{Pilot}} - \tilde{\theta}'\|_F \leq r_n(\hat{k}). \tag{18.43}$$

Finally define, for $C$ a constant chosen below,

$$C_n = \left\{ v \in \Theta_+ : \|v - \hat{\vartheta}\|_{S_1} \leq C\sqrt{\hat{k}} r_n(\hat{k}) \right\}. \tag{18.44}$$

**Theorem 18.4** *Assume Condition 18.2 for some $1 \leq k \leq d$, and let $\delta > 0$ be given. Assume that with probability greater than $1 - 2\delta/3$, (a) the RIP (18.7) holds with $\tau_n(k)$ as in (18.40) and (b) there exists an estimator $\tilde{\theta}_{\text{Pilot}}$ for which (18.42) holds. Then we can choose $C = C(\delta)$ large enough so that, for $C_n$ as in the last display,*

$$\liminf_{\min(n,d)\to\infty} \inf_{\theta \in R^+(k)} \mathbb{P}_\theta(\theta \in C_n) \geq 1 - \delta.$$

*Moreover, uniformly in $R^+(k_0)$, $1 \leq k_0 \leq k$, and with $\mathbb{P}_\theta$-probability greater than $1 - \delta$,*

$$|C_n|_{S_1} \lesssim \sqrt{k_0} r_n(k_0).$$

Theorem 18.4 should mainly serve the purpose of illustrating that the quantum shape constraint allows for the construction of an optimal trace norm confidence set that adapts to the unknown low rank structure. Implementation of $C_n$ is not straightforward so Theorem 18.4 is mostly of theoretical interest. Let us also observe that in full generality a result like Theorem 18.4 cannot be proved *without* the quantum shape constraint. This follows from a careful study of certain hypothesis testing problems (combined with lower bound techniques for confidence sets as in Refs. [19, 30]). Precise results are subject of current research and will be reported elsewhere.

## 18.4 Simulation Experiments

In order to illustrate the methods from this paper, we present some numerical simulations. The setting of the experiments is as follows: A random matrix $\eta \in \mathbb{M}_d(\mathbb{C})$ of norm $\|\eta\|_F = R^{1/2}$ is generated according to two distinct procedures that we will specify later, and the observations are

$$\bar{Y}_i = tr(X^i \eta) + \varepsilon_i.$$

where the $\varepsilon_i$ are i.i.d. Gaussian of mean 0 and variance 1. The observations are reparametrised so that $\eta$ represents the 'estimation error' $\theta - \hat{\theta}$, and we investigate how well the statistics

$$\hat{r}_n = \frac{1}{n}\|\bar{Y}\| - 1 \text{ and } \hat{R}_n = \frac{2}{n(n-1)} \sum_{i<j} \sum_{m,k} \bar{Y}_i X^i_{m,k} \bar{Y}_j X^j_{m,k}$$

estimate the 'accuracy of estimation' $\|\eta\|_F^2 = \|\theta - \hat{\theta}\|_F^2$, conditional on the value of $\hat{\theta}$. We will choose $\eta$ in order to illustrate two extreme cases: a first one where the nuclear norm $\|\eta\|_{S_1}$ is 'small', corresponding to a situation where the quantum constraint is fulfilled; and a second one where the nuclear norm is large, corresponding to a situation where the quantum constraint is *not* fulfilled. More precisely we generate the parameter $\eta$ in two ways:

- 'Random Dirac' case: set a single entry (with position chosen at random on the diagonal) of $\eta$ to $R^{1/2}$, and all the other coordinates equal to 0.
- 'Random Pauli' case: Set $\eta$ equal to a Pauli basis element chosen uniformly at random and then multiplied by $R^{1/2}$.

The designs that we consider are the Gaussian design, and the Pauli design, described in Condition 1. We perform experiments with $d = 32$, $R \in \{0.1, 1\}$ and

$$n \in \{100, 200, 500, 1000, 2000, 5000\}.$$

Note that $d^2 = 1024$, so that the first four choices of $n$ correspond to the important regime $n < d^2$. Our results are plotted as a function of the number $n$ of samples in Figs. 18.2, 18.3, 18.4, and 18.5. The solid red and blue curves are the median errors of the normalised estimation errors

$$\frac{\sqrt{\hat{R}_n - R}}{R^{1/2}}, \quad \text{and} \quad \frac{\sqrt{\hat{r}_n - R}}{R^{1/2}},$$

after 1000 iterations, and the dotted lines are respectively, the (two-sided) 90% quantiles. We also report (see Tables 18.1, 18.2, 18.3, and 18.4) how well the confidence sets based on these estimates of the norm perform in terms of coverage probabilities, and of diameters. The diameters are computed as

$$\left(\hat{R}_n + \frac{C_{\text{UStat}} d}{n} + \frac{C'_{\text{UStat}} \hat{R}_n^{1/2}}{\sqrt{n}}\right)^{1/2},$$

for the U-Statistic approach and

$$\left(\hat{r}_n + \frac{C_{\text{RSS}}}{\sqrt{n}} + \frac{C'_{\text{RSS}} \hat{r}_n^{1/2}}{\sqrt{n}}\right)^{1/2},$$

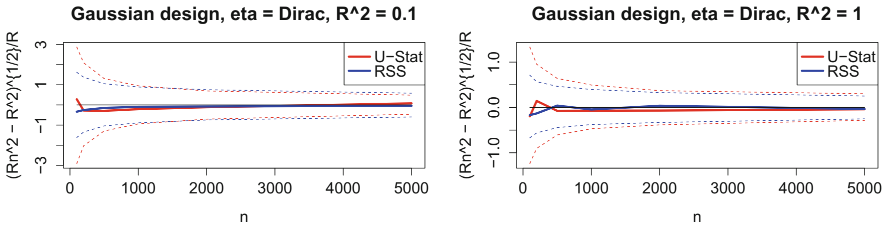


**Fig. 18.2** Gaussian design, and random Dirac (a single entry, chosen at random, is non-zero on the diagonal) $\eta$, with $R = 0.1$ (left picture) and $R = 1$ (right picture)

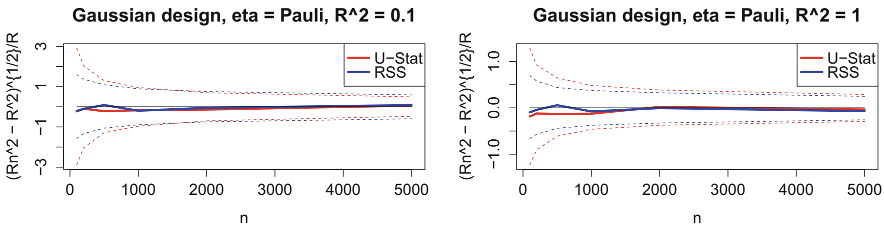


**Fig. 18.3** Gaussian design, and random Pauli $\eta$, with $R = 0.1$ (left picture) and $R = 1$ (right picture)

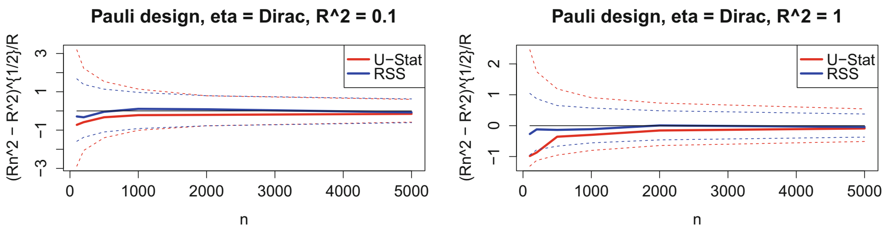


**Fig. 18.4** Pauli design, and random Dirac (a single entry, chosen at random, is non-zero on the diagonal) $\eta$, with $R = 0.1$ (left picture) and $R = 1$ (right picture)

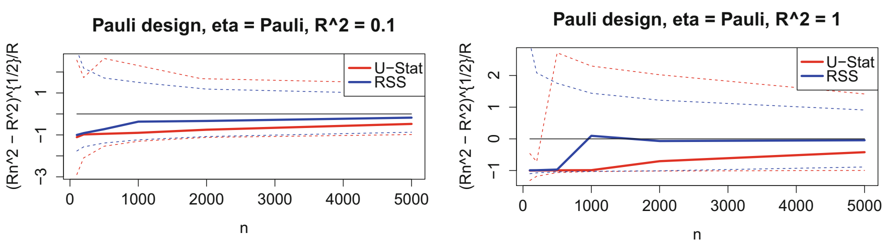


**Fig. 18.5** Pauli design, and random Pauli $\eta$, with $R = 0.1$ (left picture) and $R = 1$ (right picture)

**Table 18.1** Gaussian design, and random Dirac (a single entry, chosen at random, is non-zero on the diagonal) $\eta$, with $R = 0.1$ (left table) and $R = 1$ (right table)

| | $R = 0.1$ | | | | | | $R = 1$ | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $n$ | 100 | 200 | 500 | 1000 | 2000 | 5000 | 100 | 200 | 500 | 1000 | 2000 | 5000 |
| Coverage U-Stat | 0.97 | 0.98 | 0.99 | 1.00 | 1.00 | 1.00 | 0.93 | 0.96 | 0.97 | 0.98 | 0.98 | 0.98 |
| Diameter U-Stat | 1.10 | 0.64 | 0.34 | 0.24 | 0.18 | 0.14 | 2.43 | 1.84 | 1.44 | 1.27 | 1.17 | 1.10 |
| Coverage RSS | 0.97 | 0.97 | 0.98 | 0.98 | 0.98 | 0.98 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 |
| Diameter RSS | 0.38 | 0.31 | 0.23 | 0.19 | 0.16 | 0.14 | 1.69 | 1.49 | 1.32 | 1.22 | 1.16 | 1.10 |

**Table 18.2** Gaussian design, and random Pauli $\eta$, with $R = 0.1$ (left table) and $R = 1$ (right table)

| | $R = 0.1$ | | | | | | $R = 1$ | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $n$ | 100 | 200 | 500 | 1000 | 2000 | 5000 | 100 | 200 | 500 | 1000 | 2000 | 5000 |
| Coverage U-Stat | 0.98 | 0.98 | 0.99 | 0.99 | 1.0 | 1.0 | 0.93 | 0.95 | 0.97 | 0.98 | 0.98 | 0.98 |
| Diameter U-Stat | 1.10 | 0.62 | 0.34 | 0.24 | 0.18 | 0.14 | 2.40 | 1.83 | 1.43 | 1.27 | 1.18 | 1.10 |
| Coverage RSS | 0.98 | 0.98 | 0.97 | 0.97 | 0.97 | 0.97 | 0.99 | 0.99 | 0.99 | 0.99 | 1.00 | 1.00 |
| Diameter RSS | 0.39 | 0.31 | 0.23 | 0.19 | 0.17 | 0.14 | 1.71 | 1.49 | 1.31 | 1.22 | 1.16 | 1.10 |

**Table 18.3** Pauli design, and random Dirac (a single entry, chosen at random, is non-zero on the diagonal) $\eta$, with $R = 0.1$ (left table) and $R = 1$ (right table)

| | $R = 0.1$ | | | | | | $R = 1$ | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $n$ | 100 | 200 | 500 | 1000 | 2000 | 5000 | 100 | 200 | 500 | 1000 | 2000 | 5000 |
| Coverage U-Stat | 0.97 | 0.98 | 0.98 | 0.99 | 0.98 | 0.98 | 0.85 | 0.54 | 0.69 | 0.69 | 0.70 | 0.71 |
| Diameter U-Stat | 1.10 | 0.63 | 0.34 | 0.24 | 0.18 | 0.14 | 2.28 | 1.87 | 1.43 | 1.26 | 1.18 | 1.10 |
| Coverage RSS | 0.96 | 0.96 | 0.96 | 0.96 | 0.97 | 0.97 | 0.88 | 0.89 | 0.88 | 0.88 | 0.88 | 0.88 |
| Diameter RSS | 0.39 | 0.29 | 0.23 | 0.19 | 0.16 | 0.14 | 1.70 | 1.50 | 1.30 | 1.21 | 1.16 | 1.10 |

**Table 18.4** Pauli design, and random Pauli $\eta$, with $R = 0.1$ (left table) and $R = 1$ (right table)

| | $R = 0.1$ | | | | | | $R = 1$ | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $n$ | 100 | 200 | 500 | 1000 | 2000 | 5000 | 100 | 200 | 500 | 1000 | 2000 | 5000 |
| Coverage U-Stat | 0.97 | 0.97 | 0.96 | 0.86 | 0.65 | 0.58 | 0.82 | 0.22 | 0.25 | 0.27 | 0.30 | 0.37 |
| Diameter U-Stat | 1.09 | 0.57 | 0.34 | 0.25 | 0.18 | 0.15 | 2.45 | 2.09 | 1.33 | 1.38 | 1.19 | 1.09 |
| Coverage RSS | 0.93 | 0.86 | 0.77 | 0.77 | 0.77 | 0.77 | 0.12 | 0.19 | 0.40 | 0.63 | 0.56 | 0.53 |
| Diameter RSS | 0.38 | 0.29 | 0.22 | 0.19 | 0.16 | 0.14 | 1.71 | 1.56 | 1.31 | 1.26 | 1.14 | 1.08 |

for the RSS approach, where we have chosen $C_{\text{UStat}} = 2.5$, $C_{\text{RSS}} = 1$ and $C'_{\text{UStat}} = C_{\text{RSS}} = 6$ for all experiments—calibrated to a 95% coverage level.

From these numerical results, several observations can be made:

- In Gaussian random designs, the results are insensitive to the nature of $\eta$ (see Figs. 18.2 and 18.3 and Tables 18.1 and 18.2). This is not surprising since the Gaussian design is 'isotropic'.
- For Pauli designs with the quantum constraint (see Fig. 18.4 and Table 18.3) the RSS method works quite well even for small sample sizes. But the U-Stat method is not very reliable—indeed we see no empirical evidence that Theorem 18.3 should also hold true for Pauli design.
- For Pauli design and when the quantum shape constraint is *not* satisfied our methods cease to provide reliable results (see Fig. 18.5 and in particular Table 18.4). Indeed, when the matrix $\eta$ is chosen itself as a random Pauli (which is the hardest signal to detect under Pauli design) both the RSS and the U-Stat approach perform poorly. The confidence set are not honest anymore, which is in line with the theoretical limitations we observe in Theorem 18.2. Figure 18.5 illustrates that the methods do not detect the signal, since the norm of $\eta$ is largely under-evaluated for small sample sizes. These limitations are less pronounced when $n \geq d^2$. In this case one could use alternatively the re-averaging approach from Sect. 18.3.3.2 (not investigated in the simulations) to obtain honest results without the quantum shape constraint.

## 18.5 Proofs

### *18.5.1 Proof of Theorem 18.1*

*Proof* Before we define the algorithm and prove the result, a few preparatory remarks are required: Our sequential procedure will be implemented in $m = 1, 2, \ldots, T$ potential steps, in each of which $2 \cdot 2^m = 2^{m+1}$ measurements are taken. The arguments below will show that we can restrict the search to at most

$$T = O(\log(d/\epsilon))$$

steps. We also note that from the discussion after (18.7)—in particular since $c = c(\delta)$ from (18.8) is $O(1/\delta^2)$—a simple union bound over $m \leq T$ implies that the RIP holds with probability $\geq 1 - \delta'$, some $\delta' > 0$, *simultaneously* for every $m \leq T$ satisfying $2^m \geq c' k d \overline{\log} d$, and with $\tau_{2^m}(k) < c_0$, where $c'$ is a constant that depends on $\delta'$, $c_0$ only. The maximum over $T = O(\log(d/\epsilon))$ terms is absorbed in a slightly enlarged poly-log term. Hence, simultaneously for all such sample sizes $2^m$, $m \leq T$, a nuclear norm regulariser exists that achieves the optimal rate from (18.26) with $n = 2^m$ and for every $k \leq d$, with probability greater than $1 - \delta/3$. Projecting this estimator onto $\Theta_+$ changes the Frobenius error only by a universal multiplicative constant (arguing as in (18.43) below), and we denote by $\tilde{\theta}_{2^m} \in \Theta_+$ the resulting estimator computed from a sample of size $2^m$.

We now describe the algorithm at the $m$-th step: Split the $2^{m+1}$ observations into two halves and use the first subsample to construct $\tilde{\theta}_{2^m} \in \Theta_+$ satisfying (18.26) with $\mathbb{P}_\theta$-probability $\geq 1 - \delta/3$. Then use the other $2^m$ observations to construct a confidence set $C_{2^m}$ for $\theta$ centred at $\tilde{\theta}_{2^m}$: if $2^m < d^2$ we take $C_{2^m}$ from (18.33) and if $2^m \geq d^2$ we take $C_{2^m}$ from (18.37)—in both cases of non-asymptotic coverage at least $1-\alpha, \alpha = \delta/(3T)$. If $|C_{2^m}|_F \leq \epsilon$ we terminate the procedure ($m = \hat{m}$, $\hat{n} = 2^{\hat{m}+1}$, $\hat{\theta} = \tilde{\theta}_{2^{\hat{m}}}$), but if $|C_{2^m}|_F > \epsilon$ we repeat the above procedure with $2 \cdot 2^{m+1} = 2^{m+1+1}$ new measurements, etc., until the algorithm terminates, in which case we have used

$$\sum_{m \leq \hat{m}} 2^{m+1} \lesssim 2^{\hat{m}} \approx \hat{n}$$

measurements in total.

To analyse this algorithm, recall that the quantile constants $z, z_\alpha, \xi_\alpha$ appearing in the confidence sets (18.33) and (18.37) for our choice of $\alpha = \delta/(3T)$ grow at most as $O(\log(1/\alpha)) = O(\log T) = o(\overline{\log} d)$. In particular in view of (18.26) and (18.34) or (18.38) the algorithm necessarily stops at a 'maximal sample size' $n = 2^{T+1}$ in which the squared Frobenius risk of the maximal model ($k = d$) is controlled at level $\epsilon$. Such $T \in \mathbb{N}$ is $O(\log(d/\epsilon))$ and depends on $\sigma, d, \epsilon, \delta$, hence can be chosen by the experimenter.

