

Coarse-graining via EDP-convergence for linear fast-slow reaction systems

Alexander Mielke

*Weierstraß-Institut für Angewandte Analysis und Stochastik,
Mohrenstraße 39, 10117 Berlin*

and

*Humboldt-Universität zu Berlin, Institut für Mathematik,
Unter den Linden 6, 10099 Berlin, Germany
alexander.mielke@wias-berlin.de*

Artur Stephan

*Weierstraß-Institut für Angewandte Analysis und Stochastik,
Mohrenstraße 39, 10117 Berlin, Germany
artur.stephan@wias-berlin.de*

Received 13 November 2019

Revised 30 May 2020

Accepted 10 June 2020

Published 12 September 2020

Communicated by U. Stefanelli

We consider linear reaction systems with slow and fast reactions, which can be interpreted as master equations or Kolmogorov forward equations for Markov processes on a finite state space. We investigate their limit behavior if the fast reaction rates tend to infinity, which leads to a coarse-grained model where the fast reactions create microscopically equilibrated clusters, while the exchange mass between the clusters occurs on the slow time scale. Assuming detailed balance the reaction system can be written as a gradient flow with respect to the relative entropy. Focusing on the physically relevant cosh-type gradient structure we show how an effective limit gradient structure can be rigorously derived and that the coarse-grained equation again has a cosh-type gradient structure. We obtain the strongest version of convergence in the sense of the Energy-Dissipation Principle (EDP), namely EDP-convergence with tilting.

Keywords: Markov process with detailed balance; entropic gradient structure; energy-dissipation balance; EDP-convergence; coarse graining; reconstruction operators.

AMS Subject Classification 2020: 49S05, 47D07, 47J30, 92E20, 60J20

1. Introduction

Considering $I \in \mathbb{N}$ particles that interact linearly with each other with given rates A_{ik} , the evolution of the probability or concentration $c_i \in [0, 1]$ of a species $i \in \{1, \dots, I\} =: \mathcal{I}$ can be described by the master equation

$$\dot{c} = Ac, \tag{1.1}$$

where A is the adjoint of the Markov generator $\mathcal{L} : \mathbb{R}^I \rightarrow \mathbb{R}^I$ of the underlying Markov process, i.e. $A = \mathcal{L}^*$, see e.g. Refs. 9, 3 and 8 for more information. In particular, this means $A_{ki} \geq 0$ for $i \neq k$ and $\sum_{k=1}^I A_{ki} = 0$ for all $i \in \mathcal{I}$. We interpret the master equation as a *rate equation* defined on the state space

$$\mathbf{Q} = \text{Prob}(\mathcal{I}) := \left\{ c \in [0, 1]^I \mid \sum_{i=1}^I c_i = 1 \right\} \subset \mathbb{R}^I.$$

In many applications, the number I of particles can be huge and the reaction coefficients A_{ik} may vary in a huge range. In such cases, the analysis or the numerical treatment of system (1.1) is out of reach, and hence suitable simplifications are necessary. One natural assumption is that reactions can happen with different speeds. We will consider the case that the slow and fast reactions are distinguished, the slow ones of order 1 and the fast ones of order $1/\varepsilon$ for a small parameter $\varepsilon \rightarrow 0$. Hence, we decompose $A = A^\varepsilon$ into $A^\varepsilon = A^S + \frac{1}{\varepsilon}A^F$, “S” for slow and “F” for fast reactions. Our equation then is ε -dependent and reads

$$\dot{c}^\varepsilon = A^\varepsilon c^\varepsilon = \left(A^S + \frac{1}{\varepsilon}A^F \right) c^\varepsilon. \tag{1.2}$$

The limit passage for $\varepsilon \rightarrow 0$ in linear and nonlinear slow-fast reaction systems is a well-established field starting from pioneering work by Tikhonov⁴¹ and Fenichel.¹² We refer to Refs. 4, 6 and 38 for modern approaches and to Ref. 19 for nonlinear fast-slow reaction systems under the influence of stochastic fluctuations, see e.g. Example 6.1 there for an mRNA-DNA system for $I = 6$ species with eight slow reactions and two fast reactions.

While we repeat some of these arguments in Sec. 2, the main goal of this paper is quite different. Our study is devoted to the associated *gradient structures* for (1.2) and their limiting behavior for $\varepsilon \rightarrow 0$. Gradient structures exist under the additional assumption that the *detailed-balance condition* holds, means that there exists a positive equilibrium state $w^\varepsilon = (w_i^\varepsilon)_{i \in \mathcal{I}} \in \mathbf{Q}$ such that

$$\text{detailed-balance condition (DBC): } \forall i, k \in \mathcal{I} : A_{ik}^\varepsilon w_k^\varepsilon = A_{ki}^\varepsilon w_i^\varepsilon. \tag{1.3}$$

Following Refs. 25, 34 and 27, a gradient structure for a rate equation $\dot{c} = V_\varepsilon(c)$ on the state space \mathbf{Q} means that there exist a differentiable energy functional \mathcal{E}_ε and a dissipation potential \mathcal{R}_ε such that the rate equation can be generated as the associated gradient-flow equation, namely

$$\dot{c} = V_\varepsilon(c) = D_\xi \mathcal{R}_\varepsilon^*(c, -D_\xi \mathcal{E}_\varepsilon(c)) \quad \text{or equivalently } 0 = D_\xi \mathcal{R}_\varepsilon(c, \dot{c}) + D\mathcal{E}_\varepsilon(c). \tag{1.4}$$

Here \mathcal{R}_ε is called a dissipation potential if $\mathcal{R}_\varepsilon(c, \cdot) : T_c \mathbf{Q} \rightarrow [0, \infty]$ is lower semicontinuous and convex and satisfies $\mathcal{R}_\varepsilon(c, 0) = 0$. Then, $\mathcal{R}_\varepsilon^*$ is the (partial) Legendre–Fenchel transform

$$\mathcal{R}_\varepsilon^*(c, \xi) := \sup \{ \langle \xi, v \rangle - \mathcal{R}_\varepsilon(c, v) \mid v \in T_c \mathbf{Q} \}.$$

For reaction systems of mass-action type (which includes all linear systems) satisfying detailed balance, it was shown in Ref. 25 that an entropic gradient structure

exists, i.e. \mathcal{E}_ε is the relative Boltzmann entropy $\mathcal{E}_{\text{Bz}}^\varepsilon(c) := \mathcal{H}(c|w^\varepsilon)$ of c with respect to w^ε , see Sec. 4.3.2. However, this fact was used implicitly in earlier works, see e.g. Eq. (113) in Ref. 32 and Sec. VII in Ref. 42. For linear reaction systems, which are master equations for Markov processes, a more general theory was developed in Refs. 22 and 5 leading to a large class of possible gradient structures, see Sec. 3 and Sec. 2.5 in Ref. 23.

Here, we use the physically most natural gradient structure that has its origin in the theory of large deviation, see Refs. 30 and 29. The dual dissipation potentials $\mathcal{R}_\varepsilon^*(c, \cdot) : \text{T}_c\mathbb{Q} \rightarrow \mathbb{R}$ are not quadratic but rather exponential due to cosh terms, namely

$$\mathcal{R}_\varepsilon^*(c, \xi) = \frac{1}{2} \sum_{i < k} \kappa_{ik}^\varepsilon \sqrt{c_i c_k} C^*(\xi_i - \xi_k) \text{ with } C^*(\zeta) = 4 \cosh(\zeta/2) - 4 \quad (1.5)$$

and $\kappa_{ik}^\varepsilon = A_{ik}^\varepsilon \sqrt{w_k^\varepsilon/w_i^\varepsilon}$. The gradient structure $(\mathbb{Q}, \mathcal{E}_{\text{Bz}}^\varepsilon, \mathcal{R}_\varepsilon^*)$ exactly generates the gradient-flow evolution (1.2), and we call it simply the *cosh gradient structure*. Note that the dissipation potential $v \mapsto \mathcal{R}_\varepsilon(c, v)$ is still superlinear, but grows only like $|v| \log(1 + |v|)$. In particular, \mathcal{R}_ε does not induce a metric on \mathbb{Q} .

This gradient structure is also in line with the first derivation of exponential kinetic relations by Marcellin in 1915, see Ref. 24. Moreover, it arises as effective gradient structure in EDP converging systems, see Refs. 20 and 14. In Ref. 15, it is shown that the exponential function “cosh” arises due to the Boltzmann entropy as inverse of the logarithm. For L^p -type entropies, the dual dissipation functional \mathcal{R}^* will have a growth like $|\xi|^{c_0/(p-1)}$.

Instead of passing to the limit $\varepsilon \rightarrow 0$ in Eq. (1.2), our goal is to perform the limit passage in the gradient system $(\mathbb{Q}, \mathcal{E}_{\text{Bz}}^\varepsilon, \mathcal{R}_\varepsilon^*)$ to obtain directly an effective gradient system $(\mathbb{Q}, \mathcal{E}_0, \mathcal{R}_{\text{eff}}^*)$ via the notion of EDP-convergence as introduced in Refs. 20, 7 and 28. Roughly spoken this convergence asked for the Γ -convergence of the energies, namely $\mathcal{E}_{\text{Bz}}^\varepsilon \xrightarrow{\Gamma} \mathcal{E}_0$ on \mathbb{Q} , and for the dissipation functionals $\mathcal{D}_\varepsilon \xrightarrow{\Gamma} \mathcal{D}_0$ on $L^2([0, T]; \mathbb{Q})$ with

$$\begin{aligned} \mathcal{D}_\varepsilon(c) &= \int_0^T (\mathcal{R}_\varepsilon(c, \dot{c}) + \mathcal{R}_\varepsilon^*(c, -D\mathcal{E}_\varepsilon(c))) dt \quad \text{and} \\ \mathcal{D}_0(c) &= \int_0^T (\mathcal{R}_{\text{eff}}(c, \dot{c}) + \mathcal{R}_{\text{eff}}^*(c, -D\mathcal{E}_0(c))) dt. \end{aligned}$$

The notion of EDP-convergence produces a *unique limit gradient system*, and we may have $\mathcal{R}_\varepsilon \xrightarrow{\Gamma} \mathcal{R}_0$ while $\mathcal{R}_{\text{eff}} \neq \mathcal{R}_0$, see Refs. 20 and 7. As a trivial consequence of EDP-convergence we then find that $0 = D\mathcal{R}_{\text{eff}}(c, \dot{c}) + D\mathcal{E}_0(c)$ is the limit equation, cf. Lemma 3.4.

We emphasize that constructing the limit equation $\dot{c} = V^0(c)$ for a family of evolution equations $\dot{c} = V^\varepsilon(c)$ depending on a small parameter $\varepsilon \rightarrow 0$ is quite different from our goal. In Fig. 1, this would mean to concentrate on the two downward arrows on the right only. In general, an evolution equation may have

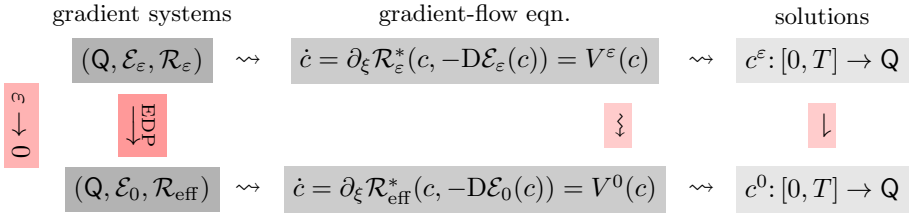


Fig. 1. EDP-convergence leads to a commuting diagram, in particular EDP-convergence generates the correct limit equation $\dot{c} = V^0(c)$ and (subsequences of) the solutions c^ϵ converge to solutions c^0 of the limit equation. However, \mathcal{R}_{eff} provides information not contained in the limit equation.

many gradient structures, and this is certainly true for the linear master equations studied here. Thus, from knowing the limiting equation $\dot{c} = V^0(c)$ we cannot recover a unique gradient structure $(\mathbf{Q}, \mathcal{E}_0, \mathcal{R}_{\text{eff}})$.

Our philosophy is opposite: We consider the gradient structure $(\mathbf{Q}, \mathcal{E}_\epsilon, \mathcal{R}_\epsilon)$ associated to $\dot{c} = V^\epsilon(c)$ as additional information that is not contained in the equation, but of course they are compatible. Typically, the additional information is of thermodynamical nature and reflects the underlying microscopic properties of the model that are no longer seen in the macroscopic model, see Refs. 30 and 29. In Ref. 35, it is shown that the parabolic equation $\dot{u} = u_{xx}$ associates with different gradient structures if one models diffusion or if one models heat transfer.

Thus, we turn around the usual limit analysis where one first works on the gradient-flow equations (1.4) and the solutions $c^\epsilon : [0, T] \rightarrow \mathbf{Q}$, and then studies gradient structures for the limit equations. As indicated in Fig. 1, EDP-convergence works solely on the gradient systems and produces \mathcal{R}_{eff} as a nontrivial result, which then gives the limit equation and the accumulation points $c^0 : [0, T] \rightarrow \mathbf{Q}$ of the solutions $c^\epsilon : [0, T] \rightarrow \mathbf{Q}$.

The choice of gradient structure for a given family of equations $\dot{c} = V^\epsilon(c)$ may even be relevant for deriving the effective of limiting equation. Choosing the gradient structures $(\mathbf{Q}, \mathcal{E}_\epsilon^{(1)}, \mathcal{R}_\epsilon^{(1)})$ and $(\mathbf{Q}, \mathcal{E}_\epsilon^{(2)}, \mathcal{R}_\epsilon^{(2)})$ we may have EDP convergence to effective limits $(\mathbf{Q}, \mathcal{E}_0^{(1)}, \mathcal{R}_{\text{eff}}^{(1)})$ and $(\mathbf{Q}, \mathcal{E}_0^{(2)}, \mathcal{R}_{\text{eff}}^{(2)})$, such that the effective equations

$$\dot{c} = V_{(1)}^0(c) = D_\xi \mathcal{R}_{\text{eff}}^{(1)*}(c, -D\mathcal{E}_0^{(1)}(c)) \quad \text{and} \quad \dot{c} = V_{(2)}^0 = D_\xi \mathcal{R}_{\text{eff}}^{(2)*}(c, -D\mathcal{E}_0^{(2)}(c))$$

give different dynamics. We refer to Sec. 3.3.5.2 in Ref. 27 for such a case.

In Sec. 3.3 in Ref. 20, an example of a simple linear reaction systems (with $I = 3$) is considered, where it is shown that the cosh structure is distinguished by the fact that it is the only one that is stable under EDP-convergence. It is one of our major results that in our situation the same stability is true, i.e. EDP-convergence yields a limit gradient structure of cosh-type again.

We now describe our results more precisely. We mainly work under the assumption that our system (1.2) satisfies the DBC (1.3) for w^ϵ and assume that $w^\epsilon \rightarrow w^0 \in]0, 1[^I$, i.e. all components w_i^0 are positive. Then, clearly A^F satisfies

the DBC for w^0 . As is shown in Sec. 2, the fast reactions encoded in A^F separate $\mathcal{I} = \{1, \dots, I\}$ into $J < I$ clusters, and we define a coarse graining operator $M \in \mathbb{R}^{J \times I}$ and a reconstruction operator $N \in \mathbb{R}^{I \times J}$ satisfying

$$MA^F = 0 \in \mathbb{R}^{J \times I}, \quad A^F N = 0 \in \mathbb{R}^{I \times J}, \quad \text{and} \quad MN = \text{id}_{\mathbb{R}^J}.$$

The coarse graining operator M satisfies $M_{ji} \in \{0, 1\}$ indicating whether the species i belongs to the cluster j . The limit equation, which is derived in Theorem 2.9 independently of any EDP-convergence for clarity, then reads

$$M\dot{c}(t) = MA^S c(t) \quad \text{and} \quad A^F c(t) = 0. \tag{1.6}$$

Although convergence of solutions of (1.2) is indeed well-known, we added a short proof, as it shows similarities to the proof of EDP-convergence in using complementary information to derive compactness. Using the coarse-grained states $\hat{c}(t) = Mc(t) \in \hat{\mathcal{Q}} \subset \mathbb{R}^J$ with probabilities $\hat{c}_j(t)$ for the cluster $j \in \mathcal{J}$ one obtains the coarse-grained linear reaction systems

$$\dot{\hat{c}}(t) = \hat{A} \hat{c}(t) \quad \text{with} \quad \hat{A} = MA^S N \in \mathbb{R}^{J \times J}. \tag{1.7}$$

We refer to Sec. 2.4 for a detailed description and an interpretation of the coarse-grained equation.

