Z. Anal. Anwend. 44 (2025), 251–275
DOI 10.4171/ZAA/1796

© 2025 European Mathematical Society
Published by EMS Press
This work is licensed under a CC BY 4.0 license

Coarse-graining and reconstruction for Markov matrices

Artur Stephan

Abstract. We present an operator theoretic coarse-graining (or model order reduction) procedure for stochastic matrices by clustering. The method is consistent with the natural structure of Markov theory, preserving positivity and mass, and does not rely on any tools from Hilbert space theory. The reconstruction is provided by a generalized Penrose–Moore inverse of the coarse-graining operator incorporating the inhomogeneous invariant measure of the Markov matrix. As we will show, the method provides coarse-graining and reconstruction also on the level of tensor spaces, which is consistent with the notion of an incidence matrix and quotient graphs, and, moreover, allows to coarse-grain and reconstruct fluxes. Furthermore, we investigate the connection with functional inequalities and Poincaré-type constants.

1. Introduction

Coarse-graining or model reduction is a fundamental procedure that reduces the complexity of a physical model. It is a well-established tool used in many branches of applied mathematics including analysis, modeling, and numerics. In this paper, we are interested in coarse-graining for physical systems on a finite state space described by Markov matrices.

Let us first describe the mathematical setting. Fixing a finite state spaces $\mathcal{Z}=\{1,\dots,n\}$, $n\in\mathbb{N}$, the statistical states are given by the set of probability vectors

$$\mathrm{Prob}(\mathcal{Z})=\left\{p\in\mathbb{R}^n : p_i\geq 0, \sum_{i=1}^n p_i = 1\right\}\subset\mathbb{R}^n := X^*.$$

The distinction between primal X and dual spaces X^* (although both are isomorphic to the $\mathbb{R}^n$ as real-vector spaces) is important, where the first contains functions on $\mathcal{Z}$ equipped with the supremum-norm, and the second contains probabilities on $\mathcal{Z}$ equipped with the 1-norm. Dual pairing is denoted by $\langle\cdot,\cdot\rangle$. Apart from X being a vector space, it has a natural order, i.e., $x\geq y$ if the pointwise estimate $x_i\geq y_i$ holds for all $i\in\mathcal{Z}$. Moreover, it is an algebra, i.e., the product of two elements x and y is given by $(x\cdot y)_i = x_i\cdot y_i$ by pointwise multiplication.

Mathematics Subject Classification 2020: 60J10 (primary); 15A42, 39B62, 05B20, 60J28, 15B52 (secondary).
Keywords: model-order reduction, stochastic matrix, Markov matrix, generalized Penrose–Moore inverse, coarse-graining and reconstruction, functional inequalities.

The change of statistical states is described by a Markov matrix (or operator) $K : X \to X$, which, by definition, satisfies $K_{ij} \geq 0$ and $\sum_{j \in \mathcal{Z}} K_{ij} = 1$ for all $i \in \mathcal{Z}$. Equivalently, K maps non-negative elements in X to non-negative, and $K\mathbb{1}_X = \mathbb{1}_X$, where $\mathbb{1}_X = (1, \ldots, 1)^{\mathrm{T}}$ is the constant one-vector. The invariant vector $\pi \in \mathrm{Prob}(\mathcal{Z}) \subset X^*$ of the adjoint matrix K^* is defined by satisfying $\sum_i \pi_i K_{ij} = \pi_j$, or, equivalently, by $\pi^T K = \pi^T$, or $K^*\pi = \pi$. Throughout the paper, we assume that the invariant vector is unique and positive. (See, e.g., [19] for introductory reading on Markov matrices.) We remark that for further generalizations to infinite and continuous state spaces (in which case X and X^* are infinite dimensional) we denote by $B^* : Y^* \to X^*$ the dual (or adjoint) operator of $B : X \to Y$, which is just the (real) transpose in matrix representation.

The aim of the paper is to recall a well-known coarse-graining procedure for Markov matrices by introducing an operator-theoretic formalism, that, by preserving the underlying structure, fits nicely into the theory of Markov matrices. Here, the coarse-graining procedure is based on clustering of the states. Clustering (often also called lumping or aggregation) is an old and well-known technique for model-order reduction for Markov matrices [13, 25]. It can be used for computing the invariant measure of large Markov matrices (see, e.g., [7, 16, 26] and references therein) or solving large linear equations [5, 6, 16]. Recently, the construction of the coarse-graining map via stochastic methods has got increased attention [1, 18].

The idea in that paper is to understand the clustering as an action of a deterministic Markov matrix, which will be called the *coarse-graining operator*. Together with a reconstruction operator, which can be understood as the inverse transformation, they define a projection in the state space. Implicitly, the involved operators (maybe in a modified form) have been used in literature (see, e.g., [5, 11, 20]), and we recall technical results from [17] in Section 2. However, to the author's knowledge, no structural study has been done so far. In applications, the original system is often called *microscopic*, and the coarse-grained system *macroscopic*, and we will also use that notion.

In contrast to classical model order reduction procedures, this construction does not rely on tools from Hilbert space theory like orthonormal projections or symmetry, as for example, Krylov subspace projection methods and so on (see, e.g., [23]). In general, Hilbert space projections will not preserve positivity and mass of the measures, and they are therefore unphysical. Here, the coarse-graining procedure is based on the structural duality between X and X^*. Heuristically, the reconstruction rebuilds the information from the coarser system back to the finer system using the local information of the invariant probability vector π. With that, the theory is not based on homogeneous Euclidean spaces, which are not canonical for Markov matrices as the invariant measure $\pi \in \mathrm{Prob}(\mathcal{Z})$ is in general not homogeneous. However, the reconstruction operator can be understood as generalized Penrose–Moore inverse of the coarse-graining operator respecting the (in general inhomogeneous) invariant measure π (see Proposition 2.5).

In Section 2, it is shown how to define a macroscopic (or coarse-grained) Markov process from a microscopic process using the coarse-graining operator. In particular, Theorem 2.6 shows that the coarse-grained process is obtained exactly if an equilibration of

the microscopic states occurs. In practical applications, where coarse-graining results from a multi-scale problem and scale-separation, the difference between the coarse-grained process and the reduced process (see Section 2.3) is relevant. Difficulties occur, because the latter is in general not a Markov process. The so-defined error provides a measure of how good the coarse-graining model is. Different approaches to estimate the error have been developed in recent years. We refer, e.g., to [15] in the context of diffusion, and to [11] for the discrete counterpart, where the comparison of microscopic and macroscopic functional inequalities and their constants is used. A recent approach is the so-called resolvent approach from [14] in the context of metastability (see also [24]). In Section 2.4, we connect the operator-theoretic framework to the resolvent approach and show their consistency. Because metastability is not the main focus of the paper, we consider only one fixed microscopic system. In Section 2.5, we show how to control and estimate the coarse-graining error. In contrast to [11] where the error is measured in terms of relative entropies and the estimate is derived by functional inequalities, we consider directly the Markov chain and derive a bound using spectral properties, which fits with the operator-theoretic approach.

Despite practical applications, the operator-theoretic coarse-graining procedure has many different mathematical advantages. First, it is consistent with the graph-theoretic notion of the incidence matrix (or operator) and the quotient graph as explained in Section 3. This property is particularly interesting for detailed balance Markov matrices, which define gradient-flow equations written in a continuity equation form (see, e.g., [17, 22]). As it turns out, the corresponding fluxes can be reconstructed, too (see Section 4.2 for more details).

These technical preparations are finally used for the consistency result between coarse-graining and functional inequalities. The first nontrivial eigenvalue defines the spectral gap and provides useful information regarding asymptotical decay of the process (see, e.g., [2] and references therein). However, a direct calculation of the spectrum is often not feasible and functional inequalities are used instead to provide at least bounds on the decay rate. In the last decades, functional inequalities for Markov processes on discrete states spaces have been studied intensively [2,8–10,12]. In Section 5, we connect the operator-theoretic coarse-graining procedure to energy functionals. In particular, we derive coarse-grained functional inequalities and estimate the associated discrete Poincaré-type constants (like the Poincaré constant, or log-Sobolev constant). As a consequence, we can derive the folklore result that coarse-graining always increases Poincaré-type constants by a simple calculation showing the consistency of the operator-theoretic coarse-graining framework.

2. Coarse-graining

We present the operator theoretic framework for capturing the collection of states, which has also been introduced in [17].

2.1. Operator theoretic coarse-graining

For two finite state spaces $\mathcal{Z} = \{1, \dots, n\}$, $\widehat{\mathcal{Z}} = \{1, \dots, \hat{n}\}$ with $\hat{n} < n$, we assume that there is a given surjective function $\phi : \mathcal{Z} \to \widehat{\mathcal{Z}}$, which plays the role of a coarse-graining or clustering map. We define the *coarse-graining operator* $M : \widehat{X} \to X$ by $(M\hat{x})_i = \hat{x}_{\phi(i)}$ for all $\hat{x} \in \widehat{X}$. One easily sees that M is a deterministic Markov matrix since the adjoint (or dual) matrix $M^* : X^* \to \widehat{X}^*$ maps pure states (or Dirac-measures) to pure states. In fact, the dual operator M^* should be called *coarse-graining operator* because it maps statistical states in X^* to coarser states in $\widehat{X}^*$. Since M is a deterministic Markov matrix we have that for all $\hat{x}, \hat{y} \in \widehat{X}$ it holds $M(\hat{x} \cdot \hat{y}) = M\hat{x} \cdot M\hat{y}$, where the multiplication is meant pointwise. By the way, this characterizes all deterministic Markov matrices.

Fixing a positive probability vector $\pi \in X^*$, we may define the multiplication operator given by the diagonal matrix $Q_\pi : X \to X^*$, i.e., $(Q_\pi x)_i = \pi_i x_i$. We observe that Q_π is symmetric and its inverse is given by $Q_\pi^{-1} : X^* \to X$, $p \mapsto \rho = (p_i/\pi_i)_i$. One easily sees that the multiplication operator satisfies $\langle x, Q_\pi y\rangle = \langle x \cdot y, \pi\rangle$ for all $x, y \in X$. Although the spaces are finite-dimensional and isomorphic, we remark that the parameter π of Q_π is an element of the dual space and the inverse $Q_\pi^{-1} : p \mapsto \rho$ maps a probability vector p to the relative density ρ of p with respect to π as a discrete analog of the Radon–Nikodym derivative.

We define a new coarse-grained measure $\hat{\pi}$ by $\hat{\pi} = M^*\pi$. We easily observe that $\hat{\pi}$ is also positive. Moreover, we observe the following.

