Calc. Var. (2013) 48:1–31
DOI 10.1007/s00526-012-0538-8

Calculus of Variations

Geodesic convexity of the relative entropy in reversible Markov chains

Alexander Mielke

Received: 25 April 2012 / Accepted: 31 May 2012 / Published online: 6 July 2012

Abstract We consider finite-dimensional, time-continuous Markov chains satisfying the detailed balance condition as gradient systems with the relative entropy E as driving functional. The Riemannian metric is defined via its inverse matrix called the Onsager matrix K. We provide methods for establishing geodesic λ-convexity of the entropy and treat several examples including some discretizations of one-dimensional Fokker–Planck equations.

Keywords Time-continuous · Markov chain · Detailed balance · Relative entropy · Onsager matrix · Logarithmic mean · Geodesic convexity

Mathematics Subject Classification 60J27 · 53C21 · 53C23 · 82B35

1 Introduction

In this work we consider reversible Markov chains with a finite state space and with continuous time. The starting point is that the *reversibility condition*, also called *detailed balance condition*, for Markov chains provides a *gradient structure* with the relative entropy as the driving functional. The associated metric gives a discrete counterpart to the Wasserstein metric used for the Fokker-Planck equation in [15,26]. The present work was motivated by a generalization in [21] of the gradient structure for the Fokker-Planck equation to general reaction-diffusion systems, where the reactions satisfy a detailed-balance condition. The point is that the diffusion terms and the reaction terms can be written as a gradient system with

Communicated by L. Ambrosio.

A. Mielke (✉)
WIAS Berlin, Mohrenstraße 39, 10117 Berlin, Germany
e-mail: alexander.mielke@wias-berlin.de

A. Mielke
Institut für Mathematik, Humboldt-Universität zu Berlin, Rudower Chaussee 25,
12489 Berlin, Germany

respect to the same relative entropy. It is even possible to keep the gradient structure when adding the physically proper energy equations for the temperature, see [21, Sect. 3.6] and [22].

The Markov chains discussed in this paper are special cases of reversible reactions, namely “exchange reactions” that lead to a linear ODE system instead of the more general polynomial right-hand side in the mass-action type reactions. Similarly, the linear Fokker-Planck equation can be seen as a special case of more general diffusion systems. The gradient structure, which follows from [21, Sect. 3.1] as a special case of more general reaction-diffusion systems, was found independently in [8,19]. It was also used in [2] to show convergence from a Fokker-Planck equation to a simple Markov chain in a certain scaling limit.

To be more precise, we say that an ODE $\dot u = -f(u)$ has a gradient structure on the open set $X \subset \mathbb{R}^m$, if there exists an C^1 functional $E : X \to R$ and a symmetric, positive definite tensor $K : X \to \mathbb{R}^{m\times m}$ such that

$$\dot u = -f(u) = -K(u)\mathrm{D}E(u) = -\nabla_G E(u) \quad \Longleftrightarrow \quad G(u)\dot u = -\mathrm{D}E(u),$$

where $G(u) = K(u)^{-1}$ is the metric tensor and ∇_G the metric gradient. To explain our gradient structure for Markov chains, we consider the discrete state space $\{1, \ldots, n\}$ and

$$u = (u_1, \ldots, u_n) \in X_n \overset{\text{def}}{=} \{\, u \in \mathbb{R}^n \mid u_j > 0,\ \sum_{i=1}^n u_i = 1\,\}$$

is the vector of the probabilities on the state space. The ODE system reads

$$\dot u = Qu \quad \text{with } Q = (Q_{ij})_{i,j=1,\ldots n} \in \mathbb{R}^{n\times n},$$

where $Q_{ij} \geq 0$ is the rate for a particle moving from state j to i, and $Q_{jj} = -\sum_{i\neq j} Q_{ij} < 0$.

We call the Markov chain *reversible* if there exists a *unique positive* steady state $w \in X_n$ (i.e. $w_i > 0$) such that

$$\pi_{ij} \overset{\text{def}}{=} Q_{ij}w_j = Q_{ji}w_i = \pi_{ji} \quad \text{for all } i, j \in \{1, \ldots, n\}. \tag{1.1}$$

Note that, without loss of generality, we include the irreducibility (i.e. the uniqueness of w) into the definition of reversibility. The gradient structure is given in terms of the *relative entropy* E and the *Onsager matrix* K:

$$\begin{aligned} E(u) &= \sum_{i=1}^n u_i \log\Big(\frac{u_i}{w_i}\Big) \ \text{ and} \\ K(u) &= \sum_{i<j} \pi_{ij}\, \Lambda\Big(\frac{u_i}{w_i}, \frac{u_j}{w_j}\Big)\ (e_i - e_j)\otimes(e_i - e_j) \in \mathbb{R}^{n\times n}_{\mathrm{sym},\geq 0}. \end{aligned} \tag{1.2}$$

We say that the Markov chain $\dot u = Qu$ is given by the gradient system (X_n, E, K), since

$$\dot u = Qu = -K(u)\mathrm{D}E(u),$$

see Proposition 3.1, where also more general gradient structures are given. Here K is the inverse of the Riemannian tensor $G(u) = K(u)^{-1}$ defined on $\mathbb{R}^n_{\mathrm{av}} = \{\, v \in \mathbb{R}^n \mid v \cdot \overline{e} = 0\,\}$.

The function $\Lambda : [0, \infty[^2 \to [0, \infty[$ used above plays a central role in the present theory. It is the logarithmic mean of a and b and is given by

$$\Lambda(a, b) = \frac{a-b}{\log a - \log b} \quad \text{for } a \neq b \quad \text{and} \quad \Lambda(a, a) = a, \tag{1.3}$$

and hence is analytic. All its relevant properties are discussed in Appendix A. Some specific properties are encoded in the function $\ell :]0, \infty[\to]0, \infty[$ given by

$$\ell(\xi) \stackrel{\text{def}}{=} \max\{\, \Lambda(1, r) - \xi r \mid r > 0\,\}. \tag{1.4}$$

As $r \mapsto \Lambda(1, r)$ is increasing and concave, ℓ is decreasing and convex. Moreover, it satisfies the surprising relation

$$\ell\,(\partial_a \Lambda(a, b)) = \partial_b \Lambda(a, b) \quad \text{for all } a, b > 0.$$

The focus of this work is to provide conditions on the matrix Q such that the relative entropy E is geodesically λ-convex with respect to the Riemannian tensor $G(u) = K(u)^{-1}$. This means that $s \mapsto E(\gamma(s))$ is λ-convex for all arc-length parametrized geodesics $\gamma : [s_\mathrm{a}, s_\mathrm{b}] \to X$, i.e.

$$E(\gamma(s_\theta)) \le (1-\theta)E(\gamma(s_0)) + \theta E(\gamma(s_1)) - \lambda \frac{\theta(1-\theta)}{2}(s_1 - s_0)^2$$

for all $\theta \in [0, 1]$ and $s_0, s_1 \in [s_\mathrm{a}, s_\mathrm{b}]$, where $s_\theta = (1-\theta)s_0 + \theta s_1$. Of course, geodesic λ-convexity implies geodesic μ-convexity for all $\mu \le \lambda$. The supremum of all possible λ will be denoted by λ_Q, which is justified, since according to our definition for reversible Markov chains the equilibrium density w is uniquely determined, whence E and K are determined as well. While in most cases it is not possible to calculate λ_Q explicitly, it is the purpose of this work to establish methods for estimating λ_Q from below.

Since the Onsager matrix K is given explicitly and there is no easy representation of its inverse, the Riemannian tensor G, nor for the Riemannian distance function d_K, it is advantageous to reformulate geodesic λ-convexity in terms of the triple (X, E, K). Here we are in the case of a smooth, finite-dimensional Riemannian manifold, so we can use classical differential geometry to give a differential characterization of geodesic λ-convexity, see Sect. 2. Using the covariant Hessian $\mathrm{H}_G E$ or the contravariant Hessian $\mathrm{H}^*_K E$ we have

$$E \text{ geodesically } \lambda\text{-convex} \quad \Longleftrightarrow \quad \mathrm{H}_G E \ge \lambda G \quad \Longleftrightarrow \quad \mathrm{H}^*_K E \ge \lambda K.$$

Following the ideas of [9,28] one can characterize geodesic λ-convexity also in terms of the evolution of infinitesimal line elements with the flow of the gradient systems. In our finite-dimensional setting this is most easily formulated by the Lie derivatives with respect to $f(u) = K\mathrm{D}E = \nabla_G E$, namely

$$E \text{ geodesically } \lambda\text{-convex} \quad \Longleftrightarrow \quad L_{-\nabla_G E} G \le -2\lambda G \quad \Longleftrightarrow \quad L_{-K\mathrm{D}E} K \ge 2\lambda K,$$

see Lemma 2.2. This method is more flexible and allows us to provides differential characterization of geodesic λ-convexity for infinite-dimensional cases such as systems of partial differential equations, cf. [9,17].

In our setting of finite-dimensional Markov chains $\dot u = Qu$ the criterion for geodesic λ-convexity yields the following characterization of the optimal λ:

$$\lambda_\mathrm{Q} \stackrel{\text{def}}{=} \inf\left\{ \frac{\langle \eta, M(u)\eta\rangle}{\langle \eta, K(u)\eta\rangle} \;\middle|\; u \in X, \eta \in \mathrm{T}^*_u X \setminus \{0\} \right\}, \tag{1.5}$$

see Proposition 2.1, where the Hessian M takes the form

$$M(u) := \mathrm{H}^*_K E(u) = \tfrac{1}{2}\Big(\mathrm{D}K(u)[Qu] - K(u)Q^\mathsf{T} - QK(u)\Big).$$

Starting in Sect. 3.2 we provide simple results on geodesic λ-convexity. In Sect. 4.1 we provide our first structural result stating that for all finite-dimensional Markov chains we have

$\lambda_Q > -\infty$. However, the construction is rather implicit and does not provide useful bounds. In Theorem 4.6 we consider the special case of reversible Markov chain with $Q_{ij} > 0$ for all $i < j$. Using a different proof we are able to provide an explicit bound for λ_Q in terms of all Q_{ij} and w_i.

In Corollary 4.4 we provide a quantitative result for special reversible Markov chains arising from a finite connected graph as follows. Denote the vertices by $\{1, \ldots, n\}$ and set $Q_{ij} = 1$ whenever i and j are connected by an edge and $Q_{ij} = 0$ otherwise. Then, Q is reversible with $w = \frac{1}{n}(1, \ldots, 1)^\mathsf{T}$. Moreover, there exists a function $f : \mathbb{N} \to \mathbb{R}$ such that $\lambda_Q \geq f(m)$, where $m := \max\{ -Q_{ii} \mid i = 1, \ldots, n \}$ is the maximum degree of the vertices.

Section 5 is devoted to Markov chains with nearest-neighbor transitions, namely

$$\dot{u} = \alpha_{i-1}u_{i-1} - (\alpha_i + \beta_{i-1})u_i + \beta_i u_{i+1}$$

with transition rates $\alpha_i, \beta_i > 0$ for $i = 1, \ldots, n-1$ and $\alpha_j = \beta_j = 0$ for $j = 0, n$. The associated tridiagonal matrix Q leads to a reversible Markov chain. Under the monotonicity condition $\alpha_i \geq \alpha_{i+1}$ and $\beta_i \leq \beta_{i+1}$ for $i = 1, \ldots, n-2$ we obtain the lower bound

$$\lambda_Q \geq \frac{1}{2} \min\{ \alpha_i - \alpha_{i+1} + \beta_i - \beta_{i-1} + \Xi(\alpha_i - \alpha_{i+1}, \beta_i - \beta_{i-1}) \mid \ \}i = 1, \ldots, n-1 \geq 0,$$

where Ξ satisfies $2\sqrt{ab} \leq \Xi(a,b) \leq 2\Lambda(a,b)$, see Theorem 5.1. Without any monotonicity assumption we have the upper bound

$$\lambda_Q \leq \min\{ \alpha_i - \frac{1}{4}\alpha_{i+1} + \beta_i - \frac{1}{4}\beta_{i-1} \mid i = 1, \ldots, n-1 \},$$

see Lemma 5.2. In fact, our lower bound is sharp enough to provide uniform estimates for discretization of the the one-dimensional Fokker-Planck equation $\partial_t U = (U_x + U V_x)_x$ on $\Omega =]0,1[$. It is well-known that for potentials $V \in \mathrm{C}^2([0,1])$ with $V''(x) \geq \widehat{\lambda}$, the continuous relative entropy $\mathcal{E}(U) = \int_0^1 U \log(U/W)\,\mathrm{d}x$ is geodesically $\widehat{\lambda}$-convex with respect to the Wasserstein distance, cf. [1,20]. We provide Markov chains arising as consistent finite-difference and finite-volume discretizations, respectively, such that the gradient system (X_n, E_n, K_n) is geodesically λ_n-convex with $\lambda_n \to \widehat{\lambda}$ for $n \to \infty$.

We end by mentioning that the techniques for estimating geodesic Λ-convexity developed for Markov chains can also be applied to nonlinear reaction systems with the gradient structure established in [21, Sect. 3.1] and [22]. In particular, using the methods established in [9] the theory of geodesic λ-convexity can be made available for reaction-diffusion systems, see [17] for first results.

Note added. After the first version of this work [WIAS Preprint 1650, October 2011] was finished, the author became aware of the recent work [11], in which geodesic convexity of the entropy is studied as well. There the focus is on Ricci curvature and general structures, while we concentrate on analytical estimates for deriving bounds for λ_Q in concrete cases.

2 Geodesic convexity

We consider an open subset X of $\mathbb{R}^m$, state vectors $u \in X$, and the gradient flow

$$G(u)\dot{u} = -\mathrm{D}E(u) \quad \Longleftrightarrow \quad \dot{u} = -\nabla_G E(u) = -K(u)\mathrm{D}E(u) = -f(u).$$

Here $E : X \to \mathbb{R}$ is an energy functional and $G(u) = G(u)^* > 0$ denotes the *Riemannian metric tensor* at the point u. We call the symmetric and positive semidefinite matrix

$K(u) = G(u)^{-1}$ the *Onsager matrix*, as it is used in thermodynamics to relate the rate $\dot{u}$ with the thermodynamic driving force $-\mathrm{D}E(u)$, which encodes the Onsager symmetry relations and the Onsager principle, see e.g. [22,25,24,27].

This section collects known results and benefits of geodesic λ-convexity of the functional E with respect to the metric G. In fact, in the present finite-dimensional situation the characterizations are easier than in the case of partial differential equations. Since in our case G and the induced distance d_K are only defined implicitly, it is desirable to characterize the geodesic convexity via K only. We do this in two different, but equivalent ways. First, we derive the defining equations for geodesic curves in terms of K and study the convexity of E along the curves, which leads to the Hessian $\mathrm{H}_G E$. Second, we use the ideas from Otto–Westdickenberg [28] and Daneri–Savaré [9] on the evolution of length elements along the gradient flow, which in our simplified ODE case, means the usage of the Lie derivative. The different approaches have different advantages: The usage of the Hessian is restricted to finite dimensions, while convexity properties along geodesic curves can be studied more generally if the geodesic curves are suitably characterized, as for instance in displacement convexity, cf. [18,20]. The second approach using Lie derivative allows for generalizations to systems of PDEs, see [9,17].

2.1 Geodesic curves, the Hessian $\mathrm{H}_G E$, and geodesic λ-convexity

Here we show how to characterize the geodesic curves in terms of the Onsager matrix K rather than of the Riemannian tensor G. Thus, constant-speed geodesics $\gamma :]s_1, s_2[\to X; s \mapsto u = \gamma(s)$ satisfy the classical Lagrange equation

$$-\frac{\mathrm{d}}{\mathrm{d}s}\left(\frac{\partial}{\partial \gamma'} L(\gamma, \gamma')\right) + \frac{\partial}{\partial \gamma} L(\gamma, \gamma') = 0, \quad \text{where } L(\gamma, \gamma') = \frac{1}{2}\langle G(\gamma)\gamma', \gamma'\rangle.$$

Since in our case G is only known implicitly, it is more convenient to use the Hamiltonian version of the Lagrange equation. Introducing the dual variable $\eta = \frac{\partial}{\partial \gamma'} L(\gamma, \gamma') = G(\gamma)\gamma'$ and the Hamiltonian $H(\gamma, \eta) = \frac{1}{2}\langle \eta, K(\gamma)\eta\rangle$ we obtain the equivalent system

$$\gamma' = \frac{\partial}{\partial p} H(\gamma, \eta) = K(\gamma)\eta, \quad \eta' = -\frac{\partial}{\partial \gamma} H(\gamma, \eta) = -\frac{1}{2}\langle \eta, \mathrm{D}K(\gamma)[\Box]\eta\rangle, \tag{2.1}$$

where $b = \langle \eta, \mathrm{D}K(\gamma)[\Box]\eta\rangle$ denotes the vector defined via $\langle b, \beta\rangle = \langle \eta, \mathrm{D}K(\gamma)[\beta]\eta\rangle$ and $\mathrm{D}K(u)[v]$ means the directional derivative.

