.

Weierstraß-Institut
für Angewandte Analysis und Stochastik
Leibniz-Institut im Forschungsverbund Berlin e. V.

Preprint ISSN 0946 – 8633

Geodesic convexity of the relative entropy in reversible Markov chains

Alexander Mielke[1, 2]

submitted: 6 October 2011

[1] Weierstraß Institut
Mohrenstr. 39
10117 Berlin
Germany
E-Mail: alexander.mielke@wias-berlin.de

[2] Institut für Mathematik
Humboldt-Universität zu Berlin
Rudower Chaussee 25
12489 Berlin
Germany

No. 1650
Berlin 2011

2010 *Mathematics Subject Classification.* 60J27, 53C21, 53C23, 82B35.

Key words and phrases. Markov chain, detailed balance, relative entropy, Onsager matrix, harmonic mean, geodesic convexity.

Partially sponsored by the European Research Council under ERC-2010-AdG 267802 *Analysis of Multiscale Systems Driven by Functionals*.

Edited by
Weierstraß-Institut für Angewandte Analysis und Stochastik (WIAS)
Leibniz-Institut im Forschungsverbund Berlin e. V.
Mohrenstraße 39
10117 Berlin
Germany

Fax: +49 30 2044975
E-Mail: `preprint@wias-berlin.de`
World Wide Web: `http://www.wias-berlin.de/`

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 more general nonlinear reaction systems.

1 Introduction

In this work we mainly 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 or for more general reaction systems 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 [JKO98, Ott01].

The present work was motivated by a generalization in [Mie11a] of the gradient structure for the Fokker-Planck equation to general reaction-diffusion systems, where the reactions satisfy a reversibility condition. The point is that the diffusion terms and the reaction terms can be written as a gradient system with 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 [Mie11a, Sect. 3.6] and [Mie11b].

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 found in [Mie11a, Sect. 3.1] as a special case of more general reaction-diffusion systems was found independently in [Maa11, CH*11]. It was also used in [AM*11] to show convergence from a Fokker-Planck equation to a simple Markov chain in a certain scaling limit.

To explain this structure, we consider a Markov chain defined on the discrete state space $\{1, ..., n\}$ via

$$\dot{u} = Pu \quad \text{with } P = (P_{ij})_{i,j=1,...n} \in \mathbb{R}^n.$$

Here $P_{ij} \geq 0$ is the rate for a particle moving from state j to i. For $j = i$ we have $P_{jj} = -\sum_{i \neq j} P_{ij} < 0$. Here

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

is the vector of the probabilities on the state space. The Markov chain is called *reversible* if there exists a positive steady state $w \in X_n$ (i.e. $w_i > 0$) such that

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

The gradient structure is given in terms of the *relative entropy*

$$E(u) = \sum_{i=1}^{n} u_i \log(u_i/w_i)$$

and the Onsager matrix

$$K(u) = \sum_{i<j} \pi_{ij}\, \Lambda\big(\tfrac{u_i}{w_i}, \tfrac{u_j}{w_j}\big)\, (e_i{-}e_j) \otimes (e_i{-}e_j) \in \mathbb{R}^{n\times n}_{\text{sym},\geq 0}.$$

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

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

see Proposition 3.1. Here K is the inverse of the Riemannian tensor $G(u) = K(u)^{-1}$ defined on $\mathbb{R}^n_{\text{av}} = \{\, v \in \mathbb{R}^n \mid v \cdot \overline{e} = 0 \,\}$.

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

$$\Lambda(a, b) = \frac{a - b}{\log a - \log b} \quad \text{for } a \neq b \quad \text{and} \quad \Lambda(a, a) = a, \tag{1.2}$$

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

$$\ell(\xi) \overset{\text{def}}{=} \max\{\, \Lambda(1, r) - \xi r \mid r > 0 \,\}. \tag{1.3}$$

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

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

The focus of this work is to provide conditions on the matrix P 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 geodesic curves $\gamma : [s_\text{a}, s_\text{b}] \to X$, i.e.

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

for all $\theta \in [0, 1]$ and $s_0, s_1 \in [s_\text{a}, s_\text{b}]$, where $s_\theta = (1{-}\theta)s_0 + \theta s_1$. We simply say that E is geodesically convex, if it is geodesically 0-convex. Of course, geodesic λ-convexity implies geodesic μ-convexity for all $\mu \leq \lambda$. Throughout this work the statement about geodesic λ-convexity never means that the corresponding λ is optimal, i.e. as large as possible.

The important point for our analysis is that the question of geodesic convexity can be analyzed in terms of the triple (X, E, K) without ever calculating the Riemannian tensor G or the Riemannian distance function d_K. This is discussed in Section 2, where we display two approaches. First, we use the implicit form of the geodesics γ and calculate the second derivative of $E \circ \gamma$. Second, we follow the approach in [OtW05, DaS08] where the length change of general curves during transport with the flow is characterized. The criterion for geodesic λ-convexity for a Markov chains $\dot u = Pu = -Au$ reduces to the estimate

$$\forall\, u \in X : \quad M(u) \geq \lambda K(u), \qquad \text{where } M(u) = \tfrac{1}{2}\big(K(u)A^{\mathsf{T}} + AK(u) - \mathrm{D}K(u)[Au]\big).$$

Starting in Section 3.2 we provide simple results on geodesic λ-convexity. In Section 4.1 we provide our first structural result stating that for all Markov chains there exists some λ such that E is geodesically λ-convex. 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 $P_{ij} > 0$ for all $i < j$. Using a different proof we are able to provide an explicit bound for λ in terms of all P_{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 $\{1, ..., n\}$ and set $P_{ij} = 1$ whenever i and j are connected by an edge and $P_{ij} = 0$ otherwise. Then, $w = \frac{1}{n}\overline{e}$ with $\overline{e} = (1, ..., 1)^{\mathsf{T}}$ is the steady state. Define $m = \max\{\, -P_{ii} \mid i = 1, ..., n\,\}$, which is the maximum possible number of neighbors of all vertices, then the relative entropy is geodesically λ_m-convex where λ_m depends only on m but not on n.

Section 5 is devoted to simple ordered Markov chains with nearest-neighbor transitions, as they occur in discretizations of a one-dimensional Fokker-Planck equation $\dot u = (u_x + u\widehat{V}_x)_x$. We are motivated by the geodesic $\widehat{\lambda}$-convexity of $\mathcal{E}(u) = \int_0^1 u \log(u/w)\,\mathrm{d}x$, where $w(x) = \mathrm{e}^{-\widehat{V}(x)}$ and $\widehat{\lambda} = \inf\{\, \widehat{V}''(x) \mid x \in [0,1]\,\}$, see [AGS05]. For the case $\widehat{\lambda} \geq 0$ we are able to construct discretizations in the form of linear Markov chains with gradient structure (X_n, E_n, K_n) in such a way that E_n is geodesically λ_n-convex with $\lambda_n = 2n^2(1 - \mathrm{e}^{-\widehat{\lambda}/(2n^2)}) \to \widehat{\lambda}$ for $n \to \infty$.

In Section 6 we show 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 [Mie11a, Sect. 3.1].

2 Geodesic convexity

We consider for $u \in X \subset \mathbb{R}^n$ the gradient flow

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

Here $E : X \to \mathbb{R}$ is an energy functional and $G(u) = G(u)^* > 0$ denotes the metric tensor at the point u. We consistently use the inverse $K(u) = G(u)^{-1}$. We call the symmetric and positive semidefinite matrix $K(u)$ 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. [Ons31, OnM53, Ött05, Mie11b].

We are interested in geodesic λ-convexity of the functional E with respect to the metric G. 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 equation for the geodesic curves in terms of K and then study the convexity of E along the curves. Second, we use the ideas from Otto-Westdickenberg [OtW05] and Daneri-Savaré [DaS08] on the evolution of length elements along the gradient flow in our simplified ODE case, where everything is smooth.

2.1 Geodesic curves 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. Throughout we assume that the space X is an open subset of $w + X_0$, where X_0 is a finite-dimensional linear space. We assume that for all $u \in X$ the matrix $K(u) : X_0 \to X_0$ is invertible, defining the metric tensor $G(u) : X_0 \to X_0$. Thus, geodesic curves $u = \gamma(s)$ in $X \subset w + X_0$ satisfy the classical Lagrange equation

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

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

$$\begin{aligned} \gamma' &= \frac{\partial}{\partial p} H(\gamma,p) = K(\gamma)p \\ p' &= -\frac{\partial}{\partial \gamma} H(\gamma,p) = -\frac{1}{2} p \cdot \mathrm{D}K(\gamma)[\Box]p, \end{aligned} \tag{2.1}$$

where $b = p \cdot \mathrm{D}K(\gamma)[\Box]p$ denotes the vector defined via $b \cdot \beta = p \cdot \mathrm{D}K(\gamma)[\beta]p$. The elimination of γ' via p is in fact the finite-dimensional counterpart to the famous approach by Benamou and Brenier [BeB00] for characterizing the Wasserstein distance via geodesics.