Lemma 2.1 ([17, Lemma 2.4]). *We have $\hat{\pi} = M^*\pi$ if and only if $Q_{\hat{\pi}} = M^* Q_\pi M$.*

Proof. Evaluating $Q_{\hat{\pi}} = M^* Q_\pi M$ at $\hat{\mathbb{1}}$, we have $\hat{\pi} = Q_{\hat{\pi}}\hat{\mathbb{1}} = M^* Q_\pi M\hat{\mathbb{1}} = M^* Q_\pi \mathbb{1} = M^*\pi$, which is one direction of the claim. For the other claim, we introduce the multiplication operator $\Pi_{\hat{y}} : \widehat{X} \to \widehat{X}$ by $(\Pi_{\hat{y}}\hat{x})_j = \hat{y}_j \hat{x}_j$ with the dual operator $\Pi_{\hat{y}}^* : \widehat{X}^* \to \widehat{X}^*$ given by $\Pi_{\hat{y}}^* \hat{c} = Q_{\hat{c}}\hat{y}$. Using that M is a deterministic operator we have $M\Pi_{\hat{y}} = \Pi_{M\hat{y}} M$, which yields, by dualizing, $\Pi_{\hat{y}}^* M^* = M^* \Pi_{M\hat{y}}^*$. So, we get for any $\hat{x}$ that $Q_{\hat{\pi}}\hat{x} = \Pi_{\hat{x}}^*\hat{\pi} = \Pi_{\hat{x}}^* M^*\pi = M^*\Pi_{M\hat{x}}^*\pi = M^* Q_\pi M\hat{x}$. ∎

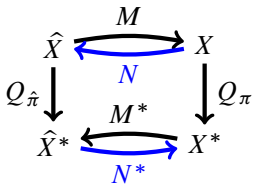
This important relation does not hold if M is not a deterministic operator. Moreover, it implies that the following diagram commutes:

$$\begin{array}{ccc} \widehat{X} & \xrightarrow{M} & X \\ {\scriptstyle Q_{\hat{\pi}}}\downarrow & & \downarrow{\scriptstyle Q_\pi} \\ \widehat{X}^* & \xleftarrow{M^*} & X^* \end{array}$$

Since $Q_{\hat{\pi}}$ is invertible, it is now possible to "invert" the coarse-graining operator M by defining the so-called *reconstruction operator* $N : X \to \widehat{X}$:

$$N = Q_{\hat{\pi}}^{-1} M^* Q_\pi : X \to \widehat{X}, \quad N^* = Q_\pi M Q_{\hat{\pi}}^{-1} : \widehat{X}^* \to X^*. \tag{2.1}$$

The operator N^* reconstructs the coarser statistical states in $\widehat{X}^*$ respecting the measure $\pi \in X^*$:



Before summarizing properties of N in the next proposition, we introduce the notion of *detailed balance*.

Definition 2.2. A Markov matrix K is said to satisfy the *detailed balance condition* with respect to its positive invariant measure π if $K^* Q_\pi = Q_\pi K$.

Proposition 2.3 ([17, Lemma 2.5, Proposition 2.7]). *Let $M : \widehat{X} \to X$ be a deterministic Markov matrix, let $\pi \in X^*$ be a given positive probability vector and let $\hat{\pi} = M^*\pi$. Let N and N^* be defined by (2.1). Then, the following holds:*

(1) *N is a Markov matrix and $N^*\hat{\pi} = \pi$, i.e., N^* inverts with respect to π,*

(2) *$NM = \mathrm{id}_{\widehat{X}}$ and $MN =: P$ is a (Markov) projection on X. We have the splitting $X = \mathrm{Range}(P) + \mathrm{Ker}(P) = \mathrm{Range}(M) + \mathrm{Ker}(N)$. The adjoint P^* has π as its stationary measure and satisfies detailed balance.*

Proof. Clearly, N is nonnegative and $N\mathbb{1} = Q_{\hat{\pi}}^{-1} M^* Q_\pi \mathbb{1} = Q_{\hat{\pi}}^{-1} M^* \pi = Q_{\hat{\pi}}^{-1} \hat{\pi} = \hat{\mathbb{1}}$. Hence, N is a Markov matrix. Moreover, we have $N^*\hat{\pi} = Q_\pi M Q_{\hat{\pi}}^{-1} \hat{\pi} = Q_\pi M \hat{\mathbb{1}} = \pi$.

For the second claim, we use Lemma 2.1, which implies $M^* N^* = M^* Q_\pi M Q_{\hat{\pi}}^{-1} = \mathrm{id}_{\widehat{X}^*}$. Hence, $NM = \mathrm{id}_{\widehat{X}}$ and $P = MN$ is a projection. This provides the decomposition of X since N is surjective and M is injective. Since it is the composition of Markov matrices P is again a Markov matrix and obviously $P^*\pi = \pi$. To see that P satisfies detailed balance, we observe

$$Q_\pi P = Q_\pi MN = Q_\pi M Q_{\hat{\pi}}^{-1} M^* Q_\pi = N^* M^* Q_\pi = P^* Q_\pi. \qquad \blacksquare$$

Remark 2.4. We remark that in [29] inverse operators for general Markov operators (not necessarily deterministic) have been introduced and their relation to the direction of time has been investigated.

Finally, we investigate the connection between the 'inverse' operator N and the Penrose–Moore inverse of linear algebra. First, we see that $N : X \to \widehat{X}$ is a pseudo inverse of $M : \widehat{X} \to X$, because $MNM = M$ and $NMN = N$ by Proposition 2.3. Recall that for an injective $M : \widehat{X} \to X$, the Penrose–Moore inverse of M can be defined by

$$M^+ = (M^* M)^{-1} M^*.$$

The next proposition shows, that this formula provides exactly N, if the adjoint operator M^* is understood in the space $L^2(\pi)$. In particular, if $\pi = \frac{1}{n}(1, \ldots, 1)^{\mathrm{T}}$, we have that

$N = M^+$. To see this, we define the $L^2(\pi)$-inner product in X by

$$(x, y)_\pi := \langle x, Q_\pi y\rangle = \langle x \cdot y, \pi\rangle.$$

Proposition 2.5. *The reconstruction operator $N = Q_{\hat\pi}^{-1} M^* Q_\pi$ is the $L^2(\pi)$-adjoint of M. In particular, N is a generalized Penrose–Moore inverse of M in $L^2(\pi)$.*

Proof. We have that

$$(M\hat x, y)_\pi = \langle \hat x, M^* Q_\pi y\rangle = \langle \hat x, Q_{\hat\pi} N y\rangle = (\hat x, Ny)_{\hat\pi}.$$ ∎

We note that the notion of *detailed balance* from Definition 2.2 means that the Markov matrix K is symmetric in $L^2(\pi)$.

2.2. Example

For an example, we consider $Z = \{1, 2, 3\}$ and $\hat Z = \{\hat 1, \hat 2\}$ and define $\phi(1) = \hat 1$, $\phi(2) = \phi(3) = \hat 2$. In matrix representation, the coarse-graining operator has the form

$$M = \begin{pmatrix} 1 & \\ & 1 \\ & 1 \end{pmatrix}, \quad M^* = \begin{pmatrix} 1 & & \\ & 1 & 1 \end{pmatrix}.$$

Setting $\pi = (\pi_1, \pi_2, \pi_3)^{\mathrm T}$, we obtain $\hat\pi = (\pi_1, \pi_2 + \pi_3)^{\mathrm T}$, and hence,

$$N = \begin{pmatrix} 1 & & \\ & \frac{\pi_2}{\pi_2+\pi_3} & \frac{\pi_3}{\pi_2+\pi_3} \end{pmatrix}, \quad P = MN = \begin{pmatrix} 1 & & \\ & \frac{\pi_2}{\pi_2+\pi_3} & \frac{\pi_3}{\pi_2+\pi_3} \\ & \frac{\pi_2}{\pi_2+\pi_3} & \frac{\pi_3}{\pi_2+\pi_3} \end{pmatrix}.$$

2.3. Coarse-graining for Markov matrices

Let a Markov matrix $K : X \to X$ be given. We assume that its adjoint K^* has a unique invariant measure π, i.e., $K^*\pi = \pi$. We define the coarse-grained Markov matrix $\widehat K$ by contracting K via

$$\widehat K = NKM : \hat X \to \hat X.$$

The next theorem shows that $\widehat K$ can indeed be understood as a coarse-grained version of K.

Theorem 2.6. *Let a Markov matrix $K : X \to X$ with an invariant measure π be given. Let M be a deterministic Markov matrix, and reconstruction operator N be defined by (2.1). Let $\widehat K := NKM : \hat X \to \hat X$. Then, we have the following:*

(1) *$\widehat K$ is a Markov matrix on $\hat X$;*

(2) *$\widehat K^*$ has $\hat\pi$ as its invariant measure;*

(3) *Define the Markov chain* $p_{k+1} = K^* p_k$, $p_0 \in X^*$, $k \in \mathbb{N}$. *If there is an equilibration of the form* $p_k = N^* \hat{p}_k$ *for* $\hat{p}_k \in \hat{X}^*$ *and all* $k \geq 0$, *then the probability vectors* $\hat{p}_k$ *satisfy the coarse-grained Markov chain* $\hat{p}_k = \hat{K}^* \hat{p}_{k-1}$;

(4) *If* K *satisfies the detailed balance condition with respect to* π, *then does* $\hat{K}$ *with respect to* $\hat{\pi}$.

Important for exact coarse-graining is the equilibration $p_k = N^* \hat{p}_k$ which means that the densities of p_k and $\hat{p}_k$ with respect to π or $\hat{\pi}$, respectively, are equilibrated, i.e.,

$$p_k = Q_\pi M Q_{\hat{\pi}}^{-1} \hat{p}_k \iff \rho_k = M \hat{\rho}_k, \tag{2.2}$$

where

$$\rho_k = Q_\pi^{-1} p_n, \quad \hat{\rho}_k = Q_{\hat{\pi}}^{-1} \hat{p}_k.$$

Of course, we have the reduced process $M^* p_k = \hat{p}_k$ and the above result shows, that the reduced process is exactly the coarse-grained process. In particular, this makes clear why $\hat{K}$ is the natural coarse-graining Markov matrix of K.

Proof. Since $\hat{K}$ is the composition of Markov matrices it is itself a Markov matrix. Moreover, we see that $\hat{\pi}$ is the invariant measure of $\hat{K}$, because

$$\hat{K}^* \hat{\pi} = M^* K^* N^* \hat{\pi} = M^* K^* \pi = M^* \pi = \hat{\pi}.$$

Considering the Markov chain, let $p_k = K^* p_{k-1}$ be given. Assuming that $p_k = N^* \hat{p}_k$, we conclude that $\hat{p}_k = M^* N^* \hat{p}_k = M^* p_k$. Hence,

$$\hat{p}_k = M^* p_k = M^* K^* p_{k-1} = M^* K^* N^* \hat{p}_{k-1} = \hat{K}^* \hat{p}_{k-1}.$$

Finally, if K satisfies detailed balance with respect to π then $Q_\pi K = K^* Q_\pi$, and hence, we have

$$Q_{\hat{\pi}} \hat{K} = Q_{\hat{\pi}} NKM = M^* Q_\pi KM = M^* K^* Q_\pi M = M^* K^* N^* Q_{\hat{\pi}} = \hat{K}^* Q_{\hat{\pi}}. \quad \blacksquare$$

Conversely, if $\hat{p}_{k+1} = \hat{K}^* \hat{p}_k$, and $\hat{K} = NKM$, then a direct computation shows that $p_k := N^* \hat{p}_k$ solves the projected Markov chain $p_{k+1} = P^* K^* P^* p_k$, which in general is different to $p_{k+1} = K^* p_k$. In this sense, the projection P describes the information loss going from a coarser system to a finer system.