Thus, we may characterize geodesic λ-convexity of a function $E : X \to \mathbb{R}$ easily by asking that the composition $s \mapsto E(\gamma(s))$ is λ-convex for all constant-speed geodesics γ. This property can be characterized by local expressions using the second derivative in the form

$$\frac{\mathrm{d}^2}{\mathrm{d}s^2} E(\gamma(s)) \geq \lambda \, \langle G(\gamma(s))\gamma'(s), \gamma'(s)\rangle.$$

In fact, the Hessian $\mathrm{H}_G E(u) : \mathrm{T}_u X \to \mathrm{T}_u^* X$ can be defined as the symmetric tensor obtained by taking the second derivative of $E(\gamma(s))$ along geodesics, viz.

$$\begin{aligned}\langle \gamma'(s), \mathrm{H}_G E(\gamma(s))\gamma'(s)\rangle &:= \frac{\mathrm{d}^2}{\mathrm{d}s^2} E(\gamma(s)) = \frac{\mathrm{d}}{\mathrm{d}s}\big\langle \mathrm{D}E(\gamma(s)), \gamma'(s)\big\rangle \\ &= \langle \mathrm{D}^2 E(\gamma)\gamma', \gamma'\rangle + \big\langle \mathrm{D}E(\gamma), \mathrm{D}K(\gamma)[\gamma']\eta + K(\gamma)\eta'\big\rangle,\end{aligned}$$

and, using (2.1) to eliminate η and η', we find the relation

$$\begin{aligned}\langle v, \mathrm{H}_G E(u)v\rangle &= \langle \mathrm{D}^2E(u)v, v\rangle + \langle \mathrm{D}E(u), \mathrm{D}K(u)[v]G(u)v\rangle \\ &\quad - \frac{1}{2}\langle G(u)v, \mathrm{D}K(u)[K(u)\mathrm{D}E(u)]G(u)v\rangle.\end{aligned} \tag{2.2}$$

The same formula for $\mathrm{H}_G E$ can be obtained by classical differential geometry using the Levi-Civita connection ∇ associated with G. Since in our applications the matrix K is given explicitly, it is advantageous to use the contravariant representation of the Hessian $\mathrm{H}_K^* E(u) := K(u)^*\mathrm{H}_G E(u)K(u) : \mathrm{T}_u^*X \to \mathrm{T}_uX$. Clearly, we have $\mathrm{H}_G E \geq \lambda G$ if and only if $\mathrm{H}_K^* E \geq \lambda K$. We will use the letter M to denote $\mathrm{H}_K^* E$ and find

$$\begin{aligned}\langle \eta, M(u)\eta\rangle &\overset{\text{def}}{=} \langle \eta, K(u)\mathrm{D}^2E(u)K(u)\eta\rangle + \langle \mathrm{D}E(u), \mathrm{D}K(u)[K(u)\eta]\eta\rangle \\ &\quad - \tfrac{1}{2}\langle \eta, \mathrm{D}K(u)[K(u)\mathrm{D}E(u)]\eta\rangle.\end{aligned} \tag{2.3}$$

We arrive at the following characterization of geodesic λ-convexity.

Proposition 2.1 *Given an open $X \subset \mathbb{R}^n$, $E \in \mathrm{C}^2(X;\mathbb{R})$, and an Onsager matrix $K \in \mathrm{C}^1(X;\mathbb{R}^{n\times n}_{\text{spd}})$, then E is geodesically λ-convex with respect to the Riemannian metric induced by $G = K^{-1}$ if and only if*

$$\forall\, u \in X: \quad M(u) \geq \lambda K(u) \tag{2.4}$$

in the ordering sense of symmetric matrices, i.e. $M(u) - \lambda K(u)$ is positive semidefinite. Here M is given via K and the vector field $u \mapsto f(u) = K(u)\mathrm{D}E(u)$ as

$$M(u) = \tfrac{1}{2}\Big(K(u)\mathrm{D}f(u)^{\mathsf{T}} + \mathrm{D}f(u)K(u) - \mathrm{D}K(u)[f(u)]\Big). \tag{2.5}$$

Proof The definition of f yields $\mathrm{D}f(u)[v] = \mathrm{D}K(u)[v]\mathrm{D}E(u) + K(u)\mathrm{D}^2E(u)v$. Choosing $v = K(u)\eta$ and inserting this into the definition (2.3) of M gives (2.5). □

The formula (2.5) is especially simple for linear vector fields $f : u \mapsto -Qu$, namely

$$M(u) = \frac{1}{2}\Big(\mathrm{D}K(u)[Qu] - K(u)Q^{\mathsf{T}} - QK(u)\Big). \tag{2.6}$$

This formula is most useful for Markov chains and, hence, will be used subsequently.

2.2 Lie derivatives and the Otto–Westdickenberg characterization

The idea of Otto–Westdickenberg [28] (see also [9]) to prove geodesic λ-convexity is based on the rate of change of infinitesimal line elements. For this we consider the semiflow $\mathcal{S} : [0,T] \times X \to X$ such that $u(t) = \mathcal{S}_t(u(0))$ is the solution of $\dot{u} = -f(u) = -K(u)\mathrm{D}E(u)$. For a general vector $v \in \mathrm{T}_uX$ the transported infinitesimal line element is $\sigma(t) = \langle G(\mathcal{S}_t(u))\mathrm{D}\mathcal{S}_t(u)v, \mathrm{D}\mathcal{S}_t(u)v\rangle$. The statement of [28] is that

$$\dot{\sigma} \leq -2\lambda\sigma \quad \text{for all } u \in X \text{ and } v \in \mathrm{T}_uX \tag{2.7}$$

is sufficient for geodesic λ-convexity of E, while the necessity is proved in [9].

This transport of line elements is best formulated in terms of the Lie derivative of G with respect to the vector field $-f$, namely

$$\langle L_{-f}G(u)v, v\rangle = \frac{\mathrm{d}}{\mathrm{d}t}\langle G(\mathcal{S}_t(u))\mathrm{D}\mathcal{S}_t(u)v, \mathrm{D}\mathcal{S}_t(u)v\rangle\big|_{t=0}$$

Similarly, we may define the Lie derivative of the K via

$$\langle \eta, L_{-f}K(u)\eta\rangle = \frac{\mathrm{d}}{\mathrm{d}t}\langle \mathrm{D}\mathcal{S}_t(u)^{-\mathsf{T}}\eta, K(\mathcal{S}_t(u))\mathrm{D}\mathcal{S}_t(u)^{-\mathsf{T}}\eta\rangle\big|_{t=0},$$

where $\mathrm{D}\mathcal{S}_t(u)^{-\mathsf{T}} : \mathrm{T}_u^*X \to \mathrm{T}^*_{\mathcal{S}_t(u)}X$ denotes the adjoint of the inverse of $\mathrm{D}\mathcal{S}_t(u)$.

The following result explains the equivalence of the geodesic λ-convexity and the contraction property of the associated gradient flow.

Lemma 2.2 *Let the smooth and finite-dimensional gradient system* (X, E, K) *generate the vector field* $f(u) = K(u)\mathrm{D}E(u)$. *Then, the Hessians and Lie derivatives are related as follows:*

$$\mathrm{H}_G E(u) = -\frac{1}{2}L_{-f}G \quad \textit{and} \quad M(u) = \mathrm{H}_K^* E(u) = \frac{1}{2}L_{-f}K.$$

Proof Using $\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{D}\mathcal{S}_t(u)|_{t=0} = -\mathrm{D}f(u)$ the definition of L_{-f} gives

$$\langle L_{-f}G(u)v, v\rangle = -\langle \mathrm{D}G(u)[f(u)]v, v\rangle - 2\langle G(u)v, \mathrm{D}f(u)v\rangle.$$

Inserting $f(u) = K(u)\mathrm{D}E(u)$, using $\mathrm{D}G(u)[v] = -G(u)\mathrm{D}K(u)[v]G(u)$, and comparing with (2.2) shows the identity $L_{-f}G = -2\mathrm{H}_G E$. Similarly $\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{D}\mathcal{S}_t(u)^{-\mathsf{T}}|_{t=0} = \mathrm{D}f(u)^{\mathsf{T}}$ yields

$$\langle \eta, L_{-f}K(u)\eta\rangle = -\langle \eta, \mathrm{D}K(u)[f(u)]\eta\rangle + 2\langle \mathrm{D}f(u)^{\mathsf{T}}\eta, K(u)\eta\rangle = 2\langle \eta, M(u)\eta\rangle$$

by using (2.5). Hence, the assertion is established. □

Remark 2.3 (Bakry–Émery conditions) Our condition $M \geq \lambda K$ has some similarities with the conditions of Bakry and Émery [4,3] for hypercontractivity. There, two symmetric bilinear mappings Γ_1 and Γ_2 are defined via

$$\Gamma_1(f, g) = \frac{1}{2}\left(\mathcal{Q}(fg) - f\mathcal{Q}g - g\mathcal{Q}f\right) \text{ and}$$
$$\Gamma_2(f, g) = \frac{1}{2}\left(\mathcal{Q}\Gamma_1(f, g) - \Gamma_1(\mathcal{Q}f, g) - \Gamma_1(f, \mathcal{Q}g)\right),$$

where $\mathcal{Q}$ is the generator of a diffusion semigroup. The analogy of the pair (Γ_1, Γ_2) with the pair (K, M) is seen in (2.6). The condition of λ-hypercontractivity reads

$$2\Gamma_2(f, f) \succcurlyeq \lambda\Gamma_1(f, f) \text{ for all sufficiently smooth } f, \tag{2.8}$$

which is analogous to (2.4), see [9,17] for more discussion on this.

2.3 Benefits from geodesic convexity

So far we have concentrated on the triple (X, E, K) as a gradient system. However, the metric tensor $G = K^{-1}$ generates a distance $d_K : X \times X \to [0, \infty[$ in the usual way:

$$d_K(u_0, u_1) = \inf\{\int_0^1 \langle G(\gamma)\gamma', \gamma'\rangle^{1/2} \mid \gamma \in \mathrm{C}^1([0, 1]; X),\ \gamma(0) = u_0,\ \gamma(1) = u_1\}.$$

Thus, we may consider also the metric gradient system (X, E, d_K) in the sense of [1,10]. The theory there clearly shows that systems with geodesic λ-convexity have a series of good properties. First, we have a Lipschitz continuous dependence of the solutions u_j on the initial data, namely

$$d_K(u_1(t), u_2(t)) \leq \mathrm{e}^{-\lambda t}\, d_K(u_1(0), u_2(0)) \quad \text{for all } t \geq 0.$$

In particular, for $\lambda \geq 0$ we have a contraction semigroup. If $\lambda > 0$ we obtain exponential decay towards the unique equilibrium state w, which minimizes E, i.e.

$$d_K(u(t), w) \leq \mathrm{e}^{-\lambda t} d_K(u(0), w).$$

Second, the time-continuous solutions $u : [0, \infty[\to X$ can be well approximated by interpolants obtained by incremental minimizations. Fixing a time step $\tau > 0$ we define iteratively

$$u^\tau_{k+1} = \operatorname*{Arg\,min}_{u\in X} \left(E(u) + \tfrac{1}{2\tau} d_K(u_k, u)^2\right).$$

For geodesically λ-convex E the minimizers are unique for $\tau \in\,]0, \tau_0[$ if $1/\tau_0 + \lambda \geq 0$. Moreover, if u is the time-continuous solution with $u(0) = u_0$ and if $\overline{u}^\tau$ is the left-continuous piecewise constant interpolant of $(u^\tau_k)_{k\in\mathbb{N}}$, then

$$d_K(u(t), \overline{u}^\tau(t)) \leq C(u_0)\sqrt{\tau}\; \mathrm{e}^{-\lambda_\tau t} \quad \text{for } t \geq 0,$$

see [1, Thms. 4.0.9+4.0.10], where $\lambda_\tau = \lambda$ for $\lambda < 0$ and $\lambda_\tau = \frac{1}{\tau}\log(1+\lambda\tau)$ for $\lambda > 0$.

Another important reason for studying geodesic λ-convexity is the recently established connections between the Ricci curvature, optimal transport, Wasserstein diffusion, and geodesic λ-convexity of the relative entropy, see [7,11,18,19,29,30]. A coarser definition of curvature for general Markov chains is given in [23].

3 Reversible Markov chains

3.1 An entropic gradient structure for Markov chains

We consider general Markov chains on n states and set

$$X_n \overset{\text{def}}{=} \{\, u = (u_1, \ldots, u_n) \in \mathbb{R}^n \mid u_i > 0, \sum_{j=1}^n u_j = 1\,\} \;\subset\; \tfrac{1}{n}\overline{e} + \mathbb{R}^n_{\text{av}},$$

where $\overline{e} = (1, \ldots, 1)^\mathsf{T}$ and $\mathbb{R}^n_{\text{av}} = \{\, v \in \mathbb{R}^n \mid v \cdot \overline{e} = 0\,\}$. The ODE system is given by

$$\dot{u} = Qu, \quad \text{where } Q_{ij} \geq 0 \text{ for } i \neq j \text{ and } Q_{ii} = -\sum_{j:j\neq i} Q_{ji}. \tag{3.1}$$

We assume that there exists a unique positive steady state $w \in X_n$ and that the crucial assumption of *reversibility*, also called the *condition of detailed balance* holds, namely

$$Q_{ij} w_j = Q_{ji} w_i \text{ for } i, j = 1, \ldots, n. \tag{3.2}$$

With $W = \operatorname{diag}(w)$ this means $QW = (QW)^\mathsf{T} = WQ^\mathsf{T}$.

Obviously, the Markov chain (3.1) has two different *linear gradient structures*, namely

$$\begin{aligned} &G_1\dot{u} = -\mathrm{D}E_1(u), \quad G_2\dot{u} = -\mathrm{D}E_2(u), \quad \text{or } \dot{u} = -K_1\mathrm{D}E_1(u) = -K_2\mathrm{D}E_2(u)\\ &\text{with } E_1(u) = \tfrac{1}{2}\langle -W^{-1}Qu, u\rangle,\; K_1 = W,\; E_2(u) = \tfrac{1}{2}\langle W^{-1}u, u\rangle, \text{ and } K_2 = -QW. \end{aligned}$$

For these systems we obviously have geodesic convexity, as E_1 and E_2 are convex and G_1 and G_2 are constant.

However, we are interested in the Wasserstein-type gradient structure where the Onsager matrix $K(u)$ is homogeneous of degree 1 in u and the driving functional is the relative entropy. This gradient structure was introduced in [21, Sect. 3.1] in a more general nonlinear context of reaction systems and independently in [8,19]. This is the special case with $\phi(a) = a\log a$ in the following result.

Theorem 3.1 *Consider* $\phi \in \mathrm{C}([0,\infty[) \cap \mathrm{C}^2(]0,\infty[)$ *satisfying* $\phi''(a) > 0$ *for all* $a > 0$*. If the Markov chain* (3.1) *satisfies the reversibility* (3.2) *for the steady state* $w \in X_n$*, then it has the gradient structure* (X_n, E^ϕ, K^ϕ) *with*

$$E^\phi(u) = \sum_{i=1}^n w_i\, \phi\left(\frac{u_i}{w_i}\right),$$

$$K^\phi(u) = \sum_{j=2}^n \sum_{i=1}^{j-1} Q_{ij} w_j\, \Phi\left(\frac{u_i}{w_i}, \frac{u_j}{w_j}\right) (e_i - e_j) \otimes (e_i - e_j), \tag{3.3}$$

where $e_i \in \mathbb{R}^n$ *denotes the* i*-th unit vector, and* $\Phi(a,b) = (a-b)/(\phi'(a) - \phi'(b))$ *for* $0 < a \neq b$ *and* $\Phi(a,a) = 1/\phi''(a)$.

In the special case that all ϕ are equal to $a \mapsto a\log a$, we obtain the classical logarithmic entropy relative entropy E and the Onsager matrix K as given in (1.2) and $\Phi = \Lambda$ in (1.3) and discussed in Appendix A.

Proof Clearly we have $\mathrm{D}E(u) = (\phi'(u_i/w_i))_{i=1,\ldots,n}$, and multiplying this vector by $e_i - e_j \in \mathbb{R}^n_{\text{av}}$ we obtain the denominator of $\Phi\big(\frac{u_i}{w_i}, \frac{u_j}{w_j}\big)$. Hence,

$$K^\phi(u)\mathrm{D}E^\phi(u) = \sum_{j=2}^n \sum_{i=1}^{j-1} Q_{ij} w_j \left(\frac{u_i}{w_i} - \frac{u_j}{w_j}\right) (e_i - e_j) = -Qu,$$

where we used $\sum_{i=1}^n Q_{ij} = 0$ and the detailed balance condition (3.2) in the last equality. Thus, the assertion is established. □

Note that (E_2, K_2) can be obtained by choosing $\phi(a) = \frac{1}{2}a^2$ or by linearization of (E, K), namely $E_2(u) = \frac{1}{2}\mathrm{D}^2E(w)[u,u]$ and $K_2 = K(w)$. The choice $\phi(\rho) = c\rho\log\rho + d\rho$ is singled out by the fact that it is the only one giving the 1-homogeneity

$$\widetilde{K}(\gamma u) = \gamma \widetilde{K}(u) \quad \text{for all } \gamma > 0 \text{ and } u \in X_n,$$

which is a specific feature of the distances related to optimal transport problems. In fact, for 1-homogeneity of K we need $\Phi(\sigma a, \sigma b) = \sigma\Phi(a,b)$ for all σ, a, and b. Using the definition $\Phi(a,b) = (a-b)/(\phi'(a) - \phi'(b))$ this leads to the condition $\phi'(\sigma a) - \phi'(\sigma b) = \phi'(a) - \phi'(b)$, which implies $a\phi''(a) = c = \text{const}$, whence $\phi_i(\rho) = c\rho\log\rho + d\rho$.

Remark 3.2 A similar gradient structure can be defined for jump processes on a continuous state space $\Omega \subset \mathbb{R}^n$. By $U(t,\cdot) : \Omega \to [0,\infty[$ one denotes the probability density which satisfies the evolution equation

$$\dot{U}(t,x) = (\mathcal{Q}U(t,\cdot))(x) := \int_\Omega q(x,y)U(t,y)\,\mathrm{d}y - \int_\Omega q(z,x)\,\mathrm{d}z\, U(t,x) \tag{3.4}$$

for a suitable transition kernel $q : \Omega\times\Omega \to [0, \infty[$. We assume that (3.4) has a unique steady state $W \in \mathrm{L}^\infty(\Omega) \cap \mathrm{Prob}(\Omega)$ with $0 < c_0 \le W(x)$ and that q satisfies the detailed balance condition $\kappa(x, y) := q(x, y)W(y) = q(y, x)W(x)$. Now we define the relative entropy $\mathcal{E}$ and the Onsager operator $\mathcal{K}$ via

$$\mathcal{E}(U) = \int_\Omega U \log(U/W)\,\mathrm{d}x \quad \text{and}$$

$$\langle \Xi, \mathcal{K}(U)\Xi\rangle = \int_\Omega\int_\Omega \frac{\kappa(x, y)}{2} \Lambda\left(\tfrac{U(x)}{W(x)}, \tfrac{U(y)}{W(y)}\right) (\Xi(x) - \Xi(y))^2\,\mathrm{d}y\,\mathrm{d}x.$$

Using the definition of Λ and detailed balance it is not difficult to show that $\mathcal{Q}U = -\mathcal{K}(U)\mathrm{D}\mathcal{E}(U)$ for $U \in \mathrm{L}^2(\Omega)$ with $0 < c_0 \le U(x)$ a.e. This can even be generalized to general measure spaces and to general strictly convex Caratheodory functions $(x, U) \mapsto \phi(x, U)$ for the relative entropy. Moreover, it is expected that this approach can be applied to general subclasses of Dirichlet forms.