From the solutions \hat{c} we obtain all solutions of the limit equation (1.6) via $c(t) = N\hat{c}(t)$. In fact, setting $\hat{w} := Mw^0 \in]0, 1[^J$ and defining the diagonal mappings $\mathbb{D}_{w^0} = \text{diag}(w_i^0)_{i \in \mathcal{I}}$ and $\mathbb{D}_{\hat{w}} = \text{diag}(\hat{w}_j)_{j \in \mathcal{J}}$ the reconstruction operator N is given via $N = \mathbb{D}_{w^0} M^* \mathbb{D}_{\hat{w}}^{-1}$. The intrinsic definition of N becomes clear from duality theory as \mathbb{D}_{w^0} can be seen as a duality mapping from relative densities $\varrho \in (\mathbb{R}^I)^*$ to concentrations $c \in \mathbb{R}^I$.

$$\begin{array}{ccc}
 c \in \mathbb{R}^I & \xrightarrow{\mathbb{D}_{w^0}^{-1}} & \varrho \in (\mathbb{R}^I)^* \supset M^*(\mathbb{R}^J)^* \\
 \downarrow M & \uparrow N & \uparrow M^* \\
 \hat{c} \in \mathbb{R}^J & \xrightarrow{\mathbb{D}_{\hat{w}}^{-1}} & \hat{\varrho} \in (\mathbb{R}^J)^*
 \end{array}$$

In Sec. 3, we discuss general gradient systems and define different notions of EDP-convergence as in Refs. 7 and 28, while Sec. 4 recalls the different possible gradient structures for linear reaction systems satisfying the DBC (1.3). In Sec. 4.4, we address the important notion of tilting of Markov processes which means the change of the equilibrium measure w into $w^\eta = \frac{1}{Z} (e^{-\eta_i w_i})_{i \in \mathcal{I}}$. It is another remarkable feature of the cosh gradient structure that it is invariant under tilting (see Proposition 4.1). Thus, the strong notion of EDP-convergence with tilting introduced in Refs. 7 and 28 can only be shown for the cosh gradient structure.

In Sec. 5, we present our main result on the EDP-convergence with tilting of the cosh-gradient systems $(\mathcal{Q}, \mathcal{E}_{\text{Bz}}^\varepsilon, \mathcal{R}_\varepsilon^*)$ defined via (1.5). While the Γ -convergence $\mathcal{E}_{\text{Bz}}^\varepsilon \xrightarrow{\Gamma} \mathcal{E}_{\text{Bz}}^0$ follows trivially from $w^\varepsilon \rightarrow w^0$, the Γ -convergence $\mathcal{Q}_\varepsilon \xrightarrow{\Gamma} \mathcal{Q}_0$ in

$L^2([0, T], \mathbb{Q})$ is much more delicate. In fact, Theorem 5.3 even provides the Mosco-convergence of $\mathfrak{D}_\varepsilon \xrightarrow{M} \mathfrak{D}_0$, i.e. (i) the liminf estimate $\liminf_{\varepsilon \rightarrow 0} \mathfrak{D}_\varepsilon(c^\varepsilon) \geq \mathfrak{D}_0(c^0)$ holds even under the weak convergence $c^\varepsilon \rightharpoonup c^0$ in $L^2([0, T]; \mathbb{Q})$ and (ii) for each $c^0 \in L^2([0, T]; \mathbb{Q})$ there exists a recovery sequence $c^\varepsilon \rightarrow c^0$ strongly(!) in $L^2([0, T]; \mathbb{Q})$ such that $\limsup_{\varepsilon \rightarrow 0} \mathfrak{D}_\varepsilon(c^\varepsilon) \leq \mathfrak{D}_0(c^0)$.

The main point of the result is the exact characterization of \mathcal{R}_{eff} . Indeed, we have

$$\mathfrak{D}_0(c) = \begin{cases} \int_0^T (\mathcal{R}_{\text{eff}}(c, \dot{c}) + \mathcal{R}_{\text{eff}}^*(c, -D\mathcal{E}_{\text{Bz}}^0(c))) \, dt & \text{for } c \in W^{1,1}([0, 1]; PQ), \\ \infty & \text{otherwise in } L^2([0, 1]; \mathbb{Q}), \end{cases}$$

where, for $c \in PQ$ the effective dissipation potential \mathcal{R}_{eff} is given by

$$\begin{aligned} \mathcal{R}_{\text{eff}}^*(c, \xi) &= \mathcal{R}_S^*(c, \xi) + \chi_{M^*(\mathbb{R}^J)^*}(\xi) \text{ or equivalently} \\ \mathcal{R}_{\text{eff}}(c, v) &= \inf_{z \in \mathbb{R}^I : Mz = Mv} \mathcal{R}_S(c, z). \end{aligned}$$

Here $P = NM$ is the projection mapping general $c \in \mathbb{Q}$ into microscopically equilibrated reactions $c = N\hat{c}$ with $\hat{c} = Mc$, and \mathcal{R}_S^* is the dual dissipation potential defined as in (1.5) but using only the slow reactions. Finally, the characteristic function χ_Ξ is 0 for $\xi \in \Xi$ and ∞ else. The condition $\chi_\Xi(-D\mathcal{E}_{\text{Bz}}^0(c)) < \infty$ is in fact equivalent to $c \in PQ$, see Sec. 5.2.

It is easy to see that the degenerate gradient system $(\mathbb{Q}, \mathcal{E}_{\text{Bz}}^0, \mathcal{R}_{\text{eff}}^*)$ generates exactly the limit equation (1.6). Moreover, using the bijective linear mapping $M : PQ \rightarrow \hat{\mathbb{Q}} := \{\hat{c} \in [0, 1]^J \mid \hat{c}_1 + \dots + \hat{c}_J = 1\} \subset \mathbb{R}^J$ with inverse $N : \hat{\mathbb{Q}} \rightarrow PQ \subset \mathbb{R}^I$ we can define the coarse-grained gradient system $(\hat{\mathbb{Q}}, \hat{\mathcal{E}}, \hat{\mathcal{R}})$ for the coarse-grained states $\hat{c} = Mc$ via

$$\hat{\mathcal{E}}(\hat{c}) = \mathcal{E}_{\text{Bz}}^0(N\hat{c}), \quad \hat{\mathcal{R}}(\hat{c}, \hat{v}) = \mathcal{R}_{\text{eff}}(N\hat{c}, N\hat{v}), \quad \hat{\mathcal{R}}^*(\hat{c}, \hat{\xi}) = \mathcal{R}_{\text{eff}}^*(N\hat{c}, M^*\hat{\xi}).$$

The construction and the explicit formula for $\mathcal{R}_{\text{eff}}^*$ yield that $(\hat{\mathbb{Q}}, \hat{\mathcal{E}}, \hat{\mathcal{R}})$ is again a cosh gradient structure and the associated gradient-flow equation is the coarse-grained equation (1.7), see Proposition 5.7.

This is indeed a rigorous coarse-graining in the sense of Sec. 6.1 in Ref. 23. This paper is intended to be an easy-to-understand first approach to more general results of EDP-convergence. In Ref. 31, we will cover nonlinear reaction systems, for which the coarse-graining procedure based on Markov operators cannot work and where the structure of the limiting equations will be more involved because of nonlinear algebraic constraints. Moreover, reaction–diffusion systems are discussed in Refs. 14 and 15. As already shown in Ref. 20, the cosh gradient structure may appear automatically from quadratic diffusion models, and we expect that the cosh gradient structure will also be stable in the more general situations in Ref. 39, where also the reaction fluxes are coarse-grained and reconstructed.

2. Fast-Slow Reaction Network

On $Q := \text{Prob}(\mathcal{I}) := \{c \in [0, 1]^I \mid \sum_{i \in \mathcal{I}} c_i = 1\} \subset X := \mathbb{R}^I$ we consider the Kolmogorov forward equation or master equation

$$\dot{c} = Ac \quad \text{with } A \in \mathbb{R}^{I \times I},$$

where A is the adjoint of a Markov generator, i.e.

$$A_{ik} \geq 0 \text{ for all } i \neq k \quad \text{and} \quad \forall k \in \mathcal{I} : 0 = \sum_{i=1}^I A_{ik}.$$

Some comments on the notation are in order. Usually, in the theory of Markov operators and stochastic processes the state space is the set of probability measures which is a subset of the dual space of continuous functions. So it would be more convenient to denote the space of interest by X^* and not X . Certainly, since we are dealing with finite dimensional spaces, both are isomorphic and the notation is just a question of manner. In that paper, the master equation is understood as a *rate equation* of a gradient system in the sense of Sec. 3 which is an equation in X . Strictly speaking, the operator A is the adjoint of a Markov generator \mathcal{L} which generates a semigroup of Markov operators $e^{t\mathcal{L}} : X^* \rightarrow X^*$. By definition, a Markov operator $M^* : X^* \rightarrow Y^*$ on a finite-dimensional state space maps positive vectors on positive vectors and the constant one vector $\mathbb{1}_{X^*}$ to a constant one vector $\mathbb{1}_{Y^*}$. Its adjoint maps the set of probability vectors onto the set of probability vectors.

The linear reactions given by A , naturally define a graph or reaction network, where edges e_{ik} from node x_i to node x_k correspond to the entries $A_{ik} > 0$. The graph is directed, i.e. edges e_{ik} and e_{ki} are different and have an orientation. We assume that A is irreducible, which means that the corresponding graph is irreducible, or in other words, that any two nodes are connected via a directed path. This implies that there is a unique steady state $w \in \text{Prob}(\mathcal{I})$ which is positive, i.e. $w_j > 0$ for all $j \in \mathcal{I}$, see e.g. Ref. 8.

The crucial assumption for our systems is the following symmetry condition. The Markov process is said to satisfy the *detailed-balance condition* (DBC) with respect to its stationary measure $w > 0$, if $A_{ik}w_k = A_{ki}w_i$ for all $i, k \in \mathcal{I}$. Assuming detailed balance, the evolution equation $\dot{c} = Ac$, which is an equation on X , can also be written in another form. Let us introduce the duality operator

$$\mathbb{D}_w = \text{diag}(w) : \begin{cases} X^* \rightarrow X, \\ \varrho \mapsto c = \mathbb{D}_w \varrho \end{cases} \quad \text{and} \quad X \ni c \xrightarrow{\mathbb{D}_w^{-1}} \varrho \in X^*.$$

Hence, \mathbb{D}_w maps the relative densities ϱ to the concentrations c , i.e. $c_i = \varrho_i w_i$. The linear master equation can now be written as

$$\dot{c} = B\varrho \quad \text{with } B = A\mathbb{D}_w.$$

Because of the DBC, $B = A\mathbb{D}_w : X^* \rightarrow X$ is a symmetric operator on X , i.e. $B^* = B$.

For our slow–fast systems, we introduce a scaling parameter $1/\varepsilon$ for $\varepsilon > 0$ and the rates A_{ik} on the right-hand side decompose into $A = A^\varepsilon = A^S + \frac{1}{\varepsilon}A^F$, where “S” stands for slow and “F” for fast reactions. Our equation is ε -dependent and reads

$$\dot{c}^\varepsilon = A^\varepsilon c^\varepsilon = \left(A^S + \frac{1}{\varepsilon}A^F \right) c^\varepsilon. \tag{2.1}$$

The aim of the paper is to investigate the system in the limit $\varepsilon \rightarrow 0$. To do this, some assumptions on the ε -dependent reaction network are needed.

2.1. Assumptions on the ε -dependency of the network

Our paper will be restricted to the case where the stationary measure $w^\varepsilon \in \mathbb{Q}$ converges to a positive limit measure $w^\varepsilon \rightarrow w^0 \in]0, 1[^{\mathcal{I}}$:

For all $\varepsilon > 0$, the reaction graph defined by A^ε is connected.

Moreover, if there is a transition from state i to k (i.e. $A_{ki} > 0$), (2.2a)
then there is also a transition backwards from k to i .

For all $\varepsilon > 0$ there is a unique and positive stationary measure $w^\varepsilon \in \mathbb{Q}$, and the stationary measure converges $w^\varepsilon \rightarrow w^0$, where w^0 (2.2b)
is positive.

(DBC): For all $\varepsilon > 0$ the detailed-balance condition with respect (2.2c)
to w^ε holds, i.e. $A_{ik}^\varepsilon w_k^\varepsilon = A_{ki}^\varepsilon w_i^\varepsilon$ for all $i, k \in \mathcal{I}$.

These three conditions are not independent of each other, but it is practical to state them as above. In particular, if (2.2a) and the DBC (2.2c) hold, then (2.2b) follow, which is the content of the following results. See Ref. 37 and the references therein for generalizations.

Proposition 2.1. *Let the reaction network satisfy (2.2a) and (2.2c) and define, for transitions according (2.2a), the transition quotients*

$$q_{ik}^\varepsilon = \frac{A_{ik}^\varepsilon}{A_{ki}^\varepsilon} = \frac{A_{ik}^S + \frac{1}{\varepsilon}A_{ik}^F}{A_{ki}^S + \frac{1}{\varepsilon}A_{ki}^F}.$$

If there is a (universal) bound $q^ < \infty$ such that for all transitions from i to k and for all $\varepsilon \geq 0$ the transition quotients q_{ik}^ε satisfy $1/q^* \leq q_{ik}^\varepsilon \leq q^*$, then w^ε converges and its limit w^0 is positive, i.e. (2.2b) holds.*

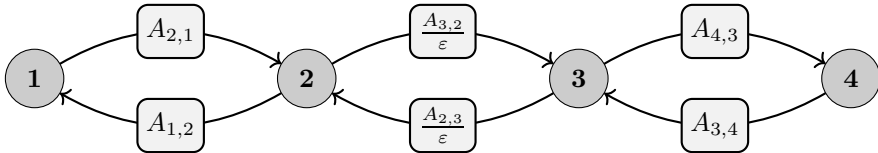
Proof. Using the DBC (2.2c), the stationary measure w^ε only depends on the transition quotient q_{ik}^ε . Hence, each $\varepsilon \mapsto w_i^\varepsilon \in [0, 1]$ is a rational polynomial in ε and thus converges to w_i^0 with $w^0 \in \mathbb{Q} = \text{Prob}(\mathcal{I})$ with polynomial dependency on $\varepsilon > 0$. Moreover, $q_{ik}^\varepsilon = 1/q_{ki}^\varepsilon$ converges to $q_{ik}^0 \in [1/q^*, q^*]$. Since the limit w^0 again depends only q_{ik}^0 , we conclude that it is positive. □

We now comment on the relevance of the above assumptions and give two non-trivial examples.