Of course, Theorem 2.6 naturally generalizes to continuous time Markov processes (see, e.g., [17]). Let $p(t) = \mathrm{e}^{tA^*} p_0$ or equivalently p solving $\dot{p} = A^* p$ be given, where A is a Markov generator such that e^{tA} is a semigroup of Markov matrices. If $p(t) = N^* \hat{p}(t)$, then the coarse-grained probability vectors $\hat{p}$ satisfies the coarse-grained Markov process $\dot{\hat{p}} = \hat{A}^* \hat{p}$, with the coarse-grained Markov generator

$$\hat{A} = NAM.$$

2.4. Connection to the resolvent

From Theorem 2.6, it is now clear to compare three different processes.

(1) The microscopic process: $p_{n+1} := K^* p_n$, $p_0 \in X^*$.

(2) The reduced process: $\hat{p}_n := M^* p_n \in \widehat{X}^*$ for $n \geq 1$ and $\hat{p}_0 = M^* p_0$.

(3) The coarse-grained process: $\hat{q}_{n+1} := \widehat{K}^* \hat{q}_n$, $\hat{q}_0 \in \widehat{X}^*$,

where the coarse-grained Markov oprator is defined as above by $\widehat{K}^* := M^* K^* N^*$ on $\widehat{X}^*$. Note that in general the reduced process is not described by a Markov operator.

Theorem 2.6 shows that the reduced process is exactly given by the coarse-grained process if the equilibration $p \in \mathrm{range}(N^*)$ holds. In application, this property can be derived from time-scaling arguments [11, 15, 17, 27].

Recently, an equivalent characterization is derived in terms of the resolvent of the generator [14] in the context of metastability, which we translate here in our operator-theoretic framework. In contrast to [14], we are not considering a sequence of Markov process, but only one Markov process and a fixed coarse-graining map $M : \widehat{X} \to X$. Note that for the resolvent approach, we consider the time-continuous process.

The resolvent condition from [14] can be stated as follows: for the Markov process defined by the generator $A : X \to X$, let $\hat{g} \in \widehat{X}$ and $g = M\hat{g} \in X$ be given, and let $f_\lambda = f_{\lambda,g}$ be the unique solution of the resolvent equation

$$(\lambda - A) f_\lambda = g.$$

Assume that f_λ is in the image of M, i.e., there is $\hat{f}_\lambda \in \widehat{X}$ such that $f_\lambda = M\hat{f}_\lambda$. Moreover, assume that there is generator $\widehat{A} : \widehat{X} \to \widehat{X}$ such that for all

$$(\lambda - \widehat{A})\hat{f}_\lambda = \hat{g}.$$

It has been shown in [14], that the reduced process $(M^* \mathrm{e}^{tA^*})_{t\geq 0}$ is given by $(\mathrm{e}^{t\widehat{A}^*})_{t\geq 0}$. Moreover, in [14] it has been also shown that the reduced resolvent condition can also be stated asymptotically for a sequence of Markov processes, and this actually characterizes metastability. Here, we are only discussing the resolvent equation in the context of the equilibration of the microscopic dynamics as stated in Theorem 2.6.

First, we see that applying N to the resolvent equation $(\lambda - A)M\hat{f}_\lambda = M\hat{g}$ and using $NM = \mathrm{id}_{\widehat{X}}$, we get that $(\lambda - NAM)\hat{f}_\lambda = \hat{g}$. Since the resolvent determines uniquely the generator, we conclude the characterization of the coarse-grained generator $\widehat{A} = NAM : \widehat{X} \to \widehat{X}$ as we have seen before.

Moreover, it is clear that also the semigroup e^{tA} keeps the subspace $\mathrm{range}(M) \subset X$ invariant. Indeed, recalling the relation of the resolvent and the semigroup by the Laplace transform (denoted by $G(\lambda) = \mathscr{L}(g(\cdot))(\lambda)$), we have

$$\begin{aligned} &g(t) = \mathrm{e}^{tA} g_0 \Leftrightarrow \dot{g}(t) = Ag(t), \\ &g(0) = g_0 \Leftrightarrow \lambda G(\lambda) - g_0 = AG(\lambda) \Leftrightarrow (\lambda - A)G(\lambda) = g_0. \end{aligned}$$

Hence, we conclude that $\mathrm{e}^{tA}\mathrm{range}(M) \subset \mathrm{range}(M)$.

To transfer the above property to the evolution of the statistical states described by the adjoint Markov generator A^*, we rely on (2.2). Indeed, assuming that e^{tA} describes the evolution of the relative densities $h(t) = Q_\pi^{-1} p(t)$, then by (2.2) we have that the corresponding measures $p(t) = e^{tA^*} p_0$ are equilibrated (i.e., are in range(N^*)), and hence, Theorem 2.6 characterizes the reduced process. We summarize the considerations in the following proposition if A satisfies detailed balance.

Proposition 2.7. *Let a coarse-graining operator $M : \widehat{X} \to X$ be given and consider for a detailed balance Markov generator A^* with $A^*\pi = 0$ and $\pi > 0$ the Markov process $p(t) = e^{tA^*} p_0$, $p_0 \in$ range(M). Assume that the resolvent of $A : X \to X$ keeps* range(M) *invariant. Then, for all $t \geq 0$, there is $\hat{p}(t) \in \widehat{X}^*$ such that $p(t) = M^* \hat{p}(t)$ and $t \mapsto \hat{p}(t)$ is the process generated by $\widehat{A}^* = N^* A^* M^*$.*

Proof. We simply observe that the relative densities $h(t) = Q_\pi^{-1} p(t)$, with $h_0 = Q_\pi^{-1} p_0$, solve

$$h(t) = Q_\pi^{-1} e^{tA^*} Q_\pi Q_\pi^{-1} p_0 = e^{tQ_\pi^{-1} A^* Q_\pi} h_0 = e^{tA} h_0,$$

where we have used that A^* satisfies detailed balance. ∎

2.5. Coarse-graining error

Now, we want to compare the distance between the coarse-grained process and the reduced process. We have seen in Theorem 2.6 that whenever the macroscopic process equilibrates, i.e., there are $\hat{p}_n \in \widehat{X}^*$ such that $p_n = N^* \hat{p}_n$, then $M^* p_n = \hat{p}_n$ is described by the coarse-grained process. However, in application this equilibration is a priori not guaranteed, and an estimate is desired.

For this, let again $K^* : X \to X$ be fixed, and assume that we have a unique invariant probability measure $\pi \in X^*$. Moreover, let the coarse-graining operator $M^* : X^* \to X^*$ be given and define the coarse-grained Markov operator $\widehat{K}^* := M^* K^* N^*$ on $\widehat{X}^*$. Using the notation from 2.4, we want to compare the processes $(\hat{p}_n)_{n\in\mathbb{N}}$ and $(\hat{q}_n)_{n\in\mathbb{N}}$. To make the difference comparable, we assume that the initial data is well-prepared, i.e., that also $\hat{q}_0 = M^* p_0$. Then, we have

$$\hat{p}_1 - \hat{q}_1 = M^* p_1 - \widehat{K}^* \hat{q}_0 = (M^* K^* - \widehat{K}^* M^*) p_0 = M^* K^* (I - P^*) p_0,$$

where we have used the projection $P^* = N^* M^*$ in X^*. This means that the difference of the coarse-grained process and the reduced process is essentially described by the distance of the initial value p_0 and the range of P^*, and how K^* acts on the range of $I - P^*$.

We compute the difference of $\hat{p}_n$ and $\hat{q}_n$. We have

$$\begin{aligned}\hat{p}_{n+1} - \hat{q}_{n+1} &= M^* p_{n+1} - (\widehat{K}^*)^{n+1} \hat{q}_0 = M^* (K^*)^{n+1} p_0 - (\widehat{K}^*)^{n+1} M^* p_0 \\ &= (M^* (K^*)^{n+1} - (\widehat{K}^*)^{n+1} M^*) p_0.\end{aligned}$$

Using the projection, we have

$$(\widehat{K}^*)^{n+1} M^* = (M^* K^* N^*)^{n+1} M^* = M^* K^* P^* K^* P^* \cdots P^* K^* P^*,$$

with in total K^* appearing $n+1$ times. Hence, we get

$$\hat{p}_{n+1} - \hat{q}_{n+1} = M^* K^* ((K^*)^n - (P^* K^* P^*)^n) p_0.$$

This means that the difference for $n \geq 1$ between the coarse-grained process and the reduced process is essentially given by

$$\Delta_n^* = (K^*)^n - (P^* K^* P^*)^n.$$

The operator Δ_n^* describes exactly the error between the coarse-grained process and the reduced process. Of course, it is bounded (because Markov operators are bounded by 1), and it is essentially given by an averaging formula between the subsequent application of the projection P^* and Markov operator K^*.

We are now going to estimate the difference by exploiting spectral properties of K^*. For this, we recall that the Markov chain $(K^*)^n$ converges to the projection P_π^* on the invariant measure $\pi \in X^*$, which we have assumed to be unique. The projection is given by

$$P_\pi^* : X^* \to X^*, \quad P_\pi^* p = \langle \mathbb{1}, p \rangle \pi,$$

and it is a classical result that $\|(K^*)^n - P_\pi^*\| \leq C \lambda_{K,2}^n$, where $\lambda_{K,2} < 1$ is the real-part of the second largest eigenvalue of K (the largest is of course $\lambda_1 = 1$). Similarly, we have $\|(\hat{K}^*)^n - P_{\hat{\pi}}^*\| \leq C \lambda_{\hat{K},2}^n$, where $\lambda_{\hat{K},2} < 1$ is the real-part of the second largest eigenvalue of $\hat{K}$.

Note that we have for all $p \in X^*$ that

$$M^* P_\pi^* p = M^* \pi \langle p, \mathbb{1} \rangle = \hat{\pi} \langle \mathbb{1}, p \rangle = \hat{\pi} \langle M \mathbb{1}, p \rangle = \hat{\pi} \langle \mathbb{1}, M^* p \rangle = P_{\hat{\pi}}^* M^* p.$$

Hence, we have

$$\begin{aligned}
&|\hat{p}_{n+1} - \hat{q}_{n+1}| \\
&= \|(M^*(K^*)^{n+1} - (\hat{K}^*)^{n+1} M^*)\| \cdot |p_0| \\
&\leq \{\|M^*(K^*)^{n+1} - M^* P_\pi^*\| + \|M^* P_\pi^* - P_{\hat{\pi}}^* M^*\| + \|P_{\hat{\pi}}^* M^* - (\hat{K}^*)^{n+1} M^*\|\} |p_0| \\
&\leq C \max\{\lambda_{K,2}, \lambda_{\hat{K},2}\}^n |p_0|.
\end{aligned}$$

Later in Section 5, the connection of coarse-graining, functional inequalities, and spectral gaps are discussed in more detail. In particular, Theorem 5.6 shows that $\lambda_{\hat{K},2} \geq \lambda_{K,2}$, which now can be summarized in the following result.

Proposition 2.8. *The difference between the coarse-grained process $\hat{q}_n$ and the reduced process $\hat{p}_n$ can be estimated by*

$$|\hat{p}_{n+1} - \hat{q}_{n+1}| \leq C \lambda_{\hat{K},2}{}^n |p_0|,$$

where $p_0 \in X^*$ is the initial value of the microscopic process, $C > 1$ is a constant, and $\lambda_{\hat{K},2}$ is the real part of the second largest eigenvalue of $\hat{K}$.