Our main concern is the geodesic convexity of the relative entropy E of (1.2) in a Markov chains $\dot{u} = Qu$ with respect to the metric defined via K given in (1.2). Since for reversible Markov chains $w \in X_n$ is uniquely determined by Q the same holds for E and K. Hence we introduce the short-hand

$$\lambda_Q := \inf\{\, \tfrac{\langle \eta, M(u)\eta\rangle}{\langle \eta, K(u)\eta\rangle} \mid u \in X_n,\ \xi \in \mathrm{T}^*_w X_n \setminus \{0\}\,\}$$

and discuss a few simple lower bounds for λ_Q. In Sect. 4 we show that for all finite-dimensional Markov chains we have $\lambda_Q > -\infty$.

3.2 A few Markov-chain examples

By definition we have $K(u)\overline{e} = 0$, and for the matrix $M(u)$ defined in (2.5) this also holds as $Q^\mathsf{T}\overline{e} = 0$, i.e. we have

$$K(u)\overline{e} = M(u)\overline{e} = 0 \quad \text{for all } u \in X_n, \quad \text{where } \overline{e} = (1, \dots, 1)^\mathsf{T}. \tag{3.5}$$

Thus, a simple criterion for positive semidefiniteness of $M(u) - \lambda K(u)$ is the following.

Lemma 3.3 *Assume that K and M are symmetric and satisfy (3.5) as well as*

$$\forall\, i \ne j\ \forall\, u \in X_n : \quad M_{ij}(u) \le \lambda K_{ij}(u) \tag{3.6}$$

for some $\lambda \in \mathbb{R}$, then $\lambda_Q \ge \lambda$.

Proof Since $K_{ij}(u) \le 0$ for $i \ne j$, all off-diagonal elements of $N(u) := M(u) - \lambda K(u)$ are nonpositive. Condition (3.5) implies that the diagonal elements satisfy

$$N_{ii}(u) = -\sum_{j\ne i} N_{ij}(u) = \sum_{j\ne i} |N_{ij}(u)|.$$

Hence N is weakly diagonal dominant and hence positive semidefinite. In fact,

$$N(u) = \sum_{i,j:\, i<j} |N_{ij}(u)|(e_i - e_j) \otimes (e_i - e_j) \ge 0.$$

This proves $M \ge \lambda K$ which is the assertion. □

Before developing a more general theory we show that this criterion can be applied in a few easy cases, where it supplies geodesic λ-convexity.

Example 3.4 A special case occurs if for the Markov chain all transition rates are the same, e.g. $Q_{ij}=1$ for $i\neq j$. The steady state is $w=\frac{1}{n}\overline{e}$, and we claim that E is geodesically $\frac{n+2}{2}$-convex.

In this case we have $Q=n\,I-\overline{e}\otimes\overline{e}$. Using $u\cdot\overline{e}=1$ and $K(u)\overline{e}=0$ we easily obtain

$$M(u)=n\,K(u)-\tfrac{1}{2}\mathrm{D}K(u)[nu-\overline{e}]. \tag{3.7}$$

In particular, for $i\neq j$ we have $K_{ij}(u)=-\Lambda_{ij}(u)$ and, with $\widetilde{u}=1-u_i-u_j\geq 0$, we find

$$\begin{aligned}2M_{ij}(u)&=-2n\Lambda_{ij}(u)+\partial_i\Lambda_{ij}(u)\left((n-1)u_i-u_j-\widetilde{u}\right)+\partial_j\Lambda_{ij}(u)\big((n-1)u_j-u_i-\widetilde{u}\big)\\&\leq-2n\Lambda_{ij}(u)+\partial_i\Lambda_{ij}(u)\left((n-1)u_i-u_j\right)+\partial_j\Lambda_{ij}(u)\big((n-1)u_j-u_i\big)\\&=-2n\Lambda_{ij}+n\Lambda_{ij}-\tfrac{u_i+u_j}{u_iu_j}\Lambda_{ij}^2,\end{aligned}$$

where the last identity follows by inserting the explicit relations (A.3) for the derivatives and using (A.4a) and (A.4d). With (A.1) we obtain $2M_{ij}(u)\leq-(n+2)\Lambda_{ij}=(n+2)K_{ij}(u)$, and conclude $\lambda_Q\geq\frac{n+2}{2}$. We expect that the result is not optimal for $n\geq 3$. However, for $u=w=\frac{1}{n}\overline{e}$ Eq. (3.7) gives $M(w)=nK(w)$ and we conclude $\lambda_Q\leq n$. Hence, we have $\lambda_Q\in[\frac{n+2}{2},n]$ and conclude $\lambda_Q=2$ for $n=2$.

Example 3.5 (Markov chains for $n=2$) For $n=2$ every nontrivial Markov chain is reversible with $w=(\theta,1-\theta)$ and $Q=\mu\left(\begin{smallmatrix}\theta&-1\ \theta\\1-\theta&-\theta\end{smallmatrix}\right)$ for $\mu>0$. We claim

$$\tfrac{\mu}{2}\leq\mu\left(\tfrac{1}{2}+\sqrt{\theta-\theta^2}\right)\;\leq\;\lambda_Q=\tfrac{\mu}{2}(1+\Xi(1-\theta,\theta))\;\leq\;\mu\left(\tfrac{1}{2}+\Lambda(1-\theta,\theta)\right)\leq\mu, \tag{3.8}$$

where Ξ is is defined in (A.6), where also the estimates $2\sqrt{ab}\leq\Xi(a,b)\leq 2\Lambda(a,b)$ are proved. In fact, using $\kappa=\mu\theta(1-\theta)$, $\Lambda_{12}=\Lambda(\rho_1,\rho_2)$ with $\rho=(u_1/\theta,u_2/(1-\theta))$ gives

$$K(u)=\kappa\Lambda_{12}\left(\begin{smallmatrix}1&-1\\-1&1\end{smallmatrix}\right)\quad\text{and}\quad M(u)=m(u)\left(\begin{smallmatrix}1&-1\\-1&1\end{smallmatrix}\right)$$
$$m(u)=\mu\kappa\Lambda_{12}-\tfrac{\mu\kappa}{2}\left((1-\theta)\partial_{\rho_1}\Lambda(\rho_1,\rho_2)-\theta\partial_{\rho_2}\Lambda(\rho_1,\rho_2)\right)(\rho_1-\rho_2).$$

Geodesic λ-convexity is equivalent to $m\geq\lambda\kappa\Lambda$ for all ρ. Using (A.3) we find

$$\tfrac{\rho_1-\rho_2}{\Lambda(\rho_1,\rho_2)}\left((1-\theta)\partial_{\rho_1}\Lambda(\rho_1,\rho_2)-\theta\partial_{\rho_2}\Lambda(\rho_1,\rho_2)\right)=1-\left(\tfrac{1-\theta}{\rho_1}+\tfrac{\theta}{\rho_2}\right)\Lambda(\rho_1,\rho_2).$$

The supremum of the last term is $1-\Xi(1-\theta,\theta)$, and the formula of λ_Q in (3.8) follows.

Taking $\mu=2$ and $\theta=1/2$ we obtain $\lambda_Q=2$ as in the case $n=2$ of Example 3.4.

The next example shows that we cannot expect $\lambda_Q\geq 0$, in general.

Example 3.6 (Geodesic λ-convexity with $\lambda_Q<0$) We consider the case that Q is a tridiagonal matrix, as will be the case in the whole of Sect. 5, namely

$$\dot{u}_i=\alpha_{i-1}u_{i-1}-(\alpha_i+\beta_{i-1})u_i+\beta_iu_{i+1},\quad\text{for } i=1,\ldots,n, \tag{3.9}$$

where $\alpha_i,\beta_i>0$ for $i=1,\ldots,n-1$ and $\alpha_k=\beta_k=0$ for $k=0$ and n. Clearly, we have a reversible Markov chain with a relative density w satisfying $w_{i+1}=\alpha_iw_i/\beta_i$.

Under the monotonicity assumption $\alpha_i \geq \alpha_{i+1}$ and $\beta_i \leq \beta_{i+1}$ for $i = 1, \ldots, n-2$ Theorem 5.1 provides the lower nonnegative bound

$$\lambda_Q \geq \min \left\{ \frac{1}{2} \left(\alpha_i - \alpha_{i+1} + \beta_i - \beta_{i-1} + \Xi(\alpha_i - \alpha_{i+1}, \beta_i - \beta_{i-1})\right) \,\middle|\, i = 1, \ldots, n-1 \right\} \geq 0.$$

For general α_i and β_i Lemma 5.2 establishes the upper bound

$$\lambda_Q \leq \min \left\{ \alpha_i - \frac{1}{4}\alpha_{i+1} + \beta_i - \frac{1}{4}\,\beta_{i-1} \,\middle|\, i = 1, \ldots, n-1 \right\}.$$

Thus, the matrix

$$Q = \begin{pmatrix} -1 & a & 0 & 0 \\ 1 & -1-a & 1 & 0 \\ 0 & 1 & -1-a & 1 \\ 0 & 0 & a & -1 \end{pmatrix}$$

satisfies $\lambda_Q \geq \min\{2-2a, a/2\}$ for $a \in \,]0, 1]$. For all $a > 0$ we have the upper bound $\lambda_Q \leq \min\{2 - a/2, a{+}3/4\}$, which implies $\lambda_Q < 0$ for $a > 4$.

3.3 The complete metric space $(\overline{X}_n, d_K)$

Above we have seen that any reversible Markov chain $\dot{u} = Qu$ can be understood as a gradient system (X_n, E, K), where the Onsager structure K is the inverse of the Riemannian metric G. As explained in Sect. 2.3 we can introduce the distance $d_K : X_n \times X_n \to [0, \infty[$. We rewrite the formula explicitly in terms of K (which is an analog of the Benamou-Brenier form [5]):

$$d_K(u_0, u_1) = \inf \left\{ \int_0^1 \langle \xi(s), K(u(s))\;\xi\;\;(s)\rangle^{1/2} \mathrm{d}s \,\middle|\, \dot{u} \in \mathrm{W}^{1,2}([0,1]; X_n), \right.$$
$$\left. u(0) = u_0,\; u(1) = u_1,\; \dot{u}(s) = K(u(s))\xi(s) \right\}.$$

So far, X_n is the open set with $u_i > 0$ for all i. In [19, Thm. 3.17] it is shown that d_K can be uniquely extended to a metric $\overline{d}_K$ on the closure $\overline{X}_n = \mathrm{Prob}(\{1, \ldots, n\})$. Moreover, this extension turns $(\overline{X}_n, d_K)$ into a complete metric space, whose topology is the same as the standard Euclidean topology on $\overline{X}_n \subset \mathbb{R}^n$.

In [9] is is shown that a geodesically λ-convex metric gradient system (X, E, d_K) can be extended in a natural way to the completion $(\overline{X}, \overline{E}, \overline{d}_K)$, which is again a geodesically λ-convex gradient system. This applies easily to our case as E and K have continuous extensions $\overline{E}$ and $\overline{K}$ on $\overline{X}_n = \mathrm{Prob}(\{1, \ldots, n\})$.

Without going into detail here, we mention that existence of geodesic curves can be obtained by the direct method in the calculus of variations. Consider the function

$$\Psi^* : X_n \times \mathbb{R}^n_{\mathrm{av}} \to \mathbb{R};\; (u, \xi) \mapsto \frac{1}{2}\langle \xi, K(u)\xi \rangle.$$

Then, $\Psi^*(u, \cdot) : \mathbb{R}^n_{\mathrm{av}} \to \mathbb{R}$ is convex while $\Psi^*(\cdot, \xi) : X_n \to \mathbb{R}$ is concave, which easily follows from the concavity of Λ and the definition of K in (3.3). Thus, by standard arguments the partial Legendre transform

$$\Psi : X_n \times \mathbb{R}^n_{\mathrm{av}} \to [0, \infty];\; (u, v) \mapsto \sup\{\, \langle \xi, v \rangle - \tfrac{1}{2}\langle \xi, K(u)\xi \rangle \mid \xi \in \mathbb{R}^n_{\mathrm{av}} \,\}$$

is (jointly) convex and lower semicontinuous. Note that Ψ may attain the value $+\infty$ for $u \in \partial X_n$. Moreover, the boundedness of K implies the coercivity of Ψ, namely $\Psi(u, v) \geq c|v|^2$. Thus, geodesics connecting u_0 and u_1 are easily obtained by minimizing $\mathcal{I}(\gamma) = \int_0^1 \Psi(\gamma(s), \gamma'(s))\, \mathrm{d}s$ in the set of absolutely continuous functions with $\gamma(0) = u_0$ and $\gamma(1) = u_1$. It can be shown there is at least one curve $\widetilde{\gamma}$ making $\mathcal{I}(\widetilde{\gamma})$ finite. By convexity of $\mathcal{I}$ the set of minimizers is also convex. We conjecture that $\mathcal{I}$ is strictly convex, i.e. there is a unique geodesic connecting any two points.

Depending on the Markov chain under investigation, there might be different cases for the geodesics when points on the boundary ∂X_n are connected. In some cases one might expect that the whole geodesics lies inside X_n except for their endpoints. In other cases, the geodesics might stay totally in ∂X_n.

4 Geodesic λ-convexity for Markov chains

4.1 A general result on geodesic λ-convexity

In this section we show that every finite-dimensional reversible Markov chain is geodesically λ-convex. Even though our theory is finite dimensional, this result is nontrivial: On the one hand the Onsager matrix K, which is formed with the entries $\Lambda(\rho_i, \rho_j)$ with $\rho_i = u_i/w_i$, is not uniformly positive definite on the state space X_n. On the other hand, the matrix $M(u)$ depends in a complicated manner on $\rho = (u_1/w_1, \ldots, u_n/w_n)$, in particular through the unbounded derivatives of $\Lambda(\rho_i, \rho_j)$. The proof uses several special properties of Λ that are discussed in Appendix A. In particular, the derivatives $\partial_{\rho_i}\Lambda(\rho_i, \rho_j)$ cannot be simply estimated by $\Lambda(\rho_i, \rho_j)$, but rather correct signs need to be used.

Theorem 4.1 *Let $\dot{u} = Qu$ be a reversible Markov chain with $Q \in \mathbb{R}^{n\times n}$, then λ_Q defined in* (1.5) *satisfies $\lambda_Q > -\infty$.*

The remainder of this subsection forms the proof of the above theorem. As the case $n = 2$ is trivial (see Example 3.5), we assume $n \geq 3$ for the rest of this section. While there is a much shorter proof for the case when all transition coefficients Q_{ij}, $i \neq j$, are strictly positive (see Sect. 4.2) we have to introduce some notation for the general result discussed here. We define the set $\mathfrak{E}$ of transition edges via

$$\mathfrak{E} = \{\, \overline{i\,j} \mid i < j,\ Q_{ij} > 0 \,\} \quad \text{and} \quad N_{\mathfrak{E}} := \#\mathfrak{E}.$$

Moreover, we define an oriented connection matrix $S \in \mathbb{R}^{N_{\mathfrak{E}}\times n}$ via

$$S_{\overline{i\,j},k} = \begin{cases} 1 & \text{if } i = k, \\ -1 & \text{if } j = k, \\ 0 & \text{else.} \end{cases}$$

Reversibility of the Markov chain $\dot{u} = Qu$ means that $W^{-1} = \mathrm{diag}(1/w_i)_{i=1,\ldots,n}$ exists and that $QW = (QW)^{\mathsf{T}} = WQ^{\mathsf{T}}$. Thus, we can rewrite the matrices Q, $K(u)$ and $M(u)$ in the form

$$Q = -S^*\mathbb{Q}SW^{-1}, \quad K(u) = S^*\mathbb{L}(u)S, \quad M(u) = S^*\mathbb{M}(u)S, \tag{4.1}$$

where, using the abbreviations $\pi_{ij} = Q_{ij}w_j = \pi_{ji} \geq 0$ for $i \neq j$ and (2.5), we have

$$\mathbb{Q} = \mathrm{diag}(\pi_{ij})_{\overline{i\,j}\in\mathfrak{E}}, \quad \mathbb{L}(u) = \mathrm{diag}(\pi_{ij}\Lambda(u_i/w_i, u_j/w_j))_{\overline{i\,j}\in\mathfrak{E}},$$
$$\mathbb{M}(u) = \tfrac{1}{2}\left(S^*\mathbb{L}SW^{-1}S^*\mathbb{Q}S + S^*\mathbb{Q}SW^{-1}S^*\mathbb{L}S - S^*\mathrm{D}\mathbb{L}(u)[S^*\mathbb{Q}SW^{-1}]S\right).$$

For the future analysis it is more convenient to express the matrices $\mathbb{L}$ and $\mathbb{M}$ in terms of the relative densities ρ_i from $u = W\rho$ via $\mathcal{L}(\rho) = \mathbb{L}(W^{-1}\rho)$ and $\mathcal{M}(\rho) = \mathbb{M}(W^{-1}\rho)$, which gives the final formulas

$$\mathcal{L}(\rho) = \mathrm{diag}(\pi_{ij}\Lambda(\rho_i, \rho_j))_{\overline{i\,j}\in\mathfrak{E}}, \quad \mathcal{M}(\rho) = \frac{1}{2}(\mathcal{L}\mathbf{S}\mathbb{Q} + \mathbb{Q}\mathbf{S}\mathcal{L} + D\mathcal{L}(\rho)[W^{-1}QW\rho]),$$

where $\mathbf{S} = SW^{-1}S^* \in \mathbb{R}^{N_\mathfrak{E}\times N_\mathfrak{E}}$. Note that in the last term there is an extra W^{-1} because of $\mathrm{D}\mathbb{L}(u)[v] = \mathrm{D}\mathcal{L}(\rho)[W^{-1}v]$.