- Remark 2.2.** (a) In the chemical literature, our assumption (2.2a) is often called (weak) *reversibility*. It implies already that the stationary measure w^ε for A^ε is unique and positive.
- (b) The assumptions in Proposition 2.1 say that the quotients q_{ij}^ε are bounded even for $\varepsilon \rightarrow 0$ and hence, they converge. In particular, this means that if there is a fast reaction $A_{ik}^F \neq 0$ then necessarily also the backward reaction is fast, i.e. $A_{ki}^F \neq 0$. So, the graph does not change its topology in the limit process $\varepsilon \rightarrow 0$. Without this assumption the mass w_i^ε may vanish for some species i , see Example 2.3(b). This case is more delicate and will be considered in subsequent work.
- (c) It was observed in Refs. 42 and 25 that reaction systems of mass-action type have an entropic gradient structure, if the DBC holds. For linear reaction systems this was independently found in Refs. 22 and 5. However, our work will not use the quadratic gradient structure derived in the latter works, but will rely on the cosh-type generalized gradient structure derived in Refs. 30 and 29, see Sec. 4.
- (d) Assuming (2.2a), (2.2c), and additionally that the reaction quotients q_{ik}^ε scale either with 1 or with $1/\varepsilon$, i.e. $A_{ik}^F \neq 0 \Rightarrow A_{ik}^S = 0$, then the transition quotients q_{ik}^ε are ε -independent. In particular, the stationary measure w_ε as well as the energy \mathcal{E}_ε (see Sec. 4.2) are independent of ε .

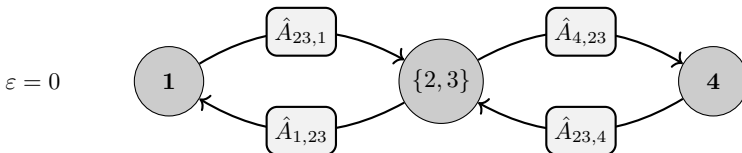
Example 2.3. We discuss two cases highlighting the relevance of our assumptions.

- (a) A prototype example is the following, where four states are involved:

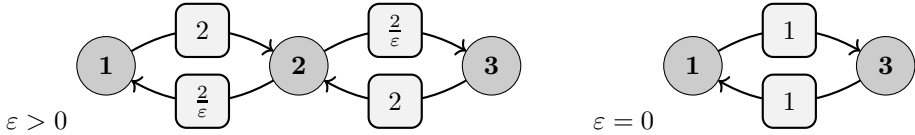


As in all reaction chains, this example satisfies the DBC (2.2c).

We observe that the reaction rates A_{ik}^ε scale either with 1 or with $1/\varepsilon$ and hence, the reaction ratios as well as the stationary measure do not depend on ε , see Remark 2.2(d). Hence, the assumptions (2.2) are satisfied. We expect that in the limit $\varepsilon \rightarrow 0$ a local equilibrium between the states 2 and 3 occurs, which means that the system can be described by only three states.



- (b) In Ref. 20, the authors considered the following reaction chain:



The DBC (2.2c) is again satisfied. The stationary measure is $w_\varepsilon = \frac{1}{2+\varepsilon}(1, \varepsilon, 1)$. The transition quotients are $q_{12}^\varepsilon = \varepsilon$ and $q_{23}^\varepsilon = \frac{1}{\varepsilon}$, which converge to 0 or ∞ , respectively. Hence assumption (2.2b) is violated. In fact the limit stationary measure is $w^0 = (\frac{1}{2}, 0, \frac{1}{2})$, which is no longer strictly positive. In Sec. 3.3 in Ref. 20, the EDP-convergence is performed for different gradient structures and only the cosh-gradient structure as defined in Sec. 4.3.3 turned out to be stable.

2.2. Capturing the states connected by fast reactions

In the limit species which are connected by fast reactions have to be treated like one large particle. Let $i_1 \sim_F i_2$ denote the relation if states i_1 and i_2 are connected via fast reactions. Assumptions (2.2a)–(2.2c) guarantee that \sim_F defines an equivalence relation on \mathcal{I} and decomposes \mathcal{I} into different equivalence classes $\mathcal{J} := \{\alpha_1, \dots, \alpha_J\}$, where the index of \sim_F , i.e. the number of (different) equivalence classes, is denoted by J . By definition all α_j are non-empty. Obviously, we have $1 \leq J \leq I$. In particular, $J = I$ means that there are no fast reactions; $J = 1$ means that each two species are connected via at least one reaction path consisting only of fast reactions. Let $\phi : \{1, \dots, I\} \rightarrow \{\alpha_1, \dots, \alpha_J\}$ be the function, which maps a state i to its equivalence class α_j , i.e. $i \mapsto \phi(i) = [i]_{\sim_F} = \alpha_j$. To make notation simpler, we denote the set of equivalence classes by $\mathcal{J} = \{1, \dots, J\}$ and further use $j \in \mathcal{J}$ and $i \in \mathcal{I}$.

The function $\phi : \mathcal{I} \rightarrow \mathcal{J}$ defines a deterministic Markov operator $M^* : Y^* \rightarrow X^*$, where Y^* is a J -dimensional real vector space, by

$$(M^* \hat{\varrho})_i := \hat{\varrho}_{\phi(i)}, \quad \hat{\varrho} \in Y^*, \quad i \in \mathcal{I}.$$

Deterministic Markov operator means that its dual $M : X \rightarrow Y$ maps pure concentrations, i.e. unit vectors e_i , to pure concentrations.

Some facts on deterministic Markov operators are in order. Clearly for a deterministic Markov operator it holds $M^*(\hat{\varrho} \cdot \hat{\psi}) = M^* \hat{\varrho} \cdot M^* \hat{\psi}$ where the multiplication is meant pointwise. (This, by the way, characterizes all deterministic Markov operator.) We want to write the multiplicative relation in form of operators. To do this let us define the multiplication by $\hat{\varrho}$ as $\Pi_{\hat{\varrho}} : Y^* \rightarrow Y^*$, with $(\Pi_{\hat{\varrho}} \hat{\psi})_j = \hat{\varrho}_j \cdot \hat{\psi}_j$. Hence, we conclude for a deterministic Markov operator that $M^* \Pi_{\hat{\varrho}} = \Pi_{M^* \hat{\varrho}} M^*$. Dualizing this equation, we get $\Pi_{\hat{\varrho}}^* M = M \Pi_{M^* \hat{\varrho}}^*$. Note, that the adjoint operator has a simple form: $\Pi_{\hat{\varrho}}^* : Y \rightarrow Y$, $\Pi_{\hat{\varrho}}^* \hat{c} = \mathbb{D}_{\hat{c}} \hat{\varrho}$. So summarizing

$$\Pi_{\hat{\varrho}}^* M = M \Pi_{M^* \hat{\varrho}}^* \quad \text{and} \quad \Pi_{\hat{\varrho}}^* \hat{c} = \mathbb{D}_{\hat{c}} \hat{\varrho}. \tag{2.3}$$

In the limit process the species connected by fast reactions are identified. This is done by a linear *coarse-graining-operator*, which is the adjoint of M^* , $M : X \rightarrow Y$.

In matrix representation induced by the canonical basis, we have

$$M : X \approx \mathbb{R}^I \rightarrow Y \approx \mathbb{R}^J, \quad M_{ji} := \begin{cases} 1, & \text{for } i \in \alpha_j, \\ 0, & \text{otherwise.} \end{cases} \tag{2.4}$$

Note that the construction is such that M maps $X \supset \text{Prob}(\mathcal{I})$ onto $Y \supset \text{Prob}(\mathcal{J})$. Since for α_j there is at least one i with $i \in \alpha_j$, the matrix of M has full rank and each column is a unit vector. Moreover, we point out that M and M^* only depend on the reaction network topology and the locations of the fast reactions, the specific reaction rates A_{ij} do not matter (see Example 2.6).

2.3. Properties of the coarse-graining operator M and the reconstruction operator N

Recall the duality map \mathbb{D}_{w^0} , which is represented by a diagonal matrix with entries $w^0 > 0$, connects the concentrations and the relative densities, i.e.

$$\varrho \in X^* \xrightarrow{\mathbb{D}_{w^0}} c \in X.$$

The subset of X^* which consists of the equilibrated densities ϱ_i is denoted by X_{eq}^* , i.e.

$$X_{\text{eq}}^* := \{ \varrho \in X^* \mid \forall i_1 \sim_F i_2 : \varrho_{i_1} = \varrho_{i_2} \}.$$

For the limit system, we define the stationary measure (denoted by \hat{w}) by $\hat{w} = Mw^0$. Since M^* is a deterministic Markov operator, we have the following characterization of the multiplication operator induced by \hat{w} .

Lemma 2.4. *Let $M^* : Y^* \rightarrow X^*$ be a deterministic Markov operator induced by a function $\phi : \{1, \dots, I\} \rightarrow \{1, \dots, J\}$ and let $w \in X$. Then $Mw = \hat{w}$ if and only if $\mathbb{D}_{\hat{w}} = M\mathbb{D}_wM^*$.*

Proof. Assume that $\mathbb{D}_{\hat{w}} = M\mathbb{D}_wM^*$ holds. Evaluating both sides at the constant vector $\mathbb{1}_{Y^*}$, we get $\mathbb{D}_{\hat{w}}\mathbb{1}_{Y^*} = \hat{w}$ and $M\mathbb{D}_wM^*\mathbb{1}_{Y^*} = M\mathbb{D}_w\mathbb{1}_{X^*} = Mw$, since M^* is a Markov operator which maps $\mathbb{1}_{Y^*} \mapsto \mathbb{1}_{X^*}$. This proves the claim in one direction.

Assume $\hat{w} = Mw$ we have to show that $\mathbb{D}_{Mw} = M\mathbb{D}_wM^*$. We use statement (2.3) for deterministic Markov operators and find $\mathbb{D}_{Mw}\hat{\varrho} = \Pi_{\hat{\varrho}}^*Mw = M\Pi_{M^*\hat{\varrho}}^*w = M\mathbb{D}_wM^*\hat{\varrho}$. □

If M^* is not a deterministic Markov operator but a general one, then the above relation will not hold.

We assumed that all equivalence classes α_j are non-empty and hence, each row of our coarse-graining operator M defined in (2.4) has at least one entry “1”. In

particular, this implies that \hat{w} is strictly positive and hence, $\mathbb{D}_{\hat{w}}$ is invertible. In particular, we proved that the following diagram commutes:

$$\begin{array}{ccc}
 c \in X & \xrightarrow{\mathbb{D}_{w^0}^{-1}} & \varrho \in X^* \supset X_{\text{eq}}^* = \{ \varrho \in X^* \mid \forall i_1 \sim_F i_2 : \varrho_{i_1} = \varrho_{i_2} \} \\
 M \downarrow & & M^* \uparrow \\
 \hat{c} \in Y & \xrightarrow{\mathbb{D}_{\hat{w}}^{-1}} & Y^*
 \end{array}$$

The crucial object is the following operator $N : Y \rightarrow X$, which “inverts” the coarse-graining operator $M : X \rightarrow Y$, by mapping coarse-grained concentrations $\hat{c} \in Y$ to concentrations $c \in X$ (see also Ref. 40, where the operator is introduced for its connection to the direction of time). We call N a *reconstruction operator* as it reconstructs the full information on the density $c \in X$ from the coarse-grained vector $\hat{c} \in Y$ assuming, of course, microscopic equilibrium. More precisely, N is defined via

$$N := \mathbb{D}_{w^0} M^* \mathbb{D}_{\hat{w}}^{-1} : Y \rightarrow X \quad \text{such that } N^* = \mathbb{D}_{\hat{w}}^{-1} M \mathbb{D}_{w^0} : X^* \rightarrow Y^*. \quad (2.5)$$

While the coarse-graining operator M simply merges the masses within the corresponding equivalence classes, the reconstruction operator N redistributes the masses in each equivalence class proportional to the equilibrium measure. As a result $P = NM$ will be a projection from general states to states in local equilibrium.

These and other important properties of the operator M and N and their adjoints M^* and N^* are summarized in the next proposition, which is independent of the generators $A^\varepsilon = A^S + \frac{1}{\varepsilon} A^F$.

Proposition 2.5. *Let $M^* : Y^* \rightarrow X^*$ be a deterministic Markov operator as in Lemma 2.4 with adjoint $M : X \rightarrow Y$ and let $\hat{w} := M w^0$ for some $w^0 \in]0, 1[^I \subset \mathbb{Q}$. Moreover, N and N^* be defined as in (2.5), then the following holds:*

- (1) N^* is a Markov operator.
- (2) $MN = \text{id}_Y$ or $N^* M^* = \text{id}_{Y^*}$, i.e. N^* is a left-inverse of the Markov operator M^* .
- (3) $P := NM$ is a projection on X , which leaves the range of $\mathbb{D}_{w^0} M^* : Y^* \rightarrow X$ invariant. The adjoint $M^* N^*$ is a projection as well, which leaves the range of M^* invariant.
- (4) $N \hat{w} = w^0$, i.e. N inverts with respect to the stationary measure.
- (5) The operator $P^* = M^* N^*$ is a Markov operator on X^* and its adjoint $P = NM$ has the stationary measure w^0 . Moreover, P^* satisfies detailed balance with respect to w^0 .

Proof. Clearly, N^* is non-negative and $N^* \mathbb{1}_{X^*} = \mathbb{D}_{\hat{w}}^{-1} M \mathbb{D}_{w^0} \mathbb{1}_{X^*} = \mathbb{D}_{\hat{w}}^{-1} M w^0 = \mathbb{1}_Y$ holds. This proves the first statement.

Lemma 2.4 implies that $MN = \text{id}_Y$ and that NM is a projection on X , which leaves the range of $\mathbb{D}_{w^0} M^* : Y^* \rightarrow X$ invariant. The fourth claim is also trivial. It

is also not hard to see that P^* is a Markov operator and that its adjoint has the stationary measure w^0 . Moreover, detailed balance holds:

$$\mathbb{D}_{w^0}P^* = \mathbb{D}_{w^0}M^*N^* = \mathbb{D}_{w^0}M^*\mathbb{D}_{\hat{w}}^{-1}M\mathbb{D}_{w^0} = NM\mathbb{D}_{w^0} = P\mathbb{D}_{w^0}.$$

This proves the result. □

The following example shows how the operators look like in a specific case.

Example 2.6. For the reaction network in Example 2.3(a) we have $I = 4$ with only one fast reaction $2 \sim_F 3$, hence $J = 3$. Using the numbering $\alpha_1 = \{1\}$, $\alpha_2 = \{2, 3\}$, and $\alpha_3 = \{4\}$ and the stationary measures $w = (w_1, w_2, w_3, w_4)^\top \in X$ and $\hat{w} = (w_1, w_2 + w_3, w_4)^\top \in Y$, respectively, we find

$$M = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad N = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{w_2}{w_2 + w_3} & 0 \\ 0 & \frac{w_3}{w_2 + w_3} & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \text{and}$$

$$P = NM = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{w_2}{w_2 + w_3} & \frac{w_2}{w_2 + w_3} & 0 \\ 0 & \frac{w_3}{w_2 + w_3} & \frac{w_3}{w_2 + w_3} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

2.4. The limit equation and the coarse-grained equation

As a direct consequence of Proposition 2.5 we obtain a decomposition of the state space $X \approx \mathbb{R}^I$ into the microscopically equilibrated states

$$c = Pc \in Q_{\text{eq}} := PQ \subset X_{\text{eq}} := PX = \{c \in X \mid A^F c = 0\},$$

which are measures having constant density with respect to w^0 , and the component $(I - P)c \in X_{\text{fast}} := (I - P)X$ that disappears exponentially on the time scale of the fast reactions. We emphasize that the following result does not use the DBC (2.2c).