To prove that this algorithms works we show that the event

$$\left\{\|\hat{\theta} - \theta\|_F^2 > \epsilon^2\right\} \cup \left\{\hat{n} > \frac{C(\delta)kd(\log d)^\gamma}{\epsilon^2}\right\} = A_1 \cup A_2$$

has probability at most $2\delta/3$ for large enough $C(\delta), \gamma$. By the union bound it suffices to bound the probability of each event separately by $\delta/3$. For the first: Since $\hat{n}$ has been selected we know $|C_{\hat{n}}|_F \leq \epsilon$ and since $\hat{\theta} = \tilde{\theta}_{\hat{n}}$ the event $A_1$ can only happen when $\theta \notin C_{\hat{n}}$. Therefore

$$\mathbb{P}_\theta(A_1) \leq \mathbb{P}_\theta(\theta \notin C_{\hat{n}}) \leq \sum_{m=1}^{T} \mathbb{P}_\theta(\theta \notin C_{2^m}) \leq \delta \frac{T}{3T} = \frac{\delta}{3}.$$

For $A_2$, whenever $\theta \in R(k)$ and for all $m \leq T$ for which $2^m \geq c' kd \overline{\log} d$, we have, as discussed above, from (18.34) or (18.38) and (18.26) that

$$\mathbb{E}_\theta |C_{2^m}|_F^2 \leq D' \frac{kd \log T}{2^m},$$

where $D'$ is a constant. In the last inequality the expectation is taken under the distribution of the sample used for the construction of $C_{2^m}$, and it holds on the event on which $\tilde{\theta}_{2^m}$ realises the risk bound (18.26). Then let $C(\delta)$, $\gamma$ be large enough so that $C(\delta)kd(\log d)^\gamma/\epsilon^2 \geq c'kd\overline{\log}d$ and let $m_0 \in \mathbb{N}$ be the smallest integer such that

$$2^{m_0} > \frac{C(\delta)kd(\log d)^\gamma}{\epsilon^2}.$$

Then, for $C(\delta)$ large enough and since $T = O(\log(d/\epsilon)$,

$$\mathbb{P}_\theta\left(\hat{n} > \frac{C(\delta)kd(\log d)^\gamma}{\epsilon^2}\right) \leq \mathbb{P}_\theta\left(|C_{2^{m_0}}|_F^2 > \epsilon^2\right) \leq \frac{\mathbb{E}_\theta|C_{2^{m_0}}|_F^2}{\epsilon^2} \leq \frac{D'\log T}{C(\delta)(\log d)^\gamma} < \delta/3,$$

by Markov's inequality, completing the proof. ■

*Remark 18.9 (Isotropic Sampling)* The proof above works analogously for isotropic designs as defined in Condition 18.1a). When $2^m \geq d^2$, we replace the confidence set (18.37) in the above proof by the confidence set from (18.36). Assuming also that $\|\theta\|_F \leq M$ for some fixed constant $M$, we can construct a similar upper bound for $T$ and the above proof applies directly (with $T$ of slighter larger but still small enough order). Instead of assuming an upper bound on $\|\theta\|_F$ one can simply continue using the confidence set (18.33) also when $2^m \geq d^2$, in which case one has the slightly worse bound

$$\hat{n} \leq C(\delta)\max\left(\frac{kd\overline{\log}d}{\epsilon^2}, \frac{1}{\epsilon^4}\right)$$

for the number of measurements required.

## 18.5.2 Proof of Theorem 18.2

*Proof* By Lemma 18.1 below with $\vartheta = \tilde{\theta} - \theta$ the $\mathbb{P}_\theta$-probability of the complement of the event

$$\mathcal{E} = \left\{\left|\frac{1}{n}\|\mathcal{X}(\tilde{\theta} - \theta)\|^2 - \|\tilde{\theta} - \theta\|_F^2\right| \leq \max\left(\frac{\|\theta - \tilde{\theta}\|_F^2}{2}, \frac{zd}{n}\right)\right\}$$

is bounded by the deviation terms $2e^{-cn}$ and $2e^{-C(K)z}$, respectively (note $z = 0$ in Case (a)). We restrict to this event in what follows. We can decompose

$$\hat{r}_n = \frac{1}{n}\|\mathcal{X}(\tilde{\theta} - \theta)\|^2 + \frac{2}{n}\langle\varepsilon, \mathcal{X}(\theta - \tilde{\theta})\rangle + \frac{1}{n}\sum_{i=1}^n(\varepsilon_i^2 - \mathbb{E}\varepsilon_i^2) = A + B + C.$$

Since $\mathbb{P}(Y + Z < 0) \le \mathbb{P}(Y < 0) + \mathbb{P}(Z < 0)$ for any random variables $Y, Z$ we can bound the probability

$$\mathbb{P}_\theta(\theta \notin C_n, \mathcal{E}) = \mathbb{P}_\theta\left(\left\{\frac{1}{2}\|\theta - \tilde{\theta}\|_F^2 > A + B + C + \frac{zd}{n} + \frac{\bar{z} + \xi_{\alpha/3,\sigma}}{\sqrt{n}}\right\}, \mathcal{E}\right)$$

by the sum of the following probabilities

$$I := \mathbb{P}_\theta\left(\left\{\frac{1}{2}\|\theta - \tilde{\theta}\|_F^2 > \frac{1}{n}\|\mathcal{X}(\tilde{\theta} - \theta)\|^2 + \frac{zd}{n}\right\}, \mathcal{E}\right),$$

$$II := \mathbb{P}_\theta\left(\left\{-\frac{1}{\sqrt{n}}\langle \varepsilon, \mathcal{X}(\theta - \tilde{\theta})\rangle > \bar{z}\right\}, \mathcal{E}\right),$$

$$III := \mathbb{P}_\theta\left(-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\varepsilon_i^2 - \mathbb{E}\varepsilon_i^2) > \xi_{\alpha/3,\sigma}\right).$$

The first probability $I$ is bounded by

$$\mathbb{P}_\theta\left(\left\{-\frac{1}{n}\|\mathcal{X}(\tilde{\theta} - \theta)\|^2 + \|\theta - \tilde{\theta}\|_F^2 > \frac{1}{2}\|\theta - \tilde{\theta}\|_F^2 + \frac{zd}{n}\right\}, \mathcal{E}\right)$$
$$\le \mathbb{P}_\theta\left(\left\{\left|\frac{1}{n}\|\mathcal{X}(\tilde{\theta} - \theta)\|^2 - \|\tilde{\theta} - \theta\|_F^2\right| > \max\left(\frac{\|\theta - \tilde{\theta}\|_F^2}{2}, \frac{zd}{n}\right)\right\}, \mathcal{E}\right) = 0$$

About term $II$: Conditional on $\mathcal{X}$ the variable $\frac{1}{\sqrt{n}}\langle \varepsilon, \mathcal{X}(\theta - \tilde{\theta})\rangle$ is centred Gaussian with variance $(\sigma^2/n)\|\mathcal{X}(\theta - \tilde{\theta})\|^2$. The standard Gaussian tail bound then gives by definition of $\bar{z}$, and conditional on $\mathcal{X}$,

$$\le \exp\{-\bar{z}^2/2(\sigma^2/n)\|\mathcal{X}(\theta - \tilde{\theta})\|^2\}$$
$$= \exp\left\{-\frac{z_{\alpha/3}\max(3\|\theta - \tilde{\theta}\|_F^2, 4zd/n)}{2\|\mathcal{X}(\theta - \tilde{\theta})\|^2/n}\right\} \le \exp\{-z_{\alpha/3}\} = \alpha/3$$

since, on the event $\mathcal{E}$,

$$\max(3\|\theta - \tilde{\theta}\|_F^2, 4zd/n) \ge (2/n)\|\mathcal{X}(\theta - \tilde{\theta})\|^2.$$

The overall bound for $II$ follows from integrating the last but one inequality over the distribution of $X$. Term $III$ is bounded by $\alpha/3$ by definition of $\xi_{\alpha,\sigma}$. ■

*Remark 18.10 (Modification of the Proof for Bernoulli Errors)* If instead of Gaussian errors we work with the error model from Sect. 18.2.2.2, we require a modified treatment of the terms $II, III$ in the above proof. For the pure noise term $III$

we modify the quantile constants slightly to $\xi_{\alpha,\sigma} = \sqrt{(1/\alpha)}$. If the number $T$ of preparations satisfies $T \geq 4d^2$ then Chebyshev's inequality and (18.24) give

$$\mathbb{P}_\theta\left(\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n(\varepsilon_i^2 - \mathbb{E}\varepsilon_i^2)\right| > \xi_{\alpha/3,\sigma}\right) \leq \frac{\alpha}{3n}\sum_{i=1}^n \mathbb{E}\varepsilon_i^4 \leq \frac{\alpha}{3}\frac{4d^2}{T} \leq \frac{\alpha}{3}.$$

For the 'cross term' we have likewise with $z_\alpha = \sqrt{1/\alpha}$ and $a_i = (\mathcal{X}(\theta - \tilde{\theta}))_i$ that, on the event $\mathcal{E}$,

$$\mathbb{P}_\varepsilon\left(\left\{-\frac{1}{\sqrt{n}}\langle\varepsilon, \mathcal{X}(\theta - \tilde{\theta})\rangle > \bar{z}\right\}, \mathcal{E}\right) \leq \frac{1}{n\bar{z}^2}\mathbb{E}_\varepsilon\left(\sum_{i=1}^n \varepsilon_i a_i 1_{\mathcal{E}}\right)^2 \leq \frac{d}{T\bar{z}^2}\frac{\|\mathcal{X}(\theta - \tilde{\theta})\|^2}{n}1_{\mathcal{E}} \leq \alpha/3,$$

just as at the end of the proof of Theorem 18.2, so that coverage follows from integrating the last inequality w.r.t. the distribution of $X$. The scaling $T \approx d^2$ is similar to the one discussed in Theorem 3 in Ref. [12].

**Lemma 18.1**

*(a) For isotropic design from Condition 18.1(a) and any fixed matrix $\vartheta \in \mathbb{H}_d(\mathbb{C})$ we have, for every $n \in \mathbb{N}$,*

$$\Pr\left(\left|\frac{1}{n}\|\mathcal{X}\vartheta\|^2 - \|\vartheta\|_F^2\right| > \frac{\|\vartheta\|_F^2}{2}\right) \leq 2e^{-cn}.$$

*In the standard Gaussian design case we can take $c = 1/24$.*

*(b) In the 'Pauli basis' case from Condition 18.1(b) we have for any fixed matrix $\vartheta \in \mathbb{H}_d(\mathbb{C})$ satisfying the Schatten-1-norm bound $\|\vartheta\|_{S_1} \leq 2$ and every $n \in \mathbb{N}$,*

$$\Pr\left(\left|\frac{1}{n}\|\mathcal{X}\vartheta\|^2 - \|\vartheta\|_F^2\right| > \max\left(\frac{\|\vartheta\|_F^2}{2}, z\frac{d}{n}\right)\right) \leq 2\exp\{-C(K)z\}$$

*where $C(K) = 1/[(16 + 8/3)K^2]$, and where $K$ is the coherence constant of the basis.*

*(c) In the 'Pauli basis' case from Condition 18.1(b) we have for any fixed matrix $\vartheta \in \mathbb{H}_d(\mathbb{C})$ such that the rank of $\vartheta$ is smaller than $2k$ and every $n \in \mathbb{N}$,*

$$\Pr\left(\left|\frac{1}{n}\|\mathcal{X}\vartheta\|^2 - \|\vartheta\|_F^2\right| > \max\left(\frac{\|\vartheta\|_F^2}{2}, z\frac{d}{n}\right)\right) \leq 2\exp\left\{-\frac{n}{17K^2k^2d}\right\}.$$

*Proof* We first prove the isotropic case. From (18.5) we see

$$\Pr\left(\left|\frac{1}{n}\|\mathcal{X}\vartheta\|^2 - \|\vartheta\|_F^2\right| > \|\vartheta\|_F^2/2\right) = \Pr\left(\left|\sum_{i=1}^n(Z_i^2 - \mathbb{E}Z_1^2)/\|\vartheta\|_F^2\right| > n/2\right)$$

where the $Z_i/\|\vartheta\|_F$ are sub-Gaussian random variables. Then the $Z_i^2/\|\vartheta\|_F^2$ are sub-exponential and we can apply Bernstein's inequality (Prop. 4.1.8 in Ref. [14]) to the last probability. We give the details for the Gaussian case and derive explicit constants. In this case $g_i := Z_i/\|\vartheta\|_F \sim N(0, 1)$ so the last probability is bounded, using Theorem 4.1.9 in Ref. [14], by

$$\Pr\left(\left|\sum_{i=1}^{n}(g_i^2 - 1)\right| > \frac{n}{2}\right) \leq 2\exp\left\{-\frac{n^2/4}{4n + 2n}\right\},$$

and the result follows.

Under Condition 18.1(b), if we write $D = \max(n\|\vartheta\|_F^2/2, zd)$ we can reduce likewise to bound the probability in question by

$$\Pr\left(\left|\sum_{i=1}^{n}(Y_i - \mathbb{E}Y_1)\right| > D\right)$$

where the $Y_i = |tr(X^i\vartheta)|^2$ are i.i.d. bounded random variables. Using $\|E_i\|_{op} \leq K/\sqrt{d}$ from Condition 18.1(b) and the quantum constraint $\|\vartheta\|_F \leq \|\vartheta\|_{S_1} \leq 2$ we can bound

$$|Y_i| \leq d^2 \max_i \|E_i\|_{op}^2 \|\vartheta\|_{S_1}^2 \leq 4K^2 d := U$$

as well as

$$\mathbb{E}Y_i^2 \leq U\mathbb{E}|Y_i| \leq 4K^2 d\|\vartheta\|_F^2 := s^2.$$

Bernstein's inequality for bounded variables (e.g., Theorem 4.1.7 in Ref. [14]) applies to give the bound

$$2\exp\left\{-\frac{D^2}{2ns^2 + \frac{2}{3}UD}\right\} \leq 2\exp\{-C(K)z\},$$

after some basic computations, by distinguishing the two regimes of $D = n\|\vartheta\|_F^2/2 \geq zd$ and $D = zd \geq n\|\vartheta\|_F^2/2$.

Finally for (c), using the same reasoning as above and using $\|E_i\|_{op} \leq K/\sqrt{d}$ from Condition 18.1(b) and the fact that the estimator is also of rank less than $k$, we have $\|\vartheta\|_F \leq \|\vartheta\|_{S_1} \leq \sqrt{2k}\|\vartheta\|_F$ we can bound

$$|Y_i| \leq d^2 \max_i \|E_i\|_{op}^2 \|\vartheta\|_{S_1}^2 \leq 2K^2 kd\|\vartheta\|_F^2 := \tilde{U}$$

as well as

$$\mathbb{E}Y_i^2 \leq \tilde{U}\mathbb{E}|Y_i| \leq 2K^2dk^2\|\vartheta\|_F^4 := \tilde{s}^2.$$

Bernstein's inequality for bounded variables (e.g., Theorem 4.1.7 in Ref. [14]) applies to give the bound

$$2\exp\left\{-\frac{D^2}{2n\tilde{s}^2 + \frac{2}{3}\tilde{U}D}\right\} \leq 2\exp\left\{-\frac{n}{17K^2k^2d}\right\},$$

after some basic computations. ■

### 18.5.3 Proof of Theorem 18.3

*Proof* Since $\mathbb{E}_\theta \hat{R}_n = \|\theta - \tilde{\theta}\|_F^2$ we have from Chebyshev's inequality

$$\mathbb{P}_\theta(\theta \notin C_n) \leq \mathbb{P}_\theta\left(|\hat{R}_n - \mathbb{E}\hat{R}_n| > z_{\alpha,n}\right)$$
$$\leq \frac{\mathrm{Var}_\theta(\hat{R}_n - \mathbb{E}\hat{R}_n)}{z_{\alpha_n}^2}.$$

Now $U_n = \hat{R}_n - \mathbb{E}_\theta\hat{R}_n$ is a centred U-statistic and has Hoeffding decomposition $U_n = 2L_n + D_n$ where

$$L_n = \frac{1}{n}\sum_{i=1}^n\sum_{m,k}(Y_iX^i_{m,k} - \mathbb{E}_\theta[Y_iX^i_{m,k}])(\Theta_{m,k} - \tilde{\Theta}_{m,k})$$

is the linear part and

$$D_n = \frac{2}{n(n-1)}\sum_{i<j}\sum_{m,k}(Y_iX^i_{m,k} - \mathbb{E}_\theta[Y_iX^i_{m,k}])(Y_jX^i_{m,k} - \mathbb{E}[Y_jX^i_{m,k}])$$

the degenerate part. We note that $L_n$ and $D_n$ are orthogonal in $L^2(\mathbb{P}_\theta)$.

The linear part can be decomposed into

$$L_n = L_n^{(1)} + L_n^{(2)}$$

where

$$L_n^{(1)} = \frac{1}{n}\sum_{i=1}^n\sum_{m,k}\left(\sum_{m',k'}X^i_{m',k'}X^i_{m,k}\Theta_{m',k'} - \Theta_{m,k}\right)(\Theta_{m,k} - \tilde{\Theta}_{m,k})$$

and

$$L_n^{(2)} = \frac{1}{n}\sum_{i=1}^{n} \varepsilon_i \sum_{m,k} X_{m,k}^i (\Theta_{m,k} - \tilde{\Theta}_{m,k}).$$

Now by the i.i.d. assumption we have

$$\mathrm{Var}_\theta(L_n^{(2)}) = \sigma^2 \frac{\|\tilde{\theta} - \theta\|_F^2}{n}.$$

Moreover, by transposing the indices $m, k$ and $m', k'$ in an arbitrary way into single indices $M = 1, \ldots, d^2, K = 1, \ldots, d^2, d^2 = p$, respectively, basic computations given before eq. (28) in Ref. [30] imply that the variance of the second term is bounded by

$$\mathrm{Var}_\theta(L_n^{(1)}) \leq \frac{c\|\theta - \tilde{\theta}\|_F^2 \|\theta\|_F^2}{n}$$

where $c$ is a constant that depends only on $\mathbb{E}X_{1,1}^4$ (which is finite since the $X_{1,1}$ are sub-Gaussian in view of Condition 18.1(a)). Moreover, the degenerate term satisfies

$$\mathrm{Var}_\theta(D_n) \leq c\frac{d}{n^2}\|\theta\|_F^4$$

in view of standard $U$-statistic computations leading to eq. (6.6) in Ref. [21], with $d^2 = p$, and using the same transposition of indices as before. This proves coverage by choosing the constants in the definition of $z_{\alpha,n}$ large enough. ■

## 18.5.4 Proof of Theorem 18.4

We prove the result for symmetric matrices with real entries—the case of Hermitian matrices requires only minor (mostly notational) adaptations.