From the special form of $M = S^*\mathcal{M}S$ and $K = S^*\mathcal{L}S$ it is obvious that it is sufficient (but by far not necessary) for geodesic λ-convexity that

$$\exists\,\lambda\ \forall \rho \in\,]0,\infty[^n: \quad \mathcal{N}(\rho) \overset{\text{def}}{=} 2\mathcal{M}(\rho) - 2\lambda\mathcal{L}(\rho) \geq 0. \tag{4.2}$$

The main point of these representations is that $\mathcal{L}$ and $\mathbb{Q}$ are diagonal matrices. All non-diagonal terms are induced by the matrix $\mathbf{S}$ only. In particular, changing λ only changes the diagonal entries of $\mathcal{N}$ in a monotone way. The structure $\mathbf{S} \in \mathbb{R}^{N_\mathfrak{E}\times N_\mathfrak{E}}$ is comparably simple, namely

$$\mathbf{S}_{\overline{i\,j}\,\overline{k\,l}} = \begin{cases} 1/w_i + 1/w_j & \text{if } \overline{i\,j} = \overline{k\,l}, \\ 1/w_m & \text{if } \overline{i\,j} \neq \overline{k\,l} \text{ and } (i=k=m \text{ or } j=l=m), \\ -1/w_m & \text{if } \overline{i\,j} \neq \overline{k\,l} \text{ and } (i=l=m \text{ or } j=k=m), \\ 0 & \text{if } \{i,j\}\cap\{k,l\} = \emptyset. \end{cases}$$

All nontrivial off-diagonal terms are associated with pairs of two edges having a common endpoint. The signs of $\mathbf{S}_{\overline{i\,j},\overline{k\,l}}$ will not matter in our estimates. Using the shorthand notations

$$\Lambda_{ij} = \Lambda(\rho_i, \rho_j) \quad \text{and} \quad \Lambda_{ij,k} = \partial_{\rho_k}\Lambda(\rho_i, \rho_j)$$

the entries $\mathcal{N}_{\overline{i\,j}\,\overline{k\,l}}$ take the form

$$\begin{aligned} \mathcal{N}_{\overline{i\,j}\,\overline{i\,j}} &= 2(\tfrac{1}{w_i}+\tfrac{1}{w_j})\pi_{ij}^2\Lambda_{ij} + \pi_{ij}\big(\Lambda_{ij,i}\tfrac{(Q\rho)_i}{w_i} + \Lambda_{ij,j}\tfrac{(Q\rho)_j}{w_j} - 2\lambda\Lambda_{ij}\big), \\ \mathcal{N}_{\overline{i\,j}\,\overline{k\,l}} &= \pi_{ij}\pi_{kl}\mathbf{S}_{\overline{i\,j}\,\overline{k\,l}}(\Lambda_{ij}+\Lambda_{kl}). \end{aligned}$$

The following lemma will be used to establish positive definiteness of $\mathcal{N}$.

Lemma 4.2 *If for a symmetric matrix $\Gamma \in \mathbb{R}^{\mu\times\mu}$ there exists $(\gamma_{\alpha\alpha}^\beta)_{\alpha,\beta=1,\ldots,\mu}$ such that*

$$\forall\,\alpha \in \{1,\ldots,\mu\}: \quad \Gamma_{\alpha\alpha} = \sum_{\beta=1}^{\mu} \gamma_{\alpha\alpha}^\beta \text{ and } \gamma_{\alpha\alpha}^\beta \geq 0, \tag{4.3a}$$

$$\forall\,\alpha \neq \beta: \quad \Gamma_{\alpha\beta}^2 \leq \gamma_{\alpha\alpha}^\beta \gamma_{\beta\beta}^\alpha, \tag{4.3b}$$

then, Γ is positive semidefinite.

Proof For all $\xi \in \mathbb{R}^\mu$ we have

$$\begin{aligned} \xi\cdot\Gamma\xi &= \sum_{\alpha}\Gamma_{\alpha\alpha}\xi_\delta^2 + \sum_{\alpha\neq\beta}\Gamma_{\alpha\beta}\xi_\alpha\xi_\beta \ \geq\ \sum_{\alpha,\beta}\gamma_{\alpha\alpha}^\beta\xi_\alpha^2 - \sum_{\alpha\neq\beta}(\gamma_{\alpha\alpha}^\beta\gamma_{\beta\beta}^\alpha)^{1/2}|\xi_\alpha\xi_\beta| \\ &\geq \sum_{\alpha\neq\beta}\gamma_{\alpha\alpha}^\beta\xi_\alpha^2 - \big(\sum_{\alpha\neq\beta}\gamma_{\alpha\alpha}^\beta\xi_\alpha^2\big)^{1/2}\big(\sum_{\alpha\neq\beta}\gamma_{\beta\beta}^\alpha\xi_\beta^2\big)^{1/2} = 0. \end{aligned}$$

This proves the desired result. □

To apply the above lemma, we need to find a splitting of $\mathcal{N}_{\overline{ij}\,\overline{ij}}$ into nonnegative parts as in (4.3a) such that the off-diagonal terms can be controlled as in (4.3b). For this we analyze the occurring terms in more detail. We first split them into three groups via

$$\begin{aligned}&\mathcal{N}_{\overline{ij}\,\overline{ij}} = N^{\mathrm{I}}_{\overline{ij}} + N^{\mathrm{II}}_{\overline{ij}} + N^{\mathrm{III}}_{\overline{ij}},\\ &\text{where } N^{\mathrm{I}}_{\overline{ij}} = 2\pi_{ij}^2\big(\tfrac{1}{w_i}+\tfrac{1}{w_j}\big)\Lambda_{ij} + \pi_{ij}\big(\Lambda_{ij,i}Q_{ii}\rho_i + \Lambda_{ij,j}Q_{jj}\rho_j\big) - 2\pi_{ij}\lambda\Lambda_{ij},\\ &N^{\mathrm{II}}_{\overline{ij}} = \pi_{ij}\sum_{l\notin\{i,j\}}\big(\Lambda_{ij,i}\tfrac{\pi_{li}}{w_i}+\Lambda_{ij,j}\tfrac{\pi_{lj}}{w_j}\big)\rho_l, \quad \text{and} \quad N^{\mathrm{III}}_{\overline{ij}} = \pi_{ij}^2\big(\Lambda_{ij,i}\tfrac{\rho_j}{w_i} + \Lambda_{ij,j}\tfrac{\rho_i}{w_j}\big).\end{aligned} \tag{4.4}$$

Note that the terms involving the derivatives $\Lambda_{ij,i}$ and $\Lambda_{ij,j}$ are distributed to the three parts according to their properties. All terms in $N^{\mathrm{I}}_{\overline{ij}}$ have upper and lower bounds in terms of Λ_{ij} by using (A.4a). In $N^{\mathrm{II}}_{\overline{ij}}$ we have collected the interaction with vertices $l \notin \{i, j\}$, while $N^{\mathrm{III}}_{\overline{ij}}$ features an important interaction term. The crucial estimate

$$N^{\mathrm{III}}_{\overline{ij}} \geq \frac{\pi_{ij}^2}{\max\{w_i, w_j\}}\big(\Lambda_{ij}^2\big(\tfrac{1}{\rho_i}+\tfrac{1}{\rho_j}\big) - \Lambda_{ij}\big) \geq \frac{\pi_{ij}^2}{\max\{w_i, w_j\}}\Lambda_{ij} \geq 0 \tag{4.5}$$

follows via (A.4b). It will be important to use the first estimate from (4.5), which is much sharper for $\rho_i \neq \rho_j$ than the lower bound by Λ_{ij} given in the second estimate.

We now define the splitting (4.3a) of the diagonal elements $\mathcal{N}_{\overline{ij}\,\overline{ij}} = \sum_{\overline{kl}\in\mathfrak{E}} N^{\overline{kl}}_{\overline{ij}\,\overline{ij}}$. If $\{i, j\} \cap \{k, l\} = \emptyset$ we simply let $N^{\overline{kl}}_{\overline{ij}\,\overline{ij}} = 0 = N^{\overline{ij}}_{\overline{kl}\,\overline{kl}}$ since the corresponding non-diagonal entry $\mathcal{N}_{\overline{ij}\,\overline{kl}}$ equals 0 as well.

Now consider $\overline{ij} \in \mathfrak{E}$ fixed and define $n_{\overline{ij}} \in \{1, \dots, 2n-2\}$ as the number of edges $\overline{kl}$ such that $\{i, j\}\cap\{k, l\} \neq \emptyset$. These edges have either the common vertex i or j. Without loss of generality we may assume $j = k$ as the ordering of the vertices does not matter here. We further define the set of all neighbors of j, namely $\mathfrak{N}_j \overset{\text{def}}{=} \{k \in \{1, \dots, n\} \,|\, \overline{jk} \in \mathfrak{E} \text{ or } \overline{kj} \in \mathfrak{E}\}$ and let $\widehat{n}_j = \#\mathfrak{N}_j$. Since $j \notin \mathfrak{N}_j$ and $i, k \in \mathfrak{N}_j$ we have $\widehat{n}_j \in \{2, \dots, n-1\}$.

Thus, we have $\overline{kl} = \overline{jl}$ for $l \in \mathfrak{N}_j \setminus \{i\}$ and can set

$$N^{\overline{jl}}_{\overline{ij}\,\overline{ij}} = \pi_{ij}\nu_{\overline{ij}\,\overline{jl}}\Lambda_{ij} + \tfrac{\pi_{ij}\pi_{jl}}{w_j}\Lambda_{ij,j}\rho_l + \tfrac{1}{\widehat{n}_j-1}N^{\mathrm{III}}_{\overline{ij}}, \tag{4.6}$$

where we followed the same splitting strategy as in (4.4) and used $n \geq 3$. The constants $\nu_{\overline{ij}\,\overline{jl}} = \nu_{\overline{jl}\,\overline{ij}} \in \mathbb{R}$ will be chosen later and we set $\nu_{\overline{ij}\,\overline{kl}} = 0$ for $\{i, j\} \cap \{k, l\} = \emptyset$.

Finally, we set $N^{\overline{ij}}_{\overline{ij}\,\overline{ij}} = \mathcal{N}_{\overline{ij}\,\overline{ij}} - \sum_{\overline{kl}\neq\overline{ij}} N^{\overline{kl}}_{\overline{ij}\,\overline{ij}}$ and obtain the lower bound

$$N^{\overline{ij}}_{\overline{ij}\,\overline{ij}} \geq \pi_{ij}\Big(2\pi_{ij}\big(\tfrac{1}{w_i}+\tfrac{1}{w_j}\big) + \min\{Q_{ii}, Q_{jj}\} - 2\lambda - \sum_{\overline{kl}\neq\overline{ij}} \nu_{\overline{ij}\,\overline{kl}}\Big)\Lambda_{ij}.$$

After having chosen all $\nu_{\overline{ij}\,\overline{kl}}$, we find a desired λ via

$$\lambda = \tfrac{1}{2}\min\Big\{\, 2\pi_{ij}\big(\tfrac{1}{w_i}+\tfrac{1}{w_j}\big) - \max\{|Q_{ii}|, |Q_{jj}|\} - \sum_{\overline{kl}\neq\overline{ij}} \nu_{\overline{ij}\,\overline{kl}} \;\Big|\; \overline{ij} \in \mathfrak{E} \,\Big\}. \tag{4.7}$$

Thus, (4.3a) is satisfied, if all $N^{\overline{kl}}_{\overline{ij}\,\overline{ij}}$ are nonnegative as well, and it remains to establish the estimate (4.3b) for the non-diagonal entries. Then, Lemma 4.2 can be applied and Theorem 4.1 follows.

To estimate the nontrivial non-diagonal entries $\mathcal{N}_{\overline{ij}\,\overline{kl}}$ as assumed in (4.3b), it again suffices to consider the case $\overline{kl} = \overline{jl}$, as the other cases are analogous. Condition (4.3) is equivalent to

$$\forall \overline{ij} \in \mathfrak{E}\colon\ \mathbf{N}^{\overline{ij}\,\overline{jl}} \overset{\text{def}}{=} \begin{pmatrix} N^{\overline{jl}}_{\overline{ij}\,\overline{ij}} & \mathcal{N}_{\overline{ij}\,\overline{jl}} \\ \mathcal{N}_{\overline{ij}\,\overline{jl}} & N^{\overline{ij}}_{\overline{jl}\,\overline{jl}} \end{pmatrix} \geq 0$$

in the sense of positive semidefiniteness of the matrices. Multiplying from left and right by the diagonal matrix $\operatorname{diag}(\pi_{ij}\Lambda_{ij}, \pi_{jl}\Lambda_{jl})^{1/2}$ this is equivalent to

$$\nu_{\overline{ij}\,\overline{jl}} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + B^{\overline{ij}\,\overline{jl}}(\rho) \geq 0, \quad \text{where } B^{\overline{ij}\,\overline{jl}} = \begin{pmatrix} B^{\overline{ij}\,\overline{jl}}_{11} & B^{\overline{ij}\,\overline{jl}}_{12} \\ B^{\overline{ij}\,\overline{jl}}_{12} & B^{\overline{ij}\,\overline{jl}}_{22} \end{pmatrix}$$

$$\begin{aligned} \text{with } B^{\overline{ij}\,\overline{jl}}_{11} &\geq \tfrac{\pi_{ij}}{(\widehat{n}_j-1)\max\{w_i,w_j\}}\big(\Lambda_{ij}(\tfrac{1}{\rho_i}+\tfrac{1}{\rho_j}) - 1\big) + \tfrac{\pi_{jl}}{w_j}\tfrac{\Lambda_{ij,j}}{\Lambda_{ij}}\rho_l, \\ B^{\overline{ij}\,\overline{jl}}_{12} &= \pm\tfrac{(\pi_{ij}\pi_{jl})^{1/2}}{w_j}\big((\Lambda_{ij}/\Lambda_{jl})^{1/2} + (\Lambda_{jl}/\Lambda_{ij})^{1/2}\big), \\ B^{\overline{ij}\,\overline{jl}}_{22} &\geq \tfrac{\pi_{jl}}{(\widehat{n}_j-1)\max\{w_j,w_l\}}\big(\Lambda_{jl}(\tfrac{1}{\rho_j}+\tfrac{1}{\rho_l}) - 1\big) + \tfrac{\pi_{ij}}{w_j}\tfrac{\Lambda_{jl,j}}{\Lambda_{jl}}\rho_i, \end{aligned}$$

where we already used the lower bound (4.5) for N^{III}.

Thus, the validity of (4.3b) follows if we are able to show that the eigenvalues of the symmetric matrices $B^{\overline{ij}\,\overline{jl}}(\rho)$ are uniformly bounded from below for all $\rho \in]0,\infty[^n$. The difficulty lies in the fact that the entries are unbounded (while being 0-homogeneous), and the task is to control the negative part of the eigenvalues.

Clearly, the lowest eigenvalue decreases if we decrease the diagonal entries or increase the off-diagonal entry of $B^{\overline{ij}\,\overline{jl}}$. Using $\Lambda_{ij}(\frac{1}{\rho_i}+\frac{1}{\rho_j}) - 1 \geq \frac{1}{2}\Lambda_{ij}(\frac{1}{\rho_i}+\frac{1}{\rho_j})$ (cf. (A.4b)), it suffices to find an estimate from below for the eigenvalues of $\alpha_{\overline{ij}\,\overline{jl}} G_{\beta_{\overline{ij}\,\overline{jl}}}(\rho_i, \rho_j, \rho_l)$ where

$$G_\beta(\rho_i,\rho_j,\rho_l) \overset{\text{def}}{=} \begin{pmatrix} \Lambda_{ij}(\frac{1}{\rho_i}+\frac{1}{\rho_j}) + \frac{\Lambda_{ij,j}}{\Lambda_{ij}}\rho_l & \beta(\Lambda_{ij}/\Lambda_{jl})^{1/2} + \beta(\Lambda_{jl}/\Lambda_{ij})^{1/2} \\ \beta(\Lambda_{ij}/\Lambda_{jl})^{1/2} + \beta(\Lambda_{jl}/\Lambda_{ij})^{1/2} & \Lambda_{jl}(\frac{1}{\rho_j}+\frac{1}{\rho_l}) + \frac{\Lambda_{jl,j}}{\Lambda_{jl}}\rho_i \end{pmatrix},$$

$$\alpha_{\overline{ij}\,\overline{jl}} = \min\{\tfrac{\pi_{ij}}{2(\widehat{n}_j-1)\max\{w_i,w_j\}}, \tfrac{\pi_{jl}}{w_j}, \tfrac{\pi_{jl}}{2(\widehat{n}_j-1)\max\{w_j,w_l\}}, \tfrac{\pi_{ij}}{w_j}\}, \quad \text{and} \quad \beta_{\overline{ij}\,\overline{jl}} = \frac{(\pi_{ij}\pi_{jl})^{1/2}}{\alpha_{\overline{ij}\,\overline{jl}}\, w_j}.$$

We now employ the following result, which is proved in Appendix B.

Proposition 4.3 *There exists a continuous, decreasing function* $\widehat{g} : [0,\infty[\ \to \mathbb{R}$ *such that for all* $\beta \geq 0$ *and all* $r,s,t > 0$ *we have* $G_\beta(r,s,t) \geq \widehat{g}(\beta)I$.

Thus, we are able to conclude that the eigenvalues of $B^{\overline{ij}\,\overline{jl}}$ are bounded uniformly from below by $\alpha_{\overline{ij}\,\overline{jl}}\widehat{g}(\beta_{\overline{ij}\,\overline{jl}})$. Hence, $\mathbf{N}^{\overline{ij}\,\overline{jl}}$ is positive semidefinite for all ρ if we choose $\nu_{\overline{ij}\,\overline{jl}} = -\alpha_{\overline{ij}\,\overline{jl}}\widehat{g}(\beta_{\overline{ij}\,\overline{jl}})$. Thus, we have established condition (4.3b) and Theorem 4.1 is proved.

In principle, the above proof for the existence of a λ for geodesic λ-convexity is constructive. However, we do not have an explicit bound for $\widehat{g}$, and the above estimate is not optimized for obtaining good lower bounds for λ_Q. At this stage we are content to establish $\lambda_Q > -\infty$. In the definition of $N^{\overline{jl}}_{\overline{ij}\,\overline{ij}}$ we did not use the term $\frac{\pi_{ij}\pi_{il}}{w_i}\Lambda_{ij,i}\rho_l$, which may indeed vanish if $\pi_{il} = 0$, because $\overline{ij},\ \overline{jl} \in \mathfrak{E}$ does not imply $\overline{il} \in \mathfrak{E}$. However, if all π_{ij} are strictly positive, this can be used to find a shorter proof for geodesic λ-convexity with a more explicit lower bound for λ_Q. This is the content of the next subsection.