Proposition 2.7. *Under the assumptions (2.2a)-(2.2b) we have*

$$PA^F = A^F P = 0 \in \mathbb{R}^{I \times I}, \quad MA^F = 0 \in \mathbb{R}^{J \times I}, \quad A^F N = 0 \in \mathbb{R}^{I \times J}, \quad (2.6a)$$

$$X = X_{\text{eq}} \oplus X_{\text{fast}} \quad \text{with} \quad (2.6b)$$

$$X_{\text{eq}} = \ker(A^F) = \text{range}(P) = \text{range}(N) \quad \text{and} \quad (2.6c)$$

$$X_{\text{fast}} = \text{range}(A^F) = \ker(P) = \ker(M). \quad (2.6d)$$

Here, X_{fast} depends on M only, i.e. only on the reaction graph of A^F , whereas X_{eq} depends on A^S and A^F through w^0 .

Proof. By construction of M from the reaction network induced by A^F we immediately obtain $\text{range}(A^F) = \ker(M)$. Indeed, the entries of M are all 0 or 1, where the j th row contains only the entry 1 exactly for $i \in \alpha(j)$. Thus, these 1s correspond to the mass conservation in the corresponding equivalence class $\alpha(j) \subset \{1, \dots, I\}$, and $MA^F = 0$ follows, which implies $\text{range}(A^F) \subset \ker(M)$. Dimension counting gives the desired equality.

Using the injectivity of N and $P = NM$ we have shown (2.6d).

To establish the relation for X_{eq} it suffices to show $\ker(A^F) = \text{range}(N)$, since the surjectivity of M and $P = NM$ gives $\text{range}(N) = \text{range}(P)$.

Using the dimension counting it is even sufficient to show $A^F N = 0$. Firstly, we use $0 = A^\varepsilon w^\varepsilon = (A^S + \frac{1}{\varepsilon}A^F)w^\varepsilon$, which gives $A^F w^\varepsilon \rightarrow 0$, and hence $A^F w^0 = 0$. Moreover, we observe that the j th column of $N = \mathbb{D}_{w^0} M^* \mathbb{D}_{\hat{w}}$ contains the unique equilibrium measure associated with the equivalence class $\alpha(j) \subset \{1, \dots, I\}$, which implies that $A^F N = 0$. □

Based on the above result we can formally pass to the limit in our linear reaction system $\dot{c}^\varepsilon = (A^S + \frac{1}{\varepsilon}A^F)c^\varepsilon$. Multiplying the equation from the left by M we can use $MA^F = 0$ and see that the term of order $\frac{1}{\varepsilon}$ disappears. Moreover, it is expected that the fast reactions equilibrate, so in the limit $\varepsilon \rightarrow 0$ we expect the microscopic equilibrium condition $A^F c^\varepsilon \rightarrow 0$. Hence, we expect that $c^\varepsilon : [0, T] \rightarrow \mathbb{Q}$ converges to a function $c^0 : [0, T] \rightarrow \mathbb{Q}$ which solves the limit equation

$$M\dot{c}(t) = MA^S c(t) \quad \text{and} \quad A^F c(t) = 0. \tag{2.7}$$

Before giving a proof for the convergence $c^\varepsilon \rightarrow c$ we want state that this system has a unique solution for each initial condition $c(0)$ that is compatible, i.e. $A^F c(0) = 0$ and that this solution is characterized by solving the so-called *coarse-grained equation*.

Theorem 2.8. (Coarse-grained equation) *For each $c_0 \in \mathbb{Q}$ with $A^F c_0 = 0$ there is a unique continuous solution $c : [0, T] \rightarrow \mathbb{Q}$ of (2.7) with $c(0) = c_0$. This solution is obtained by solving the coarse-grained ODE*

$$\dot{\hat{c}} = MA^S N\hat{c}, \quad \hat{c}(0) = M c_0 \tag{2.8}$$

and setting $c(t) = N\hat{c}(t)$. Moreover, the stationary solution is $\hat{w} = Mw^0$.

Proof. On the one hand, by (2.6c) we know that $A^F c = 0$ is equivalent to $c = Pc = NMc$. Thus, for any solution c of (2.7) the coarse-grained state $\hat{c} = Mc$ satisfies the coarse-grained equation (2.8).

On the other hand, (2.8) is a linear ODE in $\hat{\mathbb{Q}} \subset Y$ which has a unique solution satisfying $\hat{c}(t) \in \hat{\mathbb{Q}}$. This proves the first result.

To see that $\hat{w} = Mw^0$ is a stationary measure, we use $A^F w^0 = 0$ and (2.6b) implies $Pw^0 = w^0$. On the other hand using $MA^F = 0$ we can pass to the limit in

$0 = M0 = MA^\varepsilon w^\varepsilon = MASw^\varepsilon$ to obtain $MASw^0 = 0$. Combining the two results we find

$$\hat{A}\hat{w} = MASN(Mw^0) = MASPw^0 = MASw^0 = 0,$$

which is the desired result. □

We emphasize that the coarse-grained equation (2.8) is again a linear reaction system, describing the master equation for a Markov process on $\mathcal{J} = \{1, \dots, J\}$. The effective operator $\hat{A} := MASN$ can be interpreted in the following way: N divides the coarse-grained states into microscopically equilibrated states, A^S is the part of the slow reactions, and M collects the states according to their equivalence classes $\alpha(j)$.

Using $M_{ji} = \delta_{j\phi(i)}$ and $N_{ij} = \frac{w_i^0}{\hat{w}_j} \delta_{j\phi(i)}$ the coefficients of the generator $\hat{A} = MASN$ are easily obtained by a suitable average, namely

$$\hat{A}_{j_1 j_2} = \sum_{i_1 \in \alpha_{j_1}} \sum_{i_2 \in \alpha_{j_2}} A^S_{i_1 i_2} \frac{w_{i_2}^0}{\hat{w}_{j_2}}. \tag{2.9}$$

2.5. Convergence of solutions on the level of the ODE

Finally, for mathematical completeness, we provide a simple and short convergence proof. It can also be obtained as a special case of the result in Ref. 4. Of course, the convergence of solutions is also a byproduct of the EDP-convergence given below, see Lemma 3.4. The latter result, which is the main goal of this work, provides convergence of the gradient structures, which is a significantly stronger concept, because the coarse-grained equation (2.8) has many different gradient structures, while the EDP-limit is unique.

Theorem 2.9. (Convergence of c^ε to c^0) *Assume (2.2) and consider solutions $c^\varepsilon : [0, T] \rightarrow \mathbb{Q}$ of (1.2) such that $Mc^\varepsilon(0) \rightarrow \hat{c}_0$. Then, we have the convergences*

$$Mc^\varepsilon \rightarrow Mc^0 \text{ in } C^0([0, T]; X) \quad \text{and} \quad c^\varepsilon \rightarrow c^0 \text{ in } L^2([0, T]; X),$$

where c^0 is the unique solution of (2.7) with $c^0(0) = N\hat{c}_0$.

Proof. Step 1: Weak compactness. We first observe that $c^\varepsilon : [0, T] \rightarrow \mathbb{Q} \subset [0, 1]^J$ provides a trivial *a priori* bound for c^ε in $L^\infty([0, T]; \mathbb{R}^J)$. Hence, we may choose a subsequence (not relabeled) such that $c^\varepsilon \rightarrow c^0$ weakly in $L^2([0, T]; \mathbb{R}^J)$.

Step 2: Compactness of coarse-grained concentrations. With Step 1 we see that $\hat{a}^\varepsilon := Mc^\varepsilon$ is bounded in $C^{\text{Lip}}([0, T]; \mathbb{R}^J)$, because of $\hat{a}^\varepsilon = Mc^\varepsilon = MASc^\varepsilon$. Thus, there is a subsequence (not relabeled) such that $\hat{a}^\varepsilon \rightarrow \hat{a}^0$ in $C^0([0, T]; \mathbb{R}^J)$ and $\hat{a}^0(0) = \hat{c}_0$. Moreover, with Step 1 we have $\hat{a}^0 = Mc^0$.

Step 3: Generation of microscopic equilibrium. We take the dot product of the ODE with the vector of relative densities $c^\varepsilon/w^\varepsilon := (c_i^\varepsilon/w_i^\varepsilon)_{i=1,\dots,I}$. Defining the quadratic form $\mathcal{B}_\varepsilon(c) = \sum_{i=1}^I \frac{c_i^2}{2w_i^\varepsilon}$ we obtain

$$\begin{aligned} \frac{d}{dt}\mathcal{B}_\varepsilon(c^\varepsilon) &= \dot{c}^\varepsilon \cdot \frac{c^\varepsilon}{w^\varepsilon} = (A^\varepsilon c^\varepsilon) \cdot \frac{c^\varepsilon}{w^\varepsilon} \\ &= \frac{1}{\varepsilon}(B^\varepsilon c^\varepsilon) \cdot c^\varepsilon \quad \text{with } \varepsilon \mathbb{D}_{w^\varepsilon}^{-1} A^\varepsilon =: B^\varepsilon = (B^\varepsilon)^* \geq 0. \end{aligned} \tag{2.10}$$

The latter relations follow from the DBC (2.2c). Defining the quadratic functional $\mathfrak{Q}_\varepsilon(c) := \int_0^T B^\varepsilon c(t) \cdot c(t) dt$ and integrating (2.10) over $[0, T]$ gives

$$\mathfrak{Q}_\varepsilon(c^\varepsilon) = \varepsilon \mathcal{B}(c^\varepsilon(0)) - \varepsilon \mathcal{B}(c^\varepsilon(T)) \leq C_1 \varepsilon.$$

Moreover, using $|w^\varepsilon - w^0| \leq C_2 \varepsilon$ we find $|\mathfrak{Q}_\varepsilon(c) - \mathfrak{Q}_0(c)| \leq C_3 \varepsilon$. Hence $\mathfrak{Q}_0(c^\varepsilon) \leq \mathfrak{Q}_\varepsilon(c^\varepsilon) + C_3 \varepsilon \leq C_1 \varepsilon + C_3 \varepsilon$. Using the convexity of \mathfrak{Q}_0 the weak limit c^0 of c^ε satisfies

$$0 \leq \mathfrak{Q}_0(c^0) \leq \liminf_{\varepsilon \rightarrow 0} \mathfrak{Q}_0(c^\varepsilon) \leq \liminf_{\varepsilon \rightarrow 0} (C_1 + C_3)\varepsilon = 0.$$

Since $B^0 = \mathbb{D}_{w^0}^{-1} A^F$ is symmetric and positive semidefinite we conclude $A^F c^0(t) = 0$ a.e. in $[0, T]$. More precisely, by (2.6d) $c \mapsto (B^0 c \cdot c)^{1/2}$ defines a norm on X_{fast} that is equivalent to $c \mapsto |(I - P)c|$. Thus, we conclude $(I - P)c^\varepsilon \rightarrow (I - P)c^0$. Moreover, Step 2 gives $Pc^\varepsilon = NMc^\varepsilon = N\hat{a}^\varepsilon \rightarrow NMc^0 = Pc^0$ such that $c^\varepsilon \rightarrow c^0$ in $L^2([0, T]; \mathbb{R}^I)$ follows.

Step 4. Limit passage in the ODE. To see that c^0 satisfies the limit equation (2.7) we pass to the limit in

$$Mc^\varepsilon(t) = Mc^\varepsilon(0) + \int_0^t MA^S c^\varepsilon(s) ds,$$

where the left-hand side converges by Step 2, and the right-hand side converges by the assumption on the initial condition, by Step 3, and Lebesgue’s dominated convergence theorem. Thus, $Mc^0(t) = Mc^0(0) + \int_0^t MA^S c^0(s) ds$, and with $A^F c^0 = 0$ from Step 3 the desired limit equation (2.7) is established.

As we already know that the solution of (2.7) is unique, we conclude convergence of the whole family $(c^\varepsilon)_{\varepsilon>0}$, instead of a subsequence only. □

In the above proof the DBC (2.2c) is not really necessary, but it simplified our proof considerably.

3. Generalized Gradient Structures

This small section provides the general notions of gradient systems, gradient-flow equations, the energy-dissipation principle (EDP), and the three notions of EDP convergence. We follow the survey paper²⁷ and the more recent works.^{7,28}

3.1. Gradient systems and the energy-dissipation principle

A triple $(Q, \mathcal{E}, \mathcal{R})$ is called a *gradient system* if

- Q is a closed convex subset of a Banach space X ,
- $\mathcal{E} : Q \rightarrow \mathbb{R}_\infty := \mathbb{R} \cup \{\infty\}$ is a differentiable functional (e.g. free energy, negative entropy)
- $\mathcal{R} : Q \times X \rightarrow \mathbb{R}_\infty$ is a dissipation potential, i.e. for all $u \in Q$ the functional $\mathcal{R}(u, \cdot) : X \rightarrow \mathbb{R}_\infty$ is lower semicontinuous (lsc), nonnegative, convex and satisfies $\mathcal{R}(u, 0) = 0$.

(More general, Q can be a manifold, then \mathcal{R} is defined on the tangent bundle TQ , but this generalization is not needed in this work.) A gradient system $(Q, \mathcal{E}, \mathcal{R})$ is called *classical* if $\mathcal{R}(u, \cdot)$ is quadratic, i.e. if there are symmetric and positive definite operators $G(u) : X \rightarrow X^*$ such that $\mathcal{R}(u, v) = \frac{1}{2} \langle G(u)v, v \rangle$. But often $\mathcal{R}(u, \cdot)$ is not quadratic (e.g. for rate-independent processes such as elastoplasticity), see Ref. 27 and reference therein. We define the dual dissipation potential \mathcal{R}^* using the Legendre transform via

$$\mathcal{R}^*(u, \xi) = (\mathcal{R}(u, \cdot))^*(\xi) := \sup \{ \langle \xi, v \rangle - \mathcal{R}(u, v) \mid v \in X \}.$$

The gradient system is uniquely described by $(Q, \mathcal{E}, \mathcal{R})$ or, equivalently by $(Q, \mathcal{E}, \mathcal{R}^*)$ and, in particular, in this paper we prefer the second representation.

The evolution of the states $u(t)$ in a gradient system is given in terms of the so-called *gradient-flow equation* that is given in terms of \mathcal{E} and \mathcal{R} and can be formulated in three equivalent ways:

- (I) *force balance in X^** . $0 \in \partial_{\dot{u}} \mathcal{R}(u, \dot{u}) + D\mathcal{E}(u) \in X^*$,
- (II) *power balance in \mathbb{R}* . $\mathcal{R}(u, \dot{u}) + \mathcal{R}^*(u, -D\mathcal{E}(u)) = -\langle D\mathcal{E}(u), \dot{u} \rangle$, (3.1)
- (III) *rate equation in X* . $\dot{u} \in \partial_{\xi} \mathcal{R}^*(u, -D\mathcal{E}(u)) \in X$,

where ∂ is the set-valued partial subdifferential with respect to the second variable.

In general, we cannot expect that the solution of the gradient-flow equation fill the whole state space. Clearly, along solutions we want to have $\mathcal{E}(u(t)) < \infty$ for $t > 0$. Moreover, relation (III) asks that $-D\mathcal{E}(u(t))$ lies in the domain of $\partial_{\xi} \mathcal{R}^*(u(t), \cdot)$ for a.a. $t \in [0, T]$. Thus, we set

$$\text{Dom}(Q, \mathcal{E}, \mathcal{R}) := \{u \in Q \mid D\mathcal{E}(u) \text{ exists, } \partial_{\xi} \mathcal{R}^*(u, -D\mathcal{E}(u)) \text{ is nonempty}\}. \quad (3.2)$$

Typically, one expects that solutions exist for all initial conditions in the closure of $\text{Dom}(Q, \mathcal{E}, \mathcal{R})$.