Given the estimator $\tilde{\theta}_{\mathrm{Pilot}}$, we can easily transform it into another estimator $\tilde{\theta}$ for which the following is true.

**Theorem 18.5** *There exists an estimator $\tilde{\theta}$ that satisfies, uniformly in $\theta \in R(k)$, for any $k \leq d$ and with $\mathbb{P}_\theta$-probability greater than $1 - 2\delta/3$,*

$$\|\tilde{\theta} - \theta\|_F \leq r_n(k),$$

*as well as,*

$$\tilde{\theta} \in R(k),$$

*and then also*

$$\|\tilde{\theta} - \theta\|_{S_1} \leq \sqrt{2k} r_n(k).$$

*Proof* Let $\tilde{\theta}_{\text{Pilot}}$ and let $\tilde{\theta}$ be the element of $R(d)$ with smallest rank $k'$ such that

$$\|\tilde{\theta}_{\text{Pilot}} - \tilde{\theta}\|_F^2 \leq \frac{r_n^2(k')}{4}.$$

Such $\tilde{\theta}$ exists and has rank $\leq k$, with probability $\geq 1 - 2\delta/3$, since $\theta \in R(k)$ satisfies the above inequality in view of (18.41). The $\|\cdot\|_F^2$-loss of $\tilde{\theta}$ is no larger than $r_n(k)$ by the triangle inequality

$$\|\tilde{\theta} - \theta\|_F \leq \|\tilde{\theta} - \tilde{\theta}_{\text{Pilot}}\|_F + \|\tilde{\theta}_{\text{Pilot}} - \theta\|_F,$$

and this completes the proof of the third claim in view of (18.2). ■

The rest of the proof consists of three steps: The first establishes some auxiliary empirical process type results, which are then used in the second step to construct a sufficiently good simultaneous estimate of the eigenvalues of $\theta$. In Step III the coverage of the confidence set is established.

#### 18.5.4.1 Step I

Let $\theta \in R^+(k) = R(k) \cap \Theta_+$ and let $\tilde{\theta}$ be the estimator from Theorem 18.5. Then with probability $\geq 1 - 2\delta/3$, and if $\eta = \tilde{\theta} - \theta$, we have

$$\|\eta\|_F^2 \leq r_n^2(k) \quad \forall \theta \in R^+(k), \tag{18.45}$$

and that

$$\eta \in R(2k).$$

For the rest of the proof we restrict in what follows to the event of probability greater than or equal to $1 - 2\delta/3$ described by (a) and (b) in the hypothesis of the theorem.

Write $Y_i' = Y_i - \text{tr}(X^i \tilde{\theta})$ for the 'new observations'

$$Y_i' = tr(X^i \eta) + \varepsilon_i, \quad i = 1, \dots, n.$$

For any $d \times d'$ matrix $V$ we set

$$\tilde{\gamma}_\eta(V) = V^T \left( \frac{1}{n} \sum_{i=1}^{n} X^i Y_i' \right) V$$

which estimates

$$\gamma_\eta(V) = V^T \eta V.$$

Let now $U$ be any unit vector in $\mathbb{R}^d$. Then in the above notation ($d' = 1$) we can write

$$\begin{aligned}
\tilde{\gamma}_\eta(U) &= \frac{1}{n}\sum_{i=1}^{n}\sum_{m,m' \le d} U_m U_{m'} X^i_{m,m'} Y'_i \\
&= \frac{1}{n}\sum_{i=1}^{n}\sum_{m,m' \le d} U_m U_{m'} X^i_{m,m'} (tr(X^i \eta) + \varepsilon_i) \\
&= \frac{1}{n}\sum_{i=1}^{n}\sum_{m,m' \le d} U_m U_{m'} X^i_{m,m'} \left(\sum_{k,k' \le d} X^i_{k,k'} \eta_{k,k'} + \varepsilon_i\right).
\end{aligned}$$

If $\mathbb{U}$ denotes the $d \times d$ matrix $UU^T$, the last quantity can be written as

$$\frac{1}{n}\langle \mathcal{X}\mathbb{U}, \mathcal{X}\eta\rangle + \frac{1}{n}\langle \mathcal{X}\mathbb{U}, \varepsilon\rangle.$$

We can hence bound, for $\mathcal{S} = \{U \in \mathbb{R}^d : \|U\|_2 = 1\}$

$$\begin{aligned}
&\sup_{\eta \in R(2k), \|\eta\|_F \le r_n(k), U \in \mathcal{S}} |\tilde{\gamma}_\eta(U) - \gamma_\eta(U)| \\
&\le \sup_{\eta \in R(2k), \|\eta\|_F \le r_n(k), U \in \mathcal{S}} \left|\frac{1}{n}\langle \mathcal{X}\mathbb{U}, \mathcal{X}\eta\rangle - \langle \mathbb{U}, \eta\rangle\right| + \sup_{U \in \mathcal{S}} \left|\frac{1}{n}\langle \mathcal{X}\mathbb{U}, \varepsilon\rangle\right|.
\end{aligned}$$

**Lemma 18.2** *The right hand side on the last inequality is, with probability greater than* $1 - \delta$, *of order*

$$v_n := O\left(r_n(k)\tau_n(k) + \sqrt{\frac{d}{n}}\right).$$

*Proof* The first term in the bound corresponds to the first supremum on the right hand side of the last inequality, and follows directly from the matrix RIP (and Lemma 18.4). For the second term we argue conditionally on the values of $\mathcal{X}$ and on the event for which the matrix RIP is satisfied. We bound the supremum of the Gaussian process

$$\mathbb{G}_\varepsilon(U) := \frac{1}{\sqrt{n}}\langle \mathcal{X}\mathbb{U}, \varepsilon\rangle \sim N(0, \|\mathcal{X}\mathbb{U}\|^2/n)$$

indexed by elements $U$ of the unit sphere $\mathcal{S}$ of $\mathbb{R}^d$, which satisfies the metric entropy bound

$$\log N(\delta, \mathcal{S}, \|\cdot\|) \lesssim d \log(A/\delta)$$

by a standard covering argument. Moreover $\mathbb{U} = UU^T \in R(1)$ and hence for any pair of vectors $U, \bar{U} \in \mathcal{S}$ we have that $\mathbb{U} - \bar{\mathbb{U}} \in R(2)$. From the RIP we deduce for every fixed $U, \bar{U} \in \mathcal{S}$ that

$$\frac{1}{n}\|\mathcal{X}\mathbb{U} - \mathcal{X}\bar{\mathbb{U}}\|^2 = \|\mathbb{U} - \bar{\mathbb{U}}\|_F^2 \left(1 + \frac{\frac{1}{n}\|\mathcal{X}(\mathbb{U} - \bar{\mathbb{U}})\|^2 - \|\mathbb{U} - \bar{\mathbb{U}}\|_F^2}{\|\mathbb{U} - \bar{\mathbb{U}}\|_F^2}\right)$$
$$\leq (1 + \tau_n(2))\|\mathbb{U} - \bar{\mathbb{U}}\|_F^2 \leq C\|U - \bar{U}\|^2$$

since $\tau_n(2) = O(1)$ and since

$$\|\mathbb{U}-\bar{\mathbb{U}}\|_F^2 = \sum_{m,m'}(U_m U_{m'} - \bar{U}_m \bar{U}_{m'})^2 = \sum_{m,m'}(U_m U_{m'} - U_m \bar{U}_{m'} + U_m \bar{U}_{m'} - \bar{U}_m \bar{U}_{m'})^2 \leq 2\|U - \bar{U}\|^2.$$

Hence any $\delta$-covering of $\mathcal{S}$ in $\|\cdot\|$ induces a $\delta/C$ covering of $\mathcal{S}$ in the intrinsic covariance $d_{\mathbb{G}_\varepsilon}$ of the (conditional on $\mathcal{X}$) Gaussian process $\mathbb{G}_\varepsilon$, i.e.,

$$\log N(\delta, \mathcal{S}, d_{\mathbb{G}_\varepsilon}) \lesssim d \log(A'/\delta)$$

with constants independent of $X$. By Dudley's metric entropy bound (e.g., Ref. [14]) applied to the conditional Gaussian process we have for $d > 0$ some constant

$$\mathbb{E} \sup_{U \in \mathcal{S}} |\mathbb{G}_\varepsilon(U)| \lesssim \int_0^d \sqrt{\log N(\delta, \mathcal{S}, d_{\mathbb{G}_\varepsilon})} d\delta \lesssim \sqrt{d}$$

and hence we deduce that

$$\mathbb{E}_\varepsilon \sup_{U \in \mathcal{S}} \frac{1}{n} |\langle \mathcal{X}\mathbb{U}, \varepsilon \rangle| = \mathbb{E}_\varepsilon \frac{1}{\sqrt{n}} \sup_{U \in \mathcal{S}} |\mathbb{G}_\varepsilon(U)| \lesssim \sqrt{\frac{d}{n}} \tag{18.46}$$

with constants independent of $X$, so that the result follows from applying Markov's inequality. ■

#### 18.5.4.2 Step II

Define the estimator

$$\hat{\theta}' = \tilde{\theta} + \frac{1}{n}\sum_{i=1}^{n} X^i Y_i' = \tilde{\theta} + \tilde{\gamma}_\eta(I_d).$$

Then we can write, using $U^T\tilde{\gamma}_\eta(I_d)U = \tilde{\gamma}_\eta(U)$,

$$U^T\hat{\theta}'U - U^T\theta U = U^T(\tilde{\theta} + \tilde{\gamma}_\eta(I_d))U - U^T(\tilde{\theta} + \eta)U$$
$$= \tilde{\gamma}_\eta(U) - \gamma_\eta(U),$$

and from the previous lemma we conclude, for any unit vector $U$ that with probability $\geq 1 - \delta$,

$$|U^T\hat{\theta}'U - U^T\theta U| \leq v_n.$$

Let now $\hat{\theta}$ be any symmetric positive definite matrix such that

$$|U^T\hat{\theta}U - U^T\hat{\theta}'U| \leq v_n.$$

Such a matrix exists, for instance $\theta \in R^+(k)$, and by the triangle inequality we also have

$$|U^T\hat{\theta}U - U^T\theta U| \leq 2v_n. \tag{18.47}$$

**Lemma 18.3** *Let $M$ be a symmetric positive definite $d \times d$ matrix with eigenvalues $\lambda_j$'s ordered such that $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_d$. For any $j \leq d$ consider an arbitrary collection of $j$ orthonormal vectors $\mathcal{V}_j = (V^\iota : 1 \leq \iota \leq j)$ in $\mathbb{R}^d$. Then we have*

$$(a)\;\; \lambda_{j+1} \leq \sup_{U\in\mathcal{S}, U\perp span(\mathcal{V}_j)} U^T M U,$$

*and*

$$(b)\quad \sum_{\iota\leq j}\lambda_\iota \geq \sum_{\iota\leq j}(V^\iota)^T M V^\iota.$$

Let $\hat{R}$ be the rotation that diagonalises $\hat{\theta}$ such that $\hat{R}^T\hat{\theta}\hat{R} = diag(\hat{\lambda}_j : j = 1, \ldots, d)$ ordered such that $\hat{\lambda}_j \geq \hat{\lambda}_{j+1}\ \forall j$. Moreover let $R$ be the rotation that does the same for $\theta$ and its eigenvalues $\lambda_j$. We apply the previous lemma with $M = \hat{\theta}$ and $\mathcal{V}$ equal to the column vectors $r_\iota : \iota \leq l-1$ of $R$ to obtain, for any fixed $l \leq j \leq d$,

$$\hat{\lambda}_l \leq \sup_{U\in\mathcal{S}, U\perp span(r_\iota, \iota\leq l-1)} U^T\hat{\theta}U, \tag{18.48}$$

and also that

$$\sum_{l\leq j}\hat{\lambda}_l \geq \sum_{l\leq j} r_l^T\hat{\theta}r_l. \tag{18.49}$$

From (18.47) we deduce, that

$$\hat{\lambda}_l \leq \sup_{U \in \mathcal{S}, U \perp span(r_\iota, \iota \leq j-1)} U^T \theta U + 2v_n = \lambda_j + 2v_n \quad \forall\, l \leq j,$$

as well as

$$\sum_{l \leq j} \hat{\lambda}_l \geq \sum_{l \leq j} r_l^T \theta r_l - 2jv_n = \sum_{l \leq j} \lambda_l - 2jv_n,$$

with probability $\geq 1 - \delta$. Combining these bounds we obtain

$$\left| \sum_{l \leq j} \hat{\lambda}_l - \sum_{l \leq j} \lambda_l \right| \leq 2jv_n, \quad j \leq d. \tag{18.50}$$

#### 18.5.4.3 Step III

We show that the confidence sets covers the true parameter on the event of probability $\geq 1 - \delta$ on which Steps I and II are valid, and for the constant $C$ chosen large enough.

Let $\Pi = \Pi_{R^+(2\hat{k})}$ be the projection operator onto $R^+(2\hat{k})$. We have

$$\|\hat{\vartheta} - \theta\|_{S_1} \leq \|\hat{\vartheta} - \Pi\theta\|_{S_1} + \|\Pi\theta - \theta\|_{S_1}.$$

We have, using (18.50) and Lemma 18.5 below

$$\begin{aligned}\|\Pi\theta - \theta\|_{S_1} &= \sum_{J > 2\hat{k}} \lambda_J = 1 - \sum_{J \leq 2\hat{k}} \lambda_J \\ &\leq 1 - \sum_{J \leq 2\hat{k}} \hat{\lambda}_J + 4\hat{k}v_n \\ &\leq 6v_n\hat{k} \leq (C/2)\sqrt{\hat{k}}r_n(\hat{k})\end{aligned}$$

for $C$ large enough.

Moreover, using the oracle inequality (18.42) with $S = \Pi\theta$ and (18.43),

$$\begin{aligned}\|\hat{\vartheta} - \Pi\theta\|_{S_1} &\leq \sqrt{4\hat{k}}\|\hat{\vartheta} - \Pi\theta\|_F \\ &\leq \sqrt{4\hat{k}}(\|\hat{\vartheta} - \theta\|_F + \|\Pi\theta - \theta\|_F) \\ &\leq \sqrt{4\hat{k}}(\|\hat{\vartheta} - \tilde{\theta}_{\text{Pilot}}\|_F + \|\tilde{\theta}_{\text{Pilot}} - \theta\|_F + \|\Pi\theta - \theta\|_F) \\ &\lesssim \sqrt{\hat{k}}(r_n(\hat{k}) + \|\Pi\theta - \theta\|_F).\end{aligned}$$

We finally deal with the approximation error: Note

$$\|\Pi\theta - \theta\|_F^2 = \sum_{l>2\hat{k}} \lambda_l^2 \leq \max_{l>2\hat{k}} |\lambda_l| \sum_{l>2\hat{k}} |\lambda_l|.$$

By (18.50) we know that

$$\sum_{l>\hat{k}} \lambda_l = 1 - \sum_{l\leq\hat{k}} \lambda_l \leq 1 - \sum_{l\leq\hat{k}} \hat{\lambda}_l + 2v_n\hat{k} \leq 4v_n\hat{k}.$$

Hence out of the $\lambda_l$'s with indices $l > \hat{k}$ there have to be less than $\hat{k}$ coefficients which exceed $4v_n$. Since the eigenvalues are ordered this implies that the $\lambda_l$'s with indices $l > 2\hat{k}$ are all less than or equal to $4v_n$, and hence the quantity in the last but one display is bounded by (since $\hat{k} < 2\hat{k}$), using again (18.50) and the definition of $\hat{k}$,

$$4v_n\left(1 - \sum_{l\leq\hat{k}} |\lambda_l|\right) \lesssim v_n\left(1 - \sum_{l\leq\hat{k}} |\hat{\lambda}_l|\right) + \hat{k}v_n^2 \lesssim v_n^2\hat{k} \lesssim \sqrt{\hat{k}}r_n(\hat{k}).$$

Overall we get the bound

$$\|\hat{\vartheta} - \Pi\theta\|_{S_1} \lesssim \hat{k}v_n \lesssim (C/2)\sqrt{\hat{k}}r_n(\hat{k})$$

for $C$ large enough, which completes the proof of coverage of $C_n$ by collecting the above bounds. The diameter bound follows from $\hat{k} \leq k$ (in view of the defining inequalities of $\hat{k}$ being satisfied, for instance, for $\tilde{\theta}' = \theta$, whenever $\theta \in R^+(k_0)$.)