Nevertheless, we are able to derive a nontrivial quantitative result for special reversible Markov chains associated with a finite and connected graph with vertices $\{1,\ldots,n\}$. Assume that $Q_{ij}=1$ if the vertices i and j are connected by an edge and $Q_{ij}=0$ else. Then, $w=\frac{1}{n}\overline{e}$ is the unique steady state, and $\widehat{n}_j=-Q_{jj}=\sum_{i:i\neq J}Q_{ij}$ gives the number of neighboring vertices for the vertex j. Our result gives a bound on the geodesic λ-convexity in terms of $m=\max\{\widehat{n}_j \mid j=1,\ldots,n\}$, which is otherwise independent of n.

Corollary 4.4 *There exists a non-increasing function $f:\mathbb{N}\to\mathbb{R}$ such that the following holds. Consider a connected, finite graph with n vertices and the reversible Markov chain $\dot{u}=Qu\in\mathbb{R}^n$ with $Q_{ij}=1$ if i and j are connected and 0 else. Then, $\lambda_Q\geq f(m)$ with $m=\max\{-Q_{jj}\mid j=1,\ldots,n\}$.*

Proof We just go through the above proof and simplify all expressions using $w_i=1/n$ and $\pi_{ij}\in\{0,1/n\}$. We obtain $\alpha_{\overline{ij}\,\overline{jl}}=1/(\widehat{n}_j-1)$, $\beta_{\overline{ij}\,\overline{jl}}=\widehat{n}_j-1$, and note that at most for $2m-2$ edges $\overline{kl}$ we have $\nu_{\overline{ij}\,\overline{kl}}\neq 0$. The lower estimate (4.7) yields $\lambda=\frac{1}{2}(4-m+2\widehat{g}(m-1)=:f(m)$, which is the desired result.

As an example consider the infinite d-dimensional lattice of vertices $\mathbf{z}\in\mathbb{Z}^d$ with edges between $\mathbf{z}$ and $\widetilde{\mathbf{z}}$ if and only if $|\mathbf{z}-\widetilde{\mathbf{z}}|=1$ such that $\widehat{n}_{\mathbf{z}}=2d$ for all $\mathbf{z}$. Now take any connected, finite subgraph with n vertices and construct the special Markov chains as described above, then the Onsager system (X_n,E,K) is geodesically λ_Q-convex with $\lambda_Q\geq f(2d)$, independently of n and the structure of the subgraph.

4.2 Geodesic λ-convexity if all $Q_{ij}>0$

Here we give a shorter proof of a weakened version of Theorem 4.1. The point is to establish a more explicit bound and to provide a potential method for deriving sharper bounds for Markov chains with suitable additional structures. We use Lemma 3.3 for showing positive definiteness of $N(u)=M(u)-\lambda K(u)$, which cannot be strictly positive definite because of $N(u)\overline{e}=0$. Thus, we have to establish $M_{ij}(u)\leq\lambda K_{ij}(u)$ for $i<j$ and $u\in X_n$. Good estimates on M_{ij} will be obtained via the following Proposition 4.5, which replaces the more technical Proposition 4.3. In the latter the two partial derivatives $\partial_r\Lambda(r,t)$ and $\partial_t\Lambda(r,t)$ have to be collected from two different diagonal elements, while here they occur directly as sum. The result is formulated in terms of the function ℓ defined in (1.4).

Proposition 4.5 *Define $\widetilde{g}(\beta)=2\beta$ for $\beta\in[0,1/2]$ and $\widetilde{g}(\beta)=4\beta\ell(1/(4\beta))$ for $\beta\geq 1/2$. Then, for all $\beta\geq 0$ we have the estimate*

$$\forall\, r,\ s,t>0:\quad \beta\big(\Lambda(r,s)+\Lambda(s,t)\big)-\big(\partial_r\Lambda(r,t)+\partial_t\Lambda(r,t)\big)s\leq\widetilde{g}(\beta)\Lambda(r,t).$$

Proof We abbreviate $\Lambda_{rs}=\Lambda(r,s)$ and $\Lambda_{rs,r}=\partial_r\Lambda(r,s)$.

Defining $\gamma_\beta(r,s,t)=\beta\frac{\Lambda_{rs}+\Lambda_{st}}{\Lambda_{rt}}-\frac{\Lambda_{rt}}{rt}s$ we have to show $\gamma_\beta(r,s,t)\leq\widetilde{g}(\beta)$, where we used (A.4c). By the symmetry $r\leftrightarrow t$ and the 1-homogeneity we may assume $0<r\leq t=1$ giving $\Lambda_{rs}\leq\Lambda_{s1}$. Hence, it suffices to estimate

$$\sup_{0<r\leq 1,\ s>0}\Big(2\beta\frac{\Lambda_{1s}}{\Lambda_{1r}}-\frac{\Lambda_{1r}}{r}\,s\Big)=\sup_{0<r\leq 1}\frac{2\beta}{\Lambda_{1r}}\ell\Big(\frac{\Lambda_{1r}^2}{2\beta r}\Big)\leq 2\beta\sup_{0<r\leq 1}\frac{\ell(\Lambda_{1r,r}/(2\beta))}{\ell(\Lambda_{1r,r})},$$

where the last estimate follows from $\Lambda_{1r,r}\geq\Lambda_{1r}^2/r$ (cf. (A.4c)) and $\Lambda_{1r}\geq\Lambda_{ar,a}|_{a=1}=\ell(\Lambda_{1r,r})$, cf. (A.4a) and (A.8c).

Since $\xi = \Lambda_{1r,r}$ ranges through $[1/2, \infty[$ for $r \in]0,1]$, it suffices to establish

$$\overline{g}(\beta) = \sup\{\,\overline{\ell}_\beta(\xi) \mid \xi \geq 1/2\,\} = \begin{cases} 1 & \text{for } \beta \leq 1, \\ 2\ell(1/(2\beta)) & \text{for } \beta \geq 1; \end{cases} \quad \text{where } \overline{\ell}_\beta(\xi) = \tfrac{\ell(\xi/\beta)}{\ell(\xi)}.$$

Then, $\widetilde{g}(\beta) = 2\beta\,\overline{g}(2\beta)$ gives the desired result.

To calculate $\overline{g}(\beta)$ we first consider $\beta \leq 1$. Because ℓ is decreasing we easily find $\overline{\ell}_\beta(\xi) \leq 1$. Moreover, $\overline{\ell}_\beta(\xi) \to 1$ for $\xi \to \infty$ implies $\overline{g}(\beta) = 1$.

For $\beta \geq 1$ there exists a unique $\xi_\beta \in]1/2, \beta/2[$ such that $\xi_\beta = \beta\ell(\xi_\beta)$. According to (A.8a) for each $\xi \geq 1/2$ there exist $\kappa, \sigma \in \mathbb{R}$ such that

$$\widetilde{\ell}(\kappa) = \xi/\beta, \quad \widetilde{\ell}(\sigma) = \xi, \quad \widetilde{\ell}(-\kappa) = \ell(\xi/\beta), \quad \widetilde{\ell}(-\sigma) = \ell(\xi).$$

Since $\widetilde{\ell}$ is increasing and $\beta \geq 1$, we have $\sigma \geq 0$ and $\sigma \geq \kappa$. For $\xi \geq \xi_\beta$ we have $\widetilde{\ell}(\kappa) = \xi/\beta \geq \ell(\xi) = \widetilde{\ell}(-\sigma)$ yielding $\kappa \geq -\sigma$. Hence, we have

$$\overline{\ell}_\beta(\xi) = \tfrac{\ell(\xi/\beta)}{\ell(\xi)} = \tfrac{\widetilde{\ell}(-\kappa)}{\widetilde{\ell}(-\sigma)} = \beta\,\tfrac{m(\kappa)}{m(\sigma)} \leq \beta, \quad \text{where } m(\kappa) \stackrel{\text{def}}{=} \widetilde{\ell}(\kappa)\widetilde{\ell}(-\kappa).$$

For the last estimate we used that $|\kappa| \leq \sigma$ implies $m(\kappa) \leq m(\sigma)$. This follows from the fact that m is even and $m'(\kappa) > 0$ for $\kappa > 0$.

For $\xi \in [0, \xi_\beta]$ we define $\sigma_\beta > 0$ such that $\xi_\beta = \widetilde{\ell}(\sigma_\beta)$ (or $\ell(\xi_\beta) = \widetilde{\ell}(-\sigma_\beta)$) and $k_\beta : [0, \sigma_\beta] \to \mathbb{R}$ via $\widetilde{\ell}(\sigma) = \beta\widetilde{\ell}(k_\beta(\sigma))$. Hence, k_β is increasing and has range $[k_\beta(0), -\sigma_\beta]$, because of $\widetilde{\ell}(k_\beta(\sigma_\beta)) = \widetilde{\ell}(\sigma_\beta)/\beta = \xi_\beta/\beta = \ell(\xi_\beta) = \widetilde{\ell}(-\sigma_\beta)$. Using $m'(k_\beta) \leq 0$ and $m'(\sigma) \geq 0$ it follows that $\sigma \mapsto m(k_\beta(\sigma))/m(\sigma)$ is decreasing on $[0, \sigma_\beta]$ and the maximum is attained at $\sigma = 0$, which corresponds to $\xi = 1/2$:

$$\overline{\ell}_\beta(\xi) = \beta\tfrac{m(k_\beta(\sigma))}{m(\sigma)} \leq \beta\tfrac{m(k_\beta(0))}{m(0)} = 2\,\ell(1/(2\beta)) = \overline{g}(\beta).$$

From $\xi\ell(\xi) = m(\sigma) \geq 1/4$ we find $\beta \leq \overline{g}(\beta)$ for $\beta \geq 1$. Hence, $\overline{g}$ is calculated, and the desired estimate is established. □

To establish geodesic λ-convexity we use a similar notation as in Sect. 4.1, namely

$$\pi_{ij} = Q_{ij}w_j = \pi_{ji}, \quad \Lambda_{ij} = \Lambda(\rho_i, \rho_j), \quad \Lambda_{ij,k} = \partial_{\rho_k}\Lambda(\rho_i, \rho_j),$$

where $\rho_k = u_k/w_k$. Using the definition of M and the identities

$$K_{ij} = -\pi_{ij}\Lambda_{ij}, \quad K_{ii} = -\sum_{l\neq i} K_{il}, \quad -Q_{ii} = \sum_{l\neq i} Q_{li} > 0, \quad \mu_{ijl} = \frac{\pi_{il}\pi_{jl}}{w_l},$$

where $i \neq j$, we find the explicit representation

$$\begin{aligned} 2M_{ij} &= -\sum_l (K_{il}Q_{jl} + Q_{il}K_{lj}) - \pi_{ij}\big(\tfrac{1}{w_i}\Lambda_{ij,i}(Qu)_i + \tfrac{1}{w_j}\Lambda_{ij,j}(Qu)_j\big) \\ &= \sum_{l\notin\{i,j\}} \mu_{ijl}(\Lambda_{il}+\Lambda_{jl}) + \pi_{ij}\Lambda_{ij}(Q_{ii} + Q_{jj}) \\ &\quad -\pi_{ij}\Big(\tfrac{1}{w_i}\sum_{l\neq i}\pi_{il}\Lambda_{il} + \tfrac{1}{w_j}\sum_{l\neq j}\pi_{jl}\Lambda_{jl}\Big) - \pi_{ij}\big(\tfrac{1}{w_i}\Lambda_{ij,i}(Qu)_i + \tfrac{1}{w_j}\Lambda_{ij,j}(Qu)_j\big). \end{aligned}$$

For applying condition (3.6) for positive semidefiniteness, we observe that $K_{ij} = -\pi_{ij}\Lambda_{ij}$ only depends on ρ_i and ρ_j, whereas $M_{ij}(u)$ may depend on all $\rho_1, \ldots, \rho_n$. Thus, we rewrite

$M_{ij}(u)$ in a form that highlights the dependencies on (ρ_i, ρ_j) and on all the others ρ_l, namely

$$M_{ij}(u) = \tfrac{1}{2}\overline{M}_{ij}(\rho_i, \rho_j) + \tfrac{1}{2}\sum_{l\notin\{i,j\}} \widetilde{M}_{ijl}(\rho_i, \rho_j, \rho_l), \text{ where} \tag{4.8}$$
$$\overline{M}_{ij}(\rho_i, \rho_j) = -\pi_{ij}\big(Q_{ij} + Q_{ji} - Q_{ii} - Q_{jj}\big)\Lambda_{il}$$
$$-\pi_{ij}\big(\rho_i\Lambda_{ij,i}Q_{ii} + \rho_j\Lambda_{ij,j}Q_{jj} + \rho_j\Lambda_{ij,i}Q_{ji} + \rho_i\Lambda_{ij,j}Q_{ij}\big),$$
$$\widetilde{M}_{ijl}(\rho_i, \rho_j, \rho_l) = \pi_{il}(Q_{jl}-Q_{ji})\Lambda_{il}+\pi_{jl}(Q_{il}-Q_{ij})\Lambda_{jl}-\pi_{ij}\big(Q_{li}\Lambda_{ij,i}+Q_{lj}\Lambda_{ij,j}\big)\,\rho_l.$$

Using (A.4) and Proposition 4.5 both terms can be estimated in terms of Λ_{ij} via

$$\overline{M}_{ij} \le \overline{\mu}_{ij}\pi_{ij}\Lambda_{ij} \quad \text{with } \overline{\mu}_{ij} = \max\{Q_{ii}, Q_{jj}\} - Q_{ij} - Q_{ji} - \min\{Q_{ij}, Q_{ji}\},$$
$$\widetilde{M}_{ijl} \le \widetilde{\mu}_{ijl}\pi_{ij}\Lambda_{ij} \quad \text{with } \widetilde{\mu}_{ijl} = \pi_{ij}\min\{Q_{li}, Q_{lj}\}\,\widetilde{g}(\beta_{ijl}) \tag{4.9}$$
$$\text{and } \beta_{ijl} = \max\{0, \pi_{li}(Q_{jl} - Q_{ji}), \pi_{lj}(Q_{il} - Q_{ij})\}/(\pi_{ij}\min\{Q_{li}, Q_{lj}\}).$$

Thus, together with criterion (3.6) we can summarize and obtain the following result.

Theorem 4.6 *Assume that $\dot{u} = Qu$ is a reversible Markov chain where all transition rates are positive, i.e. $Q_{ij} > 0$ for all $i < j$. Then,*

$$\lambda_Q \ge -\frac{1}{2}\max\{\,\overline{\mu}_{ij} + \sum_{l\notin\{i,j\}} \widetilde{\mu}_{ijl} \mid 1 \le i < j \le n\,\},$$

where $\overline{\mu}_{ij}$ and $\widetilde{\mu}_{ijl}$ are given in (4.9).

We observe that the above arguments do not apply if $\pi_{ij} = 0$ and $\pi_{il} > 0$ for some $i \ne j$ and $l \notin \{i, j\}$. For that case, we need the more complicated and less explicit approach of Theorem 4.1.

Example 4.7 The above result allows for another simple example, where the convexity can be estimated. Take any vector $w \in X_n$ and let

$$Q = \kappa w \otimes \overline{e} - \kappa I, \quad \text{then } Q^{\mathsf{T}}\overline{e} = 0 = Qw \quad \text{and } Q_{ij}w_j = \kappa w_i w_j \text{ for } i \ne j.$$

Hence, w is the steady state of the reversible Markov chain. Applying the above theorem we see that $\widetilde{\mu}_{ijl} = 0$ as $Q_{ij} = Q_{il}$ by construction. Since $\overline{\mu}_{ij} = -\kappa - 2\kappa\min\{w_i, w_j\}$ we conclude $\lambda_Q \ge \kappa/2 + \kappa\min\{\, w_i \mid i = 1, \ldots, n\,\}$. Taking $w = \frac{1}{n}\overline{e}$ and $\kappa = n$ we recover the result of Example 3.4.

5 Chain with nearest-neighbor transitions

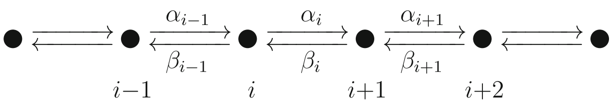
In this section we discuss the Markov chains generated by tridiagonal generators $Q \in \mathbb{R}^{n\times n}$. These Markov chains are always reversible, and under certain monotonicities of the entries on the side diagonals we obtain useful lower bounds for λ_Q. In particular, we apply them to discretizations of a one-dimensional Fokker-Planck equation and compare our results for the discretization with the well-known results on displacement convexity of the relative entropy for the Fokker-Planck equation.

5.1 Geodesic convexity for tridiagonal Markov generators

We discuss the Markov chain $\dot{u} = Qu$ for tridiagonal generators Q of the form

$$Q = \begin{pmatrix} -\alpha_1 & \beta_1 & 0 & \cdots & \cdots & 0 \\ \alpha_1 & -\alpha_2-\beta_1 & \beta_2 & 0 & & \vdots \\ 0 & \alpha_2 & -\alpha_3-\beta_2 & \ddots & & \vdots \\ \vdots & 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & \alpha_{n-2} & -\alpha_{n\text{-}1}-\beta_{n\text{-}1} & \beta_{n-1} \\ 0 & \cdots & \cdots & 0 & \alpha_{n-1} & -\beta_{n-1} \end{pmatrix} \in \mathbb{R}^{n\times n}, \qquad (5.1)$$

which is associated with the following Markov chain with nearest-neighbor transitions:



The transitions rates α_i and β_i are assumed to be positive for $i = 1, \ldots, n-1$, while $\alpha_k = \beta_k = 0$ for $k = 0$ and n. We first observe that these Markov chains are always reversible. The detailed balance condition reads $\alpha_i w_i = \kappa_i = \beta_i w_{i+1}$ which leads to the simple relation $w_{i+1} = \alpha_i w_i/\beta_i$, where $w_1 > 0$ is fixed to have $w = (w_1, \ldots, w_n) \in X$.