These three formulations are the same due to the so-called *Fenchel equivalences* (cf. Ref. 11): Let Z be a reflexive Banach space and $\Psi : Z \rightarrow \mathbb{R}_\infty$ be a proper, convex and lsc, then for every all pairs $(v, \xi) \in Z \times Z^*$ the following holds:

$$(i) \ \xi \in \partial\Psi(v) \iff (ii) \ \Psi(v) + \Psi^*(\xi) = \langle \xi, v \rangle \iff (iii) \ v \in \partial\Psi^*(\xi).$$

We emphasize that (ii) and (II) should be seen as scalar optimality conditions, because the definition of the Legendre transform easily gives the Young–Fenchel inequality, namely $\Psi(v) + \Psi^*(\xi) \geq \langle \xi, v \rangle$ for all $(v, \xi) \in Z \times Z^*$.

Integrating the power balance (II) in (3.1) over $[0, T]$ along a solution $u : [0, T] \rightarrow \mathbf{Q}$ and using the chain rule $\langle D\mathcal{E}(u(t)), \dot{u}(t) \rangle = \frac{d}{dt}\mathcal{E}(u(t))$ we find the *Energy-Dissipation Balance* (EDB):

$$\mathcal{E}(u(T)) + \int_0^T (\mathcal{R}(u(t), \dot{u}(t)) + \mathcal{R}^*(u(t), -D\mathcal{E}(u(t)))) dt = \mathcal{E}(u(0)). \tag{3.3}$$

The following *Energy-Dissipation Principle* (EDP) states that solving (3.3) is equivalent to solving the gradient-flow equation (3.1).

Theorem 3.1. (Energy-dissipation principle, see e.g. Theorem 3.2 in Ref. 27) *Assume that \mathbf{Q} is a closed convex subset of $X = \mathbb{R}^I$, that $\mathcal{E} \in C^1(\mathbf{Q}, \mathbb{R})$, and that the dissipation potential $\mathcal{R}(u, \cdot)$ is superlinear uniformly in $u \in \mathbf{Q}$. Then, a function $u \in W^{1,1}([0, T]; \mathbf{Q})$ is a solution of the gradient-flow equation (3.1) if and only if u solves the energy-dissipation balance (3.3).*

Again, the EDB is an optimality condition, because integrating the Young–Fenchel inequality for arbitrary $\tilde{u} \in W^{1,1}([0, T]; \mathbf{Q})$ and using the chain rule we obtain the estimate

$$\mathcal{E}(\tilde{u}(T)) + \int_0^T (\mathcal{R}(\tilde{u}(t), \dot{\tilde{u}}(t)) + \mathcal{R}^*(\tilde{u}(t), -D\mathcal{E}(\tilde{u}(t)))) dt \geq \mathcal{E}(\tilde{u}(0)). \tag{3.4}$$

The above considerations show that an important quantity associated with a gradient system $(\mathbf{Q}, \mathcal{E}, \mathcal{R})$ is given by the *dissipation functional*

$$\mathfrak{D}(u) := \int_0^T (\mathcal{R}(u(t), \dot{u}(t)) + \mathcal{R}^*(u(t), -D\mathcal{E}(u(t)))) dt,$$

which is defined for all curves $u \in W^{1,1}([0, T]; \mathbf{Q})$.

3.2. General gradient systems and EDP-convergence

In the following, we consider a family of gradient systems $(X, \mathcal{E}_\varepsilon, \mathcal{R}_\varepsilon)$ and define a notion of convergence on the level of gradient systems which uniquely defines the limit or effective system $(\mathbf{Q}, \mathcal{E}_0, \mathcal{R}_{\text{eff}})$. Our notion relies on the energy-dissipation principle from above and the so-called sequential Γ -convergence for functionals, which is defined as follows.

Definition 3.2. (Γ -convergence, see e.g. Ref. 1) For functionals $(I_\varepsilon)_{\varepsilon>0}$ on a Banach space Z we say I_ε (strongly) Γ -converges to I , and write $I_\varepsilon \xrightarrow{\Gamma} I$, if the following two conditions hold:

- (1) *Liminf estimate.*
if $u_\varepsilon \rightarrow u$ in Z , then $I(u) \leq \liminf_{\varepsilon \rightarrow 0} I_\varepsilon(u_\varepsilon)$,

(2) *Existence of recovery sequences.*

for all $\tilde{u} \in Z$ there exists $(\tilde{u}_\varepsilon)_{\varepsilon>0}$ such that $\tilde{u}_\varepsilon \rightarrow \tilde{u}$ and $\lim_{\varepsilon \rightarrow 0} I_\varepsilon(\tilde{u}_\varepsilon) = I(\tilde{u})$.

If the same conditions hold when the strong convergences “ \rightarrow ” are replaced by weak convergences “ \rightharpoonup ”, we say that I_ε *weakly Γ -converges to I* and write $I_\varepsilon \xrightarrow{\Gamma} I$. If $I_\varepsilon \xrightarrow{\Gamma} I$ and $I_\varepsilon \xrightarrow{\Gamma} I$ hold, we say that I_ε *Mosco converges to I* and write $I_\varepsilon \xrightarrow{M} I$. Clearly, for finite-dimensional Banach spaces Z the convergences $\xrightarrow{\Gamma}$, $\xrightarrow{\Gamma}$, and \xrightarrow{M} coincide.

The energy dissipation principle allows us to formulate the gradient-flow equation in terms of the two functionals \mathcal{E}_ε and \mathcal{D}_ε . However, to explore the full structure of gradient systems it is useful to embed the given gradient system into a family of tilted gradient systems $(\mathbf{Q}, \mathcal{E}^\eta, \mathcal{R})$, where the *tilted energies* \mathcal{E}^η are given by

$$\mathcal{E}^\eta(u) = \mathcal{E}(u) - \ell^\eta(u) \quad \text{with } \ell^\eta(u) := \langle \eta, u \rangle \tag{3.5}$$

with an arbitrary tilt $\eta \in X^*$. Moreover, introducing the *tilted dissipation functional*

$$\mathcal{D}_\varepsilon^\eta(u) := \int_0^T (\mathcal{R}_\varepsilon(u, \dot{u}) + \mathcal{R}_\varepsilon^*(u, \eta - D\mathcal{E}_\varepsilon(u))) \, dt, \tag{3.6}$$

we can now define three versions of EDP-convergence for a family $((\mathbf{Q}, \mathcal{E}_\varepsilon, \mathcal{R}_\varepsilon))_{\varepsilon>0}$ as follows.

Definition 3.3. (EDP-convergence, cf. Refs. 7 and 28) Let \mathbf{Q} be a closed convex subset of a Banach space X and let \mathcal{E}_ε be Gateaux differentiable.

(A) We say that the gradient systems $(\mathbf{Q}, \mathcal{E}_\varepsilon, \mathcal{R}_\varepsilon)_{\varepsilon>0}$ *converge in the simple EDP sense to $(\mathbf{Q}, \mathcal{E}_0, \mathcal{R}_{\text{eff}})$* , and write $(\mathbf{Q}, \mathcal{E}_\varepsilon, \mathcal{R}_\varepsilon) \xrightarrow{\text{EDP}} (\mathbf{Q}, \mathcal{E}_0, \mathcal{R}_{\text{eff}})$, if the following conditions hold:

- (i) $\mathcal{E}_\varepsilon \xrightarrow{\Gamma} \mathcal{E}_0$ on $\mathbf{Q} \subset X$, and
- (ii) $\mathcal{D}_\varepsilon \xrightarrow{\Gamma} \mathcal{D}_0$ on $L^2([0, T]; \mathbf{Q})$ with $\mathcal{D}_0(u) = \int_0^T (\mathcal{R}_{\text{eff}}(u, \dot{u}) + \mathcal{R}_{\text{eff}}^*(u, -D\mathcal{E}_0(u))) \, dt$.

(B) We say that $(\mathbf{Q}, \mathcal{E}_\varepsilon, \mathcal{R}_\varepsilon)$ *EDP-converges with tilting to $(\mathbf{Q}, \mathcal{E}_0, \mathcal{R}_{\text{eff}})$* , if for all $\eta \in X^*$ we have $(\mathbf{Q}, \mathcal{E}_\varepsilon - \ell_\eta, \mathcal{R}_\varepsilon) \xrightarrow{\text{EDP}} (\mathbf{Q}, \mathcal{E}_0 - \ell_\eta, \mathcal{R}_{\text{eff}})$.

(C) We say that $(\mathbf{Q}, \mathcal{E}_\varepsilon, \mathcal{R}_\varepsilon)$ *contact EDP-converges with tilting to $(\mathbf{Q}, \mathcal{E}_0, \mathcal{R}_{\text{eff}})$* , if (i) holds and for all $\eta \in X^*$ we have $\mathcal{D}_\varepsilon^\eta \xrightarrow{\Gamma} \mathcal{D}_0^\eta$ with $\mathcal{D}_0^\eta(u) = \int_0^T \mathcal{M}(u(t), \dot{u}(t), \eta - D\mathcal{E}_0(u(t))) \, dt$, where \mathcal{M} satisfies the contact conditions

- (c1) $\mathcal{M}(u, v, \xi) \geq \langle \xi, v \rangle$ for all $(v, \xi) \in X \times X^*$,
- (c2) $\mathcal{M}(u, v, \xi) = \langle \xi, v \rangle \iff \mathcal{R}_{\text{eff}}(u, v) + \mathcal{R}_{\text{eff}}^*(u, \xi) = \langle \xi, v \rangle$.

Clearly, “tilted EDP-convergence” is a stronger notion than “contact EDP-convergence” since the contact potential \mathcal{M} is explicitly given in $\mathcal{R} + \mathcal{R}^*$ form. We refer to Refs. 7 and 28 for a general discussions of EDP-convergence and remark that “contact EDP-convergence with tilting” was called “relaxed EDP-convergence” in Ref. 7. We emphasize that there are cases where we have the Γ (or even Mosco)

convergence $\mathcal{R}_\varepsilon \rightarrow \mathcal{R}_0$, but EDP-convergence yields $\mathcal{R}_{\text{eff}} \neq \mathcal{R}_0$. In general, EDP-convergence allows for effective dissipation potentials \mathcal{R}_{eff} that inherit properties of the family $(\mathcal{E}_\varepsilon)_{\varepsilon>0}$.

The most important feature of the three different notions of EDP-convergence is that the effective gradient system is *uniquely determined* through the family $(\mathbb{Q}, \mathcal{E}_\varepsilon, \mathcal{R}_\varepsilon)$. This is a much stronger statement than the one obtained by the classical approach, where the effective or limiting equation is derived first and then a gradient structure is constructed afterwards. Obviously, uniqueness cannot be guaranteed because one equation may have several gradient structures. We also refer to Sec. 3.3.5.2 in Ref. 27, where for one family of evolution equations two different gradient structures are considered such that the EDP-limit exists and but is different.

A further interesting observation is that the notion of EDP-convergence does not involve the solutions of the associated gradient-flow equation. This may look like an advantage, since solutions need not be characterized, however typically showing EDP-convergence is at least as difficult. Another important feature is that EDP-convergence automatically implies the convergence of the corresponding solutions u^ε of the gradient-flow equations to the solutions of the effective equation

$$0 \in \partial_v \mathcal{R}_{\text{eff}}(u(t), \dot{u}(t)) + D\mathcal{E}_0(u(t)) \quad \text{for a.a. } t \in [0, T]. \tag{3.7}$$

Lemma 3.4. *Let the assumption of Theorem 3.1 be satisfied for all $\varepsilon \geq 0$. Assume that the gradient systems $(\mathbb{Q}; \mathcal{E}_\varepsilon, \mathcal{R}_\varepsilon)$ EDP-converge to $(\mathbb{Q}, \mathcal{E}_0, \mathcal{R}_{\text{eff}})$ in one of the three senses of Definition 3.3, then the following holds. If $u^\varepsilon : [0, T] \rightarrow \mathbb{Q}$ are solutions for (3.1) and $u : [0, T] \rightarrow \mathbb{Q}$ is such that*

$$u^\varepsilon(0) \rightarrow u(0), \quad \mathcal{E}_\varepsilon(u^\varepsilon(0)) \rightarrow \mathcal{E}_0(u(0)), \quad \text{and} \quad u^\varepsilon(t) \rightarrow u(t) \text{ in } X \text{ for } t \in [0, T],$$

then $u \in W^{1,1}([0, T]; X)$ and it is a solution of the gradient-flow equation (3.7).

Proof. By Theorem 3.1 we know that the EDB (3.3) holds for u^ε as solutions for the gradient system $(\mathbb{Q}, \mathcal{E}_\varepsilon, \mathcal{R}_\varepsilon)$. Using the liminf estimates for $\mathcal{E}_\varepsilon(u^\varepsilon(t))$ and for $\mathfrak{D}_\varepsilon^0(u^\varepsilon)$ and the convergence of $\mathcal{E}_\varepsilon(u^\varepsilon(0))$, we obtain

$$\mathcal{E}_0(u(T)) + \int_0^T (\mathcal{R}_{\text{eff}}(u(t), \dot{u}(t)) + \mathcal{R}_{\text{eff}}^*(u(t), -D\mathcal{E}_0(u(t)))) dt \leq \mathcal{E}_0(u(0)). \tag{3.8}$$

Together with (3.4) and the EDP in Theorem 3.1 we see that u solves (3.7). □

4. Gradient Structures for Linear Reaction Systems

In this section, we discuss several gradient structures for linear reaction systems satisfying the detailed balance condition. Moreover, following the theory of Markov processes we define a natural way of tilting such systems in such a way that a new global equilibrium state w arises. This will show that the entropic gradient structure with cosh-type dual dissipation plays a distinguished role.

4.1. A special representation for generators

We start from a general linear reaction system with the finite index space $\mathcal{I} := \{1, \dots, I\}$. On the state space $\mathbb{Q} = \text{Prob}(\mathcal{I})$ we consider the general linear reaction system

$$\dot{c} = Ac \quad \text{where } A_{in} \geq 0 \quad \text{for } i \neq n \quad \text{and} \quad \sum_{i=1}^I A_{in} = 0 \quad \text{for all } n \in \mathcal{I}. \quad (4.1)$$

Throughout we assume that there exists a positive equilibrium state $w \in \mathbb{Q}$, i.e. $Aw = 0$ and $w_i > 0$ for all $i \in \mathcal{I}$. At this stage we do not need the detailed-balance condition.

As we later want to change the equilibrium state w (and hence also the generator A) we write A in a specific form, namely

$$A = \mathbb{D}_w^{1/2} K \mathbb{D}_w^{-1/2} - \mathbb{D}_b \quad \text{with } K = (\kappa_{in}) \in \mathbb{R}^{I \times I} \quad \text{and } b \in \mathbb{R}^I \quad \text{given by}$$

$$\kappa_{in} = A_{in} \left(\frac{w_n}{w_i}\right)^{1/2} > 0 \quad \text{for } i \neq n, \quad \kappa_{ii} = 0, \quad \text{and} \quad (4.2)$$

$$b_i = -A_{ii} = \sum_{n=1}^I \kappa_{ni} \left(\frac{w_n}{w_i}\right)^{1/2} > 0.$$

This representation is useful, because we can keep K fixed, while varying w to obtain Markov generators $A = A^{w,K}$ such that $A^{w,K}w = 0$.

Assuming the DBC again, Eq. (4.1) can be written in the symmetric form

$$\dot{c}_n = \sum_{i:i \neq n} \kappa_{ni} \left(\left(\frac{w_n}{w_i}\right)^{1/2} c_i - \left(\frac{w_i}{w_n}\right)^{1/2} c_n \right) \quad \text{for } n \in \mathcal{I}. \quad (4.3)$$

Moreover, we see that A and w satisfy the DBC $A_{in}w_n = A_{ni}w_i$ if and only if K is symmetric. Thus, fixing a symmetric K and changing w does automatically generate the DBC for $A^{K,w}$ and w .