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 geodesics $u = \gamma(s)$. This property can be characterized by local expressions using the second derivative in the form

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

Using (2.1) and the relation

$$\frac{\mathrm{d}^2}{\mathrm{d}s^2} E(\gamma(s)) = \frac{\mathrm{d}}{\mathrm{d}s}\big(\mathrm{D}E(\gamma(s)) \cdot \gamma'(s)\big) = \gamma' \cdot \mathrm{D}^2 E(\gamma)\gamma' + \mathrm{D}E(\gamma) \cdot \frac{\mathrm{d}}{\mathrm{d}s}\big(K(\gamma)p\big)$$

we find the condition

$$\gamma' \cdot \mathrm{D}^2 E(\gamma)\gamma' + \mathrm{D}E(\gamma) \cdot \big(\mathrm{D}K(\gamma)[\gamma']p + K(\gamma)p'\big) \geq \lambda \gamma' \cdot G(\gamma)\gamma'.$$

Substituting $\gamma' = K(\gamma)p$ and p' according to (2.1) we obtain a formula in terms of γ and p only, namely

$$\begin{aligned} & p \cdot M(\gamma(s))p \geq \lambda p \cdot K(\gamma(s))p \quad \text{for all } s \text{ and } p \in X_0^*, \text{ where} \\ & p \cdot M(u)p \overset{\text{def}}{=} p \cdot K(u)\mathrm{D}^2 E(u)K(u)p + \mathrm{D}E(u) \cdot \mathrm{D}K(u)[K(u)p]p \\ & \qquad\qquad - \tfrac{1}{2} p \cdot \mathrm{D}K(u)[K(u)\mathrm{D}E(u)]p. \end{aligned} \tag{2.2}$$

Because the latter condition is quadratic in p, geodesic λ-convexity of the functional $E : X \to \mathbb{R}$ with respect to the metric $G = K^{-1}$ is equivalent to

$$\forall\, u \in X : \quad M(u) \geq \lambda K(u) \tag{2.3}$$

in the ordering sense of symmetric matrices, meaning that $M(u) - \lambda K(u)$ is positive semidefinite. In fact, the form of M can be simplified when using the (negative) vector field $u \mapsto f(u) = K(u)\mathrm{D}E(u)$, namely

$$M(u) = \tfrac{1}{2}\big(K(u)\mathrm{D}f(u)^{\mathsf{T}} + \mathrm{D}f(u)K(u) - \mathrm{D}K(u)[f(u)]\big). \tag{2.4}$$

The formula is especially simple for linear vector fields $f : u \mapsto Au$, namely

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

In the general case, M can be obtained as the derivative of K along the vector field $u \mapsto f(u) = K(u)\mathrm{D}E(u)$ in the following sense. Define $\Phi_t : X \to X$ to be the (local) flow of $\dot{u} = -f(u)$. Then, the transport of $\langle \eta, K(u)\xi\rangle$ along the flow is given by $\langle \mathrm{D}\Phi_t(u)^{-\mathsf{T}}\eta, K(\Phi_t(u))\mathrm{D}\Phi_t(u)^{-\mathsf{T}}\rangle$ and we find

$$2\langle \eta, M(u)\xi\rangle = \frac{\mathrm{d}}{\mathrm{d}t}\langle \mathrm{D}\Phi_t(u)^{-\mathsf{T}}\eta, K(\Phi_t(u))\mathrm{D}\Phi_t(u)^{-\mathsf{T}}\rangle\Big|_{t=0}. \tag{2.6}$$

If G and K are constant, the previous form (2.2) is more appropriate, and we recover the standard conditions

$$\langle K\mathrm{D}^2E(u)K\phi, \phi\rangle \geq \lambda\langle K\phi, \phi\rangle \quad \Longleftrightarrow \quad \langle \mathrm{D}^2E(u)v, v\rangle \geq \lambda\langle Gv, v\rangle.$$

Remark 2.1 (Bakry-Émery conditions) *Our definition has some similarities to the conditions of Bakry and Émery [BaÉ85, Bak94]. There, two symmetric bilinear mappings Γ_1 and Γ_2 are defined via*

$$\Gamma_1(f,g) = \frac{1}{2}\big(L(fg) - fLg - gLf\big) \quad \textit{and} \quad \Gamma_2(f,g) = \frac{1}{2}\big(L\Gamma_1(f,g) - \Gamma_1(Lf,g) - \Gamma_1(f,Lg)\big),$$

where L is the generator of a diffusion semigroup. The analogy of the pair (Γ_1, Γ_2) with the pair (K, M) is seen in (2.5) *and* (2.6)*. Moreover, the condition of λ-hypercontractivity reads*

$$2\Gamma_2(f,f) \succcurlyeq \lambda\Gamma_1(f,f) \text{ for all sufficiently smooth } f, \tag{2.7}$$

which matches (2.3)*. However, there is a major difference because in the Bakry-Émery theory the state space is assumed to be an ordered algebra with a product $\Gamma_0(f,g) = fg$. Moreover, the hypercontractivity condition* (2.7) *has to be fulfilled in the sense of this ordering $\succcurlyeq$, rather than in the sense of quadratic forms as in* (2.3)*. Thus, our condition seems to be weaker but still sufficient for geodesic λ-convexity. We refer to the formal calculation in Section 5.1 where the Fokker-Planck equation is treated with the counterpart of our condition* (2.3)*.*

2.2 The Otto-Westdickenberg characterization

The idea of Otto-Westdickenberg [OtW05] (see also Daneri-Savaré [DaS08]) uses the rate of change of infinitesimal line elements. For this one needs the first variation along the flow in the form

$$\dot{v} = -\mathrm{D}f(u)v = -\mathrm{D}K(u)[v]\mathrm{D}E(u) - K(u)\mathrm{D}^2E(u)v.$$

By $\eta = \langle G(u)v, v\rangle$ we denote the square of an infinitesimal line element. The statement of [OtW05] is that geodesic λ-convexity of E is equivalent to the fact that the solutions v satisfy

$$\dot{\eta} \leq -2\lambda\eta. \tag{2.8}$$

Inserting the evolution law for u and v into $\dot{\eta} = \langle \mathrm{D}G(u)[\dot{u}]v, v\rangle + 2\langle G(u)v, \dot{v}\rangle$ we find the necessary and sufficient condition

$$\langle \mathrm{D}G(u)[-K(u)\mathrm{D}E(u)]v, v\rangle + 2\langle G(u)v, -\mathrm{D}K(u)[v]\mathrm{D}E(u) - K(u)\mathrm{D}^2E(u)v\rangle \leq -2\lambda\langle G(u)v, v\rangle$$

for all u and v.

As the Onsager matrix K is given explicitly (in contrast to the Riemannian tensor G), the Benamou-Brenier substitution $p = G(u)v$ (cf. [BeB00]) is convenient.
Employing $\mathrm{D}G(u)[\xi] = -K(u)^{-1}\mathrm{D}K(u)[\xi]K(u)^{-1}$ we find the condition $p\cdot M(u)p \geq \lambda p\cdot K(u)p$, which is exactly (2.3).

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:

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

Thus, we may consider also the metric gradient system (X, E, d_K) in the sense of [AGS05]. 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

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

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 $(u_k^\tau)_{k\in\mathbb{N}}$, 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 [AGS05, 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 [vRS05, LoV09, BoS09, Maa11].

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

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

Here $P_{ij} \geq 0$ is the transition rate from j to i.

We assume that there exists a unique positive steady state $w \in X_n$ and define the relative entropy

$$E(u) = \sum_{i=1}^n u_i \log(u_i/w_i). \tag{3.2}$$

The crucial assumption is the *reversibility*, also called the *condition of detailed balance*, namely

$$P_{ij}w_j = P_{ji}w_i \ \text{ for } i, j = 1, \ldots, n. \tag{3.3}$$

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

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

$$G_1\dot{u} = -\mathrm{D}E_1(u), \quad G_2\dot{u} = -\mathrm{D}E_2(u), \quad \text{or } \dot{u} = -K_1\mathrm{D}E_1(u) = -K_2\mathrm{D}E_2(u)$$

with

$$E_1(u) = \tfrac{1}{2}\langle -W^{-1}Pu, u\rangle,\ K_1 = W = G_1^{-1},\ E_2(u) = \tfrac{1}{2}\langle W^{-1}u, u\rangle, \text{ and } K_2 = -PW.$$

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 [Mie11a, Sect. 3.1] in a more general nonlinear context of reaction systems and independently in [Maa11, CH*11].

Proposition 3.1 *The Markov chain* (3.1) *has the gradient structure* (X_n, E, K) *with the relative entropy* E *from* (3.2) *and the Onsager matrix*

$$K(u) = \sum_{j=2}^{n} \sum_{i=1}^{j-1} P_{ij} w_j \, \Lambda\Big(\frac{u_i}{w_i}\,,\ \frac{u_j}{w_j}\Big)\ (e_i{-}e_j) \otimes (e_i{-}e_j), \tag{3.4}$$

where $e_i \in \mathbb{R}^n$ *denotes the* i*-th unit vector, and* Λ *is defined in* (1.2) *and discussed in Appendix A.*

Proof: Equation (3.1) is easily obtained by using $\mathrm{D}E(u) = (\log(u_i/w_i){+}1)_{i=1,\dots,n}$. Multiplying this vector by $e_i{-}e_j$ we obtain the denominator of $\Lambda\big(\frac{u_i}{w_i}, \frac{u_j}{w_j}\big)$. Hence,

$$K(u)\mathrm{D}E(u) = \sum_{j=2}^{n} \sum_{i=1}^{j-1} P_{ij} w_j \Big(\frac{u_i}{w_i} - \frac{u_j}{w_j}\Big)\ (e_i{-}e_j) = -Pu,$$

where we used the detailed balance condition (3.3) in the last equality. ■

Note that E_2 and K_2 can be obtained via linearization of (E, K), namely

$$E_2(u) = \tfrac{1}{2}\mathrm{D}^2 E(w)[u,u] \text{ and } K_2 = K(w). \tag{3.5}$$

We also want to mention that there are many more possible gradient structures. Taking $\widetilde{E}(u) = \sum_{i=1}^{n} \phi_i(u_i/w_i)w_i$ for some strictly convex ϕ and

$$\widetilde{K}(u) = \sum_{j=2}^{n} \sum_{i=1}^{j-1} P_{ij} w_j \Phi\big(\tfrac{u_i}{w_i}, \tfrac{u_j}{w_j}\big)(e_i{-}e_j) \otimes (e_i{-}e_j) \quad \text{with } \Phi(a,b) = \frac{a-b}{\phi'(a) - \phi'(b)} > 0,$$

it is easy to generalize Proposition 3.1. It follows that the gradient system $\dot{u} = -\widetilde{K}(u)\mathrm{D}\widetilde{E}(u)$ equals the original reversible Markov chain $\dot{u} = Pu$.

The choices $\phi(\rho) = \frac{1}{2}\rho^2$ (giving $\Phi \equiv 1$) and $\phi(\rho) = \rho \log \rho$ (giving $\Phi = \Lambda$) lead to the above gradient system (X_n, E_2, K_2) and (X_n, E, K), respectively. The case $\phi(\rho) = \rho \log \rho$ is singled out by the fact that it is the only one giving the 1-homogeneity

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

Our main concern is the geodesic convexity of the relative entropy in Markov chains with respect to the metric defined via K, which depends on all the transition rates and the equilibrium state w. We first discuss a simple criterion for geodesic convexity.