Our main result of this section is a lower bound for λ_Q for the case that α_i is decreasing and β_i is increasing for $i \in \{1, \ldots, n-1\}$. To formulate this result we introduce, for $a, b \geq 0$, the function

$$\Xi(a, b) = \inf\{\Lambda(r, s)\big(\tfrac{a}{r} + \tfrac{b}{s}\big) \,\big|\, r, s > 0\},$$

which satisfies the estimate $2\Lambda(a, b) \geq \Xi(a, b) \geq \max\{\Lambda(a, b), 2\sqrt{a\,b}\,\} \geq 0$, see (A.6).

Theorem 5.1 *Assume that Q in (5.1) satisfies the monotonicities*

$$\alpha_i \geq \alpha_{i+1} \text{ and } \beta_i \leq \beta_{i+1} \text{ for } i = 1, \ldots, n-2, \qquad (5.2)$$

then, with $G(a, b) = \frac{1}{2}\big(a + b + \Xi(a, b)\big) \geq 0$ we have the lower estimate

$$\lambda_Q \;\geq \gamma_Q := \min\{\, G(\alpha_i - \alpha_{i+1}, \beta_i - \beta_{i-1}) \,\big|\, i = 1, \ldots, n-1\} \geq 0.$$

We emphasize that the monotonicity condition (5.2) is sufficient but certainly not necessary for geodesic 0-convexity. A consequence of the monotonicity is the log-concavity of w:

$$w_{i+1} w_{i-1} \;=\; \frac{\alpha_i}{\alpha_{i-1}} \frac{\beta_{i-1}}{\beta_i}\, w_i^2 \;\leq\; w_i^2 \quad \text{for } i = 2, \ldots, n-1. \qquad (5.3)$$

Hence, this log-concavity is necessary for the applicability of our theorem, but it is not clear whether it is necessary for geodesic 0-convexity. Example 5.3 shows that the strict log-concavity is compatible with $\lambda_Q < 0$.

Proof Following the ideas of Sect. 4 we can simplify the matrices $M(u)$ and $K(u)$ by moving from the n nodes $i \in \{1, \ldots, n\}$ to the $n{-}1$ edges $\mathfrak{E} = \{\, \overline{i\ (i{+}1)}\,|\, i = 1, \ldots, n{-}1\,\}$, thus eliminating the eigenvalue 0 of $M(u)$ and $K(u)$ associated with the eigenvector $\overline{e} = (1, \ldots, 1)^{\mathsf{T}}$. The corresponding oriented connection matrix is

$$S = \begin{pmatrix} 1 & -1 & 0 & \cdots & 0 \\ 0 & 1 & -1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 & -1 \end{pmatrix} \in \mathbb{R}^{(n-1)\times n}, \tag{5.4a}$$

and we denote by $S^* \in \mathbb{R}^{n\times(n-1)}$ its transpose. We have $Q = -S^* \operatorname{diag}(\kappa) S \operatorname{diag}(w)^{-1}$ and

$$K(u) = S^*\mathbb{L}(u)S \text{ with } \mathbb{L}(u) = \operatorname{diag}\big(\kappa_i \Lambda(u_i/w_i, u_{i+1}/w_{i+1})\big). \tag{5.4b}$$

Inserting these specific forms into the definition of M we arrive

$$M(u) = \tfrac{1}{2} S^*\mathbb{M}(u)S \quad \text{with}$$
$$\mathbb{M}(u) = \mathbb{L}\ S(\operatorname{diag} w)^{-1}S^* \operatorname{diag}\kappa + \operatorname{diag}\kappa\ S(\operatorname{diag} w)^{-1}S^*\mathbb{L} + \mathrm{D}\mathbb{L}(u)[Qu].$$

By the special structures of M and K, the theorem is established if we show

$$\mathbb{M}(u) \geq 2\lambda\mathbb{L}(u) \text{ for all } u \in X_n \quad \text{with } \lambda = \gamma_Q. \tag{5.5}$$

Obviously, $\mathbb{M} \in \mathbb{R}^{(n-1)\times(n-1)}$ is symmetric and tridiagonal with

$$\mathbb{M}_{ij} = \begin{cases} a_i & \text{for } i = j, \\ b_k & \text{for } (i,j) \in \{(k,k{+}1), (k{+}1,k)\}, \\ 0 & \text{otherwise,} \end{cases}$$

where, using the abbreviations $\rho_i = u_i/w_i$, $\Lambda_i = \Lambda(\rho_i, \rho_{i+1})$, $\Lambda_{i,1} = \partial_{\rho_i}\Lambda(\rho_i, \rho_{i+1})$, and $\Lambda_{i,2} = \partial_{\rho_{i+1}}\Lambda(\rho_i, \rho_{i+1})$ we have

$$\begin{aligned} a_i &= 2\kappa_i\Lambda_i(\alpha_i{+}\beta_i) - \kappa_i\Lambda_{i,1}\big(\beta_{i-1}(\rho_i - \rho_{i-1}) + \alpha_i(\rho_i - \rho_{i+1})\big) \\ &\quad -\kappa_i\Lambda_{i,2}\big(\beta_i(\rho_{i+1} - \rho_i) + \alpha_{i+1}(\rho_{i+1} - \rho_{i+2})\big) \qquad \text{for } i = 1, \ldots, n-1; \\ b_i &= -\kappa_i\alpha_{i+1}(\Lambda_i{+}\Lambda_{i+1}) = -\kappa_{i+1}\beta_i(\Lambda_i{+}\Lambda_{i+1}) \leq 0. \end{aligned}$$

The desired positive semi-definiteness of $\mathbb{M}(u) - 2\lambda\mathbb{L}(u)$ (cf. (5.5)) will follow from diagonal dominance, which reads in this case

$$A_1 := a_1 + b_1 - 2\lambda\kappa_1\Lambda_1 \quad \geq 0, \tag{5.6a}$$
$$A_i := a_i + b_{i-1} + b_i - 2\lambda\kappa_i\Lambda_i \quad \geq 0 \text{ for } i = 2, \ldots, n-2, \tag{5.6b}$$
$$A_{n-1} = a_{n-1} + b_{n-2} - 2\lambda\kappa_{n-1}\Lambda_{n-1} \quad \geq 0. \tag{5.6c}$$

Indeed, using $b_i \leq 0$ these conditions yield the desired positive semi-definiteness

$$\mathbb{M}(u) - 2\lambda\mathbb{L}(u) = \operatorname{diag}(A_1, \ldots, A_{n-1}) + \sum_{i=1}^{n-2} |b_i|\,(e_i - e_{i+1})\otimes(e_i - e_{i+1}) \;\geq\; 0.$$

To establish the estimates (5.6b) we inspect the formula for A_i and find

$$\begin{aligned} A_i &= \kappa_i\Big(\widetilde{A}_i(\rho_i, \rho_{i+1}) - \beta_{i-1}\big(\Lambda(\rho_{i-1}, \rho_i) - \Lambda_{i,1}\rho_{i-1}\big) - \alpha_{i+1}\big(\Lambda(\rho_{i+1}, \rho_{i+2}) - \Lambda_{i,2}\rho_{i+2}\big)\Big) \\ &\quad \text{with } \widetilde{A}_i(\rho_i, \rho_{i+1}) = \Lambda_i(2\alpha_i{+}2\beta_i - \beta_{i-1} - \alpha_{i+1} - 2\lambda) \\ &\quad -\Lambda_{i,1}\big((\beta_{i-1}{+}\alpha_i)\rho_i - \alpha_i\rho_{i+1}\big) - \Lambda_{i,2}\big(-\beta_i\rho_i + (\beta_i{+}\alpha_{i+1})\rho_{i+1}\big). \end{aligned}$$

Since ρ_{i-1} and ρ_{i+2} occur only twice, the minimization with respect to ρ_{i-1} and ρ_{i+2} is easily possible. Employing the crucial estimate (A.7) for ρ_{i-1} and ρ_{i+2} separately, we find

$$A_i \geq \kappa_i\Gamma_i \quad \text{with } \Gamma_i := \widetilde{A}_i(\rho_i, \rho_{i+1}) - \beta_{i-1}\rho_i\Lambda_{i,2} - \alpha_{i+1}\rho_{i+1}\Lambda_{i,1}.$$

Reinserting the definition of $\widetilde{A}_i$ and expressing $\Lambda_{i,j}$ in terms of ρ_i, ρ_{i+1}, and Λ_i (cf. (A.3)) we obtain, after some rearrangements, cancellations, and using (A.4a), the identity

$$\Gamma_i = \Lambda_i(\alpha_i+\beta_i-\beta_{i-1}-\alpha_{i+1}-2\lambda + \Sigma_i) \text{ with } \Sigma_i := \Lambda(\rho_i, \rho_{i+1})\big(\tfrac{\alpha_i-\alpha_{i+1}}{\rho_i} + \tfrac{\beta_i-\beta_{i-1}}{\rho_{i+1}}\big).$$

Since $\Lambda(a, b)/a$ is not bounded, a lower bound for Σ_i exists if and only if the monotonicity (5.2) holds. Using this and the definition of Ξ yields $\Sigma_i \geq \Xi(\alpha_i - \alpha_{i+1}, \beta_i - \beta_{i-1})$.

Putting everything together we see that $\Gamma_i \geq 0$, and hence $A_i \geq 0$ follows from $\lambda \geq \gamma_i := G(\alpha_i - \alpha_{i+1}, \beta_i - \beta_{i-1})$, where G is defined in the statement of the theorem. This settles condition (5.6b), i.e. $i = 2, \ldots, n-2$.

For the case $i = 1$ and $i = n-1$ we proceed analogously with the only difference that the left or right neighbor are missing, respectively. All the above calculations for A_i remain valid for A_1 and A_{n-1}, if we use $\beta_0 = \kappa_0 = 0$ and $\alpha_n = \kappa_n = 0$, respectively. Thus, we obtain the additional conditions

$$\lambda \geq \gamma_1 := G(\alpha_1 - \alpha_2, \beta_1) \text{ and } \lambda \geq \gamma_{n-1} := G(\alpha_{n-1}, \beta_{n-1} - \beta_{n-2}).$$

Thus, Theorem 5.1 is established, i.e. $\lambda_Q \geq \gamma_Q = \min\{\, \gamma_i \mid i = 1, \ldots, n-1\,\}$. □

A simple first application of this result occurs in the chemical master equation for a reaction of the type $qX_a \leftrightharpoons pX_b$. On the macroscopic level the mass action law leads to the ODE system

$$\dot{a} = q\,\big(k_\mathrm{f} b^p - k_\mathrm{b} a^q\,\big), \quad \dot{b} = p\,\big(k_\mathrm{b} a^q - k_\mathrm{f} b^p\,\big),$$

where $k_\mathrm{f} > 0$ and $k_\mathrm{b} > 0$ are the forward and backward reaction rates, see e.g. [13,21]. On the microscopic level, where u_i is the probability of having exactly i atoms of species X_a, the chemical master equation gives the following Markov chain on $i \in \{0, 1, \ldots, n\}$:

$$\dot{u}_i = \alpha_{i-1}u_{i-1} - (\alpha_i+\beta_{i-1})u_i + \beta_i u_{i+1} \text{ with } \alpha_i = nk_\mathrm{f}(1 - i/n)^p \text{ and } \beta_i = nk_\mathrm{b}(i/n)^q, \tag{5.7}$$

see [16]. Clearly, the monotonicity (5.2) is always satisfied. Here the time scaling was done such that we obtain a uniform lower bound for λ_Q (i.e. independent of n) via

$$G(\alpha_i-\alpha_{i-1}, \beta_i-\beta_{i+1}) \geq \tfrac{1}{2}\big(\alpha_i-\alpha_{i-1}+\beta_i-\beta_{i+1}\big) \approx g(i/n) \geq \inf\{\, g(x) \mid x \in [0, 1]\,\} > 0,$$

where $2g(x) = pk_\mathrm{f}(1-x)^{p-1} + qk_\mathrm{b}x^{q-1}$.

The next result provides a corresponding upper bound for λ_Q that complements the lower bound given above.

Lemma 5.2 *Consider Q as in (5.1) with general α_i, $\beta_i > 0$ for $i = 1, \ldots, n-1$ and $\alpha_n = 0 = \beta_0$. Then we have the upper bound*

$$\lambda_Q \leq \min\{\, \alpha_i + \beta_i - (\alpha_{i+1}+\beta_{i-1})/4 \mid i = 1, \ldots, n-1\,\}.$$

Proof We use the same notation as in the proof of Theorem 5.1 and obtain an upper bound $\lambda_Q \leq (\eta \cdot M(u)\eta)/(\eta \cdot K(u)\eta)$ by choosing suitable η and u.

For $i = 1, \ldots, n-1$ we set $\eta^{(i)} = (1, \ldots, 1, 0, \ldots, 0) \in \mathbb{R}^n = \sum_{j=1}^{i} e_j$ and obtain the formula

$$R_i(u) = \frac{\eta^{(i)} \cdot M(u)\eta^{(i)}}{\eta^{(i)} \cdot K(u)\eta^{(i)}} = \alpha_i + \beta_i - \frac{\Lambda_{i,i}}{2\kappa_i\Lambda_i}\big[\beta_{i-1}(\rho_i - \rho_{i-1}) + \alpha_i(\rho_i - \rho_{i+1})\big]$$
$$-\frac{\Lambda_{i,i+1}}{2\kappa_i\Lambda_i}\big[\beta_i(\rho_{i+1} - \rho_i) + \alpha_{i+1}(\rho_{i+1} - \rho_{i+2})\big].$$

The expression is positively homogeneous of degree 0 in ρ. Moreover, only the four components $\rho_{i-1}, \ldots, \rho_{i+2}$ occur, where the occurrence of ρ_{i-1} and ρ_{i+2} is linear with positive prefactor. Thus, we may choose $\rho_{i-1} = \rho_{i+2} = 0$ and $\rho_i = \rho_{i+1} = 1$, i.e. $u^{(i)} = \frac{1}{w_i+w_{i+1}}(w_i e_i + w_{i+1} e_{i+1})$. Employing $\partial_a \Lambda(a,a)a = \Lambda(a,a)/2$ we obtain $R_i(u^{(i)}) = \alpha_i + \beta_i - (\alpha_{i+1}+\beta_{i-1})/4$. Since $R_i(u^{(i)}) \geq \lambda_Q$, the assertion is established.

Example 5.3 We consider the tridiagonal matrix $Q = \mathbb{R}^{3\times 3}$ as in (5.1) with $\alpha_1 = 16$, $\beta_1 = 12$, and $\alpha_2 = \beta_2 = 1$. Then, Lemma (5.2) implies $\lambda_Q \leq -1$, and the steady state $w = \frac{1}{11}(3, 4, 4)$ is strictly log-concave, i.e. (5.3) holds with strict inequality.

5.2 Geodesic convexity for the Fokker-Planck equation

To motivate this subsection on the discretization of the Fokker-Planck equation, we first consider the spatially continuous version, namely $\dot{U} = \operatorname{div}(\nabla U + U\nabla V)$ in Ω and $(\nabla U + U\nabla V)\cdot \nu = 0$ on $\partial\Omega$ for a smooth, bounded and convex domain $\Omega \subset \mathbb{R}^d$. Here we only give a formal argument motivating the geodesic $\widehat{\lambda}$-convexity of the relative entropy under the assumption that the potential V is $\widehat{\lambda}$-convex, i.e. in the smooth case we have

$$\xi \cdot \mathrm{D}^2 V(x)\xi \geq \widehat{\lambda}|\xi|^2 \text{ for all } x \in \Omega \text{ and } \xi \in \mathbb{R}^d.$$

First, we apply the approach of Sect. 2 in a formal way by assuming that all functions are sufficiently smooth and decay fast enough at infinity. The gradient structure of the Fokker-Planck equation derived in [15,26] is given via

$$\dot{U} = -\mathcal{K}(U)\mathrm{D}\mathcal{E}(U) \quad \text{with } \mathcal{E}(U) = \int_\Omega U \log U + VU \,\mathrm{d}x \text{ and } \mathcal{K}(U)\phi = -\operatorname{div}\big(U\nabla\phi\big), \tag{5.8}$$

To calculate the quadratic form $\mathcal{M}(U,\phi) = \langle \phi, M(U)\phi\rangle$ with M defined in (2.5) formally we use that the vector field $\mathcal{Q}(U) = \Delta U + \operatorname{div}(U\nabla V)$ and obtain

$$\begin{aligned}
\mathcal{M}(U,\phi) &= \tfrac{1}{2}\langle \phi, \mathrm{D}\mathcal{K}(U)[\mathcal{Q}(U)]\phi\rangle - \langle \phi, \mathrm{D}\mathcal{Q}(U)\mathcal{K}(U)\phi\rangle \\
&= \int_\Omega \tfrac{1}{2}\big(\Delta U + \operatorname{div}(U\nabla V)\big)|\nabla\phi|^2 + \phi\Big(\Delta\big(\operatorname{div}(U\nabla\phi)\big) + \operatorname{div}\big(\operatorname{div}(U\nabla\phi)\nabla V\big)\Big)\,\mathrm{d}x \\
&\overset{(*)}{=} \int_\Omega U\Big(|\mathrm{D}^2\phi|^2 + \nabla\phi\cdot \mathrm{D}^2 V\nabla\phi\Big)\,\mathrm{d}x - \int_{\partial\Omega} U\,\nabla\big(\tfrac{1}{2}|\nabla\phi|^2\big)\cdot\nu\,\mathrm{d}a \\
&\geq \int_\Omega U\,\widehat{\lambda}|\nabla\phi|^2\,\mathrm{d}x = \widehat{\lambda}\langle\phi, \mathcal{K}(U)\phi\rangle,
\end{aligned}$$

where $\overset{(*)}{=}$ is obtained by a series of integrations by parts using the no-flux boundary conditions for U and ϕ and by exploiting the relation $\Delta(\frac{1}{2}|\nabla\phi|^2) = |\mathrm{D}^2\phi|^2 + \nabla\phi\cdot\nabla(\Delta\phi)$. The final estimates follows by dropping $U|\mathrm{D}^2\phi|^2 \geq 0$, using the $\widehat{\lambda}$-convexity of V, and from the fact that $\nabla\phi\cdot\nu = 0$ on $\partial\Omega$ implies $\nabla\big(\frac{1}{2}|\nabla\phi|^2\big)\cdot\nu \leq 0$, since Ω is convex, see [17, Sect. 5]. The latter paper together with [9] provide a full proof of the geodesic $\widehat{\lambda}$-convexity that is based on a metric version of the Lie derivative $L_{\mathcal{Q}U}\mathcal{G}$ and applies to systems of PDEs.