## 18.6 Auxiliary Results

### *18.6.1 Proof of Lemma 18.3*

(a) Consider the subspaces $E = span((V^\iota)_{\iota\leq j})^\perp$ and $F = span((e_\iota)_{\iota\leq j+1})$ of $\mathbb{R}^d$, where the $e_\iota$'s are the eigenvectors of the $d \times d$ matrix $M$ corresponding to eigenvalues $\lambda_j$. Since $\dim(E) + \dim(F) = (d - j) + j + 1 = d + 1$, we know that $E \bigcap F$ is not empty and there is a vectorial sub-space of dimension 1 in the intersection. Take $U \in E \bigcap F$ such that $\|U\| = 1$. Since $U \in F$, it can be written as

$$U = \sum_{\iota=1}^{j+1} u_\iota e_\iota$$

for some coefficients $u_\iota$. Since the $e_\iota$'s are orthogonal eigenvectors of the symmetric matrix $M$ we necessarily have

$$MU = \sum_{\iota=1}^{j+1} \lambda_\iota u_\iota e_\iota,$$

and thus

$$U^T MU = \sum_{\iota=1}^{j+1} \lambda_\iota u_\iota^2.$$

Since the $\lambda_\iota$'s are all non-negative and ordered in decreasing absolute value, one has

$$U^T MU = \sum_{\iota=1}^{j+1} \lambda_\iota u_\iota^2 \geq \lambda_{j+1} \sum_{\iota=1}^{j+1} u_\iota^2 = \lambda_{j+1} \|U\|^2 = \lambda_{j+1}.$$

Taking the supremum in $U$ yields the result.

(b) For each $\iota \leq j$, let us write the decomposition of $V^\iota$ on the basis of eigenvectors $(e_l : l \leq d)$ of $M$ as

$$V^\iota = \sum_{l \leq d} v_l^\iota e_l.$$

Since the $(e_l)$ are the eigenvectors of $M$ we have

$$\sum_{\iota \leq j} (V^\iota)^T MV^\iota = \sum_{\iota \leq j} \sum_{l=1}^{d} \lambda_l (v_l^\iota)^2,$$

where $\sum_{l=1}^d (v_l^\iota)^2 = 1$ and $\sum_{\iota \leq j} (v_l^\iota)^2 \leq 1$, since the $V^\iota$ are orthonormal. The last expression is maximised in $(v_l^\iota)_{\iota \leq j, 1 \leq l \leq d}$ and under these constraints, when $v_\iota^\iota = 1$ and $v_l^\iota = 0$ if $\iota \neq l$ (since the $(\lambda_\iota)$ are in decreasing order), and this gives

$$\sum_{\iota \leq j} (V^\iota)^T MV^\iota \leq \sum_{\iota \leq j} \lambda_\iota.$$

### 18.6.2 Some Further Lemmas

**Lemma 18.4** *Under the RIP (18.7) we have for every* $1 \le k \le d$ *that, with probability at least* $1 - \delta$*,*

$$\sup_{A,B \in R(k)} \left| \frac{\frac{1}{n}\langle \mathcal{X}A, \mathcal{X}B\rangle - \langle A, B\rangle_F}{\|A\|_F \|B\|_F} \right| \le 10\tau_n(k). \tag{18.51}$$

*Proof* The matrix RIP can be written as

$$\sup_{A \in R(k)} \left| \frac{\langle \mathcal{X}A, \mathcal{X}A\rangle}{n\langle A, A\rangle_F} - 1 \right| = \frac{|\langle A, (n^{-1}M - \mathbb{I})A\rangle_F|}{\langle A, A\rangle_F} \le \tau_n(k), \tag{18.52}$$

for a suitable $M \in \mathbb{H}_{d^2}(\mathbb{C})$. The above bound then follows from applying the Cauchy-Schwarz inequality to

$$\frac{1}{n}\langle \mathcal{X}A, \mathcal{X}B\rangle - \langle A, B\rangle_F = \langle A, (n^{-1}M - \mathbb{I})B\rangle_F. \tag{18.53}$$

■

The following lemma can be proved by basic linear algebra, and is left to the reader.

**Lemma 18.5** *Let* $M \ge 0$ *with positive eigenvalues* $(\lambda_j)_j$ *ordered in decreasing order. Denote with* $\Pi_{R^+(j-1)}$ *the projection onto* $R^+(j-1) = R(j-1) \cap \Theta_+$*. Then for any* $2 \le j \le d$ *we have*

$$\sum_{j' \ge j} \lambda_{j'} = \|M - \Pi_{R^+(j-1)}M\|_{S_1}.$$

**Acknowledgements** The work of A. Carpentier was done when she was in Cambridge and is partially supported by the Deutsche Forschungsgemeinschaft (DFG) Emmy Noether grant MuSyAD (CA 1488/1-1), by the DFG-314838170, GRK 2297 MathCoRe, by the DFG GRK 2433 DAEDALUS (384950143/GRK2433), and by the DFG CRC 1294 'Data Assimilation', Project A03, and by the UFA-DFH through the French-German Doktorandenkolleg CDFA 01-18. D. Gross acknowledges support by the DFG (SPP1798 CoSIP), Germany's Excellence Strategy—Cluster of Excellence Matter and Light for Quantum Computing (ML4Q) EXC 2004/1—390534769, and the ARO under contract W911NF-14-1-0098 (Quantum Characterization, Verification, and Validation). J. Eisert was supported by the DFG, CRC 183, Project B01, by the DFG GRK DAEDALUS, the DFG SPP1798 CoSIP, the ERC, and the Templeton Foundation.

## References


1. L. Artiles, R. Gill, M. Guta, An invitation to quantum tomography. J. R. Stat. Soc. **67**, 109 (2005)
2. K. Audenaert, S. Scheel, Quantum tomographic reconstruction with error bars: a Kalman filter approach. New J. Phys. **11**(2), 023028 (2009)
3. P.J. Bickel, Y. Ritov, A.B. Tsybakov, Simultaneous analysis of lasso and Dantzig selector. Ann. Stat. **37**(4), 1705–1732 (2009)
4. R. Blume-Kohout, Robust error bars for quantum tomography (2012). arXiv:1202.5270
5. P. Bühlmann, S. van de Geer, *Statistics for High-Dimensional Data: Methods, Theory and Applications* (Springer, Berlin, 2011)
6. E.J. Candès, Y. Plan, Tight oracle inequalities for low-rank matrix recovery from a minimal number of noisy random measurements. IEEE Trans. Inform. Theory **57**(4), 2342–2359 (2011)
7. E.J. Candès, T. Tao, The Dantzig selector: statistical estimation when $p$ is much larger than $n$. Ann. Stat. **35**(6), 2313–2351 (2007)
8. A. Carpentier, A. Kim, An iterative hard thresholding estimator for low rank matrix recovery with explicit limiting distribution (2015). arxiv preprint 1502.04654
9. A. Carpentier, O. Klopp, M. Löffler, R. Nickl, et al., Adaptive confidence sets for matrix completion. Bernoulli **24**(4A), 2429–2460 (2018)
10. M. Christandl, R. Renner, Reliable quantum state tomography. Phys. Rev. Lett. **109**, 120403 (2012)
11. R. De Wolf, A brief introduction to Fourier analysis on the Boolean cube. Theory Comput. Libr. Grad. Surv. **1**, 1–20 (2008)
12. S.T. Flammia, D. Gross, Y.-K. Liu, J. Eisert, Quantum tomography via compressed sensing: error bounds, sample complexity and efficient estimators. New J. Phys. **14**(9), 095022 (2012)
13. E. Giné, R. Nickl, Confidence bands in density estimation. Ann. Stat. **38**, 1122–1170 (2010)
14. E. Giné, R. Nickl, *Mathematical Foundations of Infinite-Dimensional Statistical Models* (Cambridge University Press, Cambridge, 2015)
15. D. Gross, Recovering low-rank matrices from few coefficients in any basis. IEEE Trans. Inf. Theory **57**(3), 1548–1566 (2011)
16. D. Gross, Y.-K. Liu, S.T. Flammia, S. Becker, J. Eisert, Quantum state tomography via compressed sensing. Phys. Rev. Lett. **105**(15), 150401 (2010)
17. M. Guta, T. Kypraios, I. Dryden, Rank-based model selection for multiple ions quantum tomography. New J. Phys. **14**, 105002 (2012)
18. H. Haeffner, W. Haensel, C.F. Roos, J. Benhelm, D.C. al kar, M. Chwalla, T. Koerber, U.D. Rapol, M. Riebe, P.O. Schmidt, C. Becher, O. Gühne, W. Dür, R. Blatt, Scalable multi-particle entanglement of trapped ions. Nature **438**, 643 (2005)
19. M. Hoffmann, R. Nickl, On adaptive inference and confidence bands. Ann. Stat. **39**, 2382–2409 (2011)
20. A.S. Holevo, *Statistical Structure of Quantum Theory* (Springer, Berlin, 2001)
21. Y.I. Ingster, A.B. Tsybakov, N. Verzelen, Detection boundary in sparse regression. Electron. J. Stat. **4**, 1476–1526 (2010)
22. A. Javanmard, A. Montanari, Confidence intervals and hypothesis testing for high-dimensional regression. J. Mach. Learn. Res. **15**(1), 2869–2909 (2014)
23. A. Kalev, R.L. Kosut, I.H. Deutsch, Informationally complete measurements from compressed sensing methodology. arXiv:1502.00536
24. V. Koltchinskii, Von Neumann entropy penalization and low-rank matrix estimation. Ann. Stat. **39**(6), 2936–2973 (2011)
25. V. Koltchinskii, K. Lounici, A.B. Tsybakov, Nuclear-norm penalization and optimal rates for noisy low-rank matrix completion. Ann. Stat. **39**(5), 2302–2329 (2011)
26. H. Leeb, B.M. Pötscher, Can one estimate the conditional distribution of post-model-selection estimators? Ann. Stat. **34**(5), 2554–2591 (2006)

27. U. Leonhardt, *Measuring the Quantum State of Light* (Cambridge University Press, Cambridge, 2005)
28. Y.-K. Liu, Universal low-rank matrix recovery from Pauli measurements. Adv. Neural Inf. Process. Syst. **24**, 1638–1646 (2011)
29. V. Mnih, C. Szepesvári, J.Y. Audibert, Empirical bernstein stopping, in *Proceedings of the 25th International Conference on Machine Learning* (ACM, New York, 2008), pp. 672–679
30. R. Nickl, S. van de Geer, Confidence sets in sparse regression. Ann. Stat. **41**(6), 2852–2876 (2013)
31. M.A. Nielsen, I.L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2000)
32. A. Peres, *Quantum Theory* (Springer, Berlin, 1995)
33. B. Recht, M. Fazel, P.A. Parrilo, Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. SIAM Rev. **52**, 471 (2010)
34. J. Shang, H.K. Ng, A. Sehrawat, X. Li, B.-G. Englert, Optimal error regions for quantum state estimation. New J. Phys. **15**(12), 123026 (2013)
35. K. Temme, F. Verstraete, Quantum chi-squared and goodness of fit testing. J. Math. Phys. **56**(1), 012202 (2015)
36. S. van de Geer, P. Bühlmann, Y. Ritov, R. Dezeure, On asymptotically optimal confidence regions and tests for high-dimensional models. Ann. Stat. **42**(3), 1166–1202 (2014)

# Chapter 19
# Uniform in Bandwidth Estimation of the Gradient Lines of a Density

**David Mason and Bruno Pelletier**

*Dedicated to the memory of Jørgen Hoffmann–Jørgensen*

**Abstract** Let $X_1, \ldots, X_n$, $n \geq 1$, be independent identically distributed (i.i.d.) $\mathbb{R}^d$ valued random variables with a smooth density function $f$. We discuss how to use these $X's$ to estimate the gradient flow line of $f$ connecting a point $x_0$ to a local maxima point (mode) based on an empirical version of the gradient ascent algorithm using a kernel estimator based on a bandwidth $h$ of the gradient $\nabla f$ of $f$. Such gradient flow lines have been proposed to cluster data. We shall establish a uniform in bandwidth $h$ result for our estimator and describe its use in combination with plug in estimators for $h$.

**Keywords** Gradient lines · Density estimation · Nonparametric clustering · Uniform in bandwidth

## 19.1 Introduction

Let $f$ be a differentiable density on $\mathbb{R}^d$. Assuming that $f$ is known, consider the following iterative scheme. Fix $a > 0$ and, starting at $x_0 \in \mathbb{R}^d$, define iteratively the gradient ascent method

$$x_\ell = x_{\ell-1} + a\nabla f(x_{\ell-1}), \quad \text{for } \ell \geq 1.$$

D. Mason (✉)
Department of Applied Economics and Statistics, University of Delaware, Newark, DE, USA
e-mail: davidm@udel.edu

B. Pelletier
Département de Mathématiques, IRMAR – UMR CNRS 6625, Université Rennes II, Upper Brittany, France
e-mail: bruno.pelletier@univ-rennes2.fr



N. Gozlan et al. (eds.), *High Dimensional Probability VIII*,
Progress in Probability 74, https://doi.org/10.1007/978-3-030-26391-1_19

When it exists, define $x_\infty = \lim_{\ell\to\infty} x_\ell$. The rationale behind this iterative gradient ascent scheme is to have the sequence $(x_\ell : \ell \geq 0)$ converge to a local maxima point (mode) of $f$—representing a cluster center.

In fact, one can use this scheme to cluster a set of data by assigning to each observation the nearest mode along the direction of the gradient at the observation point (Fukunaga and Hostetler [7]), where $\nabla f$ is replaced by an estimator $\nabla \widehat{f}$ based on the data. This is close in spirit to Hartigan [9]. It would be interesting to compare this clustering scheme to other clustering methods. However such a study is far outside the scope the present paper.

In practice, the underlying density $f$ is rarely known and has to be estimated using a kernel density estimator. Let $\Phi : \mathbb{R}^d \to \mathbb{R}$ be a kernel function—a non-negative integrable function satisfying $\int_{\mathbb{R}^d} \Phi(x)\mathrm{d}x = 1$—and for a bandwidth $0 < h \leq 1$, let $\Phi_h(u) = h^{-d}\Phi(u/h)$. The corresponding kernel estimator of $f$ based on a random sample $X_1, \ldots, X_n$, i.i.d. with density $f$, is

$$\hat{f}_{n,h}(x) := \frac{1}{n}\sum_{i=1}^{n} \Phi_h(x - X_i), \tag{19.1}$$

and if $\Phi$ is differentiable, then we estimate the gradient of $f$ by the kernel type estimator

$$\nabla \hat{f}_{n,h}(x) := \frac{1}{nh}\sum_{i=1}^{n} \nabla\Phi_h(x - X_i).$$

We shall establish a general uniform in bandwidth $h$ result in a sense to be soon made precise in Sect. 19.2 for the sequence of estimators beginning with $\hat{x}_0 = x_0$

$$\hat{x}_\ell = \hat{x}_{\ell-1} + a\nabla \hat{f}_{n,h}(\hat{x}_{\ell-1}), \quad \text{for } \ell \geq 1.$$

In our results, we are not interested in studying how to choose the starting point $x_0$. Rather, given a starting point $x_0$ our aim is to consider the flow line of a function with the property that it starts at $x_0$ and ends at an isolated local maxima point $x^\star$, and estimate this flow line from a random sample.

Before we can do this we must first establish some notation and state two general results.

### *19.1.1 Two General Results*

Let $g : \mathbb{R}^d \to \mathbb{R}$ be differentiable. Starting at $x_0 \in \mathbb{R}^d$, for fixed $a > 0$ we relate the sequence

$$x_\ell = x_{\ell-1} + a\nabla g(x_{\ell-1}), \quad \text{for } \ell \geq 1, \tag{19.2}$$

with the gradient ascent line of $g$ starting at $x_0$. We study convergence of this sequence towards the gradient ascent line of $g$ starting at $x_0$. In particular, we characterize the limit $x_\infty$, providing a consistency result for the clustering algorithm based on the local maxima point of $g$. Then, given another differentiable function $\widehat{g}$, meant to approximate $g$, we compare the sequence $(x_\ell)$ to $(\hat{x}_\ell)$, where

$$\hat{x}_\ell = \hat{x}_{\ell-1} + a\nabla\widehat{g}(\hat{x}_{\ell-1}), \quad \text{for } \ell \geq 1, \tag{19.3}$$

starting at the same point $\hat{x}_0 = x_0$. In particular, when estimating the gradient ascent lines of a density $f$ based on a sample $X_1, \ldots, X_n$, $\widehat{g}$ can be taken to be some kernel estimator $\widehat{f}$ of $f$.

Recall that a *critical point* of $g$ is a point $x^*$ at which the gradient of $g$ vanishes, that is, such that $\nabla g(x^*) = 0$. A *flow line* or *integral curve* of the positive gradient flow of $g$ is a curve $x$ such that its derivative $x'(t)$ satisfies the differential equation

$$x'(t) = \nabla g(x(t)). \tag{19.4}$$

Note that, along any flow line, the value of $g$ increases, that is, the function $t \mapsto g(x(t))$ is increasing with $t$. By the theory of ordinary differential equations, through any point $x_0 \in \mathbb{R}^d$ passes a unique flow line $x(t)$ defined for $t \in [0, t_0)$, where $t_0 > 0$, such that $x(0) = x_0$ (see Section 7.2 of Hirsch et al. [10]); we say that $x(t)$ is the flow line starting at $x_0$. Let $x^\star$ be a critical point of $g$. We say that $x_0$ is in the attraction basin of $x^\star$ if the flow line $x(t)$ starting at $x_0$ is defined for all $t \geq 0$ and $\lim_{t\to\infty} x(t) = x^\star$. An accumulation point of a sequence of points through an integral curve $x(t)$, i.e., a sequence of the form $\{x(t_n) : t_1 < t_2 < \ldots\}$, $t_n \to \infty$, is called a limit point of $x(t)$. Any limit point of a gradient flow line of $g$ is necessarily a critical point of $g$.

We start by stating a general result by Arias-Castro et al. [1] (also see [2] and the remark at the end of this section) who established the convergence of the gradient ascent scheme (19.2) towards the flow lines of the underlying function $g$. Starting from a point $x_0$ in the attraction basin of an isolated local maxima point $x^\star$, under some conditions stated below, the iteration (19.2) converges to $x^\star$. By an isolated local maxima point $x^\star$ we mean that for all $\epsilon > 0$ small enough the open ball of radius $\epsilon$ around $x^\star$, $B(x^\star, \epsilon)$, contains no local maxima point other than $x^\star$. We will show that in fact, the polygonal line defined by the sequence $(x_\ell)$ is uniformly close to the flow line starting at $x_0$ and ending at $x^\star$.