3.2 A few Markov-chain examples

By definition we have $K(u)\overline{e}=0$, and for the matrix $M(u)$ defined in (2.4) this also holds as $A^\mathsf{T}\overline{e}=0$, i.e. we have

$$K(u)\overline{e}=M(u)\overline{e}=0 \quad \text{for all } u\in X_n, \quad \text{where } \overline{e}=(1,...,1)^\mathsf{T}. \tag{3.6}$$

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

Lemma 3.2 *Assume that K and M are symmetric and satisfy* (3.6) *as well as*

$$\forall\, i\neq j\ \forall\, u\in X_n: \quad M_{ij}(u)\leq \lambda K_{ij}(u), \tag{3.7}$$

then (2.3) *holds, i.e. E is geodesically λ-convex.*

Proof: The result follows simply by the fact, that all off-diagonal elements of $N(u):=M(u)-\lambda K(u)$ are nonpositive. Condition (3.6) implies that the diagonal elements satisfy

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

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

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

This proves the result. ∎

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 3.3 *A special case occurs if for the Markov chain all transition rates are the same, e.g. $P_{ij}=1$ 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 $A=-P=n\,I-\overline{e}\otimes\overline{e}$. Using $u\cdot\overline{e}=1$ and $K(u)\overline{e}=0$ we easily obtain

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

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

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

where the last identity follows by inserting the explicit relations (A.3) *for the derivatives. With* (A.1) *we obtain $2M_{ij}(u)\leq -(n{+}2)\Lambda_{ij}=(n{+}2)K_{ij}(u)$, and Lemma 3.2 yields that E is indeed geodesically $\frac{n+2}{2}$-convex. We expect that the result is not optimal for $n\geq 3$. However, for $u=w=\frac{1}{n}\overline{e}$ we can use* (3.5) *to obtain the upper bound n for the maximal λ. Hence, we have $\lambda_n^{\mathrm{opt}}\in[\frac{n+2}{2},n]$.*

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

Example 3.4 (Geodesic λ-convexity with $\lambda < 0$) *We consider a reversible Markov chain for $n = 3$ with equilibrium $w = (1, \varepsilon, 1)$ and*

$$A = -P = \begin{pmatrix} 2 & -1/\varepsilon & -1 \\ -1 & 2/\varepsilon & -1 \\ -1 & -1/\varepsilon & 2 \end{pmatrix} \quad \text{and} \quad K(u) = \begin{pmatrix} \Lambda_{12} + \Lambda_{13} & -\Lambda_{12} & -\Lambda_{13} \\ -\Lambda_{12} & \Lambda_{12} + \Lambda_{23} & -\Lambda_{23} \\ -\Lambda_{13} & -\Lambda_{23} & \Lambda_{13} + \Lambda_{23} \end{pmatrix}$$

where $\Lambda_{ij} = \Lambda(\rho_i, \rho_j)$ with $\rho_i = u_i/w_i$. We calculate $M(u) = \frac{1}{2}\big(K(u)A^T + AK(u) - \mathrm{D}K(u)[Au]\big)$ explicitly and find, for $u_* = (1, 1, 1)$ and $\varepsilon = 1/100$,

$$K(u_*) = \begin{pmatrix} 1 + \Lambda_* & -\Lambda_* & -1 \\ -\Lambda_* & 2\Lambda_* & -\Lambda_* \\ -1 & -\Lambda_* & 1 + \Lambda_* \end{pmatrix}, \quad M(u_*) \approx \begin{pmatrix} 778 & -2854 & 2076 \\ -2854 & 5708 & -2854 \\ 2076 & -2854 & 778 \end{pmatrix}$$

where $\Lambda_* = \Lambda(1, 100) \approx 21.498$. From this we see that the geodesic λ-convexity to be proved in Theorems 4.1 and 4.6 can only hold for $\lambda \leq \lambda_* \approx -55.23$, as $M(u_*) - \lambda K(u_*)$ is not positive semidefinite for $\lambda > \lambda_*$.

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

$$K(u) = \kappa\Lambda_{12}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \quad \text{and} \quad M(u) = m(u)\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$$

where $\kappa = \mu\theta(1{-}\theta)$, $\Lambda_{12} = \Lambda(\rho_1, \rho_2)$ with $\rho = (\rho_1, \rho_2) = (u_1/\theta, u_2/(1{-}\theta))$ and

$$m(u) = \mu\kappa\Lambda_{12} - \frac{\mu\kappa}{2}\big((1{-}\theta)\partial_{\rho_1}\Lambda(\rho_1, \rho_2) - \theta\partial_{\rho_2}\Lambda(\rho_1, \rho_2)\big)(\rho_1 - \rho_2).$$

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

$$\tfrac{\rho_1 - \rho_2}{\Lambda(\rho_1, \rho_2)}\big((1{-}\theta)\partial_{\rho_1}\Lambda(\rho_1, \rho_2) - \theta\partial_{\rho_2}\Lambda(\rho_1, \rho_2)\big) = 1 - \big(\tfrac{1-\theta}{\rho_1} + \tfrac{\theta}{\rho_2}\big)\Lambda(\rho_1, \rho_2) \leq c(\theta),$$

where we denote by $c(\theta)$ the maximum over all $\rho_j > 0$. Using (A.1) *we easily obtain $1 > c(\theta) \geq 1{-}2\sqrt{\theta(1{-}\theta)} \geq 0 = c(1/2)$. Thus, the two-dimensional Markov chain is geodesically λ-convex for $\lambda = \mu(1 - c(\theta)/2) \geq \mu/2$. Taking $\mu = 2$ and $\theta = 1/2$ we obtain $\lambda = 2$ as in the case $n = 2$ of Example 3.3.*

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

Above we have seen that any reversible Markov chain $\dot{u} = Pu$ 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 Section 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 (i.e. in the Benamou-Brenier form [BeB00]):

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

So far, X_n is the open set with $u_i > 0$ for all i. We want to show that d_K can be uniquely extended to the closure $\overline{X}_n = \mathrm{Prob}(\{1, ..., n\})$, i.e. there is a unique continuous extension of d_K to $\overline{X}_n \times \overline{X}_n$. 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 $\overline{X}_n \subset \mathbb{R}^n$.

We need to estimate the metric G from above near the boundary of X_n. Hence, K has to be estimated from below, and we set

$$\underline{\pi} \overset{\text{def}}{=} \min\{\, P_{ij} w_j \mid P_{ij} > 0 \,\} > 0 \quad \text{and} \quad \underline{\lambda}(u) \overset{\text{def}}{=} \min\{\, \Lambda(u_i/w_i, u_j/w_j) \mid P_{ij} > 0 \,\}.$$

Because the cotangent space to X_n is given by $\xi \in \mathbb{R}^n$ with $\xi \cdot \overline{e} = 0$, we have

$$|\xi|^2 = \tfrac{1}{2n^2} \textstyle\sum_{i=1}^n \sum_{j=1}^n (\xi_i - \xi_j)^2. \tag{3.8}$$

Recall that the Markov chain (X_n, E, K) is assumed to be irreducible. Hence, for every pair (i, j) there exists a chain $i = k_0, k_1, ..., k_{\overline{m}} = j$ with $\overline{m} \le n-1$ such that $P_{k_{m-1}k_m} \ge \underline{\pi}$ for $m = 1, ..., \overline{m}$. We now estimate

$$\begin{aligned}(\xi_i - \xi_j)^2 \;&\le\; \overline{m} \sum_{m=1}^{\overline{m}} (\xi_{k_{m-1}} - \xi_{k_m})^2 \\ &\le (n-1) \sum_{m=1}^{\overline{m}} \frac{P_{k_{m-1}k_m} w_{k_m} \Lambda(\rho_{k_{m-1}}, \rho_{k_m})}{\underline{\pi}\,\underline{\lambda}(u)} (\xi_{k_{m-1}} - \xi_{k_m})^2 \;\le\; \frac{n-1}{\underline{\pi}\,\underline{\lambda}(u)} \langle \xi, K(u)\xi \rangle.\end{aligned}$$

Inserting this into (3.8) gives $\langle \xi, K(u)\xi\rangle \ge 2n^2 \underline{\pi}\,\underline{\lambda}(u) |\xi|^2/(n-1)$. Using the lower estimate $\Lambda(a, b) \ge \sqrt{ab}$ from (A.1) and the abbreviations $\underline{u} = \min\{\, u_j \,|\, j = 1, ..., n \,\}$ and $\overline{w} = \max\{\, w_j \,|\, j = 1, ..., n \,\}$ we have $\underline{\lambda}(u) \ge \underline{u}/\overline{w}$. As a consequence, for a given Markov generator P there exists a constant $C > 0$ such that the Riemannian metric $G = K^{-1}$ satisfies

$$\langle v, G(u)v\rangle \le C|v|^2/\underline{u} \quad \text{for all } v \text{ with } v \cdot \overline{e} = 0 \text{ and all } u \in X_n. \tag{3.9}$$

Consider any sequences $(u_m)_{m\in\mathbb{N}}$ and $(\widetilde{u}_m)_{m\in\mathbb{N}}$ with

$$u_m \to u \quad \text{and} \quad \widetilde{u}_m \to \widetilde{u} \quad \text{in } \overline{X}_n,$$

where convergence is meant in the Euclidean sense. We will show that $(d_K(u_m, \widetilde{u}_m))_{m\in\mathbb{N}}$ is a Cauchy sequence such that $d_K(u, \widetilde{u})$ will be defined as $\lim_{m\to\infty} d_K(u_m, \widetilde{u}_m)$. Then, $d_K : \overline{X}_n \times \overline{X}_n \to \mathbb{R}$ is continuous.