We finally remark that since the time-continuous semigroup e^{tA^*} also converges to the projection with the same speed, an analogous result holds for the reduced and coarse-grained time-continuous process.

3. Coarse-grained network

Graph theoretically, a Markov matrix K defines a directed graph $G = G(V, E)$, with vertices given by the state space $\mathcal{Z}$ and edges between states z_i and z_j whenever $K_{ij} > 0$. An equivalence relation given by the coarse-graining map $\phi : \mathcal{Z} \to \widehat{\mathcal{Z}}$ (i.e., $z_i \sim z_j$ iff $\phi(z_i) = \phi(z_j)$), defines a partition of the graph into blocks. By definition, these blocks define the vertices of the so-called *quotient graph* $\widehat{G}$. The edges in the quotient graph are defined as follows: two blocks B_1 and B_2 are adjacent if some vertex in B_1 is adjacent to some vertex in B_2 with respect to the edges in the starting graph. That means if on $G = G(V, E)$ there is an equivalence relation $\sim$, then $\widehat{G}$ has vertices $\widehat{V} = V/\sim$ and edges $\{([u]_\sim, [v]_\sim) : (u, v) \in E\}$. In particular, the edges in each equivalence class (or block) vanish (see, e.g., [3]).

Since we are interested in functions that are defined on the edges (e.g., fluxes), we translate the above state-based coarse-graining procedure to edges. For this, it is convenient to introduce tensor spaces.

3.1. Coarse-graining in tensor spaces

Naturally, the space over the edges can be identified by matrices or equivalently by the tensor product space $X \otimes X \simeq L(X^*, X)$, where the latter is the space of linear maps from X^* to X. In particular, we use both formulations and switch between them whenever necessary. In principle, also multi-tensor spaces can be considered, for example, to capture cycles between several states. However, we restrict ourselves to tensors of second order.

Importantly, the tensor space $X \otimes X$, which is the state space in that section, is consistent with the concept of Markov matrices capturing positivity and duality. Positivity is again defined pointwise. The constant 1-element in $X \otimes X$ is given by $\mathbb{1} \otimes \mathbb{1}$ and will be denoted by $\mathbb{1}_\otimes$. Recall from linear algebra that the dual space of $L(X^*, X)$ is given by $L(X^*, X)^* \simeq L(X^{**}, X^*) \simeq L(X, X^*) \simeq X^* \otimes X^*$. The duality mapping is given by

$$A \in L(X, X^*) \mapsto \mathrm{Tr}(A^* \cdot) \in L(X^*, X)^*.$$

In the following, we will denote the dual paring between $L(X^*, X)$ and $L(X, X^*)$ by

$$\langle\!\langle A, B \rangle\!\rangle := \mathrm{Tr}(A^* B) = \mathrm{Tr}(A B^*) = \sum_{i,j} A_{ij} B_{ij},$$

which is just the usual dual paring by pointwise multiplication if the matrices are understood as $n \times n$-vectors.

In Section 2, the notation of a multiplication operator was important to define the reconstruction map. Analogously, we may also define multiplication operators with elements of the dual space here. For a given matrix $m \in L(X, X^*) \simeq X^* \otimes X^*$, we define the (diagonal) multiplication operator by

$$Q_m : L(X^*, X) \to L(X, X^*), \quad Q_m b = (m_{ij} b_{ij})_{ij}$$

by pointwise multiplication. Clearly, we have that $Q_m \mathbb{1}_\otimes = m$. Indeed, the target space of Q_m makes sense which can been seen from the following observation that Q_m is symmetric:

$$\langle\!\langle c, Q_m b \rangle\!\rangle = \sum_{i,j} m_{ij} b_{ij} c_{ij} = \sum_{i,j} m_{ij} c_{ij} b_{ij} = \langle\!\langle b, Q_m c \rangle\!\rangle.$$

Of great importance for us is the element $m = Q_\pi K \in L(X, X^*) \simeq X^* \otimes X^*$, which can be understood as a weight function defined on the edges.

For the coarse-grained state space $\hat{X}$, we analogously define $\hat{X} \otimes \hat{X} \simeq L(\hat{X}^*, \hat{X})$ and $\hat{X}^* \otimes \hat{X}^* \simeq L(\hat{X}, \hat{X}^*)$ by replacing X by $\hat{X}$. We define for a given coarse-graining operator $M : \hat{X} \to X$ a coarse-graining operator on $L(X^*, X)$ by

$$\begin{aligned} \widetilde{M} : L(\hat{X}^*, \hat{X}) \simeq \hat{X} \otimes \hat{X} &\to L(X^*, X) \simeq X \otimes X, \\ \hat{b} &\mapsto \widetilde{M}\hat{b} := M \hat{b} M^*. \end{aligned}$$

Proposition 3.1. *The operator $\widetilde{M}$ has the following properties:*

(1) *$\widetilde{M}$ is again a deterministic Markov operator,*

(2) *the adjoint operator is given by*

$$\widetilde{M}^* : L(X, X^*) \to L(\hat{X}, \hat{X}^*), \quad \widetilde{M}^* b = M^* b M,$$

(3) *let $m := Q_\pi K$. Then, $\hat{m} := \widetilde{M}^* m = Q_{\hat{\pi}} \hat{K}$.*

Proof. Clearly, $\widetilde{M}$ is again positive. Moreover, it maps to constant 1-function $\hat{\mathbb{1}}_\otimes = \hat{\mathbb{1}} \otimes \hat{\mathbb{1}}$ in $\hat{X} \otimes \hat{X}$ to the constant 1-function $\mathbb{1}_\otimes = \mathbb{1} \otimes \mathbb{1}$ in $X \otimes X$ because we have

$$\widetilde{M}(\hat{\mathbb{1}} \otimes \hat{\mathbb{1}}^*) = M(\hat{\mathbb{1}} \otimes \hat{\mathbb{1}}) M^* = M\hat{\mathbb{1}} \otimes M\hat{\mathbb{1}} = \mathbb{1} \otimes \mathbb{1}.$$

Hence, $\widetilde{M}$ is a Markov operator. To see that it is deterministic, we use the representation of the adjoint operator $\widetilde{M}^*$ which is the second claim and proved below. Using that, we have $\widetilde{M}^*(e_i \otimes e_j) = M^* e_i \otimes M^* e_j$ which is again a pure state in $\hat{X}^* \otimes \hat{X}^*$

To compute the adjoint operator $\widetilde{M}^*$, we have for $\hat{b} \in L(\hat{X}^*, \hat{X})$ and $c \in L(X, X^*)$ that

$$\langle\!\langle \widetilde{M}\hat{b}, c \rangle\!\rangle = \mathrm{Tr}((M\hat{b}M^*)^* c) = \mathrm{Tr}(M\hat{b}^* M^* c) = \mathrm{Tr}(M^* c M \hat{b}^*) = \langle\!\langle M^* c M, \hat{b} \rangle\!\rangle,$$

where we used that the trace is invariant under commuting matrices. Hence, $\widetilde{M}^* c = M^* c M$.

For the last claim, we observe that $\hat{m} = \widetilde{M}^* m = M^* Q_\pi K M = Q_{\hat{\pi}} N K M = Q_{\hat{\pi}} \hat{K}$. ∎

Proposition 3.1 shows that the (dual) deterministic coarse-graining operator $\widetilde{M}^*$ maps the weights $m = Q_\pi K$ on the coarse-grained weight $\hat{m} = Q_{\hat{\pi}} \hat{K}$. This allows to define a reconstruction operator $\widetilde{N}$ as the inverse operator of $\widetilde{M}$ with respect to m as in Section 2. We define

$$\widetilde{N} : L(X^*, X) \to L(\hat{X}^*, \hat{X}), \quad \widetilde{N} = Q_{\hat{m}}^{-1} \widetilde{M}^* Q_m. \tag{3.1}$$

Again, we have that $\widetilde{N} \mathbb{1}_\otimes = \hat{\mathbb{1}}_\otimes$. Its adjoint (with respect to $\langle\!\langle \cdot, \cdot \rangle\!\rangle$) is given by

$$\widetilde{N}^* : L(\hat{X}, \hat{X}^*) \to L(X, X^*), \quad \widetilde{N}^* = Q_m \widetilde{M} Q_{\hat{m}}^{-1}.$$

Clearly, the operator $\widetilde{N}^*$ maps $\hat{m}$ to m. The definition of reconstruction operator $\widetilde{N}$ has two advantages. First, it allows to define a coarse-grained incidence matrix as we will see next. The incidence matrix will be crucial for estimating Poincaré-type constants in Section 5. Moreover, it can be used to reconstruct fluxes, which are functions on edges (see Section 4).

3.2. Coarse-graining of the incidence matrix

The canonical connection between X and $X \otimes X$, i.e., between vertices and edges, is given by the incidence matrix (or operator) $D : X \to L(X^*, X) \approx X \otimes X$. In coordinates, it is defined by

$$De_i = \sum_j (e_i \otimes e_j - e_j \otimes e_i) \in X \otimes X.$$

We remark that this definition distinguishes between outgoing and ingoing edges.

To define the adjoint operator, we fix the canonical basis in the dual space $e_k^* \in X^*$ with $\langle e_k^*, e_i \rangle = \delta_{ik}$, the Kronecker symbol. This relation also defines a basis $\{e_k^* \otimes e_l^*\}_{k,l}$ in the tensor space $X^* \otimes X^*$ such that it holds $\langle\!\langle e_k^* \otimes e_l^*, e_i \otimes e_j \rangle\!\rangle = \delta_{ijkl}$. The adjoint operator D^* is given by

$$\begin{aligned} D^* &: X^* \otimes X^* \to X^*, \\ D^*(e_i^* \otimes e_j^*)(e_l) &= \langle\!\langle De_l, e_i^* \otimes e_j^* \rangle\!\rangle = \sum_k \langle\!\langle e_l \otimes e_k - e_k \otimes e_l, e_i^* \otimes e_j^* \rangle\!\rangle \\ &= \sum_k \langle\!\langle e_l \otimes e_k - e_k \otimes e_l, e_i^* \otimes e_j^* \rangle\!\rangle = \begin{cases} 0 & l \neq i, l \neq j, \\ 1 & l = i, \\ -1 & l = j, \end{cases} \end{aligned} \tag{3.2}$$

whenever $i \neq j$, and otherwise it is zero.

The next result shows that the coarse-graining procedure is indeed consistent with the definition of the incidence matrix and the quotient graph.

Theorem 3.2. *Let $M : \hat{X} \to X$ be given as above, which is in local coordinates $M\hat{e}_k = \sum_{i \in \phi^{-1}(k)} e_i$. Let $m \in X^* \otimes X^*$ be arbitrary and fixed, and let $\widetilde{N}$ be defined by (3.1). The operator $\hat{D} : \hat{X} \to L(\hat{X}^*, \hat{X})$ defined by*

$$\hat{D} = \widetilde{N} D M : \hat{X} \to L(\hat{X}^*, \hat{X})$$

is an incidence matrix, i.e., we have

$$\hat{D}\hat{e}_k = \sum_l (\hat{e}_k \otimes \hat{e}_l - \hat{e}_l \otimes \hat{e}_k).$$

Moreover, it holds $\tilde{M}\hat{D} = DM$.