**Theorem 19.1 (Convergence of Gradient Ascent Method)** *Let $g$ be a function of class $C^3$. Let $(x(t) : t \geq 0)$ denote the flow line of $g$ starting at $x_0$ and ending at an isolated local maxima point $x^\star$ of $g$. Let $(x_\ell)$ be the sequence defined in* (19.2) *starting at $x_0$. Then there exists $A = A(x_0, g) > 0$ such that, whenever $a < A$,*

$$\lim_{\ell\to+\infty} x_\ell = x^\star. \tag{19.5}$$

*Denote by* $x_a(t)$ *the following polygonal line*

$$x_a(t) = x_{\ell-1} + (t/a - \ell + 1)(x_\ell - x_{\ell-1}), \quad \forall t \in [(\ell-1)a, \ell a).$$

*Assume* $H_g(x^\star)$*, the Hessian of* $g$ *evaluated at* $x^\star$*, has all eigenvalues in* $(-\overline{\nu}, -\underline{\nu})$ *for some* $0 < \underline{\nu} < \overline{\nu}$*. Then, there exists a* $C_0 = C(x_0, g, \underline{\nu}, \overline{\nu}) > 0$ *such that, for any* $0 < a < A$*,*

$$\sup_{t \geq 0} \|x_a(t) - x(t)\| \leq C_0 a^\delta, \quad \textit{with } \delta := \underline{\nu} / \left(\underline{\nu} + \overline{\nu}\right). \tag{19.6}$$

Next, we state a version of a stability result of [1] for flows of smooth functions. Under some conditions, when $g$ and $\widehat{g}$ are close as $C^2$ functions, then their flow lines are also close. First we need some notation.

For a function $\varphi : \mathbb{R}^d \to \mathbb{R}$, we let $\varphi^{(\ell)}(x)$, $\ell \geq 1$, denote the differential form of $\varphi$ of order $\ell$ at a point $x \in \mathbb{R}^d$, and let $H_\varphi(x)$ denote the Hessian matrix of $\varphi$ evaluated at $x$ when they exist. The differential form $\varphi^{(\ell)}(x)$ of $\varphi$ at $x$ is the multilinear map from $\mathbb{R}^d \times \cdots \times \mathbb{R}^d$ ($\ell$ times) to $\mathbb{R}$ defined for $\ell \geq 1$ by

$$\varphi^{(\ell)}(x)[u_1, \ldots, u_\ell] = \sum_{i_1, \ldots, i_\ell = 1}^{d} \frac{\partial^\ell \varphi(x)}{\partial x_{i_1} \ldots \partial x_{i_\ell}} u_{1,i_1} \ldots u_{\ell, i_\ell},$$

where, for each $1 \leq i \leq \ell$, $u_i$ has components $u_i = (u_{i,1}, \ldots, u_{i,d})$. We write

$$\varphi^{(0)}(x) = \varphi(x), \ x \in \mathbb{R}^d.$$

Given a multilinear map $L$ of order $\ell \geq 1$ from $\mathbb{R}^d \times \cdots \times \mathbb{R}^d$ to $\mathbb{R}$, which we write as

$$L[u_1, \ldots, u_\ell] = \sum_{i_1, \ldots, i_\ell = 1}^{d} L_{i_1, \ldots, i_\ell} u_{1,i_1} \ldots u_{\ell, i_\ell}.$$

we denote by $\|L\|$ its operator norm defined by

$$\|L\| = \sup \{ |L[u_1, \ldots, u_\ell]| : \|u_1\| = \cdots = \|u_\ell\| = 1 \}. \tag{19.7}$$

Note that when $\ell = 1$, $\|L\| = \sqrt{\sum_{i=1}^d L_i^2}$, and when $\ell = 2$

$$\|L\| = \sup_{\|u\| = \|v\| = 1} \left| v' L u \right| = \sup_{\|u\| = 1} |Lu|,$$

where $L$ is the $d \times d$ matrix $\{L_{i,j} : 1 \le i, j \le d\}$, (cf. page 7 of Bhatia [3]), which implies that for any $x \in \mathbb{R}^d$

$$|Lx| \le \|L\| \|x\|. \tag{19.8}$$

When $\ell = 0$ we set $\|L\| = |L|$.

We denote by $\|L\|_{\max}$ the norm defined by

$$\|L\|_{\max} = \max\{|L_{i_1 \ldots i_\ell}| : 1 \le i_1, \ldots, i_\ell \le d\}. \tag{19.9}$$

We note for future reference that easy calculations show that

$$\|L\|_{\max} \le \|L\| \le d^{\frac{\ell}{2}} \|L\|_{\max}. \tag{19.10}$$

For a set $S \subset \mathbb{R}^d$, we define

$$\kappa_\ell(\varphi, S) = \sup_{x \in S} \left\| \varphi^{(\ell)}(x) \right\|. \tag{19.11}$$

Note that $\kappa_\ell(\varphi, S)$ is well-defined and is finite when $\varphi$ is of class $C^\ell$ and $S$ is compact.

The *upper level set* of a function $\varphi : \mathbb{R}^d \to \mathbb{R}$ at $b \in \mathbb{R}$ is defined as

$$\mathcal{L}_\varphi(b) = \{x \in \mathbb{R}^d : \varphi(x) \ge b\}. \tag{19.12}$$

We suppress the dependence on $\varphi$ whenever no confusion is possible. For any $x \in \mathbb{R}^d$ and $r > 0$ denote the open ball

$$B\,(x, r) = \{y : \|x - y\| < r\}$$

and the closed ball

$$\overline{B}\,(x, r) = \{y : \|x - y\| \le r\}\,.$$

Here is our stability result. It is a version of Theorem 2 of [1] designed to prove our uniform in bandwidth result stated as Theorem 19.3 in the next section.

**Theorem 19.2 (Stability of Smooth Flows)** *Suppose $g$ and $\widehat{g}$ are of class $C^3$. Let $(x(t) : t \ge 0)$ be a flow line of $g$ starting at $x_0$, with $g(x_0) > 0$, and ending at an isolated local maxima point $x^\star$ where $H_g(x^\star)$ has all eigenvalues in $(-\overline{\nu}, -\underline{\nu})$ for some $0 < \underline{\nu} < \overline{\nu}$. Let $\hat{x}(t)$ be the flow line of $\widehat{g}$ starting at $x_0$. Let $S = \mathcal{L}(g(x_0)/2) \cap \overline{B}(x_0, 3r_0)$, where*

$$r_0 = \max_t \|x(t) - x_0\|, \tag{19.13}$$

*and define*

$$\eta_m = \sup_{x \in S} \|g^{(m)}(x) - \widehat{g}^{(m)}(x)\|.$$

*Then for all* $D > 0$ *there exists a constant* $C := C(g, x_0, \underline{\nu}, \bar{\nu}, D) \geq 1$ *and a function* $F(g, x_0, \underline{\nu}, \bar{\nu}, 1/C, D)$ *of* $D$ *such that, whenever* $\max\{\eta_0, \eta_1, \eta_2\} \leq 1/C$ *and* $\eta_3 \leq D$, $\hat{x}(t)$ *is defined for all* $t \geq 0$ *and*

$$\sup_{t \geq 0} \|x(t) - \hat{x}(t)\| \leq F(g, x_0, \underline{\nu}, \bar{\nu}, 1/C, D) \max\left\{\sqrt{\eta_0}, \eta_1^{\delta}\right\}, \tag{19.14}$$

*where* $\delta = \underline{\nu}/\left(\underline{\nu} + \overline{\nu}\right)$.

Combining Theorems 19.1 and 19.2, we arrive at the following bound for approximating the flow lines of a function $g$ with the polygonal line obtained from the gradient ascent algorithm (19.3) based on an approximation $\widehat{g}$ to $g$.

**Corollary 19.1** *In the context of Theorem 19.2, for* $a > 0$, *define*

$$\hat{x}_a(t) = \hat{x}_{\ell-1} + (t/a - \ell + 1)(\hat{x}_\ell - \hat{x}_{\ell-1}), \quad \forall t \in [(\ell-1)a, \ell a), \tag{19.15}$$

*where* $(\hat{x}_\ell)$ *is defined in* (19.3). *Then for all* $D > 0$ *there exists a constant* $C := C(g, x_0, \underline{\nu}, \bar{\nu}, D) \geq 1$ *and a function* $F(g, x_0, \underline{\nu}, \bar{\nu}, 1/C, D)$ *of* $D$ *such that, whenever* $\max\{\eta_0, \eta_1, \eta_2\} \leq 1/C$ *and* $\eta_3 \leq D$,

$$\sup_{t \geq 0} \|\hat{x}_a(t) - x(t)\| \leq F(g, x_0, \underline{\nu}, \bar{\nu}, 1/C, D)\left[a^{\delta} + \max\left\{\sqrt{\eta_0}, \eta_1^{\delta}\right\}\right], \tag{19.16}$$

*where* $\delta = \underline{\nu}/\left(\underline{\nu} + \overline{\nu}\right)$.

In applications, the requirement that $g(x_0) > 0$ can be sidestepped.

*Remark 19.1* In Claim C on page 3 of [2], recall that $t_\ell$ is defined as $t_\ell = a\ell$. Therefore in the sentence "Let $\ell_\epsilon$ be such that $\|x(t_{\ell_\epsilon}) - x^\star\| \leq \epsilon/2$", in fact $\ell_\epsilon$ depends on $a$. But since the conclusion is for any small enough $a$, the conclusion $\|x_{\ell_\epsilon} - x^\star\| \leq \epsilon$ is potentially incorrect.

The argument requires the following modification starting after the second sentence in Claim C of [2]:

Since $x(t) \to x^\star$ as $t \to \infty$, there is $t_\epsilon$ such that $\|x(t) - x^\star\| \leq \epsilon/2$ for all $t \geq t_\epsilon$. With $a \leq A_1$, let $\ell_{\epsilon,a}$ be the smallest integer such that $a\ell_{\epsilon,a} \geq t_\epsilon$. Note that $a\ell_{\epsilon,a} \leq t_\epsilon + 1$ for all $a \leq A_1$. By equation (33) of [2], for all $a \leq A_1$, we have

$$\|x(a\ell_{\epsilon,a}) - x_{\ell_{\epsilon,a}}\| \leq \left[e^{a\ell_{\epsilon,a}\kappa_2\sqrt{d}} - 1\right]\kappa_1 a \leq \left[e^{(t_\epsilon+1)\kappa_2\sqrt{d}} - 1\right]\kappa_1 a.$$

Let $a_\epsilon$ be such that

$$\left[e^{(t_\epsilon+1)\kappa_2\sqrt{d}} - 1\right]\kappa_1 a_\epsilon = \epsilon/2.$$

Assume now that $a \le A_1 \wedge a_\epsilon$. Then, by the triangle inequality, we get

$$\|x_{\ell_{\epsilon,a}} - x^\star\| \le \|x_{\ell_{\epsilon,a}} - x(a\ell_{\epsilon,a})\| + \|x(a\ell_{\epsilon,a}) - x^\star\| \le \epsilon.$$

The remainder of the proof remains unchanged with $\ell_\epsilon$ replaced by $\ell_{\epsilon,a}$.

## 19.2 The Estimation of Gradient Lines of a Density

Let $\hat{f}_{n,h}$ be the kernel density estimator of $f$ in (19.1) with kernel $\Phi$ and bandwidth $h$. Sharp almost-sure convergence rates in the uniform norm of kernel density estimators have been obtained by several authors, for example Einmahl and Mason [5], Giné and Guillou [8], Einmahl and Mason [6], Mason and Swanepoel [12] (also see [13]) and Mason [11].

We first state a bias bound from [1].

**Lemma 19.1** *Assume $\Phi$ is nonnegative, $C^3$ on $\mathbb{R}^d$ with all partial derivatives up to order 3 vanishing at infinity, and satisfies*

$$\int_{\mathbb{R}^d} \Phi(x)\mathrm{d}x = 1, \quad \int_{\mathbb{R}^d} x\Phi(x)\mathrm{d}x = 0 \quad \textit{and} \quad \int_{\mathbb{R}^d} \|x\|^2\Phi(x)\mathrm{d}x < \infty. \tag{19.17}$$

*Then for any $C^3$ density $f$ on $\mathbb{R}^d$ with bounded derivatives up to order 3, there is a constant $C > 0$ such that for all $0 \le \ell \le 3$*

$$\sup_{x\in\mathbb{R}^d} \left\|\mathbb{E}\big[\hat{f}^{(\ell)}_{n,h}(x)\big] - f^{(\ell)}(x)\right\| \le Ch^{(3-\ell)\wedge 2}. \tag{19.18}$$

Next, by applying the main result of [12] (also see [13] and Theorem 4.1 with Remark 4.2 in [11]), [1] derive the following uniform in bandwidth result for $\hat{f}_{n,h}$ and its derivatives.

**Lemma 19.2** *Suppose that $\Phi$ is of the form $\Phi : (x_1, \dots, x_d) \mapsto \prod_{k=1}^d \phi_k(x_k)$, and that each $\phi_k$ is nonnegative, integrates to 1, and is $C^3$ on $\mathbb{R}$ with derivatives up to order 3 being of bounded variation and in $L_1(\mathbb{R}^d)$. Then, for any bounded density $f$ on $\mathbb{R}^d$, there exists a $0 < b_0 < 1$ such that almost surely*

$$\limsup_{n\to\infty} \sup_{\frac{\log n}{n} \le h^d \le b_0} \sup_{x\in\mathbb{R}^d} \sqrt{\frac{nh^{d+2\ell}}{\log n}} \left\|\hat{f}^{(\ell)}_{n,h}(x) - \mathbb{E}\left[\hat{f}^{(\ell)}_{n,h}(x)\right]\right\| < \infty, \quad \forall 0 \le \ell \le 3. \tag{19.19}$$

It is straightforward to design a kernel that satisfies the conditions of Lemmas 19.1 and 19.2. In fact, the Gaussian kernel $\Phi(x) = (2\pi)^{-d/2} \exp(-\|x\|^2/2)$ is such a kernel.

**Theorem 19.3** *Consider a density $f$ satisfying the conditions of Lemma 19.1. Suppose $\hat{f}_{n,h}$ is a kernel estimator of $f$ of the form (19.1), where $\Phi$ satisfies the conditions of Lemmas 19.1 and 19.2. Let $(x(t) \ : \ t \geq 0)$ be the flow line of $f$ starting at a point $x_0$ with $f(x_0) > 0$, ending at an isolated local maxima point $x^\star$ where $H_f(x^\star)$ has all eigenvalues in $(-\overline{\nu}, -\underline{\nu})$ for some $0 < \underline{\nu} < \overline{\nu}$. For fixed $a > 0$, $0 < h \leq 1$ and $n \geq 1$ define $(\hat{x}_a(t, n, h) \ : \ t \geq 0)$ as in (19.15) with $\hat{f}$ taken as $\hat{f}_{n,h}$ in (19.3). i.e. for $t \in [(\ell-1)a, \ell a)$, $\ell \geq 1$,*

$$\hat{x}_{\ell,n}(h) = \hat{x}_{\ell-1,n}(h) + a\nabla \hat{f}_{n,h}(\hat{x}_{\ell-1.n}(h)),$$

*with $\hat{x}_{0,n}(h) = x_0$. Suppose that*

$$c_n \to 0, \ \frac{nc_n^{1+6/d}}{\log n} \to \infty \text{ and } c_n < b_n, \text{ with } b_n \to 0, \tag{19.20}$$

*then there exists a constant $C > 0$ such that, with probability one, for all $n$ large enough, uniformly in $c_n \leq h^d \leq b_n$,*

$$\sup_{t \geq 0} \|\hat{x}_a(t, n, h) - x(t)\| \leq C\left(a^\delta + h^{2\delta}\right), \tag{19.21}$$

*where $\delta = \underline{\nu}/\left(\underline{\nu} + \overline{\nu}\right)$.*

*Remark 19.1* Let

$$\hat{h}_n = H_n(X_1, \ldots, X_n)$$

be a bandwidth estimator so that with probability 1

$$\hat{h}_n \to 0 \text{ and } \liminf_n \frac{\hat{h}_n^d}{c_n} > 0,$$

where $c_n$ satisfies the conditions in (19.20). Notice that under the assumptions and notation of Theorem 19.3 we have, with probability 1, for the *plug in* estimator $\hat{x}_a(t, n, \hat{h}_n)$, for all large enough $n$,

$$\sup_{t \geq 0} \|\hat{x}_a(t, n, \hat{h}_n) - x(t)\| \leq C\left(a^\delta + \hat{h}_n^{2\delta}\right). \tag{19.22}$$

Note that in Theorem 3, $a$ denotes the (fixed) step size of the gradient ascent scheme while $(c_n)$ is the lower bound imposed on the bandwidth sequence.

For a general treatment of bandwidth selection and data-driven bandwidths consult Sections 2.3 and 2.4 of Deheuvels and Mason [4], as well as the references therein.

*Remark 19.2* This nonparametric approach to the estimation of the flow line will converge no faster than the convergence rate of the estimators of third order derivatives of the density function. So it may potentially suffer from some curse of dimensionality for which the usual approaches could apply, e.g. structural assumptions (that could allow efficient dimension reduction as suggested by the AE), or smoothness assumptions.

## 19.3 Proofs of Theorems 19.2 and 19.3

To show the reader how all of these results fit together, we shall prove Theorem 19.3 first.

### *19.3.1 Proof of Theorem 19.3*

As in the proof of Theorem 19.2 in the next subsection, we may assume without loss of generality that $\mathcal{L}_g(f(x_0/2)) \subset \overline{B}(x_0, 3r_0)$, with $r_0 = \sup_{t\geq 0} \|x(t) - x_0\|$, which implies that $\mathcal{L}(f(x_0/2)$ is compact.