To show convergence of $(d_K(u_m, \widetilde{u}_m))_{m\in\mathbb{N}}$ we use the triangle inequality

$$|d_K(u_m, \widetilde{u}_m) - d_K(u_l, \widetilde{u}_l)| \le d_K(\widetilde{u}_l, \widetilde{u}_m) + d_K(u_m, u_l).$$

Thus, it remains to show that $d_K(u_m, u_l)$ converges to 0. For $u \in X_n$ this is trivial, as $d_K(u_m, u_l) \le d_K(u_m, u) + d_K(u, u_l) \to 0$ as the Euclidean metric is equivalent to d_K on compact subsets of X_n. For $u \in \partial X_n$ choose $\varepsilon > 0$ and take $m, l \in \mathbb{N}$ so big, that $|u_m - u|,\ |u_l - u| \le \varepsilon$. Using the point $u^\varepsilon = (1-\varepsilon)u + \frac{\varepsilon}{n}\overline{e} \in X_n$ we can estimate

$$d_K(u_m, u_l) \le d_K(u_m, u^\varepsilon) + d_K(u_l, u^\varepsilon).$$

Defining the linear path $U(s) = (1-s)u_m + su^\varepsilon$ we have $\dot U = u^\varepsilon - u_m$ giving $|\dot U| \leq 2\varepsilon$. Moreover, for all j we have $U_j(s) \geq s\varepsilon/n$, and using (3.9) we obtain

$$d_K(u_m, u^\varepsilon) \leq \int_0^1 \langle \dot U, G(U(s))\dot U\rangle^{1/2} \mathrm{d}s \leq \int_0^1 \Big(\frac{nC}{\varepsilon s} 4\varepsilon^2\Big)^{1/2} \mathrm{d}s = 4(n\, C\, \varepsilon)^{1/2}.$$

As $\varepsilon > 0$ was arbitrary we obtain the desired result $d_K(u_m, u_l) \to 0$ and conclude that d_K has a well-defined continuous extension to $\overline{X}_n \times \overline{X}_n$. We note that points on the boundary ∂X_n may be connected with geodesics that lie inside X_n except for their endpoints.

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. Enven 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 derivatives of $\Lambda(\rho_i, \rho_j)$. The proof uses several special properties of the function 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 $\dot u = Pu$ be a reversible Markov chain with $P \in \mathbb{R}^{n\times n}$, i.e. it is reversible with respect to the strictly positive steady state $w \in \,]0,1[^n$. Then, there exists a $\lambda \in \mathbb{R}$ such that the entropy $E(u) = \sum_{i=1}^N u_i \log(u_i/w_i)$ is geodesically λ-convex with respect to the metric defined by K given in* (3.4).

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 P_{ij}, $i \neq j$, are strictly positive (see Section 4.2) we have to introduce some notation for the general result discussed here. We define the set $\mathfrak{E}$ of transition edges via

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

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

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

Thus, we can rewrite the matrices $A = -P$, $K(u)$ and $M(u)$ in the form

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

where we use the abbreviations

$$W = \operatorname{diag} w, \quad \pi_{ij} = P_{ij} w_j = \pi_{ji} \geq 0 \text{ for } i \neq j,$$
$$\mathbb{P} = \operatorname{diag}(\Pi_{ij})_{\overline{ij}\in\mathfrak{E}}, \quad \mathbb{L}(u) = \operatorname{diag}(\pi_{ij}\Lambda(u_i/w_i, u_j/w_j))_{\overline{ij}\in\mathfrak{E}}.$$

Reversibility of the Markov chain $\dot{u} = Pu$ means that $W^{-1} = \operatorname{diag}(1/w_i)_{i=1,\dots,n}$ exists and that $AW = (AW)^\mathsf{T} = WA^\mathsf{T}$.

For the future analysis it is more convenient to express the matrices $\mathbb{K}$, $\mathbb{L}$, and $\mathbb{M}$ in terms of the relative densities ρ_i from $u = W\rho$ via

$$\mathcal{K}(\rho) = \mathbb{K}(W^{-1}\rho), \quad \mathcal{L}(\rho) = \mathbb{L}(W^{-1}\rho), \quad \mathcal{M}(\rho) = \mathbb{M}(W^{-1}\rho),$$

which gives the final formulas

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

where $\mathbf{S} = SW^{-1}S^* \in \mathbb{R}^{N_\mathfrak{E}\times N_\mathfrak{E}}$. Note that in the last term we have $Au = AW\rho$ (for the formula (2.4)) and that 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 $\mathbb{P}$ 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

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

Thus, all nontrivial off-diagonal terms are associated with a pair of two edges having one common endpoint. The signs of $\mathbf{S}_{\overline{ij},\overline{kl}}$ 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 $\mathcal{N}_{\overline{ij}\,\overline{kl}}$ take the form

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

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

Lemma 4.2 *Consider a symmetric matrix $\Gamma \in \mathbb{R}^{\mu\times\mu}$. If*

$$\forall\, \alpha \in \{1,\dots,\mu\}: \quad \Gamma_{\alpha\alpha} = \sum_{\beta=1}^{\mu} \gamma^{\beta}_{\alpha\alpha} \text{ with } \gamma^{\beta}_{\alpha\alpha} \geq 0, \tag{4.3a}$$

$$\forall\, \alpha \neq \beta: \quad \Gamma^2_{\alpha\beta} \leq \gamma^{\beta}_{\alpha\alpha}\gamma^{\alpha}_{\beta\beta}, \tag{4.3b}$$

then, Γ is positive semidefinite.

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

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

This proves the desired result. ∎

To apply the above lemma, we need to find a proper 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 have to analyze the occurring terms in more detail. We first split them into three groups via

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

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

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

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 $n_{\overline{ij}} \in \{1,...,2n-2\}$ 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 $\mathfrak{N}_j \overset{\text{def}}{=} \{\, k \in \{1,...,n\} \mid \overline{jk} \in \mathfrak{E} \text{ or } \overline{kj} \in \mathfrak{E}\,\}$ and let $\widehat{n}_j = \#\mathfrak{N}_j$. Since $j \notin \mathfrak{N}_j$ and $i,k \in \mathfrak{N}_j$ we have $\widehat{n}_j \in \{2,...,n-1\}$.

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

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

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

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

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. The conditions in (4.3) are equivalent to

$$\mathbf{N}^{\overline{ij}\,\overline{jl}} \stackrel{\text{def}}{=} \begin{pmatrix} N^{\overline{jl}}_{\overline{ij}\,\overline{ij}} & \mathcal{N}_{\overline{ij}\,\overline{jl}} \\ \mathcal{N}_{\overline{ij}\,\overline{jl}} & N^{\overline{ij}}_{\overline{jl}\,\overline{jl}} \end{pmatrix} \geq 0$$

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

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

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

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

Thus, the validity of (4.3b) is shown 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 $\Lambda_{ij}\big(\frac{1}{\rho_i}+\frac{1}{\rho_j}\big) - 1 \geq \frac{1}{2}\Lambda_{ij}\big(\frac{1}{\rho_i}+\frac{1}{\rho_j}\big)$ (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 values for λ in the geodesic λ-convexity of the relative entropy E. At this stage we are content to establish the existence of one $\lambda \in \mathbb{R}$. Note that 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 $\pi_{il} = 0$. Note that $\overline{ij}\,,\ \overline{jl} \in \mathfrak{E}$ 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 an explicit bound. 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 $\{1,...,n\}$. Assume that $P_{ij} = 1$ if the vertices i and j are connected by an edge and $P_{ij} = 0$ else. Then, $w = \frac{1}{n}\overline{e}$ is the unique steady state, and $\widehat{n}_j = -P_{jj} = \sum_{i:i\neq J} P_{ij}$ gives the number of neighboring vertices for the vertex j. Our result gives a bound on the geodesic λ-convexity in terms of $m = \max\{\,\widehat{n}_j \mid j = 1,...,n\,\}$, 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 finite graph with n vertices and the special reversible Markov chain $\dot{u} = Pu \in \mathbb{R}^n$ with $P_{ij} = 1$ if i and j are connected and 0 else. Let $w = \frac{1}{n}\overline{e}^{\mathsf{T}}$ and*

$$m = \max\{\,-P_{jj} \mid j = 1,...,n\,\},$$

then the relative entropy E is geodesically λ-convex for some $\lambda = f(m)$.

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 $\nu_{\overline{ij}\,\overline{kl}} \neq 0$. With Proposition 4.3 the lower estimate (4.7) yields

$$\lambda = f(m) \overset{\text{def}}{=} \frac{1}{2}\Big(4 - m + 2\,\widehat{g}(m{-}1)\Big),$$

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 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-convex with $\lambda = f(2d)$, independently of n.

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

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 3.2 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 (1.3).

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

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

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.7c).

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

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

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.7a) 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

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

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 $\xi = 1/2$:

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

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 Section 4.1, namely

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

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

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

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

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

For applying condition (3.7) 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 the form that highlights the dependencies on (ρ_i, ρ_j) and on all the others ρ_l, namely

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

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

$$\begin{aligned}\overline{M}_{ij} &\leq \overline{\mu}_{ij}\pi_{ij}\Lambda_{ij} \quad \text{with } \overline{\mu}_{ij} = -\min\{A_{ii}, A_{jj}\} - P_{ij} - P_{ji} - \min\{P_{ij}, P_{ji}\},\\ \widetilde{M}_{ijl} &\leq \widetilde{\mu}_{ijl}\pi_{ij}\Lambda_{ij} \quad \text{with } \widetilde{\mu}_{ijl} = \pi_{ij}\min\{P_{li}, P_{lj}\}\,\widetilde{g}(\beta_{ijl})\\ &\qquad \text{and } \beta_{ijl} = \max\{0, \pi_{li}(P_{jl}-P_{ji}), \pi_{lj}(P_{il}-P_{ij})\}/(\pi_{ij}\min\{P_{li}, P_{lj}\}).\end{aligned} \tag{4.9}$$

Thus, together with criterion (3.7) we can summarize and obtain the following result.