4.2. A general class of gradient structures

We now assume the DBC $A\mathbb{D}_w = (A\mathbb{D}_w)^*$ or equivalently $K = K^*$ in (4.2) and discuss a general class of gradient structures for (4.1) following the general approach in Sec. 2.5 in Ref. 23.

Let $\Phi : [0, \infty[\rightarrow [0, \infty[$ and $\Psi_{in} : \mathbb{R} \rightarrow [0, \infty[$ for $1 \leq i < n \leq I$ be lower semi-continuous and strictly convex C^2 functions such that $\Psi_{in}(0) = 0$ and $\Psi''_{in}(0) > 0$. We search for a gradient system $(\mathbb{Q}, \mathcal{E}, \mathcal{R}^*)$ with an energy functional \mathcal{E} and a dual dissipation potential in the form

$$\mathcal{E}(c) = \sum_{i=1}^I w_i \Phi \left(\frac{c_i}{w_i}\right) \quad \text{and} \quad \mathcal{R}^*(c, \xi) = \sum_{i=1}^{I-1} \sum_{n=i+1}^I a_{in}(c) \Psi_{in}(\xi_i - \xi_n),$$

where the coefficient functions a_{in} must be chosen appropriately, but need to be nonnegative to guarantee that $\mathcal{R}^*(c, \cdot)$ is a dissipation potential.

With $\partial_{\xi_n} \mathcal{R}^*(c, \xi) = \sum_{k=n+1}^I a_{nk}(c) \Psi'_{nk}(\xi_n - \xi_k) - \sum_{i=1}^{n-1} a_{in}(c) \Psi'_{in}(\xi_i - \xi_n)$ and $D\mathcal{E}(c) = (\Phi'(\frac{c_k}{w_k}))_k$ we find the relation

$$\begin{aligned} \partial_{\xi_n} \mathcal{R}^*(c, -D\mathcal{E}(c)) &= \sum_{i=n+1}^I a_{ni}(c) \Psi'_{ni} \left(\Phi' \left(\frac{c_i}{w_i} \right) - \Phi' \left(\frac{c_n}{w_n} \right) \right) \\ &\quad - \sum_{i=1}^{n-1} a_{in}(c) \Psi'_{in} \left(\Phi' \left(\frac{c_n}{w_n} \right) - \Phi' \left(\frac{c_i}{w_i} \right) \right). \end{aligned}$$

Thus, the equations $\dot{c}_n = \partial_{\xi_n} \mathcal{R}^*(c, -D\mathcal{E}(c))$ are the same as in (4.3), provided we choose the coefficient functions a_{in} as

$$\begin{aligned} a_{ni}(c) &:= \frac{\kappa_{ni} \sqrt{w_n w_i} \left(\frac{c_i}{w_i} - \frac{c_n}{w_n} \right)}{\Psi'_{ni} \left(\Phi' \left(\frac{c_i}{w_i} \right) - \Phi' \left(\frac{c_n}{w_n} \right) \right)} \quad \text{for } \frac{c_i}{w_i} \neq \frac{c_n}{w_n} \quad \text{and} \\ a_{ni}(c) &:= \frac{\kappa_{ni} \sqrt{w_n w_i}}{\Psi''_{ni}(0) \Phi'' \left(\frac{c_i}{w_i} \right)} \quad \text{for } \frac{c_i}{w_i} = \frac{c_n}{w_n} \end{aligned} \tag{4.4}$$

and exploit the DBC $\kappa_{in} = \kappa_{ni}$. We also emphasize that Φ' is strictly increasing such that $\frac{c_i}{w_i} - \frac{c_n}{w_n}$ and $\Phi'(\frac{c_i}{w_i}) - \Phi'(\frac{c_n}{w_n})$ always have the same sign. Since $\Psi'(\zeta)$ and ζ also always have the same sign, we conclude that $a_{in}(c) \geq 0$ as desired for dissipation potentials.

As the choice of entropy functional density Φ and of the dual dissipation potentials Ψ_{in} is general quite arbitrary we see that we can generate a whole zoo of different gradient structures for (4.1) or (4.3). The following choices relate to situation where all Ψ_{in} are given by one function Ψ , but more general cases are possible.

From the construction it is clear that \mathcal{R}^* is linear in the generator A , i.e. if $A = A^1 + A^2$ and the equilibrium w is fixed, then $\mathcal{R}^* = \mathcal{R}^*_{A^1} + \mathcal{R}^*_{A^2}$ where $\mathcal{R}^*_{A^m}$ is constructed as above.

4.3. Some specific gradient structures for linear reaction systems

We now realize special choices for the general gradient structures in the previous subsection. These choices are singled out because they lead to natural entropy functionals and relatively simple coefficient functions a_{in} in (4.4).

4.3.1. Quadratic energy and dissipation

The quadratic gradient structure is given by quadratic energy and dissipation, i.e.

$$\Phi_{\text{quad}}(\varrho) = \frac{1}{2} \varrho^2 \quad \text{and} \quad \Psi_{\text{quad}}(\zeta) = \frac{1}{2} \zeta^2.$$

The coefficient functions are constant and read $a_{in}(c) = \kappa_{in}\sqrt{w_i w_n}$. Thus, we find

$$\mathcal{E}_{\text{quad}}(c) = \frac{1}{2} \sum_{i=1}^I \frac{c_i^2}{w_i} \quad \text{and}$$

$$\mathcal{R}_{\text{quad}}^*(c, \xi) = \frac{1}{2} \sum_{i=1}^{I-1} \sum_{n=i+1}^I \kappa_{in} \sqrt{w_i w_n} (\xi_i - \xi_n)^2 = \frac{1}{2} \langle \xi, \mathbb{K}_{\text{quad}} \xi \rangle.$$

In this case the dual dissipation functional does not depend on the concentration $c \in \mathbb{Q}$, which means that the equation $\dot{c} = Ac = -\mathbb{K}D\mathcal{E}(c)$ can be treated as a self-adjoint linear evolution problem in the Hilbert space with the norm induced by \mathcal{R} . This leads to the classical Hilbert space approach for reversible Markov operators.

4.3.2. Boltzmann entropy and quadratic dissipation

The *quadratic-entropic* gradient structure is defined by the choices

$$\Phi_{\text{Boltzmann}}(\varrho) = \lambda_{\text{Bz}}(\varrho) := \varrho \log \varrho - \varrho + 1 \quad \text{and} \quad \Psi_{\text{quad}}(\zeta) = \frac{1}{2} \zeta^2.$$

This gradient structure was first introduced in Refs. 25, 22, 10, 5 and 26 as a possible generalization of Otto's gradient structure for the Fokker–Planck and more general diffusion equations equation, cf. Refs. 18 and 33. However, similar structures also appear earlier in the physics literature, see e.g. Eq. (113) in Ref. 32.

The associated entropy is Boltzmann's relative entropy and, using the logarithmic mean $\Lambda(a, b) = \int_0^1 a^s b^{1-s} \, ds = \frac{a-b}{\log a - \log b}$, the dual dissipation potential \mathcal{R}^* reads

$$\mathcal{E}_{\text{Bz}}(c) = \sum_{i=1}^I w_i \lambda_{\text{Bz}}\left(\frac{c_i}{w_i}\right) \quad \text{and}$$

$$\mathcal{R}^*(c, \xi) = \frac{1}{2} \sum_{i=1}^{I-1} \sum_{n=i+1}^I \kappa_{in} \sqrt{w_i w_n} \Lambda\left(\frac{c_i}{w_i}, \frac{c_n}{w_n}\right) (\xi_i - \xi_n)^2.$$

Again \mathcal{R}^* is quadratic in ξ but now also depends nontrivially on $c \in \mathbb{Q}$, viz. $\mathcal{R}^*(c, \xi) = \frac{1}{2} \langle \xi, \mathbb{K}_{\text{Bz}}(c) \xi \rangle$. This means that \mathbb{Q} can be equipped with the Riemannian metric induced by \mathcal{R} , see Ref. 22.

Note that $\mathbb{K}_{\text{Bz}}(w) = \mathbb{K}_{\text{quad}}$ and $\mathcal{E}_{\text{quad}}(c) = \frac{1}{2} D^2 \mathcal{E}_{\text{Bz}}(w)[c, c]$, which is the desired compatibility under linearization at $c = w$.

4.3.3. Boltzmann entropy and cosh-type dissipation

The following, so-called *entropic cosh-type gradient structure*, was derived via a large-deviation principle from an interacting particle system in Refs. 30 and 29. We refer to Marcellin's PhD thesis²⁴ from 1915 for a historical, first physical derivation of exponential kinetic relations in the context of Boltzmann statistics. Only little of

this important result penetrated into the main stream thermomechanical modeling of reaction systems, see Item iii on p. 77 and Eq. (69) in Ref. 16 for a discussion.

For this gradient structure the choices are

$$\begin{aligned} \Phi_{\text{Boltzmann}}(\varrho) &= \lambda_{\text{Bz}}(\varrho) := \varrho \log \varrho - \varrho + 1 \quad \text{and} \\ \Psi_{\text{cosh}}(\zeta) &= \text{C}^*(\zeta) := 4 \cosh\left(\frac{\zeta}{2}\right) - 4, \end{aligned}$$

giving Boltzmann’s relative entropy \mathcal{E}_{Bz} and the cosh-type dual dissipation potential:

$$\begin{aligned} \mathcal{E}_{\text{Bz}}(c) &:= \sum_{i=1}^I w_i \lambda_{\text{Bz}}\left(\frac{c_i}{w_i}\right) \quad \text{and} \\ \mathcal{R}_{\text{cosh}}^*(c, \xi) &= \sum_{i=1}^{I-1} \sum_{n=i+1}^I \kappa_{in} \sqrt{c_i c_n} \text{C}^*(\xi_i - \xi_n). \end{aligned} \tag{4.5}$$

The especially simple form of the coefficient functions arises from the interaction of the cosh function with the Boltzmann function λ_{Bz} , namely

$$\text{C}^{*'}(\lambda'_{\text{Bz}}(p) - \lambda'_{\text{Bz}}(q)) = 2 \sinh(\log \sqrt{p/q}) = \sqrt{p/q} - \sqrt{q/p} = \frac{p-q}{\sqrt{pq}}.$$

With this we easily find the simple formula $a_{in}(c) = \kappa_{in} \sqrt{c_i c_n}$.

Because of the close connection between the cosh-type function C^* and the Boltzmann function λ_{Bz} , it is obvious that using C^* means that we also use the Boltzmann entropy. Hence, it will not lead to confusion if we simply call $(\mathbb{Q}, \mathcal{E}_{\text{Bz}}, \mathcal{R}_{\text{cosh}})$ the *cosh gradient structure*. Again, the quadratic gradient structure in Sec. 4.3.1 is obtained by linearization:

$$\mathcal{E}_{\text{quad}}(c) = \frac{1}{2} \text{D}^2 \mathcal{E}_{\text{Bz}}(w)[c, c] \quad \text{and} \quad \mathbb{K}_{\text{quad}} = \text{D}_{\xi}^2 \mathcal{R}_{\text{cosh}}^*(w, 0).$$

4.4. *Tilting of Markov processes*

Tilting, also called exponential tilting, is a standard procedure in stochastics (in particular in the theory of large deviations) to change the dynamics of a Markov process in a controlled way. In particular, the equilibrium measure w is changed into another one, let us say \tilde{w} . For more motivation and theory we refer to Ref. 28 and the references therein.

Defining two entropy functionals, namely the Boltzmann entropies for w and \tilde{w} ,

$$\mathcal{E}_{\text{Bz}}(c) = \sum_{i=1}^I w_i \lambda_{\text{Bz}}\left(\frac{c_i}{w_i}\right) \quad \text{and} \quad \tilde{\mathcal{E}}_{\text{Bz}}(c) = \sum_{i=1}^I \tilde{w}_i \lambda_{\text{Bz}}\left(\frac{c_i}{\tilde{w}_i}\right)$$

the special structure of λ_{Bz} leads to the relation

$$\tilde{\mathcal{E}}_{\text{Bz}}(c) = \mathcal{E}_{\text{Bz}}(c) - \langle \eta, c \rangle \quad \text{with} \quad \eta = (\log(\tilde{w}_i/w_i))_{i \in \mathcal{I}}.$$

Thus, we see that a change of the equilibrium measure leads to a tilt in the sense of (3.5) for the entropy. Moreover, for every tilt $\eta \in X^*$ there is a unique new equilibrium state w^η , namely the minimizer of $c \mapsto \mathcal{E}^\eta(c) = \mathcal{E}_{Bz}(c) - \langle \eta, c \rangle$. We easily find

$$w_i^\eta = \frac{1}{Z} e^{-\eta_i} w_i \quad \text{with } Z = \sum_{n=1}^I e^{-\eta_n} w_n.$$

This explains the name “exponential tilting”.

For a time-dependent linear reaction systems the tilting is defined in a consistent way, namely using the representation (4.2). Given $\dot{c} = Ac$ with positive equilibrium w and a tilt η we first construct the equilibrium w^η and then, using $K = (\kappa_{in})$ from (4.2), we define the evolution

$$\dot{c} = A^\eta c \quad \text{with } A^\eta := \mathbb{D}_{w^\eta}^{1/2} K \mathbb{D}_{w^\eta}^{-1/2} - \mathbb{D}_{b^\eta}. \tag{4.6}$$

One of the important observations in Ref. 28 is that the cosh gradient structure is invariant under tilting, i.e. the dissipation potential does not change if the Boltzmann entropy is tilted. This can now be formulated as follows:

$$A^\eta c = D_\xi \mathcal{R}_{\cosh}^*(c, -D\mathcal{E}^\eta(c)). \tag{4.7}$$

This relation can easily checked by noting that (4.6) has the form (4.3), where now w is replaced by w^η . But \mathcal{E}^η is exactly the relative entropy with respect to w^η such that the results in Sec. 4.3.3 yield identity (4.7).

Using the formula (4.4) for $a_{in}(c)$ we can find all possible gradient structures in terms of Φ and Ψ_{in} such that the $a_{in}(c)$ is independent for w . The result shows that, up to a trivial scaling, the only tilt-invariant gradient structures in the form of Sec. 4.2 are given by the cosh gradient structure. Indeed, in Ref. 30 the case $\gamma = 1/2$ is obtained from the theory of large deviations.