For any integer $0 \leq \ell \leq 3$, $n \geq 1$ and $0 < h \leq 1$, let

$$\eta_{\ell,n}(h) = \sup_{x\in S} \|\hat{f}^{(\ell)}_{n,h}(x) - f^{\ell}(x)\|,$$

where the norm used is defined in (19.7). From (19.18) and (19.19), we see from the triangle inequality that for some constant $A_\ell > 0$, uniformly in $c_n \leq h^d \leq b_n$, for all large $n$

$$\eta_{\ell,n}(h) \leq A_\ell \left( h^{(3-\ell)\wedge 2} + \sqrt{\frac{\log n}{nh^{d+2\ell}}} \right)$$

$$\leq A_\ell \left( b_n^{(3-\ell)\wedge 2} + \sqrt{\frac{\log n}{nc_n^{1+2\ell/d}}} \right).$$

It is easily checked using (19.20) that for any $0 \leq \ell \leq 2$

$$\sup_{c_n \leq h^d \leq b_n} \eta_{\ell,n}(h) \to 0, \text{ a.s.},$$

while

$$\limsup_{n\to\infty} \sup_{c_n \le h^d \le b_n} \eta_{3,n}(h) \le A_3, \text{ a.s.}$$

Also one finds that uniformly in $c_n \le h^d \le b_n$ for all large $n$ for some constant $B > 0$

$$h^{(3-\ell)\wedge 2} + \sqrt{\frac{\log n}{nh^{d+2\ell}}} \le Bh^2, \text{ for } \ell = 0, 1.$$

Thus since $\delta < 1/2$, uniformly in $c_n \le h^d \le b_n$ for all $n$ large enough,

$$\max\{\sqrt{\eta_{0,n}(h)}, \eta_{1,n}^{\delta}(h)\} \le Ah^{2\delta},$$

with $A = \max\{\sqrt{A_0 B}, (A_1 B)^{\delta}\}$. We finish the proof by applying Corollary 19.1. □

### *19.3.2 Proof of Theorem 19.2*

Our proof will follow that of Theorem 2 of [1], however with some major modifications and clarifications needed to obtain the present result. We shall require the following two lemmas, which we state here without proof. They are respectively Lemma 5 and 6 of Theorem 2 of [1].

**Lemma 19.3** *Suppose that $g$ is of class $C^3$. Let $x^\star$ be an isolated local maxima point of $g$ where $H_g(x^\star)$ has all eigenvalues in $(-\overline{\nu}, -\underline{\nu})$ with $\overline{\nu} > \underline{\nu} > 0$. For $\epsilon > 0$, let $\mathcal{C}(\epsilon)$ be the connected component of $\mathcal{L}_g(g(x^\star) - \epsilon)$ that contains $x^\star$. Then there is a constant $C_3 = C_3(g, x^\star)$ such that*

$$\overline{B}(x^\star, \sqrt{(2\epsilon/\overline{\nu})}) \subset \mathcal{C}(\epsilon) \subset \overline{B}(x^\star, \sqrt{2\epsilon/\underline{\nu}}), \quad \textit{for all } \epsilon \le C_3, \tag{19.23}$$

*and*

$$g(x^\star) - g(x) \le \frac{\overline{\nu}}{2}\|x - x^\star\|^2, \quad \textit{for all } x \textit{ such that } \|x - x^\star\| \le \sqrt{C_3/\overline{\nu}}. \tag{19.24}$$

**Lemma 19.4** *Suppose that $g$ is of class $C^3$. Let $(x(t) : t \ge 0)$ be the flow line of $g$ starting at $x_0$ and ending at $x^\star$ where $H_g(x^\star)$ has all its eigenvalues in $(-\infty, -\underline{\nu})$, with $\underline{\nu} > 0$. Then, there is $C_4 = C_4(g, x_0)$ such that, for all $t \ge 0$,*

$$\|x(t) - x^\star\| \le C_4 e^{-\underline{\nu} t}, \tag{19.25}$$

*and*

$$g(x^\star) - g(x(t)) \le C_4 e^{-2\underline{\nu} t}. \tag{19.26}$$

The following, adapted from Hirsch et al. [10, Section 17.5], is a stability result for autonomous gradient flows.

**Lemma 19.5** *Suppose $\varphi$ and $\psi$ are of class $C^1$ and for a measurable subset $\mathcal{S} \subset \mathbb{R}^d$*

$$\|\nabla\varphi(x) - \nabla\psi(x)\| < \varepsilon, \quad \forall x \in \mathcal{S}.$$

*Let $K$ be a Lipschitz constant for $\nabla\varphi$ on $\mathcal{S}$. Let $(x(t) : t \ge t_0)$ and $(y(t) : t \ge t_0)$ with $t_0 \ge 0$, be the flow lines of $\varphi$ and $\psi$ starting at $x_1$ and $y_1$, respectively, i.e. $x(t_0) = x_1$ and $y(t_0) = y_1$, and*

$$x'(t) = \nabla\varphi(x(t)) \text{ and } y'(t) = \nabla\psi(y(t)), \text{ for } t \ge t_0.$$

*Assume that the flow lines $x(t)$ and $y(t)$ are in $\mathcal{S}$. Then,*

$$\|x(t) - y(t) - (x_1 - y_1)\| \le \frac{\varepsilon}{K}\left[e^{Kt} - 1\right], \quad \forall t \ge t_0.$$

For the convenience of the reader we state here the Weyl Perturbation Theorem (see Corollary III.2.6 of Bhatia [3]).

**Weyl Perturbation Theorem** Let $M$ and $H$ be $n$ by $n$ Hermitian matrices, where $M$ has eigenvalues $\mu_1 \ge \cdots \ge \mu_n$ and $H$ has eigenvalues $\nu_1 \ge \cdots \ge \nu_n$. If $\|M - H\| \le \varepsilon$, then $|\mu_i - \nu_i| \le \varepsilon$ for $i = 1, \dots, n$.

Next is a result on the stability of local maxima points.

**Lemma 19.6** *Suppose $f$ and $g$ are of class $C^3$, and have local maxima points at $x$ and $y$, respectively, with $H_f(x)$ having all eigenvalues in $(-\infty, -\nu]$ for some $\nu > 0$. Then for any $0 < b \le 1$ and $\kappa \ge \max\left(\kappa_3(f, \overline{B}(x,b)), \kappa_3(g, \overline{B}(x,b))\right)$,*

$$\|x - y\| \le \min\left\{\frac{3\nu}{4\kappa}, b\right\} \quad \Rightarrow \quad \|x - y\| \le \frac{2}{\sqrt{\nu}}\left(|f(x) - g(x)| + |f(y) - g(y)|\right)^{1/2}. \tag{19.27}$$

*Proof* Let $\mathbf{H}_f$ and $\mathbf{H}_g$ be short for the Hessian matrices $H_f(x)$ and $H_g(y)$, respectively. We develop $f$ and $g$ around $x$ and $y$, respectively. Assuming $\|x - y\| \le \min\left\{\frac{3\nu}{4\kappa}, b\right\}$, which implies that $y \in \overline{B}(x,b)$, we have

$$f(y) = f(x) + \frac{1}{2}\mathbf{H}_f[x-y, x-y] + R_f(x,y), \quad \text{with} \quad |R_f(x,y)| \le \frac{\kappa}{6}\|x-y\|^3;$$

$$g(x) = g(y) + \frac{1}{2}\mathbf{H}_g[x-y, x-y] + R_g(x,y), \quad \text{with} \quad |R_g(x,y)| \le \frac{\kappa}{6}\|x-y\|^3.$$

Summing these two equalities, we obtain

$$\frac{1}{2}(\mathbf{H}_f + \mathbf{H}_g)[x-y, x-y] = f(y) - g(y) + g(x) - f(x) - R_f(x, y) - R_g(x, y).$$

Let $\nu > 0$ be such that the largest eigenvalue of $\mathbf{H}_f$ is bounded by $-\nu$. By the triangle inequality and the fact that $\mathbf{H}_g$ is negative semidefinite,

$$\nu\|x-y\|^2 \leq \|(\mathbf{H}_f + \mathbf{H}_g)[x-y, x-y]\| \leq 2\,|f(x) - g(x)| + 2\,|f(y) - g(y)| + \frac{2\kappa}{3}\|x-y\|^3.$$

Thus, when $\|x-y\| \leq \min\left\{\frac{3\nu}{4\kappa}, b\right\}$, we have $\nu\|x-y\|^2 - \frac{2\kappa}{3}\|x-y\|^3 \geq \frac{\nu}{2}\|x-y\|^2$, so that

$$\|x-y\|^2 \leq \frac{4}{\nu}\left(|f(x) - g(x)| + |f(y) - g(y)|\right),$$

and from this we conclude (19.27). □

*It would help the reader to make his or her way through the intricate arguments that follow to always keep in mind that $\eta_0, \eta_1, \eta_2$ and $\epsilon > 0$ are assumed to be sufficiently small and $t_\epsilon > 0$ sufficiently large as needed, and $\eta_3 \leq D$, where $D > 0$ is a pre-chosen constant.*

**Bound on** $\|\hat{x}^\star - x^\star\|$

Our first goal is to derive a bound on $\|\hat{x}^\star - x^\star\|$. Arguing as in the proof of Theorem 1 of [1], we may assume, without loss of generality [WLOG], that $\mathcal{L}_g(g(x_0)/2) \subset \overline{B}(x_0, 3r_0)$, where $r_0$ is as in (19.13). So from now on, we assume that $\mathcal{L}_g(g(x_0)/2)$ is compact and we set

$$S = \mathcal{L}_g(g(x_0)/2). \tag{19.28}$$

Note that since $g\,(x\,(t))$ increases along $t \geq 0$, $x\,(t) \in S$ for all $t \geq 0$.

We also let $\kappa_\ell$ be short for $\kappa_\ell(g, S)$, as defined in (19.11).

*Claim 19.1* For $\eta_0$ sufficiently small, $\hat{x}(t) \in S$, for all $t \geq 0$, with $S$ as in (19.28). Indeed, suppose there is $t > 0$ such that $\hat{x}(t) \notin S$. Fix $\varrho = g(x_0)/2$. Then, by continuity, there is $0 \leq t' < t$ such that $g(\hat{x}(t')) = g(x_0) - \varrho$. Since both $\hat{x}(t')$ and $x_0 \in S$, we have

$$\begin{aligned}
\widehat{g}(\hat{x}(t')) &= \widehat{g}(\hat{x}(t')) - g(\hat{x}(t')) + g(\hat{x}(t')) \\
&\leq \eta_0 + g(x_0) - \varrho \\
&= \eta_0 + \widehat{g}(x_0) + g(x_0) - \widehat{g}(x_0) - \varrho \\
&\leq \widehat{g}(x_0) + 2\eta_0 - \varrho,
\end{aligned}$$

by the triangle inequality, applied twice. Since $\widehat{g}(\hat{x}(t')) \geq \widehat{g}(x_0)$, we see that this situation does not arise when $\eta_0 < \varrho/2$. This establishes Claim 19.1.

From now on we shall assume that $\eta_0$ is sufficiently small, so that

$$\hat{x}(t) \in S, \text{ for all } t \geq 0. \tag{19.29}$$

*Claim 19.2* For all $\eta_0$, $\eta_1$ *and* $\eta_2$ sufficiently small, $\hat{x}^\star = \lim_{t\to\infty} \hat{x}(t)$ is well defined and is close to $x^\star$. Since $\widehat{g}$ is of class $C^3$ by assumption, the map $x \mapsto \nabla\widehat{g}(x)$ is $C^1$, and since by Claim 19.1 for all $\eta_0$ sufficiently small $\hat{x}(t)$ stays in $S$ and $S$ is compact, $\hat{x}(t)$ is defined for all $t \geq 0$ by the first corollary to the first theorem in [10, Section 17.5].

Applying Lemma 19.5 with $t_0 = 0$ and $x_1 = y_1 = x_0$ we get

$$\|\hat{x}(t) - x(t)\| \leq \frac{\eta_1}{\sqrt{d}\kappa_2} e^{\sqrt{d}\kappa_2 t}, \quad \forall t \geq 0, \tag{19.30}$$

For $\epsilon \in (0, C_3)$, where $C_3$ is as in Lemma 19.3, let $t_\epsilon$ be such that $x(t) \in B(x^\star, \sqrt{(2\epsilon/\overline{\nu})})$ for all $t \geq t_\epsilon$, which is well-defined since $x(t) \to x^\star$ as $t \to \infty$. Hence

$$\begin{aligned}\|\hat{x}(t_\epsilon) - x^\star\| &\leq \|\hat{x}(t_\epsilon) - x(t_\epsilon)\| + \|x(t_\epsilon) - x^\star\| \\ &\leq \frac{\eta_1}{\sqrt{d}\kappa_2} e^{\sqrt{d}\kappa_2 t_\epsilon} + \sqrt{\frac{2\epsilon}{\overline{\nu}}} =: \delta_1.\end{aligned} \tag{19.31}$$

Assume that $\eta_1$ and $\epsilon$ are small enough so that $\delta_1 < \sqrt{C_3/\overline{\nu}}$. Letting $\mathcal{C}(\epsilon)$ be as in Lemma 19.3, by (19.23) we have

$$\overline{B}\left(x^\star, \delta_1\right) \subset \mathcal{C}(\epsilon_1), \text{ with } \epsilon_1 = \frac{\overline{\nu}}{2}\delta_1^2,$$

noting that $\sqrt{\epsilon_1 2/\overline{\nu}} = \delta_1$ and $\epsilon_1 < C_3/2$. Thus $\hat{x}(t_\epsilon)$ belongs to $\mathcal{C}(\epsilon_1)$ and in particular $g(\hat{x}(t_\epsilon)) \geq g(x^\star) - \epsilon_1$. Using this last inequality, we deduce from the triangle inequality and the fact that $t \mapsto \widehat{g}\left(\hat{x}(t)\right)$ is increasing that for $t \geq t_\epsilon$,

$$\begin{aligned}g(\hat{x}(t)) &\geq \widehat{g}(\hat{x}(t)) - \eta_0 \geq \widehat{g}(\hat{x}(t_\epsilon)) - \eta_0 \\ &\geq g(\hat{x}(t_\epsilon)) - 2\eta_0 \geq g(x^\star) - \epsilon_2,\end{aligned}$$

where

$$\epsilon_2 := \epsilon_1 + 2\eta_0. \tag{19.32}$$

Since $\hat{x}(t_\epsilon) \in \mathcal{C}(\epsilon_1) \subset \mathcal{C}(\epsilon_2)$ and $\left(\hat{x}(t) : t \geq t_\epsilon\right)$ is connected and in $\mathcal{L}_g(g(x^\star) - \epsilon_2)$, we necessarily have $\left(\hat{x}(t) : t \geq t_\epsilon\right) \subset \mathcal{C}(\epsilon_2)$. Assume that $\epsilon$, $\eta_0$ and $\eta_1$ are small enough so that $\epsilon_2 \leq C_3$. Then, by Lemma 19.3, $\mathcal{C}(\epsilon_2) \subset \overline{B}\left(x^\star, \sqrt{2\epsilon_2/\underline{\nu}}\right)$, and so

$$\|\hat{x}(t) - x^\star\| \leq \epsilon_3 := \sqrt{2\epsilon_2/\underline{\nu}}, \text{ for all } t \geq t_\epsilon. \tag{19.33}$$

Assume $\epsilon, \eta_0, \eta_1$ are small enough so that $\overline{B}(x^\star, \epsilon_3) \subset S$. For any $x$ and $y$ in $\overline{B}(x^\star, \epsilon_3)$ we get by (19.10) that

$$\|H_g(x) - H_g(y)\| \leq d\|H_g(x) - H_g(y)\|_{\max} \leq d^{3/2}\kappa_3\|x - y\|. \tag{19.34}$$

Using (19.34) and (19.33), for any $x \in \overline{B}(x^\star, \epsilon_3)$

$$\|H_{\widehat{g}}(x) - H_g(x^\star)\| \leq \|H_{\widehat{g}}(x) - H_g(x)\| + \|H_g(x) - H_g(x^\star)\| \tag{19.35}$$

$$\leq \eta_2 + d^{3/2}\kappa_3\|x - x^\star\| \leq \eta_2 + d^{3/2}\kappa_3\epsilon_3. \tag{19.36}$$

Let $\nu > \underline{\nu}$, but close enough such that all the eigenvalues of $\mathbf{H}$ are still in $(-\infty, -\nu)$. We then apply the Weyl Perturbation Theorem, cited above, to conclude that for all $\eta_2$ and $\epsilon_3$ small enough and $x \in \overline{B}(x^\star, \epsilon_3)$ so that

$$\eta_2 + d^{3/2}\kappa_3\epsilon_3 \leq \nu - \underline{\nu} \tag{19.37}$$

the eigenvalues of $H_{\widehat{g}}(x)$ are all in $(-\infty, -\underline{\nu})$. We shall assume that $\epsilon, \eta_0, \eta_1, \eta_2$ are small enough so that this is the case. Using (19.33) and compactness of $\overline{B}(x^\star, \epsilon_3)$, we get by Cantor's intersection theorem that

$$K := \cap_{t \geq t_\epsilon} \overline{\{\widehat{x}(u) : u \geq t\}}$$

is nonempty. In addition $K$ is composed of critical points of $\widehat{g}$. (See [10], Section 9.3, Proposition, p. 206 and Theorem p. 205). Therefore we conclude that $K$ is a singleton, which we denote $\hat{x}^\star$. This is a critical point of $\widehat{g}$ in $\overline{B}(x^\star, \epsilon_3)$ and is the limit of $\widehat{x}(t)$ as $t \to \infty$. Moreover, $\hat{x}^\star$ is a local maxima point of $\widehat{g}$. This proves Claim 19.2.