Theorem 4.6 *Assume that $\dot{u} = Pu$ is a reversible Markov chain where all transition rates are positive, i.e. $P_{ij} > 0$ for all $i < j$. Consider the gradient structure (X_n, E, K) given in Proposition 3.1, then E is geodesically λ-convex for*

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

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 $i \neq j$ 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 $w \in X_n$ with $w \cdot \overline{e} = 1$ and let*

$$P = \kappa w \otimes \overline{e} - \kappa I, \quad \textit{then } P^{\mathsf{T}}\overline{e} = 0 = Pw \textit{ and } P_{ij}w_j = \kappa w_i w_j.$$

Hence, w is the steady state of the reversible Markov chain. Applying the above theorem we see that $\widetilde{\mu}_{ijl} = 0$ as $P_{ij} = P_{il}$ by construction. Since $\overline{\mu}_{ij} = -\kappa - 2\kappa\min\{w_i, w_j\}$ we conclude geodesic λ-convexity for E in (X_n, E, K) with $\lambda = \kappa/2 + \kappa\min\{\,w_i \mid i = 1, ..., n\,\}$.

Taking $w = \frac{1}{n}\overline{e}$ and $\kappa = n$ we recover the result of Example 3.3.

5 Discretization of a 1D Fokker-Planck equation

In this section we discuss a special Markov chain, which occurs as a discretization of a one-dimensional Fokker-Planck equation. The points $1, ..., n$ are aligned and transitions only occur to the nearest neighbor, i.e. P is a tridiagonal matrix. Before investigating this situation we show how for the associated diffusion equation the geodesic convexity can be established. The proof involves several integrations by part that are a guideline for the discrete setting.

5.1 Geodesic convexity for the Fokker-Planck equation

We consider the Fokker-Planck equation $\dot{u} = \mathrm{div}(\nabla u + u\nabla\widehat{V})$ on $\Omega = \mathbb{R}^d$. Here we only give a formal argument motivating the geodesic $\widehat{\lambda}$-convexity of the relative entropy under the assumption that the potential $\widehat{V}$ is $\widehat{\lambda}$-convex, i.e. in the smooth case we have

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

First, we apply the approach of Section 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 is given via

$$\dot u = -\mathcal{K}(u)\mathrm{D}\mathcal{E}(u) \quad \text{with } \mathcal{E}(u) = \int_\Omega u \log u + \widehat{V} u \,\mathrm{d}x \text{ and } \mathcal{K}(u)\xi = -\operatorname{div}\big(u\nabla\xi\big), \tag{5.1}$$

see [JKO98, Ott01]. To calculate the corresponding quadratic form $\mathcal{M}$ we use that the vector field $\mathcal{F}(u)$ is linear with

$$\mathcal{F}(u) = \mathcal{A}u = -\Delta u - \operatorname{div}(u\nabla\widehat{V}) \quad \text{and} \quad \mathrm{D}\mathcal{F}(u)^*\phi = \mathcal{A}^*\phi = -\Delta\phi + \nabla\phi\cdot\nabla\widehat{V}.$$

Hence, using (2.4) and $\Delta(\frac12|\nabla\phi|^2) = |\mathrm{D}^2\phi|^2 + \nabla\phi\cdot\nabla(\Delta\phi)$ we obtain

$$\begin{aligned}
&\langle\mathcal{M}(u)\phi,\phi\rangle = \langle \mathrm{D}\mathcal{F}(u)^*\phi, \mathcal{K}(u)\phi\rangle - \tfrac12\langle\phi, \mathrm{D}\mathcal{K}(u)[\mathcal{F}(u)]\phi\rangle\\
&= \int_\Omega \big(-\Delta\phi+\nabla\phi\cdot\nabla\widehat{V}\big)\big(-\operatorname{div}(u\nabla\phi)\big) - \tfrac12\big(-\Delta u - \operatorname{div}(u\nabla\widehat{V})\big)|\nabla\phi|^2\,\mathrm{d}x\\
&= \int_\Omega u\Big(\nabla\big(-\Delta\phi+\nabla\phi\cdot\nabla\widehat{V}\big)\cdot\nabla\phi + \Delta(\tfrac12|\nabla\phi|^2) - \nabla\widehat{V}\cdot\nabla(\tfrac12|\nabla\phi|^2)\Big)\,\mathrm{d}x\\
&= \int_\Omega u\Big(|\mathrm{D}^2\phi|^2 + \nabla\phi\cdot\mathrm{D}^2\widehat{V}\nabla\phi\Big)\,\mathrm{d}x \geq \int_\Omega u\,\widehat{\lambda}|\nabla\phi|^2\,\mathrm{d}x = \widehat{\lambda}\langle\phi,\mathcal{K}(u)\phi\rangle.
\end{aligned}$$

Second, we emphasize that the above calculation can be turned into a full proof of geodesic $\widehat{\lambda}$-convexity following the analysis developed in [DaS08]. However, there are several other approaches to geodesic $\widehat{\lambda}$-convexity, also called displacement convexity in this context, see [McC97, AGS05]. We also refer to Remark 2.1 to compare with the stronger pointwise Bakry-Émery condition (2.7).

In the next section we will use the above formal calculation as a guide line for arranging terms. In particular, the fact that $\langle\mathcal{M}(u)\phi,\phi\rangle$ and $\langle\mathcal{K}(u)\phi,\phi\rangle$ depend on $\nabla\phi$ only, will be mirrored in the fact that in the discrete Markov chain the corresponding forms depend on the differences $u_{j+1}-u_j$ only.

5.2 Uniform geodesic λ-convexity for the discretization

We now return to the one-dimensional case with constant mobility. Our aim is to find a spatial discretization of the corresponding Fokker-Planck equation that keeps the geodesic convexity-properties. In particular, the discretization will be geodesically λ_n-convex with $\lambda_n \to \widehat{\lambda}$, and $\widehat{\lambda}$ is the best value for geodesic λ-convexity of the Fokker-Planck equation

$$\dot u = \big(u_x + u\widehat{V}_x(x)\big)_x, \quad u_x(t,0) = u_x(t,1) = 0, \quad \int_0^1 u(t,x)\,\mathrm{d}x = 1. \tag{5.2}$$

This equation is given by the gradient system $(\mathcal{X}_n, \mathcal{E}, \mathcal{K})$ with

$$\mathcal{E}(u) = \int_0^1 u\log(u/\widehat{w})\,\mathrm{d}x \quad \text{and} \quad \mathcal{K}(u) = -\operatorname{div}(u\nabla\Box), \tag{5.3}$$

where $\widehat{w}(x) = c\mathrm{e}^{-\widehat{V}(x)}$ with $1/c = \int_0^1 \mathrm{e}^{-\widehat{V}(x)}\,\mathrm{d}x$. It is geodesically $\widehat{\lambda}$-convex where $\widehat{\lambda} = \inf\{\,\widehat{V}''(x) \mid x \in\,]0,1[\,\}$.

Our discretization should be again a gradient system with a discrete relative entropy and an Onsager matrix K such the gradient system is again a Markov process. On the state space $X_n \subset \mathbb{R}^n$ we define the energy functional E and the Onsager matrix K as

$$E(u) = \sum_{i=1}^{n} u_i \log(u_i/w_i) \quad \text{and} \quad K(u) = S^*\mathbb{L}(u)S, \tag{5.4a}$$

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

and $S^* \in \mathbb{R}^{n\times(n-1)}$ is the transposed of S. The diagonal matrix $\mathbb{L} = \operatorname{diag}(\mathfrak{L}) \in \mathbb{R}^{(n-1)\times(n-1)}$ is given via

$$\mathfrak{L}(u) = (\mathfrak{L}_i(u))_{i=1,\dots,n-1} \quad \text{with } \mathfrak{L}_i(u) = \kappa_i \Lambda(\tfrac{u_i}{w_i}, \tfrac{u_{i+1}}{w_{i+1}}), \tag{5.4c}$$

where the vector $\kappa = (\kappa_i)_{i=1,\dots,n-1}$ is still to be chosen and $\Lambda : \,]0,\infty[^2 \to \,]0,\infty[$ is given in (1.2). As in Proposition 3.1 the gradient system (X_n, E, K) leads to

$$\dot u = -S^*\mathbb{L}(u)S\mathrm{D}E(u) = -S^* \operatorname{diag}(\kappa) S (u_i/w_i)_{i=1,\dots,n} = -Au,$$

where $A \in \mathbb{R}^{n\times n}$ has the form

$$A = S^* \operatorname{diag}(\kappa) S \operatorname{diag}(w)^{-1} = \begin{pmatrix} \frac{\kappa_1}{w_1} & -\frac{\kappa_1}{w_2} & 0 & \cdots & 0 \\ -\frac{\kappa_1}{w_1} & \frac{\kappa_1+\kappa_2}{w_2} & -\frac{\kappa_2}{w_3} & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \frac{\kappa_{n-2}+\kappa_{n-1}}{w_{n-1}} & -\frac{\kappa_{n-1}}{w_n} \\ 0 & \cdots & 0 & -\frac{\kappa_{n-1}}{w_{n-1}} & \frac{\kappa_{n-1}}{w_n} \end{pmatrix} \in \mathbb{R}^{n\times n}.$$

Clearly, we have the equilibrium condition $Aw = 0$ and the conservation of total mass $A^\mathsf{T}\overline{e}^\mathsf{T} = 0$. We recall that in the continuous case $\mathcal{K}(u)$ is independent of the steady state w, which is not the case for the discrete system. Nevertheless, we want to choose $K(u)$ such that it converges to $\mathcal{K}$ in the discrete-to-continuous limit. For this, one has to choose κ_i suitably. The 1-homogeneity (A.5) of Λ suggests to choose κ_i such that it is 1-homogeneous in w. Then, $K(u)$ will be 0-homogenous in w. Since κ_i corresponds to the transfer between the nodes i and $i{+}1$, it is natural to choose the symmetric variant

$$\kappa_i = \sqrt{w_i\, w_{i+1}}. \tag{5.5}$$

Theorem 5.1 *If* (5.5) *is satisfied and if for some $\beta \geq 0$ the concavity condition*

$$w_i \geq \mathrm{e}^{\beta}\sqrt{w_{i-1}w_{i+1}} \quad \text{for } i = 2, \dots, n-1, \tag{5.6}$$

holds, then the gradient system (X_n, E, K) given in (5.4) *and leading to the Markov chain $\dot u = -Au$ is geodesically λ-convex with $\lambda = 2(1{-}\mathrm{e}^{-\beta}) \geq 0$.*