Remarkably, the form of $\hat{D}$ is independent of the $m \in X \otimes X$. We note that the second claim does not follow immediately from $\hat{D} = \tilde{N}DM$, which would imply $\tilde{M}\hat{D} = \tilde{P}DM$ with the projection $\tilde{P} = \tilde{M}\tilde{N}$ on $X \otimes X$. The relation $\tilde{M}\hat{D} = DM$ is finer and provides that the coarse-graining procedure is consistent with the definition of the quotient graph.

Proof. For the proof, we compute $\hat{D}\hat{e}_k$ explicitly. We have $M\hat{e}_k = \sum_{i\in\phi^{-1}(k)} e_i$, and hence,

$$DM\hat{e}_k = \sum_j (M\hat{e}_k \otimes e_j - e_j \otimes M\hat{e}_k) = \sum_j \sum_{i\in\phi^{-1}(k)} (e_i \otimes e_j - e_j \otimes e_i).$$

Moreover, we have that $\tilde{N} = Q_{\hat{m}}^{-1}\tilde{M}^* Q_m$. To evaluate $\tilde{N}(e_i \otimes e_j)$, we first observe that $\tilde{M}^*(e_i \otimes e_j) = \hat{e}_{\phi(i)} \otimes \hat{e}_{\phi(j)}$ which implies that

$$\begin{aligned}
\tilde{M}^* Q_m DM\hat{e}_k &= \sum_j \sum_{i\in\phi^{-1}(k)} (m_{ij}\tilde{M}^*(e_i \otimes e_j) - m_{ji}\tilde{M}^*(e_j \otimes e_i)) \\
&= \sum_j \sum_{i\in\phi^{-1}(k)} (m_{ij}\hat{e}_{\phi(i)} \otimes \hat{e}_{\phi(j)} - m_{ji}\hat{e}_{\phi(j)} \otimes \hat{e}_{\phi(i)}) \\
&= \sum_{l_j} \sum_{j\in\phi^{-1}(l_j)} \sum_{i\in\phi^{-1}(k)} (m_{ij}\hat{e}_{\phi(i)} \otimes \hat{e}_{\phi(j)} - m_{ji}\hat{e}_{\phi(j)} \otimes \hat{e}_{\phi(i)}) \\
&= \sum_{l_j} (\hat{m}_{kl_j}\hat{e}_k \otimes \hat{e}_{l_j} - \hat{m}_{l_j k}\hat{e}_{l_j} \otimes \hat{e}_k),
\end{aligned}$$

where we have used the definition of $\hat{m} = \tilde{M}^* m$ given by $\hat{m}_{kl} = \sum_{j\in\phi^{-1}(l)} \sum_{i\in\phi^{-1}(k)} m_{ij}$. Hence, we conclude that

$$\hat{D}\hat{e}_k = \tilde{N}DM\hat{e}_k = Q_{\hat{m}}^{-1}\tilde{M}^* Q_m DM\hat{e}_k = \sum_l \hat{e}_k \otimes \hat{e}_l - \hat{e}_l \otimes \hat{e}_k,$$

which is the desired formula.

Now, we prove $\tilde{M}\hat{D} = DM$ again by direct calculation. We have $\tilde{M}(\hat{e}_k \otimes \hat{e}_l) = M\hat{e}_k \otimes M\hat{e}_l$, and hence,

$$\tilde{M}\hat{D}\hat{e}_k = \sum_l M\hat{e}_k \otimes M\hat{e}_l - M\hat{e}_l \otimes M\hat{e}_k = \sum_l \sum_{j\in\phi^{-1}(l)} M\hat{e}_k \otimes e_j - e_j \otimes M\hat{e}_k = DM\hat{e}_k,$$

which we wanted to show. ∎

Remark 3.3. The coarse-graining procedure can also be applied to undirected graphs. Introducing the space $X \odot X \simeq L_{\mathrm{sym}}(X^*, X)$ containing the symmetric tensors (or matrices), the coarse-graining operator $\widetilde{M}$ respects the symmetric structure. Moreover, if $m = Q_\pi K$ is symmetric then also $\widetilde{N}$ maps into symmetric tensors. In the following, we will treat the case of symmetric m with more details but we will not consider undirected graphs.

4. Detailed balance Markov matrices

We now fix $m = Q_\pi K$. A special situation occurs if the operator K satisfies detailed balance meaning that $m_{ij} = \pi_i K_{ij} = \pi_j K_{ji} = m_{ji}$. We define the associated Markov generator A by $A = K - \mathrm{id}$ and investigate the following evolution system in X^*:

$$\dot{c} = A^* c,$$

where $c \in \mathrm{Prob}(\mathcal{Z})$ is understood as a probability distribution or a concentration. In the next lemma, we recall that assuming that K (or equivalently A) satisfies detailed balance, the system $\dot{c} = A^* c$ can be written as a gradient flow expressed via the incidence operator (see, e.g., [17, 21, 22]):

$$\begin{aligned} \dot{c} &= -D^* b, \\ b &= \frac{1}{2} Q_m D\rho, \\ \rho &= Q_\pi^{-1} c, \end{aligned}$$

where the first equation is a continuity equation between the fluxes b and the concentrations c, the last equation defines the relative densities ρ of c with respect to π and the second equation is the constitutive relation between the relative densities ρ and the fluxes b, which uses the tensor valued diagonal operator Q_m. Note that there is a factor $\frac{1}{2}$ because the incidence operator D counts every edge twice. The above continuity equation form is a structural decomposition of the generator A^* as the next lemma shows.

Lemma 4.1. *Let K (or equivalently A) satisfy the detailed balance condition. Then, $A^* = -\frac{1}{2} D^* Q_m D Q_\pi^{-1}$.*

Proof. By direct computation, we have for $\rho = \sum_i \rho_i e_i$ that

$$D\rho = \sum_i \rho_i D e_i = \sum_{i,j} \rho_i (e_i \otimes e_j - e_j \otimes e_i).$$

Hence, we get that $Q_m D\rho = \sum_{i,j} \rho_i m_{ij} e_i^* \otimes e_j^* - \rho_i m_{ji} e_j^* \otimes e_i^*$, which implies

$$-D^* Q_m D\rho = -\sum_{i,j} (\rho_i m_{ij} D^*(e_i^* \otimes e_j^*) + \rho_i m_{ji} D^*(e_j^* \otimes e_i^*)).$$

Evaluating both sides at e_l and using the explicit formula (3.2), we get that

$$\begin{aligned}(-D^*Q_m D\rho)e_l &= -\sum_{i,j}(\rho_i m_{ij} D^*(e_i^* \otimes e_j^*)e_l) + \sum_{i,j}(\rho_i m_{ji} D^*(e_j^* \otimes e_i^*)e_l)\\ &= -\sum_j \rho_l m_{lj} + \sum_i \rho_i m_{il} + \sum_i \rho_i m_{li} - \sum_j \rho_l m_{jl}\\ &= -\rho_l \sum_j (m_{lj} + m_{jl}) + \sum_i \rho_i (m_{il} + m_{li}).\end{aligned}$$

Using that $m_{il} = m_{li} = \pi_i A_{il} = \pi_l A_{li}$ and $\rho_i = c_i/\pi_i$, we get

$$\Big(-\frac{1}{2}D^*Q_m D\rho\Big)e_l = -c_l \sum_j m_{lj}/\pi_l + \sum_i c_i m_{il}/\pi_i = -c_l \sum_j A_{lj} + \sum_i c_i A_{il},$$

which implies that $A^*c = -\frac{1}{2}D^*Q_m DQ_\pi^{-1}c$ for all $c \in X$. ∎

4.1. Coarse-graining for detailed balance Markov operators

We are going to show that the above gradient flow decomposition is also consistent with the coarse-graining procedure. Crucial for that result is the relation $DM = \widetilde{M}\widehat{D}$.

Theorem 4.2. *Let us assume that there is an equilibration of the concentrations $c = N^*\hat{c}$. Then, with the above notation, the coarse-grained concentrations $\hat{c}$ solve the coarse-grained evolution equation of the form*

$$\begin{aligned}\dot{\hat{c}} &= -\widehat{D}^*\hat{b},\\ \hat{b} &= \frac{1}{2}Q_{\hat{m}}\widehat{D}\hat{\rho},\\ \hat{\rho} &= Q_{\hat{\pi}}^{-1}\hat{c},\end{aligned}$$

which is again a gradient flow of continuity equation form. In particular, we have an equilibration of the fluxes $b = \widetilde{N}^*\hat{b}$.

Proof. Since $M^*N^* = \mathrm{id}_{X^*}$, we observe that $M^*c = \hat{c}$. By Theorem 3.2, we have that $DM = \widetilde{M}\widehat{D}$. Hence, we obtain

$$\dot{\hat{c}} = M^*\dot{c} = -M^*D^*b = -\widehat{D}^*\widetilde{M}^*b.$$

Defining $\hat{b} := \widetilde{M}^*b$ (which implies that $b = \widetilde{N}^*\hat{b}$) and using that $\widetilde{M}^*Q_m = Q_{\hat{m}}\widetilde{N}$ and $Q_\pi^{-1}N^* = MQ_{\hat{\pi}}^{-1}$, we get

$$\hat{b} = \widetilde{M}^*b = \frac{1}{2}\widetilde{M}^*Q_m DQ_\pi^{-1}N^*\hat{c} = \frac{1}{2}Q_{\hat{m}}\widetilde{N}DMQ_{\hat{\pi}}^{-1}\hat{c} = \frac{1}{2}Q_{\hat{m}}\widehat{D}\hat{\rho},$$

where we have introduced the coarse-grained relative density $\hat{\rho} = Q_{\hat{\pi}}^{-1}\hat{c}$. ∎

The above theorem can be understood as a structural refinement of the form derived in Section 2.3, which just says that the coarse-grained generator is given by $\widehat{A}^* = M^*A^*N^*$.

4.2. Flux reconstruction

Theorem 4.2 provides that an equilibration of the fluxes necessarily occurs if concentrations equilibrate. In practice, often the converse question arises, namely, how fluxes on the large graph can be reconstructed out of the coarse-grained quantities like the coarse-grained concentrations and coarse-grained fluxes. As one would guess $b = \widetilde{N}^* \hat{b}$ is not the desired flux because additional fluxes between coarse-grained states are needed.

The next proposition provides an affirmative answer and shows that for a given pair $(\hat{c}, \hat{b})$ satisfying the continuity equation $\dot{\hat{c}} = -\widehat{D}^* \hat{b}$, there is a microscopic flux b such that $c = N^* \hat{c}$ solves $\dot{c} = -D^* b$ and the microscopic flux consists of two parts $b = b_1 + b_2$ such that $b_1 = \widetilde{N}^* \hat{b}$ and $b_2 \in \mathrm{Ker} \widehat{D}^* \widetilde{M}^*$. Moreover, b depends linearly on $\hat{b}$.