We have just shown that for $\epsilon > 0$, $\eta_0, \eta_1$ and $\eta_2$ sufficiently small

$$\|\hat{x}^\star - x^\star\| \leq \epsilon_3.$$

To summarize, the analysis from Eqs. (19.30) through (19.37) shows that for all $\epsilon > 0$, $\eta_0, \eta_1$ and $\eta_2$ small enough, $\overline{B}(x^*, \epsilon_3) \subset S$, $\hat{x}^* \in \overline{B}(x^*, \epsilon_3)$, $\eta_2 + d^{3/2}\kappa_3\epsilon_3 \leq \nu - \underline{\nu}$ and (19.33) holds, where

$$\delta_1 = \frac{\eta_1}{\sqrt{d}\kappa_2} e^{\sqrt{d}\kappa_2 t_\epsilon} + \sqrt{\frac{2\epsilon}{\overline{\nu}}},\ \epsilon_1 = \frac{\overline{\nu}}{2}\delta_1^2, \epsilon_2 = \epsilon_1 + 2\eta_0, \tag{19.38}$$

and

$$\epsilon_3 = \sqrt{2\epsilon_2/\overline{\nu}}. \tag{19.39}$$

Notice that $\epsilon_3$ is a function of $(\epsilon, \eta_0, \eta_1, \eta_2)$ and

$$\frac{\nu - \underline{\nu} - \eta_2}{d^{3/2}\kappa_3} \geq \epsilon_3 = \sqrt{\frac{2\,(\epsilon_1 + 2\eta_0)}{\overline{\nu}}} = \sqrt{\frac{2\left(\frac{\overline{\nu}}{2}\delta_1^2 + 2\eta_0\right)}{\overline{\nu}}}.$$

Letting $\kappa = \kappa_3 + \eta_3$ and $b = \epsilon_3$ in Lemma 19.6 we see by (19.27) that whenever

$$\|\hat{x}^\star - x^\star\| \leq \min\left\{\epsilon_3, \frac{3\underline{\nu}}{4\,(\kappa_3 + \eta_3)}\right\},$$

then

$$\|\hat{x}^\star - x^\star\| \leq \frac{2\sqrt{2\eta_0}}{\sqrt{\underline{\nu}}}. \tag{19.40}$$

Clearly when $\eta_3 \leq D$ for some $D > 0$ and $\epsilon_3 \leq \frac{3}{4}\underline{\nu}/\,(\kappa_3 + D)$ then

$$\min\left\{\epsilon_3, \frac{3\underline{\nu}}{4\,(\kappa_3 + \eta_3)}\right\} \geq \min\left\{\epsilon_3, \frac{3\underline{\nu}}{4\,(\kappa_3 + D)}\right\} = \epsilon_3.$$

Putting everything together, we can conclude for every $D > 0$ there exists a constant

$$q_0 := q_0(g, x_0, \underline{\nu}, \bar{\nu}, D) \geq 1$$

such that whenever $\max\{\epsilon, \eta_0, \eta_1, \eta_2\} \leq 1/q_0$ and $\eta_3 \leq D$

$$\|\hat{x}^\star - x^\star\| \leq \frac{2\sqrt{2\eta_0}}{\sqrt{\underline{\nu}}} =: Q_0\sqrt{\eta_0}. \tag{19.41}$$

*Throughout the remainder of the proof, we shall assume $\max\{\epsilon, \eta_0, \eta_1, \eta_2\} \leq 1/q_0$ and $\eta_3 \leq D$ so that (19.41) holds.

**Bound on $\|x(t) - \hat{x}(t)\|$ for Large $t$**

Next we obtain a bound on $\|x(t) - \hat{x}(t)\|$ for large $t > 0$. Let $\mathbf{H}$ and $\hat{\mathbf{H}}$ be short for $H_g(x^\star)$ and $H_{\widehat{g}}(\hat{x}^\star)$, respectively. We proceed with a linearization of the flows near the critical points. Let $\nu > \underline{\nu}$, but close enough such that all the eigenvalues of $\mathbf{H}$ are still in $(-\infty, -\nu)$. By combining (19.36) and (19.41)

$$\|\hat{\mathbf{H}} - \mathbf{H}\| \leq \eta_2 + d^{\frac{3}{2}}\kappa_3 Q_0\sqrt{\eta_0}. \tag{19.42}$$

Choose $\nu > \nu_2 > \nu_1 > \underline{\nu}$. Clearly the eigenvalues of $\mathbf{H}$ are also in $(-\infty, -\nu_2)$. Suppose that $\eta_0$ and $\eta_2$ are small enough that

$$\eta_2 + d^{\frac{3}{2}}\kappa_3 Q_0\sqrt{\eta_0} < \nu - \nu_2.$$

Thus $\|\hat{\mathbf{H}} - \mathbf{H}\| \leq \nu - \nu_2$ and by Weyl's inequality the eigenvalues of $\hat{\mathbf{H}}$ are in

$$(-\infty, -\nu + (\nu - \nu_2)) = (-\infty, -\nu_2). \tag{19.43}$$

Recall that WLOG we assume that $S = \mathcal{L}_g(g(x_0)/2)$. By the definition of $S$, clearly there is an $r_+ > 0$ such that $\bar{B}(x^\star, r_+) \subset S$. Note that for any $D > 0$ fixed the constant $q_0 \geq 1$ can be taken large enough so that (19.29), (19.31), (19.33), (19.34), (19.36) and (19.41) hold simultaneously. Fix an $\epsilon > 0$ small enough so that this is the case, and also such that $\sqrt{\epsilon} < (\sqrt{\underline{\nu}/2})r_+/2$. Recall the constants (19.38) and note that $\epsilon_2 \geq \epsilon$. Then recall by (19.33) there is a $t_\epsilon$ (depending on $\epsilon$ and the trajectory $x(t)$) such that

$$\|\hat{x}(t) - x^\star\| \leq \sqrt{2\epsilon_2/\underline{\nu}}, \qquad \text{for all } t \geq t_\epsilon,$$

which in combination with (19.41) gives

$$\|\hat{x}(t) - \hat{x}^\star\| \leq \sqrt{2\epsilon_2/\underline{\nu}} + Q_0\sqrt{\eta_0}, \qquad \text{for all } t \geq t_\epsilon. \tag{19.44}$$

Also by (19.25) for all $t \geq t_\epsilon$, where $t_\epsilon > 0$ is large enough,

$$\|x(t) - x^\star\| \leq r_+/2. \tag{19.45}$$

We see by (19.41) that when $\eta_0$ and $\eta_1$ are small enough we get $\bar{B}(\hat{x}^\star, r_+/2) \subset \bar{B}(x^\star, r_+)$ and we see by (19.44) that when $\eta_0$ and $\eta_1$ are small enough, $\|\hat{x}(t) - \hat{x}^\star\| \leq r_+/2$ (note that this is possible since we have fixed $\sqrt{\epsilon} < (\sqrt{\underline{\nu}/2})r_+/2$). Setting $r_\ddagger = r_+/2$ and

$$t_\ddagger = t_\epsilon, \tag{19.46}$$

we get that

$$\bar{B}(x^\star, r_\ddagger) \subset S \quad \text{and} \quad \bar{B}(\hat{x}^\star, r_\ddagger) \subset S,$$

and

$$x(t) \in \bar{B}(x^\star, r_\ddagger) \quad \text{and} \quad \hat{x}(t) \in \bar{B}(\hat{x}^\star, r_\ddagger), \qquad \text{for any } t \geq t_\ddagger, \tag{19.47}$$

when $\eta_0$, $\eta_1$, and $\eta_2$ are small enough and $\eta_3 \leq D$, and also keeping (19.45) in mind. (Note that $t_\ddagger$ depends only on $g$ and the trajectory $x(t)$).

Letting

$$x_{\ddagger}(t) = x(t) - x^{\star} \text{ and } \hat{x}_{\ddagger}(t) = \hat{x}(t) - \hat{x}^{\star},$$

by a Taylor expansion, for all $t \geq t_{\ddagger}$ we have

$$x'_{\ddagger}(t) = \nabla f(x(t)) = \mathbf{H}\, x_{\ddagger}(t) + R(t), \quad \text{with} \quad \|R(t)\| \leq \frac{\sqrt{d}\kappa_3}{2} \|x_{\ddagger}(t)\|^2\,; \tag{19.48}$$

$$\hat{x}'_{\ddagger}(t) = \nabla \hat{f}(\hat{x}(t)) = \hat{\mathbf{H}}\, \hat{x}_{\ddagger}(t) + \hat{R}(t), \quad \text{with} \quad \|\hat{R}(t)\| \leq \frac{\sqrt{d}(\kappa_3 + \eta_3)}{2} \|\hat{x}_{\ddagger}(t)\|^2\,. \tag{19.49}$$

The difference gives

$$\begin{aligned} x'_{\ddagger}(t) - \hat{x}'_{\ddagger}(t) &= \mathbf{H}x_{\ddagger}(t) - \widehat{\mathbf{H}}\hat{x}_{\ddagger}(t)) + R(t) - \hat{R}(t) \\ &= \mathbf{H}(x_{\ddagger}(t) - \hat{x}_{\ddagger}(t)) + (\mathbf{H} - \hat{\mathbf{H}})\hat{x}_{\ddagger}(t) + R(t) - \hat{R}(t). \end{aligned} \tag{19.50}$$

*Claim 19.3* We get after integrating (19.50),

$$x_{\ddagger}(t) - \hat{x}_{\ddagger}(t) = -e^{t\mathbf{H}}(x^{\star} - \hat{x}^{\star}) + \int_0^t e^{(t-s)\mathbf{H}}\big[(\mathbf{H} - \hat{\mathbf{H}})\hat{x}_{\ddagger}(s) + R(s) - \hat{R}(s)\big]\mathrm{d}s. \tag{19.51}$$

To check this note that $x_{\ddagger}(0) - \hat{x}_{\ddagger}(0) = x^{\star} - \hat{x}^{\star}$, and differentiating (19.51), we get

$$\begin{aligned} x'_{\ddagger}(t) - \hat{x}'_{\ddagger}(t) = -\mathbf{H}e^{t\mathbf{H}}(x^{\star} - \hat{x}^{\star}) + \mathbf{H}e^{t\mathbf{H}} \int_0^t e^{-s\mathbf{H}} \left[(\mathbf{H} - \hat{\mathbf{H}})\hat{x}_{\ddagger}(s) + R(s) - \hat{R}(s)\right] \mathrm{d}s \\ + (\mathbf{H} - \hat{\mathbf{H}})\hat{x}_{\ddagger}(t) + R(t) - \hat{R}(t). \end{aligned} \tag{19.52}$$

From (19.51), $e^{t\mathbf{H}}(x^{\star} - \hat{x}^{\star})$ may be expressed as

$$e^{t\mathbf{H}}(x^{\star} - \hat{x}^{\star}) = -\left(x'_{\ddagger}(t) - \hat{x}'_{\ddagger}(t)\right) + \int_0^t e^{(t-s)\mathbf{H}}\big[(\mathbf{H} - \hat{\mathbf{H}})\hat{x}_{\ddagger}(s) + R(s) - \hat{R}(s)\big]\mathrm{d}s. \tag{19.53}$$

Putting (19.53) in (19.52) we get (19.50). This verifies Claim 19.3.
Now since all of the eigenvalues of $\mathbf{H}$ are in $(-\infty, -\nu)$ we have

$$\left\| e^{\alpha \mathbf{H}} \right\| \leq e^{-\nu\alpha}, \quad \text{for all } \alpha > 0.$$

Using this fact with the triangle inequality along with (19.8), (19.42) and the inequalities in (19.48) and (19.49) we get

$$\|x_{\ddagger}(t) - \hat{x}_{\ddagger}(t)\|$$

$$\leq e^{-\nu t}\|x^{\star} - \hat{x}^{\star}\| + \int_0^t e^{-\nu(t-s)}\left[\Delta\|\hat{x}_{\ddagger}(s)\| + \sqrt{d}\left(\frac{\kappa_3}{2}\|x_{\ddagger}(s)\|^2 + \frac{\kappa_3+\eta_3}{2}\|\hat{x}_{\ddagger}(s)\|^2\right)\right]\mathrm{d}s, \tag{19.54}$$

where

$$\Delta = \eta_2 + d^{\frac{3}{2}}\kappa_3 Q_0\sqrt{\eta_0}.$$

Recall that by Lemma 19.4, for some $C_4 = C_4(g, x_0)$,

$$\|x_{\ddagger}(t)\| \leq C_4 e^{-\nu_1 t} \text{ for all } t \geq 0. \tag{19.55}$$

*Claim 19.4* For $\epsilon > 0$, $\eta_0$, $\eta_1$, and $\eta_2$ small enough and that $\eta_3 \leq D$ so that (19.41), (19.43) and (19.47) hold, there is a constant $C_4' := C_4'(g, x_0, \underline{\nu}, \bar{\nu}, \epsilon, D)$ such that

$$\|\hat{x}_{\ddagger}(t)\| \leq \max C_4' e^{-\nu_1 t}, \quad \text{for all } t \geq 0. \tag{19.56}$$

*Proof* We assume WLOG that $S = \mathcal{L}_g\,(g\,(x_0)\,/2)$ and is compact. Thus

$$\sup_{x,y\in S}\|x - y\| = L < \infty. \tag{19.57}$$

Let $\hat{\kappa}_3$ be short for $\kappa_3(\widehat{g}, S)$. We have that,

$$\hat{\kappa}_3 \leq \kappa_3 + \eta_3 \leq \kappa_3 + D.$$

We assume that $\epsilon > 0$, $\eta_0$, $\eta_1$, and $\eta_2$ are small enough and that $\eta_3 \leq D$ so that (19.41) and (19.47) hold.

A Taylor expansion of $\nabla\widehat{g}$ at $x \in \bar{B}(\widehat{x}^{\star}, r_0)$ gives

$$\nabla\widehat{g}(x) = \widehat{\mathbf{H}}(x - \widehat{x}^{\star}) + \widehat{R}(x, \widehat{x}^{\star}), \tag{19.58}$$

with

$$\|\widehat{R}(x, \widehat{x}^{\star})\| \leq \widehat{\kappa}_3\frac{\sqrt{d}}{2}\|x - \widehat{x}^{\star}\|^2.$$

Therefore by (19.58) and $\widehat{x}'(t) = \nabla \widehat{g}(\widehat{x}(t))$, we have,

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\widehat{x}(t) - \widehat{x}^\star\right) - \widehat{\mathbf{H}}\left(\widehat{x}(t) - \widehat{x}^\star\right) = \widehat{R}\left(\widehat{x}(t), \widehat{x}^\star\right), \tag{19.59}$$

and since $\widehat{x}(0) = x_0$ and $\widehat{x}(t)$ satisfies the differential Eq. (19.59) it is readily checked that

$$\widehat{x}(t) - \widehat{x}^\star = e^{t\widehat{\mathbf{H}}}(x_0 - \widehat{x}^\star) + \int_0^t e^{(t-s)\widehat{\mathbf{H}}}\widehat{R}\left(\widehat{x}(s), \widehat{x}^\star\right)\mathrm{d}s.$$

Since all the eigenvalues of $\widehat{\mathbf{H}}$ are in $(-\infty, -\nu_2)$ we have

$$\left\| e^{\alpha\widehat{\mathbf{H}}} \right\| \le e^{-\nu_2\alpha}, \quad \text{for all } \alpha > 0.$$

Then,

$$\|\widehat{x}(t) - \widehat{x}^\star\| \le e^{-\nu_2 t}\|\widehat{x}_0 - \widehat{x}^\star\| + \widehat{\kappa}_3\frac{\sqrt{d}}{2}\int_0^t e^{-\nu_2(t-s)}\|\widehat{x}(s) - \widehat{x}^\star\|^2\mathrm{d}s. \tag{19.60}$$

Set

$$\widehat{u}(t) = e^{\nu_2 t}\|\widehat{x}(t) - \widehat{x}^\star\|$$

and

$$\widehat{U}(t) = \|x_0 - \widehat{x}^\star\| + \widehat{\kappa}_3\frac{\sqrt{d}}{2}\int_0^t e^{\nu_2 s}\|\widehat{x}(s) - \widehat{x}^\star\|^2\mathrm{d}s. \tag{19.61}$$

Thus by (19.60), $\widehat{u}(t) \le \widehat{U}(t)$ and $\widehat{U}'(t) = \widehat{\kappa}_3\frac{\sqrt{d}}{2}e^{-\nu_2 t}\widehat{u}^2(t)$, so

$$\begin{aligned}\frac{\widehat{U}'(t)}{\widehat{U}(t)} &= \widehat{\kappa}_3\frac{\sqrt{d}}{2}e^{-\nu_2 t}\widehat{u}(t)\frac{\widehat{u}(t)}{\widehat{U}(t)}\\ &\le \widehat{\kappa}_3\frac{\sqrt{d}}{2}e^{-\nu_2 t}\widehat{u}(t) = \widehat{\kappa}_3\frac{\sqrt{d}}{2}\|\widehat{x}(t) - \widehat{x}^\star\|\\ &\le \frac{\sqrt{d}}{2}(\kappa_3 + D)\|\hat{x}(t) - \hat{x}^\star\|.\end{aligned} \tag{19.62}$$

Recall that $\nu_2 > \nu_1 > \underline{\nu}$. We can choose WLOG $r_\ddagger$ in (19.47) small enough so that

$$r_\ddagger \le \left[\frac{\sqrt{d}}{2}(\kappa_3 + D)\right]^{-1}(\nu_2 - \nu_1).$$

Assuming that this is the case, we get from (19.62)

$$\frac{\hat{U}'(t)}{\hat{U}(t)} \leq \nu_2 - \nu_1, \quad \text{for all } t \geq t_\ddagger.$$

By integrating between $t_\ddagger$ and $t$, we deduce that

$$\log \widehat{U}(t) \leq \log \widehat{U}(t_\ddagger) + (\nu_2 - \nu_1)(t - t_\ddagger),$$

and so

$$\|\hat{x}(t) - \hat{x}^\star\| = e^{-\nu_2 t}\hat{u}(t) \leq e^{-\nu_2 t}\hat{U}(t) \leq c_1 e^{-\nu_1 t}, \quad \text{for all } t \geq t_\ddagger,$$

with

$$c_1 := \widehat{U}(t_\ddagger)e^{-(\nu_2-\nu_1)t_\ddagger}.$$

For $t < t_\ddagger$, we simply have

$$\|\widehat{x}(t) - \widehat{x}^\star\| \leq c_2 e^{-\nu_1 t},$$

where

$$c_2 = \max_{0 \leq t \leq t_\ddagger} \|\widehat{x}(t) - \widehat{x}^\star\| e^{\nu_1 t}.$$

Notice that by (19.57) and (19.61), keeping in mind that we always assume by Claim 19.1 that $\eta_0$ is sufficiently small so that $\hat{x}(t) \in S$, for all $t \geq 0$,

$$\widehat{U}(t_\ddagger) = \|x_0 - \widehat{x}^\star\| + \widehat{\kappa}_3 \frac{\sqrt{d}}{2} \int_0^{t_\ddagger} e^{\nu_2 s} \|\widehat{x}(s) - \widehat{x}^\star\|^2 \mathrm{d}s$$

$$\leq L + (\kappa_3 + D)\, \frac{\sqrt{d}L^2}{2\nu} e^{\nu_2 t_\ddagger}$$

and thus

$$c_1 \leq \left(L + (\kappa_3 + D)\, \frac{\sqrt{d}L^2}{2\nu} e^{\nu t_\ddagger}\right) e^{-(\nu_2-\nu_1)t_\ddagger} =: \overline{c}_1$$

and

$$c_2 \leq L e^{\nu_1 t_\ddagger} =: \overline{c}_2.$$

Hence (19.56) holds with the constant $C_4' = \max(\overline{c}_1, \overline{c}_2)$, which proves Claim 19.4.