Proposition 4.1. (Characterization of tilt-invariant gradient structures) *If Φ and Ψ_{in} are such that a_{in} in (4.4) is independent of w , then there exists $\varphi_0, \varphi_1 \in \mathbb{R}$ and $\psi_{in}, \gamma > 0$ such that*

$$\Phi(c) = \gamma \lambda_{Bz}(c) + \varphi_0 + \varphi_1 c \quad \text{and} \quad \Psi_{in}(\zeta) = \gamma \psi_{in} C^* \left(\frac{\zeta}{\gamma} \right).$$

In particular, we always obtain $a_{in}(c) = \frac{\kappa_{in}}{\psi_{in}} \sqrt{c_i c_n}$. Since ψ_{in} can be integrated into κ_{in} , all tilt-invariant gradient structures are given by scaled cosh gradient structures

$$\mathcal{E}(c) = \gamma \mathcal{E}_{Bz}(c) + \varphi_0 I + \varphi_1 \quad \text{and} \quad \mathcal{R}^*(c, \xi) = \gamma \mathcal{R}_{\cosh}^* \left(c, \frac{1}{\gamma} \xi \right).$$

Proof. We rewrite a_{in} in the form

$$a_{in}(c) = \kappa_{in} \sqrt{c_i c_n} \frac{\varrho_i - \varrho_n}{\sqrt{\varrho_i \varrho_n} \Psi'_{in}(\Phi'(\varrho_i) - \Phi'(\varrho_n))}, \quad \text{where } \varrho_k = \frac{c_k}{w_k}$$

Because the expression has to be independent of w_i and w_n for all $c, w \in \mathbb{Q}$, the fraction involving ϱ_i and ϱ_n has to be a constant, which we set $1/\psi_{in}$, i.e.

$$(i) \quad \Phi'(\varrho_i) - \Phi'(\varrho_n) = G\left(\frac{\varrho_i}{\varrho_n}\right), \quad (ii) \quad G(\sigma) = (\Psi'_{in})^{-1}\left(\psi_{in}\left(\sqrt{\sigma} - \frac{1}{\sqrt{\sigma}}\right)\right).$$

Setting $r_k = \log \varrho_k$, $f(r) = \Phi'(e^r)$, and $g(s) = G(e^s)$ in (i), we arrive at the relation

$$f(r_i) - f(r_n) = g(r_i - r_n) \quad \text{for all } r_i, r_n \in \mathbb{R}.$$

As f and g are continuous the only solutions of this functional relation are $f(r) = \varphi_1 + \gamma r$ and $g(s) = \gamma s$ with $\varphi_1, \gamma \in \mathbb{R}$. This implies $\Phi'(\varrho) = \varphi_1 + \gamma \log \varrho$ and, hence, $\Phi(\varrho) = \varphi_0 + \varphi_1 \varrho + \gamma \lambda_{Bz}(\varrho)$. Strict convexity of Φ leads to the restriction $\gamma > 0$.

Solving (ii) with $G(\sigma) = \gamma \log \sigma =: \zeta$ yields

$$\Psi'_{in}(\zeta) = \psi_{in}(e^{\zeta/(2\gamma)} - e^{-\zeta/(2\gamma)}) = \psi_{in} 2 \sinh\left(\frac{\zeta}{2\gamma}\right) = \psi_{in} C^{*'}\left(\frac{\zeta}{\gamma}\right).$$

Because of $\Psi_{in}(0) = 0$ this determines Ψ_{in} uniquely, and the result is established. □

We also refer to Ref. 17 for the connections of the cosh gradient structure to the SQRA-discretization scheme for drift-diffusion systems.

5. EDP-Convergence and the Effective Gradient Structure

In this section, we fully concentrate on the cosh gradient structure, because only this gradient structure allows to prove EDP convergence with tilting.

Our energy functionals \mathcal{E}_ε are the relative Boltzmann entropies, while the dual dissipation potentials $\mathcal{R}_\varepsilon^*$ is the sum of a slow and a fast part:

$$\mathcal{E}_\varepsilon(c) = \sum_{i=1}^I w_i^\varepsilon \lambda_{Bz}\left(\frac{c_i}{w_i^\varepsilon}\right) \quad \text{and}$$

$$\mathcal{R}_\varepsilon^*(c, \xi) = \mathcal{R}_{S,\varepsilon}^*(c, \xi) + \frac{1}{\varepsilon} \mathcal{R}_{F,\varepsilon}^*(c, \xi), \quad \text{where}$$

$$\mathcal{R}_{Z,\varepsilon}^*(c, \xi) := \sum_{i=1}^{I-1} \sum_{n=i+1}^I \kappa_{in}^{Z,\varepsilon} \sqrt{c_i c_n} C^*(\xi_i - \xi_n) \quad \text{with}$$

$$\kappa_{in}^{Z,\varepsilon} = A_{in}^Z \sqrt{w_n^\varepsilon / w_i^\varepsilon}, \quad Z \in \{S, F\}.$$

Here, the ε -dependencies of the coefficients $\kappa_{in}^{S,\varepsilon}$ and $\kappa_{in}^{F,\varepsilon}$ is trivial in the sense that the limits for $\varepsilon \rightarrow 0$ exist. The real important term is the factor $1/\varepsilon$ in front of $\mathcal{R}_{F,\varepsilon}^*$.

The structure of this section is as follows. In Sec. 5.1, we present the main results concerning the Γ -convergence of \mathcal{E}_ε and \mathcal{D}_ε which then imply the EDP-convergence with tilting of $(\mathbb{Q}, \mathcal{E}_\varepsilon, \mathcal{R}_\varepsilon)$ to the limit system $(\mathbb{Q}, \mathcal{E}, \mathcal{R}_{\text{eff}})$. In Sec. 5.2, we show

that this provides a gradient structure for the limit equation (2.7), and moreover that we obtain the natural cosh gradient structure $(\hat{\mathbf{Q}}, \hat{\mathcal{E}}, \hat{\mathcal{R}})$ for the coarse-grained equation (2.8).

The remaining part of this section then provides the proof of the convergence $\mathfrak{D}_\varepsilon \xrightarrow{M} \mathfrak{D}_0$, namely the *a priori* estimates in Sec. 5.3, the liminf estimate in Sec. 5.4, and the construction of recovery sequences in Sec. 5.5.

5.1. Main theorem on EDP-convergence

We now study the limit for $\varepsilon \rightarrow 0$ of the family of gradient systems $((\mathbf{Q}, \mathcal{E}_\varepsilon, \mathcal{R}_\varepsilon^*))_{\varepsilon>0}$ by showing EDP-convergence with tilting for a suitable limit.

As a first, and trivial result we state the Mosco convergence of the energies, which follows immediately from our assumption (2.2b), i.e. $w^\varepsilon \rightarrow w^0$.

Proposition 5.1. *On $\mathbf{Q} = \text{Prob}(\mathcal{I})$, we have the uniform convergence $\mathcal{E}_\varepsilon \rightarrow \mathcal{E}_0$, where $\mathcal{E}_0(c) = \sum_{i=1}^I w_i^0 \lambda_{\text{Bz}}(c_i/w_i^0)$. In particular, we have $\mathcal{E}_\varepsilon \xrightarrow{M} \mathcal{E}_0$ on X .*

To have a proper functional analytic setting we let

$$L^2([0, T]; \mathbf{Q}) = \{c \in L^2([0, T]; \mathbb{R}^I) \mid c(t) \in \mathbf{Q} \text{ a.e. in } [0, T]\}$$

and use the weak and strong topology induced by $L^2([0, T]; \mathbb{R}^I)$. The dissipation functional \mathfrak{D}_ε is now defined via

$$\mathfrak{D}_\varepsilon(c) := \begin{cases} \int_0^T (\mathcal{R}_\varepsilon(c, \dot{c}) + \mathcal{R}_\varepsilon^*(c, -D\mathcal{E}_\varepsilon(c))) dt & \text{for } c \in W^{1,1}([0, T]; \mathbf{Q}), \\ \infty & \text{otherwise on } L^2([0, T]; \mathbf{Q}), \end{cases}$$

where $\mathcal{R}_\varepsilon(c, \cdot)$ is defined implicitly as Legendre transform of $\mathcal{R}_\varepsilon^*(c, \cdot)$. To see that \mathfrak{D}_ε is well defined, we derive suitable properties for \mathcal{R}_ε .

Proposition 5.2. (Properties of \mathcal{R}_ε) *Let $\mathcal{R}_\varepsilon : \mathbf{Q} \times X \rightarrow [0, \infty]$ be defined by $\mathcal{R}_\varepsilon(c, \cdot) = (\mathcal{R}_\varepsilon^*(c, \cdot))^*$. Then, $\mathcal{R}_\varepsilon : \mathbf{Q} \times X \rightarrow [0, \infty]$ is lower semicontinuous and jointly convex.*

Proof. Since $(c_i, c_n) \mapsto \sqrt{c_i c_n}$ is concave and $\xi \mapsto C(\xi_i - \xi_n)$ is convex, the mapping $\mathcal{R}^* : \mathbf{Q} \times X^* \rightarrow [0, \infty]$ is concave-convex and thus its partial conjugate is convex in (c, v) .

For the lower semicontinuity consider $(c_k, v_k) \rightarrow (c, v)$. Then, for all $\delta > 0$ there exist ξ_δ with $\mathcal{R}_\varepsilon(c, v) \leq \langle \xi_\delta, v \rangle - \mathcal{R}_\varepsilon^*(c_k, \xi_\delta) + \delta$. The definition of the Legendre transform yields

$$\mathcal{R}_\varepsilon(c_k, v_k) \geq \langle \xi_\delta, v_k \rangle - \mathcal{R}_\varepsilon^*(c_k, \xi_\delta) \xrightarrow{k \rightarrow \infty} \langle \xi_\delta, v \rangle - \mathcal{R}_\varepsilon^*(c, \xi_\delta) \geq \mathcal{R}_\varepsilon(c, v) - \delta,$$

where we used the continuity of $c \mapsto \mathcal{R}_\varepsilon^*(c, \xi)$. Since $\delta > 0$ was arbitrary, we find $\liminf_{k \rightarrow \infty} \mathcal{R}_\varepsilon(c_k, v_k) \geq \mathcal{R}_\varepsilon(c, v)$ as desired. □

To formulate the main Γ -convergence result for \mathfrak{D}_ε we define the effective dissipation $\mathcal{R}_{\text{eff}}^*$ beforehand. It can be understood as the formal limit of $\mathcal{R}_\varepsilon^*$ when taking $\varepsilon \rightarrow 0$. The slow part $\mathcal{R}_{S,\varepsilon}^*$ simply converges to its limit

$$\mathcal{R}_S^*(c, \xi) := \sum_{i=1}^{I-1} \sum_{n=i+1}^I \kappa_{in}^{S,0} \sqrt{c_i c_n} C^*(\xi_i - \xi_j) \quad \text{with}$$

$$\kappa_{in}^{S,0} = A_{in}^S \sqrt{w_n^0/w_i^0} = \lim_{\varepsilon \rightarrow 0} \kappa_{in}^{S,\varepsilon}.$$

For the fast part $\frac{1}{\varepsilon} \mathcal{R}_{F,\varepsilon}^*$ we obtain blow up, except for those ξ that lie in the subspace that is not affected by fast reactions. For this we set

$$\Xi = M^* Y^* = \text{range}(M^*) = \ker(M)^\perp := \{ \xi \in X^* \mid \forall v \in \ker(M) : \langle \xi, v \rangle = 0 \}.$$

and observe that by construction for all $\varepsilon > 0$ we have

$$\mathcal{R}_{F,\varepsilon}^*(c, \xi) = 0 \quad \text{for all } \xi \in \Xi. \tag{5.1}$$

Indeed, $\mathcal{R}_{F,\varepsilon}^*(c, \xi)$ contains $C^*(\xi_i - \xi_n)$ with a positive prefactor only if $i \sim_F n$, while $\xi \in \Xi$ implies $\xi_i = \xi_n$ in that case. Together we set

$$\mathcal{R}_{\text{eff}}^*(c, \xi) := \mathcal{R}_S^*(c, \xi) + \chi_\Xi(\xi), \quad \text{where } \chi_A(a) = \begin{cases} 0 & \text{for } a \in A, \\ \infty & \text{for } a \notin A. \end{cases} \tag{5.2}$$

The dual dissipation potential $\mathcal{R}_{\text{eff}}^*$ consists of two terms: The first term \mathcal{R}_S^* contains the information of the slow reactions in the limit $\varepsilon \rightarrow 0$. The second term χ_Ξ restricts the vector of chemical potentials $\xi = D\mathcal{E}_0(c)$ exactly in such a way that the microscopic equilibria of the fast reactions hold, i.e. $A^F c = 0$ or equivalently $Pc = c$, see below.

Because of this constraint, it is actually irrelevant how $\mathcal{R}_{\text{eff}}^*(c, \cdot) : \Xi \rightarrow [0, \infty]$ is defined for $c \notin \mathbf{Q}_{\text{eq}} = \mathbf{Q} \cap PX$.

We note that $\mathcal{R}_\varepsilon^*(c, \cdot)$ has a Mosco limit $\mathcal{R}_0^*(c, \cdot)$ that is not necessarily equal to $\mathcal{R}_{\text{eff}}^*(c, \cdot)$. For c on the boundary of \mathbf{Q} , where some of the c_i are 0, we may have $\mathcal{R}_{F,\varepsilon}^*(c, \xi) = 0$ for all ξ , which implies $\mathcal{R}_0^*(c, \xi) = \mathcal{R}_S^*(c, \xi)$ for these c and all $\xi \in \mathbb{R}^I$. However, the Γ -limit of \mathfrak{D}_ε yields $\mathcal{R}_{\text{eff}}^* \geq \mathcal{R}_0^*$.

Theorem 5.3. (Mosco convergence of \mathfrak{D}_ε) *On $L^2([0, T]; \mathbf{Q})$ we have $\mathfrak{D}_\varepsilon \xrightarrow{M} \mathfrak{D}_0$ with*

$$\mathfrak{D}_0(c) := \begin{cases} \int_0^T (\mathcal{R}_{\text{eff}}(c, \dot{c}) + \mathcal{R}_{\text{eff}}^*(c, -D\mathcal{E}_0(c))) dt & \text{for } c \in W^{1,1}([0, T]; \mathbf{Q}), \\ \infty & \text{otherwise in } L^2([0, T]; \mathbf{Q}), \end{cases} \tag{5.3}$$

where $\mathcal{R}_{\text{eff}}^*$ is given in (5.2) and leads to the primal dissipation potential

$$\mathcal{R}_{\text{eff}}(c, v) = \inf \{ \mathcal{R}_S(c, z) \mid z \in \mathbb{R}^I \text{ with } Mz = Mv \} \quad \text{for all } c \in \mathbf{Q}_{\text{eq}} = P\mathbf{Q}.$$

The proof of this theorem is the main part of this section and will be given in Secs. 5.3–5.5. Now, we want to use the above result to conclude the EDP-convergence with tilting. For this result, it is essential to study the dependence of the limit \mathfrak{D}_0 on the limit equilibrium measure w^0 . On the one hand, $\mathcal{E}_0(c)$ is the relative Boltzmann entropy of c with respect to w^0 , which provides a simple and well-behaved dependence on w^0 . On the other hand, $\mathcal{R}_{\text{eff}}^*$ is given through \mathcal{R}_S^* and χ_{Ξ} . The former only depends on $(\kappa_{in}^{S,0})_{i,n \in \mathcal{I}}$ and the latter depends only on $M \in \{0, 1\}^{J \times I}$. Thus, there is no dependence on w^0 at all. The proof relies on the fact that the two processes of (i) tilting with driving forces η and of (ii) taking the limit $\varepsilon \rightarrow 0$ commute.

Theorem 5.4. (EDP-convergence with tilting) *The gradient systems $(\mathbb{Q}, \mathcal{E}_\varepsilon, \mathcal{R}_\varepsilon)$ EDP-converge with tilting to the limit gradient structure $(\mathbb{Q}, \mathcal{E}_0, \mathcal{R}_{\text{eff}}^*)$. The closure of the domain of the limit gradient system in the sense of (3.2) is \mathbb{Q}_{eq} .*

Proof. Proposition 5.1 and Theorem 5.3 already provide the simple EDP convergence $(\mathbb{Q}, \mathcal{E}_\varepsilon, \mathcal{R}_\varepsilon^*) \xrightarrow{\text{EDP}} (\mathbb{Q}, \mathcal{E}_0, \mathcal{R}_{\text{eff}}^*)$. The domain is restricted by the conditions (i) that $D\mathcal{E}_0(c)$ exists, which means that $c_i > 0$ for all i , and (ii) that $D\mathcal{E}_0(c)$ lies in the domain of $\partial_\xi \mathcal{R}_{\text{eff}}^*(c, \cdot)$. The latter condition is equivalent to $D\mathcal{E}_0(c) \in \Xi$ or equivalently $c \in X_{\text{eq}}$.