Proposition 4.3. *Let $(\hat{c}, \hat{b})$ satisfy the continuity equation $\dot{\hat{c}} = -\widehat{D}^* \hat{b}$. Define the microscopic concentrations by $c = N^* \hat{c}$. Then, there is a microscopic flux b, which solves the microscopic continuity equation $\dot{c} = -D^* b$ and b is given by $b = b_1 + b_2$ with $b_1 = \widetilde{N}^* \hat{b}$ and $b_2 \in \mathrm{Ker} \widehat{D}^* \widetilde{M}^*$, which depends linearly on $\hat{b}$.*

Proof. Let $b_1 = \widetilde{N}^* \hat{b}$. Then, we have to construct $b_2 \in X^* \otimes X^*$ such that

$$\dot{c} = -D^*(b_1 + b_2) = -D^*(\widetilde{N}^* \hat{b} + b_2).$$

Since $c = N^* \hat{c}$, which implies $\dot{c} = N^* \dot{\hat{c}} = -N^* \widehat{D}^* \hat{b}$, we get that b_2 has to satisfy the linear equation

$$D^* b_2 = (N^* \widehat{D}^* - D^* \widetilde{N}^*) \hat{b} =: x^* \in X^*. \tag{4.1}$$

Using Fredholm's alternative, this equation is solvable if for all $x \in \mathrm{Ker}(D)$ we have that $\langle x, x^* \rangle = 0$.

So, let $x \in \mathrm{Ker}(D) \subset X$. Hence,

$$\langle x, x^* \rangle = \langle x, (N^* \widehat{D}^* - D^* \widetilde{N}^*) \hat{b} \rangle = \langle \widehat{D} N x, \hat{b} \rangle.$$

Using Proposition 2.3, we decompose $x \in X = \mathrm{Ker}(N) + \mathrm{Range}(M)$. Hence, for proving that $\langle x, x^* \rangle = 0$, we may assume that $x \in \mathrm{Range}(M)$, i.e., $M \hat{x} = x$. Using that $\hat{b} = \widetilde{M}^* b_1$ and that $NM = \mathrm{id}_{\widehat{X}}$, we compute

$$\langle x, x^* \rangle = \langle \widehat{D} N M \hat{x}, \widetilde{M}^* b_1 \rangle = \langle \widetilde{M} \widehat{D} \hat{x}, b_1 \rangle = \langle DM \hat{x}, b_1 \rangle = \langle Dx, b_1 \rangle = 0,$$

where we have used that $\widetilde{M} \widehat{D} = DM$. Hence, there exists b_2 such that $D^* b_2 = (N^* \widehat{D}^* - D^* \widetilde{N}^*) \hat{b}$, which implies that $b = b_1 + b_2$ solves $\dot{c} = -D^* b$. By (4.1), we see that b_2 depends linearly on $\hat{b}$.

To see that $b_2 \in \mathrm{Ker}(\widehat{D}^* \widetilde{M}^*)$, we observe that

$$\begin{aligned} \widehat{D}^* \widetilde{M}^* b_2 &= M^* D^* b_2 = M^* (N^* \widehat{D}^* - D^* \widetilde{N}^*) \hat{b} \\ &= (M^* N^* M^* D^* \widetilde{N}^* - M^* D^* \widetilde{N}^*) \hat{b} = 0. \end{aligned}$$

■

Remark 4.4. We note that the existence of b_2 as well as the linear dependence on $\hat{b}$ has been used in [28] to coarse-grain fast-slow linear reaction-diffusion systems.

5. Functional inequalities and Poincaré constants

Computing the first non-trivial eigenvalue of a given Markov matrix is a challenging analytic as well as numerical task, see, e.g., [4, 10]. Although eigenvalues and eigenfunctions do not fit into the Markov theory with its $X - X^*$ duality and positivity (they might even be complex), the coarse-graining procedure can be used to derive estimates for associated functionals. With the help of functional inequalities, insights into spectral properties can be obtained, see, e.g., [2].

In this section, we apply the coarse-graining procedure to derive estimates between functionals on X and $\widehat{X}$. First, we observe the following for the expectations:

$$\mathbb{E}_\pi(x) := \sum_{i\in\mathcal{Z}} \pi_i x_i = \langle x, \pi\rangle, \quad \mathbb{E}_{\hat\pi}(\hat x) := \sum_{j\in\hat{\mathcal{Z}}} \hat\pi_j \hat x_j = \langle \hat x, \hat\pi\rangle.$$

Lemma 5.1. *If $\hat x = Nx$ or $x = M\hat x$, then $\mathbb{E}_\pi(x) = \mathbb{E}_{\hat\pi}(\hat x)$.*

Proof. If $\hat x = Nx$, we have $\mathbb{E}_{\hat\pi}(\hat x) = \langle \hat\pi, \hat x\rangle = \langle \hat\pi, Nx\rangle = \langle N^*\hat\pi, x\rangle = \langle \pi, x\rangle = \mathbb{E}_\pi(x)$. If $x = M\hat x$, we conclude that $Nx = \hat x$. ∎

5.1. Coarse-graining for energy functionals

For any strictly convex and non-negative function $\Phi : \mathbb{R} \to [0, \infty[$, we define the associated energy functional on X by

$$\mathcal{E}_\Phi(x) = \mathbb{E}_\pi \Phi(x) - \Phi(\mathbb{E}_\pi x) = \langle \Phi(x), \pi\rangle - \Phi(\langle x, \pi\rangle),$$

where for a vector $x \in X \simeq \mathbb{R}^n$ the function $\Phi(x) \in X$ is defined componentwise, i.e., $\Phi(x)_i = \Phi(x_i)$. Note, that the measure $\pi \in X^*$ is fixed and implicitly given in the definition of $\mathcal{E}_\Phi$.

Lemma 5.2. *The functional $\mathcal{E}_\Phi$ is non-negative and its minimum is attained on constant vectors. Moreover, we have for the rescaled function $\widetilde\Phi(r) = \Phi(r) + cr + d$ that $\mathcal{E}_{\widetilde\Phi} = \mathcal{E}_\Phi$.*

Proof. The first claim follows directly be Jensen's inequality since Φ is strictly convex. Moreover, a direct computation shows that

$$\begin{aligned}\mathcal{E}_{\widetilde\Phi}(x) &= \mathbb{E}_\pi \widetilde\Phi(x) - \widetilde\Phi(\mathbb{E}_\pi x) = \langle \Phi(x) + cx + d, \pi\rangle - \Phi(\mathbb{E}_\pi x) - c\mathbb{E}_\pi x - d = \\ &= \langle \Phi(x), \pi\rangle + c\langle x, \pi\rangle + d\langle \mathbb{1}, \pi\rangle - \Phi(\mathbb{E}_\pi x) - c\mathbb{E}_\pi x - d = \mathcal{E}_\Phi(x).\end{aligned}$$ ∎

Typical examples for the function Φ are the following.

(1) $\Phi(r) = \frac{1}{2}r^2$. Then, $\mathcal{E}_\Phi(x) = \frac{1}{2}(\langle x^2, \pi\rangle - \langle x, \pi\rangle^2)$ corresponds to the quadratic energy or statistical variance.

(2) $\Phi(r) = r\log r - r + 1$. Then, $\mathcal{E}_\Phi$ corresponds to the entropy or free energy of Boltzmann type, which will be denoted by Ent_π in the following.

Remark 5.3. There are several remarks in order.

(1) Often functionals of the form $\mathcal{E}^*(p) = \langle \Psi(p/\pi), \pi \rangle$ as relative energies or entropies are considered. In contrast to $\mathcal{E}_\Phi$, which is defined on X, functionals of the latter form are defined on probability vectors as elements of the dual space X^*. However, they are related via the Legendre transform. To see this, forgetting about the normalization term $-\Phi(\mathbb{E}_\pi x)$, the Legendre transform of the functional $\widetilde{\mathcal{E}_\Phi}$, $\widetilde{\mathcal{E}_\Phi}(x) := \langle \Phi(x), \pi \rangle$ is given by

$$\widetilde{\mathcal{E}_\Phi}^*(p) = \sup_{x \in X} (\langle p, x \rangle - \widetilde{\mathcal{E}_\Phi}(x)) = \sup_{x \in X} (\langle p, x \rangle - \langle \Phi(x), \pi \rangle).$$

Introducing the relative density g of p with respect to the positive probability vector π, we get that

$$\begin{aligned} \widetilde{\mathcal{E}_\Phi}^*(p) &= \sup_{x \in X} (\langle g\pi, x \rangle - \langle \Phi(x), \pi \rangle) = \sup_{x \in X} (\langle g \cdot x, \pi \rangle - \langle \Phi(x), \pi \rangle) \\ &= \sup_{x \in X} \langle g \cdot x - \Phi(x), \pi \rangle = \langle \sup_{x \in X} (g \cdot x - \Phi(x)), \pi \rangle \\ &= \langle \Phi^*(g), \pi \rangle = \left\langle \Phi^*\left(\frac{p}{\pi}\right), \pi \right\rangle, \end{aligned}$$

which is exactly the desired form.

(2) We could also investigate functionals of the form $\widetilde{\mathcal{E}}_\Phi(x) = \langle \Phi(x - \mathbb{E}_\pi x), \pi \rangle$ as a generalization of the variance $\Phi(r) = r^2$. These functionals have the property that they are always convex. However, we will restrict to the above form.

Analogously, we define $\widehat{\mathcal{E}}_\Phi$ on $\widehat{X}$ by replacing x by $\hat{x}$ and π by $\hat{\pi}$. The functionals $\mathcal{E}_\Phi$ on X and $\widehat{\mathcal{E}}_\Phi$ on $\widehat{X}$ can be estimated as follows.

Proposition 5.4. *We have the following relation for the functionals regarding coarse-graining and reconstruction.*

(1) *For all $\hat{x} \in \widehat{X}$, $x = M\hat{x}$ implies $\mathcal{E}_\Phi(x) = \widehat{\mathcal{E}}_\Phi(\hat{x})$. This holds even for all functions $\Phi : \mathbb{R} \to \mathbb{R}$ not necessarily convex.*

(2) *For all $x \in X$, $\hat{x} = Nx$ implies $\widehat{\mathcal{E}}_\Phi(\hat{x}) \leq \mathcal{E}_\Phi(x)$.*

Proof. For the first claim, take any $\hat{x} \in \widehat{X}$. Then, we have with Lemma 5.1 that

$$\begin{aligned} \mathcal{E}_\Phi(x) = \mathcal{E}_\Phi(M\hat{x}) &= \sum_i \pi_i \Phi((M\hat{x})_i) - \Phi(\mathbb{E}_\pi(M\hat{x})) = \sum_{j \in \hat{Z}} \sum_{i = \phi^{-1}(j)} \pi_i \Phi((M\hat{x})_i) - \mathbb{E}_{\hat{\pi}} \hat{x} \\ &= \sum_{j \in \hat{Z}} \sum_{i = \phi^{-1}(j)} \pi_i \Phi(\hat{x}_{\phi(i)}) - \mathbb{E}_{\hat{\pi}} \hat{x} = \sum_{j \in \hat{Z}} \Phi(\hat{x}_j) \sum_{i = \phi^{-1}(j)} \pi_i - \mathbb{E}_{\hat{\pi}} \hat{x} \\ &= \sum_{j \in \hat{Z}} \hat{\pi}_j \Phi(\hat{x}_j) - \mathbb{E}_{\hat{\pi}} \hat{x} = \widehat{\mathcal{E}}_\Phi(\hat{x}), \end{aligned}$$

where we used that $\hat{\pi}_j = \sum_{i = \phi^{-1}(j)} \pi_i$.