5.3 Uniform geodesic λ-convexity for the discretization

We now turn our attention to the one-dimensional Fokker-Planck equation

$$\dot U = \big(U' + UV'\big)' \text{ in } \Omega = \,]0,1[\quad \text{and} \quad U'(t,x) + U(t,x)V'(x) = 0 \text{ for } x \in \{0,1\}. \tag{5.9}$$

Using an equidistant partition $x_i^n = i/n$ we may consider $u_i(t)$ as an approximation of $\int_{(i-1)/n}^{i/n} U(t,x)\,\mathrm{d}x$ and a simple *finite-difference discretization* gives the system of ODEs

$$\begin{aligned}
\dot u_1 &= n^2(u_2 - u_1) + \tfrac{n}{2}\big(u_1 V'(x_{1/2}^n) + u_2 V'(x_{3/2}^n)\big),\\
\dot u_i &= n^2(u_{i-1} - 2u_i + u_{i+1}) + \tfrac{n}{2}\big(u_{i+1} V'(x_{i+1/2}^n) - u_{i-1} V'(x_{i-3/2}^n)\big),\\
\dot u_n &= n^2(u_{n-1} - u_n) - \tfrac{n}{2}\big(u_{n-1} V'(x_{n-3/2}^n) + u_n V'(x_{n-1/2}^n)\big),
\end{aligned} \tag{5.10}$$

where the first terms correspond to the diffusion part, while the second term contains the drift induced by the potential. Thus, we have

$$\alpha_i = n^2 - \tfrac{n}{2} V'(x_{i-1/2}) \quad \text{and} \quad \beta_i = n^2 + \tfrac{n}{2} V'(x_{i+1/2}).$$

Thus, assuming $V \in \mathrm{C}^1([0,1])$ we have $\alpha_i, \beta_i > 0$ whenever $2n > \|V'\|_{\mathrm{C}^0}$, which implies that (5.10) is a Markov chain. Assuming further that V is $\widehat\lambda$-convex with $\widehat\lambda \geq 0$, we obtain the desired monotonicity (5.2). Moreover, we obtain the quantitative estimates

$$\alpha_i - \alpha_{i+1} = \beta_i - \beta_{i-1} = \tfrac{n}{2}\big(V'(x_{i+1/2}) - V'(x_{i-1/2})\big) \geq \widehat\lambda/2. \tag{5.11}$$

Using that G satisfies $G(a,a) = 2a$ we arrive at the following result.

Corollary 5.4 *Assume* $V \in \mathrm{C}^2([0,1])$, $\inf\{\, V''(x) \,|\, x \in \,]0,1[\,\} \geq \widehat\lambda \geq 0$, *and* $\|V'\|_{\mathrm{L}^\infty} < 2n$. *Then* Q^{FD} *defined via the finite-difference scheme* (5.10) *satisfies* $\lambda_{Q^{FD}} \geq \widehat\lambda$.

The above result has the disadvantage that it only works for sufficiently high n and that it applies only for equidistant discretizations. For general partitions

$$0 = x_0^n < x_1^n < \cdots < x_{n-1}^n < x_n^n = 1 \quad \text{and} \quad x_{i-1/2}^n := \tfrac12 (x_{i-1}^n + x_i^n) \tag{5.12}$$

we can still find the consistent discretization (5.10), but now α_i and β_i are given by

$$\alpha_i = \frac{1}{x_{i+1} - x_{i-1}}\Big(\frac{2}{x_i - x_{i-1}} - V'(x_{i-\frac12}^n)\Big) \text{ and } \beta_i = \frac{1}{x_{i+1} - x_{i-1}}\Big(\frac{2}{x_{i+1} - x_i} + V'(x_{i+\frac12}^n)\Big).$$

While $\max\{\, x_i^n - x_{i-1}^n \,|\, i = 1, \ldots, n\,\}\|V'\|_{\mathrm{C}^0} < 2$ again implies the positivity $\alpha_i, \beta_i > 0$, it very difficult to satisfy the monotonicity conditions (5.2).

Finite-volume discretization schemes are better adapted to drift-diffusion equations, because they automatically preserve positivity and conserve the mass exactly. We rewrite (5.9) using the equilibrium density $W(x) = c\,\mathrm{e}^{-V(x)}$ with $\int_0^1 W\,\mathrm{d}x = 1$ and find

$$\dot U = \big(W(U/W)'\big)' \text{ in } \Omega = \,]0,1[\quad \text{and} \quad (U/W)'(t,x) = 0 \text{ for } x \in \{0,1\}. \tag{5.13}$$

For a general partition as in (5.12) we define $w_i^n = \int_{x_{i-1}^n}^{x_i^n} W(x)\,\mathrm{d}x$ and expect $u_i(t)$ to approximate $\int_{x_{i-1}^n}^{x_i^n} U(t,x)\,\mathrm{d}x$. Integrating (5.13) over $[x_{i-1}^n, x_i^n]$ gives $\dot u_i = q_i^n - q_{i-1}^n$ where q_i^n approximates $W(U/W)'$ at x_i^n and $q_0^n = q_n^n = 0$. The natural choice is $q_i^n = \kappa_i^n\big(\frac{u_{i+1}}{w_{i+1}^n} - \frac{u_i}{w_i^n}\big)$, where consistency of the discretization scheme holds if $\kappa_i^n\big(x_{i+1/2}^n - x_{i-1/2}^n\big)/W(x_i^n) \to 1$ for $n \to \infty$, uniformly in i, see e.g. [12]. Thus, the discretization takes the form

$$\dot u_i = \alpha_{i-1}^n u_{i-1} - (\alpha_i^n + \beta_{i-1}^n)u_i + \beta_i^n u_{i+1} \quad \text{with } \alpha_i^n = \frac{\kappa_i^n}{w_i^n} \text{ and } \beta_i^n = \frac{\kappa_i^n}{w_{i+1}^n}. \tag{5.14}$$

Note that the present usage of α_i, β_i, κ_i, and w_i is consistent with that in Sect. 5.1.

From the definition of the finite-volume scheme we immediately have the positivity $\alpha_i^n, \beta_i^n > 0$ independent of the fineness of the partition. To discuss the monotonicity (5.2) we first consider the equidistant case.

Corollary 5.5 *Assume* $V \in \mathrm{C}^2([0,1])$ *and* $\inf\{\, V''(x)\,|\, x \in\,]0,1[\,\} \geq \widehat\lambda \geq 0$. *If* $\dot u = Q_n^{FV} u \in \mathbb{R}^n$ *denotes the finite-volume discretization* (5.14) *with the equidistant partition* $x_i^n = i/n$,

$$w_i^n = \int_{x_{i-1}^n}^{x_i^n} W(x)\,\mathrm{d}x, \quad \text{and} \quad \kappa_i^n = n^2\sqrt{w_i^n w_{i+1}^n} \text{ for } n = 1, \ldots, n-1,$$

then we have $\lambda_{Q_n^{FV}} \geq 2n^2\Phi(\widehat\lambda/(8n^2))$ *and* $\lambda_{Q_n^{FV}} \to \widehat\lambda$ *for* $n \to \infty$, *where*

$$\Phi(\mu) = \frac{3\,\mathrm{Erf}(\sqrt{\mu}) - \mathrm{Erf}(3\sqrt{\mu})}{2\,\mathrm{Erf}(\sqrt{\mu})} = 4\mu + O(\mu^2) \text{ with } \mathrm{Erf}(s) = \frac{2}{\sqrt{\pi}}\int_0^s \mathrm{e}^{-r^2}\,\mathrm{d}r.$$

Proof To simplify the notation we drop the superscript ^{n} and set $q_i = \sqrt{w_{i+1}/w_i}$ such that $\alpha_i = q_1$ and $\beta_i = 1/q_i$. We estimate w_i from below and w_{i-1} and w_{i+1} from above by comparing V with a parabola $c + dx + \widehat\lambda x^2/2$ coinciding with V in x_{i-1} and x_i. With the definition of Φ and the abbreviation $\Psi = \Phi(\widehat\lambda/(8n^2))$ we find

$$(1-\Psi)w_i \geq \sqrt{w_{i-1}w_{i+1}} \quad \Longleftrightarrow \quad (1-\Psi)q_{i-1} \geq q_i \qquad \text{for } i = 2, \ldots, n-1.$$

To apply Theorem 5.1 we estimate as follows:

$$\alpha_i - \alpha_{i+1} = n^2(q_i - q_{i+1}) \geq n^2\Psi q_i \text{ and } \beta_i - \beta_{i-1} = n^2/q_i - n^2/q_{i-1} \geq n^2\Psi/q_i.$$

Using the monotonicity and 1-homogeneity of G as well as $G(a,b) \geq 2\sqrt{ab}$ we conclude

$$G(\alpha_i - \alpha_{i+1}, \beta_i - \beta_{i-1}) \geq G(n^2\Psi q_i, n^2\Psi/q_i) = n^2\Psi\, G(q_i, 1/q_i) \geq 2n^2\Psi,$$

and the result is established. □

The major advantage of the finite-volume discretization is that it is possible to allow for non-equidistant partitions. For $\widehat\lambda > 0$ we can borrow convexity from the potential V to accommodate variations in the lengths of the intervals of the partition. For a general partition, see (5.12), we define

$$w_i^n = c_n W(x_{i-1/2}^n)(x_i^n - x_{i-1}^n) \quad \text{and} \quad \kappa_i^n = \frac{\sqrt{w_i^n w_{i+1}^n}}{(x_i^n - x_{i-1}^n)(x_{i+1}^n - x_i^n)}, \tag{5.15}$$

where c_n is chosen such that $\sum_{i=1}^n w_i^n = 1$, which implies $c_n \to 1$. Clearly, this choice leads to a consistent finite-volume scheme. The next result shows that $\widehat\lambda$-convexity of V with $\widehat\lambda > 0$ allows for graded meshes if the allowed factor γ in

$$\frac{1}{\gamma} \le \frac{x_i^n - x_{i-1}^n}{x_{i+1}^n - x_i^n} \le \gamma \quad \text{for } i = 1, \dots, n-1, \tag{5.16}$$

is sufficiently close to 1.

Corollary 5.6 *Assume* $V \in \mathrm{C}^2([0,1])$ *and* $V'' \ge \widehat{\lambda} > 0$. *If a partition* (5.12) *satisfies* (5.16) *with* $\gamma \ge 1$ *and*

$$\Psi := 1 - \gamma^2\, \mathrm{e}^{-\widehat{\lambda} h_*^2/2} \ge 0 \quad \textit{where } h_* = \min\{\, x_i^n - x_{i-1}^n \mid i = 1, \dots, n\,\}, \tag{5.17}$$

then the Markov chain $\dot{u} = Q_n u$ *defined via* (5.15) *satisfies* $\lambda_{Q_n} \ge 2\Psi^2/H^2 \ge 0$, *where* $H = \max\{\, x_i^n - x_{i-1}^n \mid i = 1, \dots, n\,\}$.

Proof As in the previous proof we drop the superscript ^{n} and introduce the quotient $q_i = \sqrt{W(x_{i+1/2}^n)/W(x_{i-1/2}^n)}$. The $\widehat{\lambda}$-convexity of V yields $q_i \ge \mathrm{e}^{\widehat{\lambda} h_*^2/2} q_{i+1}$. Using the abbreviations $h_i = x_i - x_{i-1}$ we find the representations $\alpha_i = q_i/(h_i^{3/2} h_{i+1}^{1/2})$ and $\beta_i = 1/(q_i h_i^{1/2} h_{i+1}^{3/2})$. For α_i we obtain the estimates

$$\alpha_i - \alpha_{i+1} = \alpha_i \big(1 - \frac{h_i^{3/2}}{h_{i+1} h_{i+2}^{1/2}} \frac{q_{i+1}}{q_i}\big) \ge \alpha_i \big(1 - \gamma^2 \mathrm{e}^{-\widehat{\lambda} h_*^2/2}\big) \ge \Psi \alpha_i \ge 0$$

by (5.17). Similarly, we have $\beta_i - \beta_{i-1} \ge \Psi \beta_i \ge 0$. To apply Theorem 5.1 we use

$$G(\alpha_i - \alpha_{i+1}, \beta_i - \beta_{i-1}) \ge \Psi\, G(\alpha_i, \beta_i) \ge 2\Psi \sqrt{\alpha_i \beta_i} = 2\Psi/(h_i h_{i+1}) \ge 2\Psi/H^2,$$

which proves the assertion. □

A very similar finite-volume scheme for drift-diffusion equations is the *Scharfetter–Gummel scheme*, which in the one-dimensional case takes again the form (5.14) but now with

$$\alpha_i = \frac{B(-h_{i+1/2})V'(x_i)}{h_i h_{i+1/2}} \text{ and } \beta_i = \frac{B(h_{i+1/2})V'(x_i)}{h_{i+1} h_{i+3/2}}, \text{ where } B(s) = \frac{1}{\Lambda(1, \mathrm{e}^s)} = \frac{s}{\mathrm{e}^s - 1},$$

$h_i = x_i - x_{i-1}$, and $h_{i+1/2} = x_{i+1/2} - x_{i-1/2}$, see e.g. [6]. Here B is the Bernoulli function that is closely related to the logarithmic mean Λ. Restricting to an equidistant partition with $x_i = i/n$, assuming $V''(x) \ge \widehat{\lambda} \ge 0$, and setting $b_i = V'(i/n)$, we can use $B' < 0$ and $B'' \ge 0$ to obtain

$$\begin{aligned} \alpha_i - \alpha_{i+1} &= n^2\big(B(-b_i/n) - B(-b_{i+1}/n)\big) \overset{B''>0}{\ge} n^2\big(-B'(-b_{i+1}/n)\big)(b_{i+1} - b_i)/n \\ &\overset{B'<0}{\ge} \big(-B'(-b_i/n)\big) n (b_{i+1} - b_i) \overset{V'' \ge \widehat{\lambda}}{\ge} \widehat{\lambda}\big(-B'(-b_i/n)\big) \ge 0. \end{aligned}$$

Similarly, we obtain $\beta_i - \beta_{i-1} \ge \widehat{\lambda}\big(-B'(b_i/n)\big) \ge 0$. Using the well-known identity $B(s) + s = B(-s)$ we obtain $B'(s) + B'(-s) = -1$ and, using $G(a,b) = \frac{1}{2}(a+b+2\sqrt{ab})$ we conclude

$$\lambda_Q \ge \widehat{\lambda} \Big(\frac{1}{2} + \big|B'(-\tfrac{1}{n}\|V'\|_{\mathrm{L}^\infty})\big|\Big).$$

Thus, the Scharfetter–Gummel scheme yields a good uniform bound on the geodesic convexity even in the case that that $\|V'\|_\infty$ is huge or ∞, as long as V is convex.

Remark 5.7 In two-point finite-volume schemes the occurrence of quotients $\Phi(a,b) = (h(a) - h(b))/(\phi'(a) - \phi'(b))$ as in Proposition 3.1 (in particular $\Lambda(a,b)$) is quite common, see [6, Eq. (28)].

Remark 5.8 While we have only considered the one-dimensional case, we expect that it is possible to find suitable generalization for higher dimensions as well. In fact, the numerical finite-volume discretizations constructed in [14,13] obviously lead to reversible Markov chains, but their geodesic λ-convexity needs to be investigated.

Appendix A: Properties of the function Λ

In this section we collect the essential properties of the function Λ defined in (1.3). The value $\Lambda(a,b)$ can also be seen as the logarithmic average of a and b defined via

$$\Lambda(a,b) = \int_{\theta=0}^{1} a^\theta\, b^{1-\theta}\,\mathrm{d}\theta.$$

Other useful representations of Λ are for the inverse, namely

$$\frac{1}{\Lambda(a,b)} = \int_{\theta=0}^{1} \frac{\mathrm{d}\theta}{(1-\theta)a+\theta b} = \int_{t=0}^{\infty} \frac{\mathrm{d}t}{(a+t)\,(b+t)}.$$

We have the obvious estimates

$$\frac{2ab}{a+b} \le \sqrt{ab} \le \Lambda(a,b) \le \frac{1}{2}\,(a+b). \tag{A.1}$$

The lower estimate for Λ can be generalized to

$$\forall\,\theta\in[0,1]\;\forall\,a,b\ge 0:\quad \Lambda(a,b)\ge 2\min\{\theta,1-\theta\}\,a^\theta b^{1-\theta}. \tag{A.2}$$

This estimate follows from the convexity of $f : s \mapsto a^s b^{1-s}$ via integration of $f(s) \ge f(\theta)+f'(\theta)(s-\theta)$ over $[0,2\theta]$ or $[2-2\theta,1]$, respectively. Elementary calculations give

$$0<\partial_a\Lambda(a,b) = \frac{1}{\log a-\log b}\Big(1-\frac{\Lambda(a,b)}{a}\Big) = \frac{\Lambda(a,b)\,(a-\Lambda(a,b))}{a\,(a-b)} > 0, \tag{A.3}$$

which implies

$$a\partial_a\Lambda(a,b)+b\partial_b\Lambda(a,b) = \Lambda(a,b), \tag{A.4a}$$

$$b\partial_a\Lambda(a,b)+a\partial_b\Lambda(a,b) = \Lambda(a,b)^2\big(\tfrac{1}{a}+\tfrac{1}{b}\big)-\Lambda(a,b)\ge\Lambda(a,b), \tag{A.4b}$$

$$\partial_a\Lambda(a,b)+\partial_b\Lambda(a,b) = \tfrac{\Lambda(a,b)^2}{ab}\ge 1, \tag{A.4c}$$

$$(\partial_a\Lambda(a,b)-\partial_b\Lambda(a,b))(a-b) = \Lambda(a,b)\big(2-\tfrac{a+b}{ab}\Lambda(a,b)\big)\le 0. \tag{A.4d}$$

Note that (A.4a) is also a consequence of the following 1-homogeneity:

$$\Lambda(\gamma\,a,\gamma\,b) = \gamma\,\Lambda(a,b)\text{ for all }a,\ b,\ \gamma>0. \tag{A.5}$$

The following estimate is used in Theorem 5.1: for all $a,b>0$ we have

$$\max\{\Lambda(a,b),2\sqrt{ab}\}\le\Xi(a,b):=\inf\{\,\Lambda(r,s)(\frac{a}{r}+\frac{b}{s})\mid r,s>0\,\}\le 2\Lambda(a,b). \tag{A.6}$$

For the upper bound choose $(r, s) = (a, b)$ and for the lower estimate proceed as follows:

$$(\frac{a}{r} + \frac{b}{s})\Lambda(r, s) = \int_{\sigma=0}^{1} ar^{\sigma-1}s^{1-\sigma} + br^{\sigma}s^{-\sigma}\,\mathrm{d}\sigma \geq \int_{0}^{1} \frac{a^{\sigma}b^{1-\sigma}}{\sigma^{\sigma}(1-\sigma)^{1-\sigma}}\,\mathrm{d}\sigma \geq \Lambda(a, b),$$

$$(\frac{a}{r} + \frac{b}{s})\Lambda(r, s) \geq (\frac{a}{r} + \frac{b}{s})\sqrt{rs} \geq 2\sqrt{ab}.$$

A nontrivial estimate and identity is the following:

$$\max\{\,\Lambda(r, a) - \partial_a\Lambda(a, b)r \mid r > 0\,\} = a\partial_b\Lambda(a, b). \tag{A.7}$$

The result uses somehow hidden properties of Λ and is crucial for our lower bounds for λ_Q. Using the homogeneity (A.5), this identity follows from (A.8c), which is established below using the auxiliary function ℓ defined in (1.4).