If the special choice (5.5) holds it is useful to introduce the quotients

$$q_i = \sqrt{w_{i+1}/w_i} \quad \text{for } i = 1, \dots n-1.$$

We observe that the concavity condition (5.6) is equivalent to

$$q_i \geq \mathrm{e}^{\beta}\, q_{i+1} \quad \text{for } i = 1, \dots, n{-}2. \tag{5.7}$$

Relating w_j to a potential V via $V_j = \gamma - \log w_j$, we also see that (5.6) corresponds to a classical convexity for V, namely

$$V_i \leq \tfrac{1}{2}(V_{i-1} + V_{i+1}) - \beta \quad \text{for } i = 2, \dots, n-1. \tag{5.8}$$

Moreover, $K(u)$ and A can be expressed in terms of q_j only. For $K(u)$ simply note that $\mathfrak{L}_i(u) = \Lambda(q_i u_i, \frac{1}{q_i} u_{i+1})$. For A we find

$$A = \begin{pmatrix} 0{+}q_1 & -\frac{1}{q_1} & 0 & \cdots & 0 \\ -q_1 & \frac{1}{q_1}{+}q_2 & -\frac{1}{q_2} & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \frac{1}{q_{n-2}}{+}q_{n-1} & -\frac{1}{q_{n-1}} \\ 0 & \cdots & 0 & -q_{n-1} & \frac{1}{q_{n-1}}{+}0 \end{pmatrix} \in \mathbb{R}^{n\times n}. \tag{5.9}$$

We also see that the case $w = \frac{1}{n}(1, \dots, 1)^{\mathsf{T}}$ leads to the simple matrix A having $+1$ and $+2$ on the diagonal and -1 on the two secondary diagonals, which is the standard discretization scheme of the one-dimensional diffusion equation $u_t = u_{xx}$. In this case our ODE $\dot{u} = -Au$ satisfies (5.6) with $\beta = 0$, and we conclude geodesic convexity.

In the following proof we do the first part of the calculations for general κ_j for possible future generalizations. The proof relies on lengthy but elementary calculations and uses specific properties of the function Λ derived in Appendix A. The final proof of $M(u) \geq \lambda K(u)$ then uses the special choice (5.5) and the concavity (5.6).

To see the relation between the discretization and the Fokker-Planck equation (5.2) we consider the smooth and $\widehat{\lambda}$-convex potential $\widehat{V} : [0,1] \to \mathbb{R}$, i.e. $\widehat{V}''(x) \geq \widehat{\lambda} \geq 0$. For the discretization we let $V_i = \widehat{V}(i/n)$, which implies that (5.8) holds with $\beta_n = \widehat{\lambda}/(2n^2)$. Moreover, for a proper discretization of (5.2) we need to rescale the time or set $A_n = n^2 A$ with A from (5.9). As the matrix M_n depends quadratically on A_n whereas K_n depends linearly on A_n, we find that the properly scaled systems (X_n, E_n, K_n) are geodesically λ_n-convex with

$$\lambda_n = 2\, n^2 \left(1 - \mathrm{e}^{-\widehat{\lambda}/(2n^2)}\right).$$

Obviously, we have $\lambda_n \to \widehat{\lambda}$ for $n \to \infty$, which shows that the discretization is such that we have an asymptotically sharp lower bound for the geodesic λ-convexity for $n \to \infty$.

Remark 5.2 *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 [Gli08, GlG09] obviously lead to reversible Markov chains, but their geodesic λ-convexity needs to be investigated.*

5.3 Proof of Theorem 5.1

Inserting the specific forms of $K = S^*\mathbb{L}(u)S$ and $A = S^* \operatorname{diag} \kappa S \operatorname{diag}(w)^{-1}$ into the definition of M we arrive at

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

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

$$\mathcal{N}(u) \geq 2\lambda \mathbb{L}(u) \text{ for all } u \in X_n. \tag{5.10}$$

To shorten the following calculations we introduce the following abbreviations:

$$\rho_i = u_i/w_i, \quad \Lambda_i = \Lambda(\rho_i, \rho_{i+1}), \Lambda_{i,1} = \partial_{\rho_i}\Lambda(\rho_i, \rho_{i+1}), \text{ and } \Lambda_{i,2} = \partial_{\rho_{i+1}}\Lambda(\rho_i, \rho_{i+1})$$

Obviously, $N(u) \in \mathbb{R}^{(n-1)\times(n-1)}$ is the symmetric tridiagonal matrix given by

$$\mathcal{N}(u) = \begin{pmatrix} a_1 & b_1 & 0 & \cdots & 0 \\ b_1 & a_2 & b_2 & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & b_{n-3} & a_{n-1} & b_{n-2} \\ 0 & \cdots & 0 & b_{n-2} & a_{n-1} \end{pmatrix} \in \mathbb{R}^{(n-1)\times(n-1)}.$$

with

$$\begin{aligned}
a_1 &= 2\kappa_1 \mathfrak{L}_1 \big(\tfrac{1}{w_1} + \tfrac{1}{w_2}\big) - \tfrac{\kappa_1}{w_1}\Lambda_{1,1}\kappa_1(\rho_1 - \rho_2) - \tfrac{\kappa_1}{w_2}\Lambda_{1,2}\big(\kappa_1(\rho_2 - \rho_1) + \kappa_2(\rho_2 - \rho_3)\big), \\
a_i &= 2\kappa_i \mathfrak{L}_i \big(\tfrac{1}{w_i} + \tfrac{1}{w_{i+1}}\big) - \tfrac{\kappa_i}{w_i}\Lambda_{i,1}\big(\kappa_{i-1}(\rho_i - \rho_{i-1}) + \kappa_i(\rho_i - \rho_{i+1})\big) \\
&\quad - \tfrac{\kappa_i}{w_{i+1}}\Lambda_{i,2}\big(\kappa_i(\rho_{i+1} - \rho_i) + \kappa_{i+1}(\rho_{i+1} - \rho_{i+2})\big) \qquad \text{for } i = 2, \ldots, n-2; \\
a_{n-1} &= 2\kappa_{n-1}\mathfrak{L}_{n-1}\big(\tfrac{1}{w_{n-1}} + \tfrac{1}{w_n}\big) - \tfrac{\kappa_{n-1}}{w_{n-1}}\Lambda_{n-1,1}\big(\kappa_{n-2}(\rho_{n-1} - \rho_{n-2}) + \kappa_{n-1}(\rho_{n-1} - \rho_n)\big) \\
&\quad - \tfrac{\kappa_{n-1}}{w_n}\Lambda_{n-1,2}\kappa_{n-1}(\rho_n - \rho_{n-1}) \\
b_i &= -\tfrac{1}{w_{i+1}}(\mathfrak{L}_i\kappa_{i+1} + \mathfrak{L}_{i+1}\kappa_i) = -\tfrac{\kappa_i\kappa_{i+1}}{w_{i+1}}\big(\Lambda_i + \Lambda_{i+1}\big) \leq 0.
\end{aligned}$$

The desired positive semi-definiteness of $\mathcal{N}(u) - 2\lambda\mathbb{L}(u)$ (cf. (5.10)) will follow from diagonal dominance, which reads in this case

$$A_1 = a_1 + b_1 - 2\lambda\mathfrak{L}_1 \quad \geq 0, \tag{5.11a}$$
$$A_i := a_i + b_{i-1} + b_i - 2\lambda\mathfrak{L}_i \quad \geq 0 \text{ for } i = 2, ..., n-2, \tag{5.11b}$$
$$A_{n-1} := a_{n-1} + b_{n-2} - 2\lambda\mathfrak{L}_{n-1} \quad \geq 0. \tag{5.11c}$$

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

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

To establish the estimates (5.11) we first treat the case $i = 2, \dots, n-2$. Inspecting the formula for A_i we find

$$A_i = \kappa_i\Big(\widetilde{A}_i(\rho_i,\rho_{i+1}) - \tfrac{\kappa_{i-1}}{w_i}\big(\Lambda(\rho_{i-1},\rho_i){-}\Lambda_{i,1}\rho_{i-1}\big) - \tfrac{\kappa_{i+1}}{w_{i+1}}\big(\Lambda(\rho_{i+1},\rho_{i+2}){-}\Lambda_{i,2}\rho_{i+2}\big)\Big)$$
$$\text{with } \widetilde{A}_i(\rho_i,\rho_{i+1}) = \Lambda_i\Big(\tfrac{2\kappa_i}{w_i}{+}\tfrac{2\kappa_i}{w_{i+1}} - \tfrac{\kappa_{i-1}}{w_i} - \tfrac{\kappa_{i+1}}{w_{i+1}} - 2\lambda\Big)$$
$$\qquad - \tfrac{\Lambda_{i,1}}{w_i}\big((\kappa_{i-1}{+}\kappa_i)\rho_i - \kappa_i\rho_{i+1}\big) - \tfrac{\Lambda_{i,2}}{w_{i+1}}\big({-}\kappa_i\rho_i + (\kappa_i{+}\kappa_{i+1})\rho_{i+1}\big).$$

We note that ρ_{i-1} and ρ_{i+2} occur only twice, such that minimization with respect to ρ_{i-1} and ρ_{i+2} is easily possible. By employing the crucial estimate (A.6) for ρ_{i-1} and ρ_{i+2} separately, we find

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

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

$$\Gamma_i = \Lambda_i\Big(\frac{\kappa_i}{w_i}{+}\frac{\kappa_i}{w_{i+1}} - \frac{\kappa_{i-1}}{w_i}{-}\frac{\kappa_{i+1}}{w_{i+1}} - 2\lambda + \Sigma_i)$$
$$\text{with } \Sigma_i := \Lambda(\rho_i,\rho_{i+1})\Big(\big(\frac{\kappa_i}{w_i}{-}\frac{\kappa_{i+1}}{w_{i+1}}\big)\frac{1}{\rho_i} + \big(\frac{\kappa_i}{w_{i+1}}{-}\frac{\kappa_{i-1}}{w_i}\big)\frac{1}{\rho_{i+1}}\Big).$$