For the second claim, take any $x \in X$. Using Jensen's inequality for the convex function Φ, which means that we have the pointwise inequality $\Phi(Nx) \leq N\Phi(x)$, we obtain

$$\begin{aligned}\widehat{\mathcal{E}}_\Phi(Nx) &= \langle \hat{\pi}, \Phi(Nx)\rangle - \mathbb{E}_{\hat{\pi}}(Nx) \leq \langle \hat{\pi}, N\Phi(x)\rangle - \mathbb{E}_\pi(x) \\ &= \langle N^*\hat{\pi}, \Phi(x)\rangle - \mathbb{E}_\pi(x) = \langle \pi, \Phi(x)\rangle - \mathbb{E}_\pi(x) = \mathcal{E}_\pi(x).\end{aligned}$$

∎

The log-Sobolev constant is defined via $\mathcal{E}_\Phi(x^2)$ where $\Phi(r) = r\log r$. From the above proposition it is not clear that is possible to obtain estimates between $\mathrm{Ent}_{\hat{\pi}}((\hat{x})^2)$ and $\mathrm{Ent}_\pi((M\hat{x})^2)$. In fact, we prove that this is possible even for general convex functions not necessarily quadratic.

Proposition 5.5. *Let $g : \mathbb{R} \to [0, \infty[$ be convex and satisfy $g(x) > 0$ if $x \neq 0$ and $g(0) = 0$. Then, we have*

$$\forall \hat{x} \in \widehat{X} : \mathrm{Ent}_{\hat{\pi}}(g(\hat{x})) \leq \mathrm{Ent}_\pi(g(M\hat{x})),$$

where again with a small abuse of notation $g(x)$ is meant component-wise, i.e., $g(x)_i = g(x_i)$.

Proof. The proof is done in two steps. First, we shift the function $\Phi(r) = r\log r$ to incorporate g. Secondly, we derive the estimate.

Step 1. Clearly, we have equality for $\hat{x} = 0$. So, let us take $\hat{x} \neq 0$. Then, $g(\hat{x}) > 0$ and because $\pi > 0$, there is a constant $C > 0$ such that $\langle g(M\hat{x}), \pi\rangle \geq C > 0$. Let us define $c > 0$ by $c := \mathrm{e}^{-(1+C)}$. We define $\widetilde{\Phi}(r) = r\log r + cr$, which has its minimum at $r = C$. So, we have $\widetilde{\Phi}(r_1) \geq \widetilde{\Phi}(r_2)$ for $r_1 \geq r_2 \geq C$. Recalling that the energy functional is invariant under affine shifts (Lemma 5.2), we have that $\mathrm{Ent}_\pi(x) = \mathcal{E}_{\widetilde{\Phi}}(x)$ and analogously also for the coarse-grained states $\hat{x}$. So, we are going to show $\hat{\mathcal{E}}_{\widetilde{\Phi}}(g(\hat{x})) \leq \mathcal{E}_{\widetilde{\Phi}}(g(M\hat{x}))$, or, equivalently,

$$\langle \widetilde{\Phi}(g(\hat{x})), \hat{\pi}\rangle - \widetilde{\Phi}(\langle g(\hat{x}), \hat{\pi}\rangle) \leq \langle \widetilde{\Phi}(g(M\hat{x})), \pi\rangle - \widetilde{\Phi}(\langle g(M\hat{x}), \pi\rangle).$$

Step 2. We observe that (completely similar to the proof of Proposition 5.4)

$$\begin{aligned}\langle \widetilde{\Phi}(g(M\hat{x})), \pi\rangle &= \sum_i \widetilde{\Phi}(g(M\hat{x}))_i \pi_i \\ &= \sum_i \widetilde{\Phi}(g(M\hat{x})_i)\pi_i = \sum_{j\in\widehat{Z}} \sum_{i=\phi^{-1}(j)} \pi_i \widetilde{\Phi} \circ g(\hat{x}_{\phi(i)}) \\ &= \sum_{j\in\widehat{Z}} \hat{\pi}_j \widetilde{\Phi} \circ g(\hat{x}_j) = \langle \widetilde{\Phi}(g(\hat{x})), \hat{\pi}\rangle.\end{aligned}$$

Hence, it suffices to prove that $\widetilde{\Phi}(\langle g(\hat{x}), \hat{\pi}\rangle) \geq \widetilde{\Phi}(\langle g(M\hat{x}), \pi\rangle)$. To see this, we use Jensen's inequality which states that $g(M\hat{x}) \leq Mg(\hat{x})$ for all Markov matrices M and convex functions g. Hence, we get $\langle g(M\hat{x}), \pi\rangle \leq \langle Mg(\hat{x}), \pi\rangle = \langle g(\hat{x}), M^*\pi\rangle = \langle g(\hat{x}), \hat{\pi}\rangle$. Since $\langle g(M\hat{x}), \pi\rangle \geq C$ and the function $\widetilde{\Phi}$ is monotone for arguments larger than C by construction, we conclude $\widetilde{\Phi}(\langle g(M\hat{x}), \pi\rangle) \leq \widetilde{\Phi}(\langle g(\hat{x}), \hat{\pi}\rangle)$. Hence, the claim is proved. ∎

5.2. Dirichlet forms and Poincaré-type estimates

To estimate Poincaré-type constants, we introduce the Dirichlet form (or dissipation) for K and $\widehat{K}$ by

$$\mathcal{D}_K(x) = \frac{1}{2}\sum_{i,j} \pi_i K_{ij}(x_i - x_j)^2, \quad \mathcal{D}_{\widehat{K}}(x) = \frac{1}{2}\sum_{i,j} \hat{\pi}_i \widehat{K}_{ij}(x_i - x_j)^2.$$

Without loss of generality, we assume that K and $\widehat{K}$ satisfy detailed balance, because the Dirichlet form takes into account only the symmetric part of $Q_m K$. Using $m = Q_\pi K$ and $\hat{m} = Q_{\hat{\pi}}\widehat{K}$, the Dirichlet form is related to the generator $A = K - \mathrm{id}$ by

$$\begin{aligned}\mathcal{D}_K(x) &= \frac{1}{2}\langle\!\langle Dx, Q_m Dx\rangle\!\rangle = \frac{1}{2}\langle x, D^* Q_m Dx\rangle \\ &= -\langle x, A^* Q_\pi x\rangle = -\langle Ax, Q_\pi x\rangle = -\langle x \cdot Ax, \pi\rangle.\end{aligned}$$

Moreover, we have

$$\mathcal{D}_{\widehat{K}}(\hat{x}) = \frac{1}{2}\langle\!\langle \widehat{D}\hat{x}, Q_{\hat{m}}\widehat{D}\hat{x}\rangle\!\rangle = -\langle \hat{x}\cdot\widehat{A}\hat{x}, \hat{\pi}\rangle.$$

We are interested in estimating the spectral gap $\lambda = \lambda(K, \Phi)$, which is defined by the largest constant $c > 0$ that satisfies the discrete Poincaré-type inequality

$$\mathcal{D}_K(x) \geq c\mathcal{E}_\Phi(x), \quad \text{i.e.,} \quad \lambda(K,\Phi) = \inf\left\{\frac{\mathcal{D}_K(x)}{\mathcal{E}_\Phi(x)} : \forall x\ \mathcal{E}_\Phi(x) \neq 0\right\}.$$

Analogously, we define

$$\hat{\lambda} = \hat{\lambda}(\widehat{K}, \Phi) = \inf\left\{\frac{\mathcal{D}_{\widehat{K}}(\hat{x})}{\widehat{\mathcal{E}}_\Phi(\hat{x})} : \forall \hat{x}\ \widehat{\mathcal{E}}_\Phi(\hat{x}) \neq 0\right\}.$$

We also define the log-Sobolev constants

$$\lambda_{g,\mathrm{LS}} = \inf\left\{\frac{\mathcal{D}_K(x)}{\mathrm{Ent}_\pi(g(x))} : \mathrm{Ent}_\pi(g(x)) \neq 0\right\}$$

and analogously $\hat{\lambda}_{g,\mathrm{LS}}$.

The next theorem relates the Poincaré-type constants to their coarse-grained counter part. In particular, it states that coarse-graining always increases the log-Sobolev constants.

Theorem 5.6. *With the above notation, we have the following:*

(1) *for all $\hat{x} \in \widehat{X}$, we have that $\mathcal{D}_K(M\hat{x}) = \mathcal{D}_{\widehat{K}}(\hat{x})$,*

(2) *for all functions Φ, we have that $\lambda(K,\Phi) \leq \lambda(\widehat{K},\Phi)$,*

(3) *we have for the log-Sobolev constants that $\lambda_{g,\mathrm{LS}} \leq \hat{\lambda}_{g,\mathrm{LS}}$.*

Proof. We have

$$\mathcal{D}_K(M\hat{x}) = \frac{1}{2}\langle\!\langle DM\hat{x}, Q_m DM\hat{x}\rangle\!\rangle = -\langle M\hat{x}, A^*Q_\pi M\hat{x}\rangle$$
$$= -\langle \hat{x}, M^*A^*N^*Q_{\hat{\pi}}\hat{x}\rangle = -\langle \hat{x}, \widehat{A}^*Q_{\hat{\pi}}\hat{x}\rangle = \frac{1}{2}\langle \hat{x}, \widehat{D}^*Q_{\hat{m}}D\hat{x}\rangle = \mathcal{D}_{\widehat{K}}(\hat{x}),$$

which is the first claim. Hence, we obtain the following relations for the functional inequalities:

$$\lambda = \inf\left\{\frac{\mathcal{D}_K(x)}{\mathcal{E}_\Phi(x)} : \mathcal{E}_\Phi(x) \neq 0\right\} \leq \inf\left\{\frac{\mathcal{D}_K(M\hat{x})}{\mathcal{E}_\Phi(M\hat{x})} : \mathcal{E}_\Phi(M\hat{x}) \neq 0\right\}$$
$$= \inf\left\{\frac{\mathcal{D}_{\widehat{K}}(\hat{x})}{\mathcal{E}_\Phi(M\hat{x})} : \mathcal{E}_\Phi(M\hat{x}) \neq 0\right\}.$$

For the second claim, we use Proposition 5.4 which implies that $\mathcal{E}_\Phi(M\hat{x}) = \widehat{\mathcal{E}}_\Phi(\hat{x})$, and hence, $\lambda \leq \hat{\lambda}$. For the third claim, we use $\mathrm{Ent}_{\hat{\pi}}(g(\hat{x})) \leq \mathrm{Ent}_\pi(g(M\hat{x}))$ by Proposition 5.5 to obtain the bound $\lambda_{g,\mathrm{LS}} \leq \hat{\lambda}_{g,\mathrm{LS}}$. ∎

The above theorem answers the question how Poincaré-type constants behave under coarse-graining. So, the natural question arising is what happens by the inverse reconstruction procedure. Although Proposition 5.4 states that we have $\mathcal{E}_\pi(x) \geq \mathcal{E}_{\hat{\pi}}(Nx)$ for all $x \in X$, it is not possible to obtain uniform estimates between and $\mathcal{D}_K(x)$ and $\mathcal{D}_{\widehat{K}}(Nx)$ as the counterexample in the next section shows.