This, in combination with (19.55), shows that for all $t \geq 0$

$$\max(\|x_\ddagger(t)\|, \|\hat{x}_\ddagger(t)\|) \leq C_M e^{-\nu_1 t}, \tag{19.63}$$

where $C_M = \max(C_4, C_4')$.

We shall use (19.63) to bound the integral in (19.54). We have by (19.63) and $\nu > \nu_1 > \underline{\nu}$

$$\int_0^t e^{-\nu(t-s)} \left[ \Delta \|\hat{x}_\ddagger(s)\| + \sqrt{d} \left( \frac{\kappa_3}{2} \|x_\ddagger(s)\|^2 + \frac{\kappa_3 + \eta_3}{2} \|\hat{x}_\ddagger(s)\|^2 \right) \right] \mathrm{d}s,$$

$$\leq \int_0^t e^{-\underline{\nu}(t-s)} \left[ \Delta C_M e^{-\nu_1 s} + \sqrt{d} \left( \frac{\kappa_3}{2} C_M^2 e^{-2\nu_1 s} + \frac{\kappa_3 + \eta_3}{2} C_M^2 e^{-2\nu_1 s} \right) \right] \mathrm{d}s$$

$$\leq \int_0^t e^{-\underline{\nu}(t-s)} \left[ \Delta C_M e^{-\nu_1 s} + \sqrt{d}\,(\kappa_3 + \eta_3)\, C_M^2 e^{-2\underline{\nu} s} \right] \mathrm{d}s$$

$$\leq C_M e^{-\underline{\nu} t} \left[ \Delta \frac{1 - e^{-(\nu_1 - \underline{\nu})t}}{\nu_1 - \underline{\nu}} + \sqrt{d}\,(\kappa_3 + \eta_3)\, C_M \frac{1 - e^{-\underline{\nu} t}}{\underline{\nu}} \right].$$

Applying this bound in (19.54) we get

$$\|x_\ddagger(t) - \hat{x}_\ddagger(t)\|$$

$$\leq e^{-\underline{\nu} t} \|x^* - \hat{x}^*\| + C_M e^{-\underline{\nu} t} \left[ \Delta \frac{1 - e^{-(\nu_1 - \underline{\nu})t}}{\nu_1 - \underline{\nu}} + \sqrt{d}\,(\kappa_3 + \eta_3)\, C_M \frac{1 - e^{-\underline{\nu} t}}{\underline{\nu}} \right]. \tag{19.64}$$

By the triangle inequality

$$\|x(t) - \hat{x}(t)\| \leq \|x^* - \hat{x}^*\| + \|x_\ddagger(t) - \hat{x}_\ddagger(t)\|$$

and using (19.41) and (19.64) we deduce that for all $t \geq t_\ddagger$,

$$\|x(t) - \hat{x}(t)\|$$

$$\leq (1 + e^{-\underline{\nu} t}) Q_0 \sqrt{\eta_0} + C_M e^{-\underline{\nu} t} \left[ \Delta \frac{1 - e^{-(\nu_1 - \underline{\nu})t}}{\nu_1 - \underline{\nu}} + \sqrt{d}(\kappa_3 + \eta_3) C_M \frac{1 - e^{-\underline{\nu} t}}{\underline{\nu}} \right].$$

Keeping in mind that we assume that $\eta_3 \leq D$, $\eta_0$, $\eta_1$ and $\eta_2 \leq 1/q_0 \leq 1$, which makes $\Delta \leq 1 + d^{3/2}\kappa_3 Q_0$. Therefore for $t_\ddagger = t_\epsilon > 0$ suitably large we get that for some constant $Q_1 = Q_1(g, x_0, \underline{\nu}, \overline{\nu}, \epsilon, D) > 0$,

$$\|x(t) - \hat{x}(t)\| \leq Q_1\left(\sqrt{\eta_0} + e^{-\underline{\nu}t}\right), \quad \text{for all } t \geq t_\epsilon. \tag{19.65}$$

(Recall that in (19.46) we defined $t_\ddagger := t_\epsilon$.)

Notice that since $g$ is in $C^3$, there is an $\epsilon > 0$ such that all the eigenvalues of $H_g(x)$ exceed $-\overline{\nu}$ when $x \in \overline{B}(x^\star, \epsilon)$, $\epsilon > 0$, being fixed. Note that this implies that $\nabla g$ is Lipschitz on $\overline{B}(x^\star, \epsilon)$ with constant $\overline{\nu}$. Let $t_\epsilon$ be large enough such that for all $t \geq t_\epsilon$, $x(t) \in B(x^\star, \epsilon/2)$. Assume that $\eta_0$ is small enough so that $\|\widehat{x}^\star - x^\star\| \leq \epsilon/2$, which is possible by (19.41). Moreover by (19.65) for a suitably large $t_\epsilon > 0$ and small $\eta_0 > 0$ with $\eta_2 \leq 1/q_0 \leq 1$ and $\eta_3 \leq D$

$$\|x(t) - \hat{x}(t)\| \leq Q_1\left(\sqrt{\eta_0} + e^{-\underline{\nu}t_\epsilon}\right) \leq \epsilon/2, \quad \text{for all } t \geq t_\epsilon, \tag{19.66}$$

Then we also have $\widehat{x}(t) \in \overline{B}(x^\star, \epsilon)$ for all $t \geq t_\epsilon$. We may now apply Lemma 19.5 with $\mathcal{S} = \overline{B}(x^\star, \epsilon)$, $t_0 = t_\epsilon$, $x_1 = x(t_\epsilon)$, $y_1 = \widehat{x}(t_\epsilon)$, keeping in mind that $\nabla g$ is Lipschitz on $\overline{B}(x^\star, \epsilon)$ with constant $\overline{\nu}$, to get

$$\|x(t) - \hat{x}(t) - \left(x(t_\epsilon) - \hat{x}(t_\epsilon)\right)\| \leq \frac{\eta_1}{\overline{\nu}} e^{\overline{\nu}t}, \quad \forall t \geq t_\epsilon. \tag{19.67}$$

**Bound on $\|x(t) - \hat{x}(t)\|$ for Small $t$**

Since $\epsilon$ is fixed, by (19.30) we also get by Lemma 19.5 the following bound on $\|x(t) - \hat{x}(t)\|$ for small $t \geq 0$

$$\|x(t) - \hat{x}(t)\| \leq \frac{\eta_1}{\sqrt{d}\kappa_2} e^{\sqrt{d}\kappa_2 t} \leq \frac{\eta_1 e^{\left|\sqrt{d}\kappa_2 - \overline{\nu}\right| t_\epsilon}}{\sqrt{d}\kappa_2} e^{\overline{\nu}t}, \quad 0 \leq t \leq t_\epsilon. \tag{19.68}$$

**Completion of the Proof of Theorem 19.2**

Combining (19.67) and (19.68) we get

$$\|x(t) - \hat{x}(t)\| \leq Q_2 \eta_1 e^{\overline{\nu}t}, \quad \forall t \geq 0, \tag{19.69}$$

for some constant $Q_2 = Q_2(g, x_0, \underline{\nu}, \overline{\nu}, \epsilon, D)$. Then from (19.65) and (19.69) we arrive at

$$\|x(t) - \hat{x}(t)\| \leq Q_3 \min\left[\sqrt{\eta_0} + e^{-\underline{\nu}t}, \eta_1 e^{\overline{\nu}t}\right], \quad \forall t \geq 0, \tag{19.70}$$

for some constant $Q_3 = Q_3(g, x_0, \underline{\nu}, \overline{\nu}, \epsilon, D)$. Indeed, the curves $t \mapsto Q_1\left(\sqrt{\eta_0} + e^{-\underline{\nu}t}\right)$ and $t \mapsto Q_2\eta_1 e^{\overline{\nu}t}$ intersect at some point $t$ larger than $t_\epsilon$ if

$$Q_1\left(\sqrt{\eta_0} + e^{-\underline{\nu}t_\epsilon}\right) \geq Q_2\eta_1 e^{\overline{\nu}t_\epsilon} \iff Q_1 \geq Q_2\frac{\eta_1 e^{\overline{\nu}t_\epsilon}}{\sqrt{\eta_0} + e^{-\underline{\nu}t_\epsilon}},$$

and this is guaranteed if we choose $Q_1$ large enough that $Q_1 \geq Q_2\frac{1}{q_0}e^{(\underline{\nu}+\overline{\nu})t_\epsilon}$. (Recall the bounds in (19.41) and note that $Q_2$ does not depend on $Q_1$.)

We are now ready to finish the proof of Theorem 19.2. We shall show that the bound (19.14) follows from (19.70). To verify this, we start with

$$\min\left[\sqrt{\eta_0} + e^{-\underline{\nu}t}, \eta_1 e^{\overline{\nu}t}\right] \leq 2B(t), \quad B(t) := \min\left[\max\{\sqrt{\eta_0}, e^{-\underline{\nu}t}\}, \eta_1 e^{\overline{\nu}t}\right].$$

Set $t_0 = \frac{1}{2\underline{\nu}}\log(1/\eta_0)$ and note that

$$\max\{\sqrt{\eta_0}, e^{-\underline{\nu}t}\} = \begin{cases} e^{-\underline{\nu}t} \text{ when } t \leq t_0 \\ \sqrt{\eta_0}, \text{ when } t > t_0. \end{cases}$$

Suppose that $\eta_0$ is small enough so that $t_0 \geq t_\ddagger$.

- When $t \geq t_0$, then we simply observe that $B(t) \leq \eta_0^{1/2}$.
- When $t \leq t_0$, we have $B(t) = \min\left[e^{-\underline{\nu}t}, \eta_1 e^{\overline{\nu}t}\right]$. Let $t_1 = \frac{1}{\underline{\nu}+\overline{\nu}}\log(1/\eta_1)$. Note that the map defined on $[0, \infty)$ by $t \mapsto \min\left[e^{-\underline{\nu}t}, \eta_1 e^{\overline{\nu}t}\right]$ is increasing over $[0, t_1]$, decreasing $[t_1, \infty)$ , and that

$$\min\{\sqrt{\eta_0}, e^{-\underline{\nu}t}\} = \begin{cases} \eta_1 e^{\overline{\nu}t} \text{ when } t \leq t_1 \\ e^{-\underline{\nu}t}, \text{ when } t \geq t_1. \end{cases}$$

- When $t_1 \geq t_0$ and $t \leq t_0$, we see that $B(t) = \eta_1 e^{\overline{\nu}t_0} \leq \eta_1\eta_0^{-\frac{\overline{\nu}}{2\underline{\nu}}}$.
- When $t_1 < t_0$ and $t \leq t_0$, then $B(t) \leq B(t_1) = e^{-\underline{\nu}t_1} \leq \eta_1^{\frac{\underline{\nu}}{\underline{\nu}+\overline{\nu}}}$.

  Since $t_0 \leq t_1$ if and only if $\eta_1\eta_0^{-\frac{\overline{\nu}}{2\underline{\nu}}} \leq \eta_1^{\frac{\underline{\nu}}{\underline{\nu}+\overline{\nu}}}$, we conclude that $B(t) \leq \min\left\{\eta_1^{\frac{\underline{\nu}}{\underline{\nu}+\overline{\nu}}}, \eta_1\eta_0^{-\frac{\overline{\nu}}{2\underline{\nu}}}\right\}$ for all $t \leq t_0$.

Hence, we worked (19.70) into

$$\sup_{t\geq 0}\|x(t) - \hat{x}(t)\| = 2Q_3\max\left\{\sqrt{\eta_0}, \min\left[\eta_1^\delta, \eta_0^{\frac{\delta-1}{2\delta}}\eta_1\right]\right\},$$

where $\delta = \frac{\underline{\nu}}{\underline{\nu}+\overline{\nu}}$. We note that

$$\sqrt{\eta_0} \leq \eta_1^\delta \iff \eta_0^{\frac{1}{2\delta}} \leq \eta_1 \iff \sqrt{\eta_0} \leq \eta_1\eta_0^{\frac{1}{2}-\frac{1}{2\delta}} \iff \sqrt{\eta_0} \leq \eta_0^{\frac{\delta-1}{2\delta}}\eta_1$$

and

$$\eta_1^{\delta} \le \eta_0^{\frac{\delta-1}{2\delta}} \eta_1 \Longleftrightarrow \eta_0^{\frac{1-\delta}{2\delta}} \le \eta_1^{1-\delta} \Longleftrightarrow \sqrt{\eta_0} \le \eta_1^{\delta}.$$

Using these equivalences we deduce that

$$\max\left\{\sqrt{\eta_0}, \min\left[\eta_1^{\delta}, \eta_0^{\frac{\delta-1}{2\delta}} \eta_1\right]\right\} = \max\left\{\sqrt{\eta_0}, \eta_1^{\delta}\right\}.$$

Putting together our bounds on $\|x(t) - \hat{x}(t)\|$ for $t > 0$ large and $t \ge 0$ small, we can now conclude from (19.70) that for all $\epsilon > 0$ small enough and all $D > 0$ there exists a constant $C := C(g, x_0, \underline{\nu}, \bar{\nu}, D) \ge 1$ and a function $F(g, x_0, \underline{\nu}, \bar{\nu}, \epsilon, D)$ of $\epsilon$ and $D$ such that, whenever $\max\{\epsilon, \eta_0, \eta_1, \eta_2\} \le 1/C$ and $\eta_3 \le D$, $\hat{x}(t)$ is defined for all $t \ge 0$ and

$$\sup_{t \ge 0} \|x(t) - \hat{x}(t)\| \le F(g, x_0, \underline{\nu}, \bar{\nu}, \epsilon, D) \max\left\{\sqrt{\eta_0}, \eta_1^{\delta}\right\}, \tag{19.71}$$

holds, where $\delta := \underline{\nu}/\left(\underline{\nu} + \overline{\nu}\right)$. We now take $\epsilon = 1/C$ in (19.71). This completes the proof of Theorem 19.2. □

**Acknowledgements** David Mason thanks the editors for inviting him to contribute to this proceedings. He had intended give a talk based on the results in this paper at the 2017 High Dimensional Probability VIII conference. Unfortunately he had to cancel his participation due heath reasons. Part of this paper was written while he was a guest of Australian National University. All of the authors acknowledge the associate editor and referee for their useful comments and suggestions.

## References

1. E. Arias-Castro, D.M. Mason, B. Pelletier, On the estimation of the gradient lines of a density and the consistency of the mean-shift algorithm. J. Mach. Learn. Res. **17**, Paper No. 43, 1487–1514 (2016)
2. E. Arias-Castro, D.M. Mason, B. Pelletier, Errata: on the estimation of the gradient lines of a density and the consistency of the mean-shift algorithm. J. Mach. Learn. Res. **17**, Paper No. 206, 1487–1514 (2016)
3. R. Bhatia, *Matrix Analysis. Graduate Texts in Mathematics*, vol. 169 (Springer, New York, 1997)
4. P. Deheuvels, D.M. Mason, General asymptotic confidence bands based on kernel-type function estimators. Stat. Inference Stoch. Process. **7**, 225–277 (2004)
5. U. Einmahl, D.M. Mason, An empirical process approach to the uniform consistency of kernel-type function estimators. J. Theor. Probab. **13**, 1–37 (2000)
6. U. Einmahl, D.M. Mason, Uniform in bandwidth consistency of kernel-type function estimators. Ann. Stat. **33**, 1380–1403 (2005)
7. K. Fukunaga, L.D. Hostetler, The estimation of the gradient of a density function, with applications in pattern recognition. IEEE Trans. Inf. Theory **21**, 32–40 (1975)

8. E. Giné, A. Guillou, Rates of strong uniform consistency for multivariate kernel density estimators. Ann. Inst. Henri Poincaré Probab. Stat. **38**, 907–921 (2002)
9. J.A. Hartigan, *Clustering Algorithms* (Wiley, New York, 1975)
10. M.W. Hirsch, S. Smale, R.L. Devaney, *Differential Equations, Dynamical Systems & An Introduction to Chaos*, 2nd edn. (Academic, Cambridge, 2004)
11. D.M. Mason, Proving consistency of non-standard kernel estimators. Stoch. Inference Stoch. Process. **15**, 151–176 (2012)
12. D.M. Mason, J. Swanepoel, A general result on the uniform in bandwidth consistency of kernel-type function estimators. Test **20**, 72–94 (2011)
13. D.M. Mason, J. Swanepoel, Errata: a general result on the uniform in bandwidth consistency of kernel-type function estimators. Test **24**, 205–206 (2015)