To show $\Gamma_i \geq 0$, we need to find a lower bound on Σ_i. Since $\Lambda(a,b)/a$ is not bounded, lower bound exists if and only if

$$\frac{\kappa_i}{w_i} - \frac{\kappa_{i+1}}{w_{i+1}} \geq 0 \quad \text{and} \quad \frac{\kappa_i}{w_{i+1}} - \frac{\kappa_{i-1}}{w_i} \geq 0. \tag{5.12}$$

Under these conditions we can use (A.1) to find

$$\Sigma_i \geq \sqrt{\rho_i\rho_{i+1}}\Big(\big(\tfrac{\kappa_i}{w_i}{-}\tfrac{\kappa_{i+1}}{w_{i+1}}\big)\tfrac{1}{\rho_i} + \big(\tfrac{\kappa_i}{w_{i+1}}{-}\tfrac{\kappa_{i-1}}{w_i}\big)\tfrac{1}{\rho_{i+1}}\Big) \geq 2\Big(\big(\tfrac{\kappa_i}{w_i}{-}\tfrac{\kappa_{i+1}}{w_{i+1}}\big)\big(\tfrac{\kappa_i}{w_{i+1}}{-}\tfrac{\kappa_{i-1}}{w_i}\big)\Big)^{1/2}.$$

Putting everything together we see that $\Gamma_i \geq 0$, and hence $A_i \geq 0$ follows from

$$\lambda \leq \gamma_i := G\Big(\frac{\kappa_i}{w_i}{-}\frac{\kappa_{i+1}}{w_{i+1}}\,,\ \frac{\kappa_i}{w_{i+1}}{-}\frac{\kappa_{i-1}}{w_i}\Big) \quad \text{for } i = 2, ..., n{-}2,$$

where the function G is defined as

$$G : \mathbb{R}^2 \to [0,\infty];\ (a,b) \mapsto \begin{cases} \frac{1}{2}(a{+}b) + \sqrt{ab} & \text{for } a,b \geq 0, \\ \infty & \text{otherwise.} \end{cases}$$

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

$$\lambda \leq \gamma_1 := G\Big(\tfrac{\kappa_1}{w_1}{-}\tfrac{\kappa_2}{w_2}\,,\ \tfrac{\kappa_1}{w_2}\Big) \text{ and } \lambda \leq \gamma_{n-1} := G\Big(\tfrac{\kappa_{n-1}}{w_{n-1}}\,,\ \tfrac{\kappa_{n-1}}{w_n}{-}\tfrac{\kappa_{n-2}}{w_{n-1}}\Big).$$

All these estimates hold for general coefficients κ_i and w_i. We now reduce to the special choice (5.5) for κ_j and use $q_i = \sqrt{w_{i+1}/w_i}$ satisfying (5.7). Since the function G is monotone increasing in both arguments and 1-homogeneous, we can estimate γ_i from below as follows

$$\gamma_i = G(q_i - q_{i+1}, \tfrac{1}{q_i} - \tfrac{1}{q_{i-1}}) \geq G(q_i - \mathrm{e}^{-\beta} q_i, \tfrac{1}{q_i} - \mathrm{e}^{-\beta} \tfrac{1}{q_{i-1}}) = (1-\mathrm{e}^{-\beta}) G(q_i, \tfrac{1}{q_i}) \geq 2(1-\mathrm{e}^{-\beta}),$$

where we set $\frac{1}{q_0} = 0 = q_n$ to cover the cases $i = 1$ and $i = n-1$.

Since now (5.11) is established with $\lambda = 2(1-\mathrm{e}^{-\beta})$, we have $\mathcal{N}(u) \geq 2\lambda \mathbb{L}(u)$ for all $u \in X_n$. This implies $M(u) \geq \lambda K(u)$, and our desired result on geodesic λ-convexity follows. Thus, Theorem 5.1 is proved.

Remark 5.3 *The case $\kappa_i = \frac{1}{2}(w_i + w_{i+1})$ can also be handled. Using $q_i^2 = w_{i+1}/w_i$ we have the relation*

$$\lambda \leq \min\{ \tfrac{1}{2} G(q_{i+1}^2 - q_i^2, \tfrac{1}{q_i^2} - \tfrac{1}{q_{i-1}^2}) \mid i = 1, ..., n-1 \},$$

where $\frac{1}{q_0^2} = 0 = q_n^2$. We obtain geodesic λ-convexity with $\lambda = (1 - \mathrm{e}^{-2\beta})$, which is smaller than $2(1-\mathrm{e}^{-\beta})$ obtained above.

6 Nonlinear reaction systems

We give here some preliminary results for geodesic λ-convexity for reversible reaction systems of mass-action type. We refer to [GlG09, Mie11a, Mie11b] and the references therein for more details and motivation. Consider again a vector $u \in \,]0, \infty[^n$ of densities and a polynomial reaction system with R reactions:

$$\dot{u} = -\sum_{r=1}^{R} k_r(u) \Big(\frac{u^{\alpha^r}}{w^{\alpha_r}} - \frac{u^{\beta^r}}{w^{\beta^r}} \Big) \big(\alpha^r - \beta^r \big), \quad \text{where } u^{\alpha^r} = \Pi_{i=1}^n u_i^{\alpha_i^r}. \tag{6.1}$$

Here $w \in \,]0, \infty[^n$ is a fixed reference density, which is obviously a steady state and satisfies the detailed balance condition (reversibility), since for $u = w$ all R reactions are balanced simultaneously. The index r is the reaction number, $k_r(u) \geq 0$ is the reaction coefficient (normalized with respect to w), and the vectors $\alpha^r, \beta^r \in [0, \infty[^r$ are called the stoichiometric vectors for the forward and backward reaction. Usually the entries are assumed to be nonnegative integers, but this is not necessary here. A typical example is

$$2\mathrm{CO} + \mathrm{O}_2 \rightleftharpoons 2\mathrm{CO}_2 \quad \text{giving} \quad \dot{u} = -k(u) \Big(\frac{u_1^2 u_2}{w_1^2 w_2} - \frac{u_3^2}{w_3^2} \Big) \begin{pmatrix} 2 \\ 1 \\ -2 \end{pmatrix}, \tag{6.2}$$

where $u = (u_1, u_2, u_3) = (u_{\mathrm{CO}}, u_{\mathrm{O}_2}, u_{\mathrm{CO}_2})$, $\alpha = (2, 1, 0)$, and $\beta = (0, 0, 2)$.

It was shown in [Mie11a] that (6.1) is generated by the gradient system $(]0, \infty[^n, E, K)$ with

$$E(u) = \sum_{i=1}^{n} u_i \big(\log(u_i/w_i) - 1 \big) \quad \text{and} \quad K(u) = \sum_{r=1}^{R} k_r(u) \Lambda\big(\tfrac{u^{\alpha^r}}{w^{\alpha_r}}, \tfrac{u^{\beta^r}}{w^{\beta^r}} \big) \big(\alpha^r - \beta^r \big) \otimes \big(\alpha^r - \beta^r \big).$$

In fact, the result follows as in Proposition 3.1 by using the definition of Λ in (1.2) and

$$\mathrm{D}E(u) = (\log(u_i/w_i))_i \quad \text{and} \quad \mathrm{D}E(u)\cdot\big(\alpha^r-\beta^r\big) = \log\big(\tfrac{u^{\alpha^r}}{w^{\alpha^r}}\big) - \log\big(\tfrac{u^{\beta^r}}{w^{\beta^r}}\big).$$

Of course, the above Markov chains are special cases where all the vectors α^r and β^r are unit vectors, which corresponds to simple exchange reactions.

We want to study a few simple cases and discuss the possibility of geodesic λ-convexity. The fundamental difference to the case of the Markov chains is that the vector field $f(u)=K(u)\mathrm{D}E(u)$ is no longer linear and that the matrices $K(u)$ and $M(u)$ have no homogeneity properties any more.

For $R=1$ we drop the reaction number r and write $\gamma=\alpha-\beta$. Moreover, we write $\rho=(u_i/w_i)_i$. Then,

$$f(u)=\phi(u)\gamma \text{ with } \phi(u)=k(u)(\rho^\alpha-\rho^\beta), \quad K(u)=\kappa(u)\gamma\otimes\gamma \text{ with } \kappa(u)=k(u)\Lambda(\rho^\alpha,\rho^\beta).$$

Hence we find

$$M(u)=m(u)\gamma\otimes\gamma \text{ with } m(u)=\kappa(u)\mathrm{D}\phi(u)\cdot\gamma-\tfrac12\phi(u)\mathrm{D}\kappa(u)\cdot\gamma.$$

The general case seems too difficult to be analyzed, hence we reduce to the case $k(u)\equiv 1$. Introducing the matrix $V=\mathrm{diag}(1/u_i)_i$ we have $\mathrm{D}_u(u^\alpha)[\gamma]=u^\alpha\alpha\cdot V\gamma$, and after some elementary calculations involving (A.3) we find

$$m(u)=\tfrac12\Lambda(\rho^\alpha,\rho^\beta)\big(\rho^\alpha\alpha-\rho^\beta\beta+\Lambda(\rho^\alpha,\rho^\beta)(\alpha{-}\beta)\big)\cdot V(\alpha{-}\beta).$$

For geodesic λ-convexity we have to show $m(u)\geq\lambda\Lambda(\rho^\alpha,\rho^\beta)$, which leads to the formula

$$\lambda=\tfrac12\inf\{\textstyle\sum_{i=1}^n\frac{(\alpha_i-\beta_i)}{w_i\rho_i}\big[\rho^\alpha\alpha_i-\rho^\beta\beta_i+\Lambda(\rho^\alpha,\rho^\beta)(\alpha_i{-}\beta_i)\big]\mid\rho\in\,]0,\infty[^n\,\}.$$

To analyze the formula for λ we consider the case $\alpha_i\beta_i=0$ for all i, which holds for (6.2). Then,

$$\lambda=\tfrac12\inf\{\textstyle\sum_1^n\frac{1}{w_i\rho_i}\big(\alpha_i^2\rho^\alpha{+}\beta_i^2\rho^\beta{+}\Lambda(\rho^\alpha,\rho^\beta)(\alpha_i^2{+}\beta_i^2)\big)\mid\rho\in\,]0,\infty[^n\,\}\geq 0.$$

Thus, in the case of a single reaction with $\alpha_i\beta_i=0$ for all i we always have geodesic convexity. If additionally $\min\{|\alpha|_1,|\beta|_1\}>1$ (as for the reaction in (6.2)), we always have $\lambda=0$ by taking $\rho\to 0$. In the case $|\alpha|_1=1=|\beta|_1$ (as for Markov chains) we have homogeneity of degree 0 and may even obtain $\lambda>0$.