5.3. Counterexample

We compute and compare $\mathcal{D}_{\widehat{K}}(Nx)$ and $\mathcal{D}_K(x)$ for a fixed $x \in X$. We have

$$\mathcal{D}_{\widehat{K}}(Nx) = \frac{1}{2}\langle\!\langle \widehat{D}Nx, Q_{\hat{m}}\widehat{D}Nx\rangle\!\rangle = \frac{1}{2}\langle x, N^*\widehat{D}^*Q_{\hat{m}}\widehat{D}Nx\rangle = -\langle x, P^*A^*Q_\pi Px\rangle =$$
$$= -\langle Px, Q_\pi APx\rangle = -\langle Px \cdot APx, \pi\rangle.$$

Recall, that we have $\mathcal{D}_K(x) = -\langle x \cdot Ax, \pi\rangle$. It is clear that $\mathcal{D}_K(x) \geq \mathcal{D}_{\widehat{K}}(Nx)$ holds for all $x \in \mathrm{Range}(P)$ (because then the inequality is a trivial equality) and also for all $x \in \mathrm{Range}(\mathrm{id} - P) = \mathrm{Ker}(P)$ (because $\mathcal{D}_{\widehat{K}}(Nx) = 0$). In particular, we always have $\mathcal{D}_K(x) \geq \mathcal{D}_{\widehat{K}}(Nx)$ in the simple case of $Z = \{1, 2\}$.

For $a \geq 0$, we define on $\mathbb{R}^3$ the parameter dependent Markov generator

$$A_a = \begin{pmatrix} -8 & 4 & 4 \\ 1 & -2 & 1 \\ a & a & -2a \end{pmatrix}.$$

Then, A_a^* has the stationary measure $\pi_a = \frac{1}{5a+4}(a, 4a, 4)^{\mathrm{T}}$. One easily checks that A_a satisfies detailed balance with respect to π_a. As in the example from Section 2.2, we

define the coarse-graining function $\phi : Z \to \hat{Z}$ with $\phi(1) = \hat{1}$ and $\phi(2) = \phi(3) = \hat{2}$. The corresponding Markov operator $M : \hat{X} \to X$ is given by $M = \begin{pmatrix} 1 & & \\ & 1 & \\ & 1 & \end{pmatrix}$. The coarse-grained stationary measure is given by $\hat{\pi}_a = M^* \pi_a = \frac{1}{5a+4}(a, 4a+4)^{\mathrm{T}}$.The inverse operator $N_a : X \to \hat{X}$ and the projection $P_a : X \to X$ are given by

$$N_a = Q_{\hat{\pi}_a}^{-1} M^* Q_{\pi_a} = \begin{pmatrix} 1 & & \\ & \frac{a}{a+1} & \frac{1}{a+1} \end{pmatrix}, \quad P_a = M N_a = \begin{pmatrix} 1 & & \\ & \frac{a}{a+1} & \frac{1}{a+1} \\ & \frac{a}{a+1} & \frac{1}{a+1} \end{pmatrix}.$$

We compute $\mathcal{D}_{K_a}(x)$ and $\mathcal{D}_{\widehat{K}_a}(N_a x)$ for $x = (3, 1, 2)^{\mathrm{T}}$. We have $-\langle x \cdot A_a x, \pi_a \rangle = \frac{24a}{5a+4}$. Moreover, we have $-\langle P_a x \cdot A_a P_a x, \pi_a \rangle = \frac{8a(1+2a)^2}{(a+1)^2(5a+4)}$. Hence, we have that $\mathcal{D}_{K_a}(x) \geq \mathcal{D}_{\widehat{K}_a}(N_a x)$ is equivalent to

$$\frac{24a}{5a+4} \geq \frac{8a(1+2a)^2}{(a+1)^2(5a+4)} \Leftrightarrow (1+2a)^2 \leq 3(a+1)^2 \Leftrightarrow a \leq 1+\sqrt{3} =: a_*.$$

In particular, we have for $x = (3, 1, 2)^{\mathrm{T}}$ that $\mathcal{D}_{K_a}(x) \geq \mathcal{D}_{\widehat{K}_a}(N_a x)$ for $a \in [0, a^*]$ and that $\mathcal{D}_{K_a}(x) \leq \mathcal{D}_{\widehat{K}_a}(N_a x)$ for $a \in [a_*, \infty[$. Summarizing, it is not possible to have uniform estimates between $\mathcal{D}_{K_a}$ and $\mathcal{D}_{\widehat{K}_a}(N_a \cdot)$. Hence, no inequality for the Poincaré-type constants can be expected.

Funding. This work was partially supported by the DFG via SFB 1114 (Project no. 235221301, subproject C05).

References

[1] A. M. S. Barreto and M. D. Fragoso, Lumping the states of a finte Markov chain through stochastic factorization. In *Proceedings of the 18th World congress "the internaction federation of automatic control", Milano (Italy), August 28–September 2*, pp. 4206–4211, 44, Elsevier, 2011

[2] S. G. Bobkov and P. Tetali, Modified logarithmic Sobolev inequalities in discrete settings. *J. Theoret. Probab.* **19** (2006), no. 2, 289–336 Zbl 1113.60072 MR 2283379

[3] B. Bollobás, *Modern graph theory*. Grad. Texts in Math. 184, Springer, New York, 1998 Zbl 0902.05016 MR 1633290

[4] G.-Y. Chen, W.-W. Liu, and L. Saloff-Coste, The logarithmic Sobolev constant of some finite Markov chains. *Ann. Fac. Sci. Toulouse Math. (6)* **17** (2008), no. 2, 239–290 MR 2487855

[5] X. Cheng and J. M. A. Scherpen, Clustering-based model reduction of Laplacian dynamics with weakly connected topology. *IEEE Trans. Automat. Control* **65** (2020), no. 10, 4393–4399 Zbl 1536.93107 MR 4159136

[6] X. Cheng, L. Yu, D. Ren, and J. M. A. Scherpen, Reduced order modeling of diffusively coupled network systems: an optimal edge weighting approach. *IEEE Trans. Automat. Control* **68** (2023), no. 7, 4233–4240 Zbl 1548.93021 MR 4614303

[7] H. De Sterck, T. A. Manteuffel, S. F. McCormick, K. Miller, J. Pearson, J. Ruge, and G. Sanders, Smoothed aggregation multigrid for Markov chains. *SIAM J. Sci. Comput.* **32** (2010), no. 1, 40–61 Zbl 1209.65011 MR 2599766

[8] M. Erbar and J. Maas, Ricci curvature of finite Markov chains via convexity of the entropy. *Arch. Ration. Mech. Anal.* **206** (2012), no. 3, 997–1038 Zbl 1256.53028 MR 2989449

[9] M. Fathi and Y. Shu, Curvature and transport inequalities for Markov chains in discrete spaces. *Bernoulli* **24** (2018), no. 1, 672–698 Zbl 1396.60084 MR 3706773

[10] O. Faust and H. Fawzi, Sum-of-squares proofs of logarithmic Sobolev inequalities on finite Markov chains. *IEEE Trans. Inform. Theory* **70** (2024), no. 2, 803–819 Zbl 1547.60067 MR 4703690

[11] B. Hilder and U. Sharma, Quantitative coarse-graining of Markov chains. *SIAM J. Math. Anal.* **56** (2024), no. 1, 913–954 Zbl 1532.60157 MR 4695757

[12] O. Johnson, A discrete log-Sobolev inequality under a Bakry–émery type condition. *Ann. Inst. Henri Poincaré Probab. Stat.* **53** (2017), no. 4, 1952–1970 Zbl 1390.39077 MR 3729642

[13] J. G. Kemeny and J. L. Snell, *Finite Markov chains*. Undergrad. Texts Math., Springer, New York-Heidelberg, 1976 Zbl 0328.60035 MR 0410929

[14] C. Landim, D. Marcondes, and I. Seo, A resolvent approach to metastability. *J. Eur. Math. Soc. (JEMS)* **27** (2025), no. 4, 1563–1618 Zbl 08012230 MR 4875612

[15] F. Legoll and T. Lelièvre, Effective dynamics using conditional expectations. *Nonlinearity* **23** (2010), no. 9, 2131–2163 MR 2672640

[16] I. Marek and P. Mayer, Aggregation/disaggregation iterative methods applied to Leontev systems and Markov chains. *Appl. Math.* **47** (2002), no. 2, 139–156 Zbl 1090.15500 MR 1894666

[17] A. Mielke and A. Stephan, Coarse-graining via EDP-convergence for linear fast-slow reaction systems. *Math. Models Methods Appl. Sci.* **30** (2020), no. 9, 1765–1807 Zbl 1454.60113 MR 4151795

[18] J. S. Niedbalski, K. Deng, P. G. Mehta, and S. Meyn, Model reduction for reduced order estimation in traffic models. In *American control conference, Washington (USA) June 11–13*, IEEE, Piscataway, NJ, 2008

[19] J. R. Norris, *Markov chains*. Camb. Ser. Stat. Probab. Math. 2, Cambridge University Press, Cambridge, 1998 Zbl 0873.60043 MR 1600720

[20] S. Peles, B. Munsky, and M. Khammash, Reduction and solution of the chemical master equation usding time scale separation and finite state projection. *J. Chem. Phys.* **125** (2006), article no. 204104

[21] M. A. Peletier, Variational modelling: Energies, gradient flows, and large deviations. 2014, arXiv:1402.1990v1

[22] M. A. Peletier and A. Schlichting, Cosh gradient systems and tilting. *Nonlinear Anal.* **231** (2023), article no. 113094 Zbl 1512.35003 MR 4575702

[23] W. H. A. Schilders, H. A. van der Vorst, and J. Rommes (eds.), *Model order reduction: theory, research aspects and applications*. Math. Ind. 13, Springer, Berlin, 2008 MR 2516399

[24] A. Schlichting and M. Slowik, Poincaré and logarithmic Sobolev constants for metastable Markov chains via capacitary inequalities. *Ann. Appl. Probab.* **29** (2019), no. 6, 3438–3488 Zbl 1432.60070 MR 4047985

[25] P. J. Schweitzer, Aggregation methods for large Markov chains. In *Mathematical computer performance and reliability (Pisa, 1983)*, pp. 275–286, North-Holland, Amsterdam, 1984 MR 0785970

[26] T. Shioyama and K. Tanaka, A new aggregation-disaggregation algorithm. *Eur. J. Oper. Res.* **83** (1995), no. 3, 655–669 Zbl 0899.90088

[27] A. Stephan, On EDP-convergence for gradient systems with different time scales. *Proc. Appl. Math. Mech.* **19** (2019), no. 1, article no. e201900366

[28] A. Stephan, EDP-convergence for a linear reaction-diffusion system with fast reversible reaction. *Calc. Var. Partial Differential Equations* **60** (2021), no. 6, article no. 226 Zbl 1479.35053 MR 4314142

[29] H. Stephan, Inequalities for Markov operators, majorization and the direction of time. *WIAS preprint 1896* (2013), DOI 10.20347/WIAS.PREPRINT.1896

Received 29 June 2022; revised 8 December 2024.

Artur Stephan
Institute for Analysis and Scientific Computing, Technische Universität Wien, Wiedner Hauptstraße 8-10, 1040 Vienna, Austria; artur.stephan@tuwien.ac.at