Proposition A.1 *We define the function* $\widetilde{\ell}(\kappa) = (\mathrm{e}^{\kappa} - 1 - \kappa)/\kappa^2 > 0$. *The function* ℓ *satisfies the following properties:*

$$l = \ell(\xi) \iff \big(\exists\,\kappa \in \mathbb{R} : \; l = \widetilde{\ell}(\kappa) \text{ and } \xi = \widetilde{\ell}(-\kappa)\big), \tag{A.8a}$$

$$\forall\,\xi > 0 : \quad \ell(\ell(\xi)) = \xi, \tag{A.8b}$$

$$\forall\,a, b > 0 : \quad \ell(\partial_a\Lambda(a, b)) = \partial_b\Lambda(a, b). \tag{A.8c}$$

Proof We first observe that $\Lambda(\cdot, 1)$ is strictly concave and that it has sublinear growth as $\Lambda(r, 1) \sim r/\log r$ for $r \gg 1$. Hence, the maximum in the definition (1.4) of ℓ is attained a unique value r. We find $\ell(\xi) = \widetilde{\ell}(\kappa)$, where $\kappa = \widehat{\kappa}(\xi)$ is the unique solution of $\xi = (\kappa - 1 + \mathrm{e}^{-\kappa})/\kappa^2$ and $r = \mathrm{e}^{\kappa}$ is the maximizer of $r \mapsto \Lambda(r, 1) - \xi r$. Thus, (A.8a) is established. Identity (A.8b) follows directly from (A.8a), because l and ξ can be interchanged, when κ is multiplied by -1.

Finally, the partial derivatives $\partial_a\Lambda(a, b)$ and $\partial_b\Lambda(a, b)$ are 0-homogeneous and depend only on $\sigma = \log(a/b)$, namely $\partial_a\Lambda(a, b) = \widetilde{\ell}(-\sigma)$ and $\partial_b\Lambda(a, b) = \widetilde{\ell}(\sigma)$. Using $\kappa = -\sigma$ this gives (A.8c). □

The important identity (A.8b) follows also directly for any $\overline{\ell}$ defined via $\overline{\ell}(\xi) = \sup\{\,\overline{\lambda}(r) - \xi r \mid r > 0\,\}$ if $\overline{\lambda}(r) = r\overline{\lambda}(1/r)$, which in our case follows from $\Lambda(1, r) = r\Lambda(1/r, 1) = r\Lambda(1, 1/r)$.

Remark A.2 While the above proof of (A.7) can be adapted easily to general symmetric, concave, and 1-homogeneous functions Λ (see also [11, Lemma 5.4]), there is a short way to derive (A.7) for Λ being the logaritmic mean. By 1-homogeneity of Λ the unique solution r of $\partial_r\Lambda(r, a) = \partial_a\Lambda(a, b)$ is a^2/b, and hence the maximum in (A.7) is attained for $r = a^2/b$. Inserting this and using (A.4a) gives the result.

Appendix B: Proof of Proposition 4.3

Here we provide the lower bound for the eigenvalues of the matrix

$$G_\beta(r, s, t) \overset{\text{def}}{=} \begin{pmatrix} \Lambda_{rs}(\frac{1}{r}+\frac{1}{s}) + \frac{\Lambda_{rs,s}}{\Lambda_{rs}}\,t & \beta(\Lambda_{rs}/\Lambda_{st})^{1/2} + \beta(\Lambda_{st}/\Lambda_{rs})^{1/2} \\ \beta(\Lambda_{rs}/\Lambda_{st})^{1/2} + \beta(\Lambda_{st}/\Lambda_{rs})^{1/2} & \Lambda_{st}(\frac{1}{s}+\frac{1}{t}) + \frac{\Lambda_{st,s}}{\Lambda_{st}}\,r \end{pmatrix},$$

where again $\Lambda_{ab} = \Lambda(a, b)$ and $\Lambda_{ab,a} = \partial_a \Lambda(a, b)$. By homogeneity of degree 0 it is sufficient to consider

$$(r, s, t) \in \Delta \stackrel{\text{def}}{=} \{ (r, s, t) \in]0, 1[^3 \mid r + s + t = 1 \}.$$

Since G_β is continuous on Δ its lowest eigenvalue depends continuously on $(r, s, t) \in \Delta$ as well. To prove boundedness from below it hence suffices to show a lower bound near the boundary of Δ. In fact, we prove that G_β is positive semidefinite near the boundary of Δ. For this, it is sufficient to show that the determinant of G_β is nonnegative, as the diagonal entries are bigger than 1.

The sign of the determinant of G_β is controlled by the auxiliary function $\widehat{\gamma}$ via

$$\det G_\beta(r, s, t) \geq 0 \iff \widehat{\gamma}(r, s, t) \leq 1/\beta^2,$$

$$\text{where } \widehat{\gamma}(r, s, t) \stackrel{\text{def}}{=} \frac{\frac{\Lambda_{rs}}{\Lambda_{st}} + 2 + \frac{\Lambda_{st}}{\Lambda_{rs}}}{\big(\Lambda_{rs}(\frac{1}{r}+\frac{1}{s}) + \frac{\Lambda_{rs,s}}{\Lambda_{rs}}\, t\big)\big(\Lambda_{st}(\frac{1}{s}+\frac{1}{t}) + \frac{\Lambda_{st,s}}{\Lambda_{st}}\, r\big)}.$$

Using (A.1) it is not difficult to show $\widehat{\gamma}(r, s, t) \leq 1$ which implies that $G_\beta(r, s, t)$ is positive semidefinite for $|\beta| \leq 1$ and all (r, s, t).

To prove our statement for all $\beta \geq 0$, we have to show that $\widehat{\gamma}(r, s, t) \to 0$ if (r, s, t) approaches the boundary of the two-dimensional triangle Δ. We do this by discussing the three corners and the three sides of Δ separately. For proving convergence of $\widehat{\gamma}$ to 0, it is obviously sufficient to omit the "2" in the numerator, so that we estimate the function γ with $\widehat{\gamma} \leq 2\gamma$ and

$$\gamma(r, s, t) \stackrel{\text{def}}{=} \frac{\Lambda_{rs}^2 + \Lambda_{st}^2}{\big(\Lambda_{rs}^2(\frac{1}{r}+\frac{1}{s}) + \Lambda_{rs,s}\, t\big)\big(\Lambda_{st}^2(\frac{1}{s}+\frac{1}{t}) + \Lambda_{st,s}\, r\big)}.$$

Case 1: $s \to 1$ *and* $r, t \to 0$. We have

$$\gamma \leq \frac{\Lambda_{rs}^2+\Lambda_{st}^2}{\big(\Lambda_{rs}^2/r\big)\big(\Lambda_{st}^2/t\big)} = rt\Big(\frac{1}{\Lambda_{rs}^2} + \frac{1}{\Lambda_{rs}^2}\Big) \leq rt\Big(\frac{4}{r^{2/3}} + \frac{4}{t^{2/3}}\Big) = 4(rt)^{1/3}\big(r^{2/3}+t^{2/3}\big) \to 0,$$

where we used (A.2) in the form $\Lambda_{rs} \geq \frac{2}{3} r^{1/3} s^{2/3} \geq r^{1/3}/2$ for $s \approx 1$.

Case 2: $t \to 1$ *and* $r, s \to 0$. Using $r < t$ we have $\Lambda_{rs} < \Lambda_{st}$ and obtain

$$\gamma \leq \frac{2\Lambda_{st}^2}{\big(\Lambda_{rs}^2(\frac{1}{r}+\frac{1}{s}) + \Lambda_{rs,s}\, t\big)\big(\Lambda_{st}^2/s\big)} = \frac{2s}{\Lambda_{rs}^2(\frac{1}{r}+\frac{1}{s}) + \Lambda_{rs,s}\, t}$$

To proceed we need a good lower bound for $\Lambda_{rs,s}$, namely

$$\Lambda_{rs,s} = \Lambda_{rs} \frac{\Lambda_{rs}-s}{s(r-s)} \geq \Lambda_{rs} \frac{\Lambda_{rs}+s}{3s(r+s)} \geq \Lambda_{rs}/(3r+3s).$$

We continue via

$$\gamma \leq \frac{6rs^2}{\Lambda_{rs}^2(r+s) + \Lambda_{rs} rs/(r+s)} \leq \frac{6rs^2}{\Lambda_{rs}^2 \max\{r, s\} + \Lambda_{rs} \min\{r, s\}}.$$

Hence, for $0 < r \leq s \ll 1$ we obtain

$$\gamma \leq \frac{6rs^2}{\Lambda_{rs}^2 s + \Lambda_{rs} r} \leq 6 \min\{\tfrac{rs}{\Lambda_{rs}^2}, \tfrac{s^2}{\Lambda_{rs}}\} \leq 14 \min\{r^{1/3}s^{-1/3}, r^{-1/2}s^{3/2}\} \leq 14 s^{2/5}.$$

where we used (A.2) with $\theta = 1/3$. For $0 < s < r \ll 1$ we use (A.1) to obtain

$$\gamma \leq \frac{6rs^2}{\Lambda_{rs}^2 r + \Lambda_{rs} s} \leq 6\frac{rs}{\Lambda_{rs}} \leq 6\sqrt{rs} \leq 6r.$$

Thus, $\gamma(r, s, t) \to 0$ follows for $r, s \to 0$.

Case 3: $r \to 1$ *and* $t, s \to 0$. This case is the same as Case 2 via interchanging r and t.

Case 4: $s \to 0$*,* $r \to r_* > 0$*, and* $t \to t_* = 1 - r_* > 0$. We have

$$\gamma(r, s, t) \le \frac{\Lambda_{rs}^2 + \Lambda_{st}^2}{\big(\Lambda_{rs}^2 \frac{1}{s}\big)\big(\Lambda_{st}^2 \frac{1}{s}\big)} = s^2\Big(\frac{1}{\Lambda_{rs}^2} + \frac{1}{\Lambda_{st}^2}\Big)^2 \le s^2\Big(\frac{1}{rs} + \frac{1}{st}\Big) \le 2s(1/r_* + 1/t_*) \to 0.$$

Case 5: $r \to 0$, $s \to s_* > 0$, and $t \to t_* = 1 - s_* > 0$. Since the numerator of γ converges to $\Lambda(s_*, t_*)^2 > 0$ it suffices to show that the denominator tens to $+\infty$. Indeed,

$$\big(\Lambda_{st}^2(\tfrac{1}{s}+\tfrac{1}{t}) + \Lambda_{st,s}\, r\big) \to n_* > 0 \text{ and } \big(\Lambda_{rs}^2(\tfrac{1}{r}+\tfrac{1}{s}) + \Lambda_{rs,s}\, t\big) \ge \Lambda_{rs}^2/r \to +\infty.$$

Thus, $\gamma(r, s, t) \to 0$ follows also for $r \to 0$.

Case 6: $t \to 0$, $s \to s_* > 0$*, and* $r \to r_* = 1 - s_* > 0$. This case is the same as Case 5 via interchanging r and t.

This finishes the proof of Proposition 4.3.

Acknowledgments This research partially sponsored by the European Research Council under ERC-2010-AdG 267802. The author is grateful to Giuseppe Savaré for stimulating discussions and many helpful comments. He also thanks Jürgen Fuhrmann and Robert Eymard for help with finite-volume schemes. Moreover, he thanks an unknown referee for clearing up the relation of M with the Hessians and Lie derivatives as well as for spotting several inconsistencies. The author thanks Thomas Frenzel for pointing out Remark A.2. The research was partially supported by ERC-2010-AdG 267802 *(Analysis of Multiscale Systems Driven by Functionals)*.

References

1. Ambrosio, L., Gigli, N., Savaré, G.: Gradient Flows in Metric Spaces and in the Space of Probability Measures. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel (2005)
2. Arnrich, S., Mielke, A., Peletier, M.A., Savaré, G., Veneroni, M.: Passing to the limit in a Wasserstein gradient flow: from diffusion to reaction. Calc. Var. Partial Differ. Equ. **44**, 419–454 (2012)
3. Bakry, D.: L'hypercontractivité et son utilisation en théorie des semigroupes. In: Lectures on Probability Theory (Saint-Flour, 1992), vol. 1581 of Lecture Notes in Mathematics, pp. 1–114. Springer, Berlin (1994)
4. Bakry, D., Emery, M.: Diffusions hypercontractives. In: Séminaire de probabilité, XIX, 1983/1984, vol. 1123 of Lecture Notes in Mathematics, pp. 177–206. Springer, Berlin (1985)
5. Benamou, J.-D., Brenier, Y.: A computational fluid mechanics solution to the Monge–Kantorovich mass transfer problem. Numer. Math. **84**(3), 375–393 (2000)
6. Bessemoulin-Chatard, M.: A finite volume scheme for convection-diffusion equations with nonlinear diffusion derived from the Scharfetter–Gummel scheme. Numer. Math., to appear (2011)
7. Bonciocat, A.-I., Sturm, K.-T.: Mass transportation and rough curvature bounds for discrete spaces. J. Funct. Anal. **256**, 2944–2966 (2009)
8. Chow, S.-N., Huang, W., Li, Y., Zhou, H.: Fokker–Planck equations for a free energy functional of Markov process on a graph. Arch. Rational Mech. Anal. **203**(3), 969–1008 (2012)
9. Daneri, S., Savaré, G.: Eulerian calculus for the displacement convexity in the Wasserstein distance. SIAM J. Math. Anal. **40**, 1104–1122 (2008)
10. Daneri, S., Savaré, G.:Lecture notes on gradient flows and optimal transport. arXiv:1009.3737v1, 2010.
11. Erbar, M., Maas, J.: Ricci curvature of finite Markov chains via convexity of the entropy. arXiv: 1111.2687, 2011. Arch. Ration. Mech. Anal., to appear (2012)
12. Eymard, R., Gallouët, T., Herbin, R.: The finite volume method. In: Handbook of Mumerical Analysis, vol. VII, pp. 715–1022. North Holland, Amsterdam (2000)
13. Glitzky, A.: Exponential decay of the free energy for discretized electro-reaction-diffusion systems. Nonlinearity **21**(9), 1989–2009 (2008)
14. Glitzky, A., Gärtner, K.: Energy estimates for continuous and discretized electro-reaction-diffusion systems. Nonlinear Anal. **70**(2), 788–805 (2009)

15. Jordan, R., Kinderlehrer, D., Otto, F.: The variational formulation of the Fokker–Planck equation. SIAM J. Math. Anal. **29**(1), 1–17 (1998)
16. Kurtz, T.G.: Solutions of ordinary differential equations as limits of pure jump processes. J. Appl. Prob. **7**, 49–58 (1970)
17. Liero, M., Mielke, A.: Gradient structures and geodesic convexity for reaction-diffusion systems. Philos. Trans. R. Soc. A, submitted (2012), WIAS preprint 1701 (April 2012)
18. Lott, J., Villani, C.: Ricci curvature for metric-measure spaces via optimal transport. Ann. Math. **169**(2–3), 903–991 (2009)
19. Maas, J.: Gradient flows of the entropy for finite Markov chains. J. Funct. Anal. **261**, 2250–2292 (2011)
20. McCann, R.J.: A convexity principle for interacting gases. Adv. Math. **128**, 153–179 (1997)
21. Mielke, A.: A gradient structure for reaction-diffusion systems and for energy-drift-diffusion systems. Nonlinearity **24**, 1329–1346 (2011)
22. Mielke, A.: Thermomechanical modeling of energy-reaction-diffusion systems, including bulk–interface interactions. Discr. Contin. Dyn. Syst. Ser. S, To appear (2012). WIAS preprint 1661 (2011)
23. Ollivier, Y.: Ricci curvature of markov chains on metric spaces. J. Funct. Anal. **256**(3), 810–864 (2009)
24. Onsager, L.: Reciprocal relations in irreversible processes, I + II. Phys. Rev. **37**, 405–426 (part II, **38**, 2265–2227) (1931)
25. Onsager, S., Machlup, L.: Fluctuations and irreversible processes. Phys. Rev. **91**(6), 1505–1512 (1953)
26. Otto, F.: The geometry of dissipative evolution equations: the porous medium equation. Commun. Partial Differ. Equ. **26**, 101–174 (2001)
27. Öttinger, H.C.: Beyond Equilibrium Thermodynamics. Wiley, Hoboken (2005)
28. Otto, F., Westdickenberg, M.: Eulerian calculus for the contraction in the Wasserstein distance. SIAM J. Math. Anal. **37**, 1227–1255 (2005)
29. Sturm, K.-T.: On the geometry of metric measure spaces. I + II. Acta Math. **196**(1), 1–131 + 133–177 (2006)
30. von Renesse, M.-K., Sturm, K.-T.: Transport inequalities, gradient estimates, entropy, and Ricci curvature. Commun. Pure Appl. Math. **58**(7), 923–940 (2005)