We finally discuss the annihilation-creation reaction, which is used to model the generation and recombination of electron-hole pairs in semiconductors, cf. [Gli08, GlG09, Mie11a]. We have

$$\dot u=-\kappa\Big(\frac{u_1u_2}{w_1w_2}{-}1\Big)\begin{pmatrix}1\\1\end{pmatrix}, \quad \text{where} \quad \alpha=\begin{pmatrix}1\\1\end{pmatrix} \text{ and } \beta=\begin{pmatrix}0\\0\end{pmatrix}.$$

Using $\kappa=w_j=1$ for simplicity we obtain

$$\lambda=\tfrac12\inf\{\,\big(\tfrac{1}{u_1}{+}\tfrac{1}{u_2}\big)\big(u_1u_2{+}\Lambda(u_1u_2,1)\big)\mid u_1,u_2>0\,\}=\cosh(1)=1.53408...$$

A Properties of the function Λ

In this section we collect the essential properties of the function Λ defined in (1.2). 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

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

The lower estimate for Λ can be generalized to

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

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

$$\partial_a\Lambda(a,b) = \tfrac{1}{\log a - \log b}\big(1-\tfrac{\Lambda(a,b)}{a}\big) > 0, \quad \partial_b\Lambda(a,b) = \tfrac{1}{\log b - \log a}\big(1-\tfrac{\Lambda(a,b)}{b}\big) > 0, \tag{A.3}$$

which implies

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

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

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

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

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

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

A nontrivial estimate and identity is the following:

$$\max\{\, \Lambda(r,a) - \partial_a\Lambda(a,b)r \mid r>0 \,\} = a\partial_b\Lambda(a,b). \tag{A.6}$$

The result uses somehow hidden properties of Λ and is crucial for our analysis of geodesic λ-convexity of (X_n,E,K). Using the homogeneity (A.5), this identity follows from (A.7c), which is established below using the auxiliary function ℓ defined in (1.3).

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:*

$$l = \ell(\xi) \quad\Longleftrightarrow\quad \big(\,\exists\,\kappa\in\mathbb{R}:\ l=\widetilde\ell(\kappa) \text{ and } \xi = \widetilde\ell(-\kappa)\,\big), \tag{A.7a}$$

$$\forall\,\xi>0: \quad \ell(\ell(\xi)) = \xi, \tag{A.7b}$$

$$\forall\, a,b>0: \quad \ell(\partial_a\Lambda(a,b)) = \partial_b\Lambda(a,b). \tag{A.7c}$$

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 (1.3) 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.7a) is established.

Identity (A.7b) follows directly from (A.7a), 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.7c). ■

The important identity (A.7b) follows also directly for any $\overline{\ell}$ defined via $\overline{\ell}(\xi) = \sup\{\,\overline{\lambda}(r) - \xi r \,|\, r > 0\,\}$ if $\overline{\lambda}(r) = r\overline{\lambda}(1/r)$, which in our case follows from $\Lambda(1,r) = r\Lambda(1/r,1) = r\Lambda(1,1/r)$.

B Proof of Proposition 4.3

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

$$G_\beta(r,s,t) \overset{\text{def}}{=} \begin{pmatrix} \Lambda_{rs}(\frac1r+\frac1s) + \frac{\Lambda_{rs,s}}{\Lambda_{rs}}\,t & \beta(\Lambda_{rs}/\Lambda_{st})^{1/2} + \beta(\Lambda_{st}/\Lambda_{rs})^{1/2} \\ \beta(\Lambda_{rs}/\Lambda_{st})^{1/2} + \beta(\Lambda_{st}/\Lambda_{rs})^{1/2} & \Lambda_{st}(\frac1s+\frac1t) + \frac{\Lambda_{st,s}}{\Lambda_{st}}\,r \end{pmatrix},$$

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

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

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

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

$$\text{where } \widehat{\gamma}(r,s,t) \overset{\text{def}}{=} \frac{\frac{\Lambda_{rs}}{\Lambda_{st}} + 2 + \frac{\Lambda_{st}}{\Lambda_{rs}}}{\big(\Lambda_{rs}(\frac1r+\frac1s) + \frac{\Lambda_{rs,s}}{\Lambda_{rs}}\,t\big)\big(\Lambda_{st}(\frac1s+\frac1t) + \frac{\Lambda_{st,s}}{\Lambda_{st}}\,r\big)}.$$

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 nominator, so that we estimate the function γ with $\widehat{\gamma} \leq 2\gamma$ and

$$\gamma(r,s,t) \overset{\text{def}}{=} \frac{\Lambda_{rs}^2 + \Lambda_{st}^2}{\big(\Lambda_{rs}^2(\frac1r+\frac1s) + \Lambda_{rs,s}\,t\big)\big(\Lambda_{st}^2(\frac1s+\frac1t) + \Lambda_{st,s}\,r\big)}.$$

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

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

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

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

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

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

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

We continue via

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

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

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

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

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

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

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

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

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

<u>**Case 5:** $r \to 0$**,** $s \to s_* > 0$**, and** $t \to t_* = 1{-}s_* > 0$**.</u> Since the nominator 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 $r \to 0$.

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

This finishes the proof of Proposition 4.3.

Acknowledgment. The author is grateful to Giuseppe Savaré for stimulating discussions and many helpful comments. The research was partially supported by ERC-2010-AdG 267802 *(Analysis of Multiscale Systems Driven by Functionals)*

References

[AGS05] L. AMBROSIO, N. GIGLI, and G. SAVARÉ. *Gradient flows in metric spaces and in the space of probability measures*. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, 2005.

[AM∗11] S. ARNRICH, A. MIELKE, M. A. PELETIER, G. SAVARÉ, and M. VENERONI. Passing to the limit in a Wasserstein gradient flow: from diffusion to reaction. *Calc. Var. Part. Diff. Eqns.*, 2011. Accepted.

[BaÉ85] D. BAKRY and M. ÉMERY. Diffusions hypercontractives. In *Séminaire de probabilités, XIX, 1983/84*, volume 1123 of *Lecture Notes in Math.*, pages 177–206. Springer, Berlin, 1985.

[Bak94] D. BAKRY. L'hypercontractivité et son utilisation en théorie des semigroupes. In *Lectures on probability theory (Saint-Flour, 1992)*, volume 1581 of *Lecture Notes in Math.*, pages 1–114. Springer, Berlin, 1994.

[BeB00] J.-D. BENAMOU and Y. BRENIER. A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numer. Math.*, 84(3), 375–393, 2000.

[BoS09] A.-I. BONCIOCAT and K.-T. STURM. Mass transportation and rough curvature bounds for discrete spaces. *J. Funct. Anal.*, 256, 2944–2966, 2009.

[CH∗11] S.-N. CHOW, W. HUANG, Y. LI, and H. ZHOU. Fokker-Planck equations for a free energy functional of Markov process on a graph. *Preprint*, 2011.

[DaS08] S. DANERI and G. SAVARÉ. Eulerian calculus for the displacement convexity in the Wasserstein distance. *SIAM J. Math. Analysis*, 40, 1104–1122, 2008.

[GlG09] A. GLITZKY and K. GÄRTNER. Energy estimates for continuous and discretized electro-reaction-diffusion systems. *Nonlinear Anal.*, 70(2), 788–805, 2009.

[Gli08] A. GLITZKY. Exponential decay of the free energy for discretized electro-reaction-diffusion systems. *Nonlinearity*, 21(9), 1989–2009, 2008.

[JKO98] R. JORDAN, D. KINDERLEHRER, and F. OTTO. The variational formulation of the Fokker-Planck equation. *SIAM J. Math. Analysis*, 29(1), 1–17, 1998.

[LoV09] J. LOTT and C. VILLANI. Ricci curvature for metric-measure spaces via optimal transport. *Ann. of Math. (2)*, 169(3), 903–991, 2009.

[Maa11] J. MAAS. Gradient flows of the entropy for finite Markov chains. *arXiv:1102.5238v1*, 2011.

[McC97] R. J. MCCANN. A convexity principle for interacting gases. *Adv. Math.*, 128, 153–179, 1997.

[Mie11a] A. MIELKE. A gradient structure for reaction-diffusion systems and for energy-drift-diffusion systems. *Nonlinearity*, 24, 1329–1346, 2011.

[Mie11b] A. MIELKE. Thermomechanical modeling of bulk-interface interaction in reaction-diffusion systems. *In preparation*, 2011.

[OnM53] L. ONSAGER and S. MACHLUP. Fluctuations and irreversible processes. *Phys. Rev.*, 91(6), 1505–1512, 1953.

[Ons31] L. ONSAGER. Reciprocal relations in irreversible processes, I+II. *Physical Review*, 37, 405–426, 1931. (part II, 38:2265-227).

[Ott01] F. OTTO. The geometry of dissipative evolution equations: the porous medium equation. *Comm. Partial Differential Equations*, 26, 101–174, 2001.

[Ött05] H. C. ÖTTINGER. *Beyond Equilibrium Thermodynamics*. John Wiley, New Jersey, 2005.

[OtW05] F. OTTO and M. WESTDICKENBERG. Eulerian calculus for the contraction in the Wasserstein distance. *SIAM J. Math. Analysis*, 37, 1227–1255, 2005.

[vRS05] M.-K. VON RENESSE and K.-T. STURM. Transport inequalities, gradient estimates, entropy, and Ricci curvature. *Comm. Pure Appl. Math.*, 58(7), 923–940, 2005.