For the tilted energies $\mathcal{E}_\varepsilon^\eta = \mathcal{E}_\varepsilon - \langle \eta, \cdot \rangle$ we obviously have $\mathcal{E}_\varepsilon^\eta \xrightarrow{M} \mathcal{E}_0^\eta$. We can now apply Theorem 5.3 once again for $\mathfrak{D}_\varepsilon^\eta$. Using the fact that \mathcal{E}^η is again a relative Boltzmann entropy with respect to the exponentially tilted equilibrium state $w^{\eta, \varepsilon}$ that satisfies $w^{\eta, \varepsilon} \rightarrow w^{\eta, 0}$. Thus, the Mosco limit \mathfrak{D}_0^η of $\mathfrak{D}_\varepsilon^\eta$ again exists and has the same form as \mathfrak{D}_0 in (5.3), but with $D\mathcal{E}_0(c)$ replaced by $D\mathcal{E}(c) - \eta$. In particular, \mathcal{R}_{eff} remains unchanged and EDP-convergence with tilting is established. \square

5.2. The limit and the coarse-grained gradient structure

Before going into the proof of Theorem 5.3 we connect the limit gradient systems with the limit equation (2.7). The gradient-flow equation for the limit gradient systems reads

$$\dot{c} \in \partial_\xi \mathcal{R}_{\text{eff}}^*(c, -D\mathcal{E}_0(c)) \quad \text{a.e. on } [0, T]. \tag{5.4}$$

Since $\mathcal{R}_{\text{eff}}^*$ is no longer smooth, we use the set-valued convex subdifferential ∂_ξ that satisfies, because of the continuity of \mathcal{R}_S^* , the sum rule

$$\partial_\xi \mathcal{R}_{\text{eff}}^*(c, \xi) = D_\xi \mathcal{R}_S^*(c, \xi) + \partial \chi_{\Xi}(\xi) \quad \text{with } \partial \chi_{\Xi}(\xi) = \begin{cases} \ker(M) & \text{for } \xi \in \Xi, \\ \emptyset & \text{for } \xi \notin \Xi, \end{cases}$$

where we used the relation $\Xi = \text{range}(M^*) = \ker(M)^\perp$.

On the one hand, (5.4) implies that $D\mathcal{E}_0(c) \in \Xi$ for a.a. $t \in [0, T]$. Recalling that the rows of $M \in \{0, 1\}^{J \times I}$ consists of vectors having the entry 1 in exactly one equivalent class $\alpha(j) \subset \mathcal{I}$ for \sim_F and 0 else, we have

$$\Xi = \text{range}(M^*) = \{ \xi \in \mathbb{R}^I \mid \forall j \in \mathcal{J} \ \forall i_1, i_2 \in \alpha(j) : \xi_{i_1} = \xi_{i_2} \}$$

we conclude

$$\begin{aligned} D\mathcal{E}_0(c) \in \Xi &\iff \forall j \in \mathcal{J} \forall i_1, i_2 \in \alpha(j) : \frac{c_{i_1}}{w_{i_1}^0} = \frac{c_{i_2}}{w_{i_2}^0} \\ &\iff c \in X_{\text{eq}} \iff A^F c = 0. \end{aligned}$$

One the other hand, by construction of the gradient structure the term $D_\xi \mathcal{R}_S^*(c, -D\mathcal{E}_0(c))$ generates exactly the term $A^S c$. Thus, (5.4) is equivalent to

$$\dot{c}(t) \in A^S c(t) + \ker(M), \quad A^F c(t) = 0 \quad \text{a.e. on } [0, T]. \tag{5.5}$$

Applying M to the first equation gives the limit equation (2.7) and the following result.

Proposition 5.5. (Gradient structure for limit equation) *The limit equation (2.7) is the gradient-flow equation generated by the limit gradient system $(\mathbf{Q}, \mathcal{E}_0, \mathcal{R}_{\text{eff}}^*)$.*

As a last step, we show that the gradient structure for the limit equation also provides a gradient structure for the coarse gradient equation (2.8) $\dot{\hat{c}} = MA^S N \hat{c}$ for the coarse-grained states $\hat{c} = M c \in \hat{\mathbf{Q}}$. For this we exploit the special relations derived for coarse graining via $M : X \rightarrow Y$ and reconstruction via $N : Y \rightarrow X$.

Theorem 5.6. (Gradient structure for coarse-grained equation) *The coarse-grained equation (2.8) (viz. $\dot{\hat{c}} = MA^S N \hat{c}$) is the gradient-flow equation generated by the coarse-grained gradient system $(\hat{\mathbf{Q}}, \hat{\mathcal{E}}, \hat{\mathcal{R}})$ given by*

$$\hat{\mathcal{E}}(\hat{c}) = \mathcal{E}_0(N\hat{c}) = \mathcal{H}_J(\hat{c} | \hat{w}) \quad \text{and} \quad \hat{\mathcal{R}}(\hat{c}, \hat{v}) = \mathcal{R}_{\text{eff}}(N\hat{c}, N\hat{v}).$$

Moreover, we have $\hat{\mathcal{R}}^*(\hat{c}, \hat{\xi}) = \mathcal{R}_{\text{eff}}^*(N\hat{c}, M^* \hat{\xi}) = \mathcal{R}_S^*(N\hat{c}, M^* \hat{\xi})$.

This result can be seen as an exact coarse graining in the sense of the formal approach developed in Sec. 6.1 in Ref. 23.

Before giving the proof of this result we want to highlight its relevance. First, we emphasize that the coarse-grained equation is again a linear reaction system, now in \mathbb{R}^J , i.e. the master equation for a Markov process on $\mathcal{J} = \{1, \dots, J\}$. Second, the coarse-grained energy functional is again the relative Boltzmann entropy, now with respect to the coarse-grained equilibrium $\hat{w} = M w^0$. Third, the coarse-grained dual dissipation potential is again given in terms of the function \mathbf{C}^* , i.e. the coarse-grained gradient system is again of cosh-type. In summary, the coarse-grained gradient structure $(\hat{\mathbf{Q}}, \hat{\mathcal{E}}, \hat{\mathcal{R}})$ is again a cosh gradient structure, see Proposition 5.7 below.

We refer to Sec. 3.3 in Ref. 20 for an example that shows that other gradient structures may not be stable under EDP-convergence. All these results rely on the special properties of M and N developed in Sec. 2.3. In particular, we use that the projection $P = NM : X \rightarrow X$ is a Markov operator, i.e. it maps \mathbf{Q} onto itself.

Proof of Theorem 5.6.

Step 1: $\hat{\mathcal{E}}$ is a relative entropy. We use the special form $N = \mathbb{D}_{w^0} M^* \mathbb{D}_{\hat{w}}$, which gives $(N\hat{c})_i = w_i^0 \hat{c}_j / \hat{w}_j$, where $i \in \alpha(j)$. With this and $\hat{w}_j = \sum_{i \in \alpha(j)} w_i^0$ we obtain

$$\begin{aligned} \hat{\mathcal{E}}(\hat{c}) &= \mathcal{E}_0(N\hat{c}) = \mathcal{H}_I(Nc | w^0) = \sum_i w_i^0 \lambda_{\text{Bz}} \left(\frac{(N\hat{c})_i}{w_i^0} \right) \\ &= \sum_{j=1}^J \sum_{i \in \alpha(j)} w_i^0 \lambda_{\text{Bz}} \left(\frac{\hat{c}_j}{\hat{w}_j} \right) = \sum_{j=1}^J \hat{w}_j \lambda_{\text{Bz}} \left(\frac{\hat{c}_j}{\hat{w}_j} \right) = \mathcal{H}_J(\hat{c} | \hat{w}). \end{aligned}$$

Step 2: Legendre-conjugate pair $\hat{\mathcal{R}}$ and $\hat{\mathcal{R}}^*$. We start from the formula for $\hat{\mathcal{R}}^*$ and calculate $\hat{\mathcal{R}}$ as follows. Using $MN = \text{id}_Y$ and $\Xi = M^* Y^*$, we obtain

$$\begin{aligned} \hat{\mathcal{R}}(\hat{c}, \hat{v}) &= \sup \{ \langle \hat{\xi}, MN\hat{v} \rangle_J - \hat{\mathcal{R}}^*(\hat{c}, \hat{\xi}) \mid \hat{\xi} \in Y^* \} \\ &= \sup \{ \langle M^* \hat{\xi}, N\hat{v} \rangle_I - \hat{\mathcal{R}}^*(\hat{c}, \hat{\xi}) \mid \hat{\xi} \in Y^* \} \\ &= \sup \{ \langle \xi, N\hat{v} \rangle_I - \mathcal{R}_S^*(N\hat{c}, \xi) \mid \xi \in M^* Y^* \} \\ &= \sup \{ \langle \xi, N\hat{v} \rangle_I - \mathcal{R}_S^*(N\hat{c}, \xi) - \chi_\Xi(\xi) \mid \xi \in X^* \} = \mathcal{R}_{\text{eff}}(N\hat{c}, N\hat{v}), \end{aligned}$$

where we use the definition of $\mathcal{R}_{\text{eff}}^*$ in (5.2).

Step 3: The gradient-flow equation for $(\hat{Q}, \hat{\mathcal{E}}, \hat{\mathcal{R}})$. We first observe

$$M^* N^* D\mathcal{E}_0(N\hat{c}) = D\mathcal{E}_0(N\hat{c}). \tag{5.6}$$

Indeed, let us define the component-wise log-function on \mathbb{R}^I , $\log : x \mapsto (\log(x_i))_{i=1, \dots, I}$. We have $D\mathcal{E}_0(c) = \log(\mathbb{D}_{w^0}^{-1} c)$. Hence, for $c = N\hat{c} = \mathbb{D}_{w^0} M^* \mathbb{D}_{\hat{w}}^{-1} \hat{c}$, we conclude

$$\begin{aligned} D\mathcal{E}_0(N\hat{c}) &= \log(\mathbb{D}_{w^0}^{-1} N\hat{c}) = \log(M^* \mathbb{D}_{\hat{w}}^{-1} \hat{c}) \\ &= M^* \log(\mathbb{D}_{\hat{w}}^{-1} \hat{c}) = M^* D\hat{\mathcal{E}}(\hat{c}) = M^* N^* D\mathcal{E}_0(N\hat{c}), \end{aligned}$$

where we used that $D\hat{\mathcal{E}}(\hat{c}) = N^* D\mathcal{E}_0(N\hat{c})$.

With $D\hat{\mathcal{E}}(\hat{c}) = N^* D\mathcal{E}_0(N\hat{c})$ and (5.6) the gradient flow for $(\hat{Q}, \hat{\mathcal{E}}, \hat{\mathcal{R}})$ reads

$$\begin{aligned} \dot{\hat{c}} &= \partial_{\hat{c}} \hat{\mathcal{R}}^*(\hat{c}, -D\hat{\mathcal{E}}(\hat{c})) = M \partial_{\hat{c}} \mathcal{R}_S^*(N\hat{c}, -M^* D\hat{\mathcal{E}}(\hat{c})) \\ &= M \partial_{\hat{c}} \mathcal{R}_S^*(N\hat{c}, -M^* N^* D\mathcal{E}_0(N\hat{c})) = M \partial_{\hat{c}} \mathcal{R}_S^*(N\hat{c}, -D\mathcal{E}_0(N\hat{c})) = M A^S N \hat{c}, \end{aligned}$$

where we used the identity $D_{\hat{c}} \mathcal{R}_S^*(c, -D\mathcal{E}_0(c)) = A^S c$, which holds for all c by the construction of our gradient structure. \square

In analogy to formula (2.9) providing the coefficients $\hat{A}_{j_1 j_2}$ of the coarse-grained generator $\hat{A} = M A^S N$ we can also give a formula for the tilting-invariant reaction intensities $\kappa_{i_1 i_2}^{S,0}$ to obtain the corresponding intensities $\hat{\kappa}_{j_1 j_2}$ for the coarse-grained

equation (2.8) by a suitable averaging. In particular, the gradient systems $(\hat{Q}, \hat{E}, \hat{\mathcal{R}})$ provides again a cosh gradient structure in the sense of Sec. 4.3.3.

Proposition 5.7. (Cosh structure of $\hat{\mathcal{R}}^*$) *The coarse-grained dual dissipation potential $\hat{\mathcal{R}}^*$ reads*

$$\hat{\mathcal{R}}^*(\hat{c}, \hat{\xi}) = \sum_{1 \leq j_1 < j_2 \leq J} \hat{\kappa}_{j_1, j_2} \sqrt{\hat{c}_{j_1} \hat{c}_{j_2}} \mathbf{C}^*(\hat{\xi}_{j_1} - \hat{\xi}_{j_2}) \quad \text{with}$$

$$\hat{\kappa}_{j_1, j_2} = \sum_{i_1 \in \alpha(j_1)} \sum_{i_2 \in \alpha(j_2)} \kappa_{i_1 i_2}^{S,0} \left(\frac{w_{i_1}^0 w_{i_2}^0}{\hat{w}_{j_1} \hat{w}_{j_2}} \right)^{1/2}.$$

Proof. Theorem 5.6 provides an explicit formula for $\hat{\mathcal{R}}^*$. Inserting the definitions of M and N and grouping according to equivalence classes will provide the result. Recalling the function $\phi : \mathcal{I} \rightarrow \mathcal{J}$ giving for each i the associated equivalence class $\alpha(\phi(i)) \subset \mathcal{I}$ we have $(N\hat{c})_i = w_i^0 \hat{c}_{\phi(i)} / \hat{w}_{\phi(i)}$ and $(M^*\hat{\xi})_i = \hat{\xi}_{\phi(i)}$ and find

$$\begin{aligned} \hat{\mathcal{R}}^*(\hat{c}, \hat{\xi}) &= \mathcal{R}_S^*(N\hat{c}, M^*\hat{\xi}) \\ &= \frac{1}{2} \sum_{i_1 \in \mathcal{I}} \sum_{i_2 \in \mathcal{I}} \kappa_{i_1 i_2}^{S,0} \left(\frac{w_{i_1}^0 c_{\phi(i_1)} w_{i_2}^0 c_{\phi(i_2)}}{\hat{w}_{\phi(i_1)} \hat{w}_{\phi(i_2)}} \right)^{1/2} \mathbf{C}^*(\hat{\xi}_{\phi(i_1)} - \hat{\xi}_{\phi(i_2)}) \\ &= \frac{1}{2} \sum_{j_1 \in \mathcal{J}} \sum_{j_2 \in \mathcal{J}} \sum_{i_1 \in \alpha(i_1)} \sum_{i_2 \in \alpha(j_2)} \kappa_{i_1 i_2}^{S,0} \left(\frac{w_{i_1}^0 c_{j_1} w_{i_2}^0 c_{j_2}}{\hat{w}_{j_1} \hat{w}_{j_2}} \right)^{1/2} \mathbf{C}^*(\hat{\xi}_{j_1} - \hat{\xi}_{j_2}). \end{aligned}$$

This shows the desired result. □

5.3. A priori bounds and compactness

We start the proof of the Γ -convergence for the dissipation functional \mathfrak{D}_ε on $L^2([0, T], \mathbb{Q})$ by deriving the necessary *a priori* bounds for proving the compactness for a family $(c^\varepsilon)_{\varepsilon > 0}$ of functions satisfying $\mathfrak{D}_\varepsilon(c^\varepsilon) \leq C < \infty$.

Clearly since for all $t \in [0, T]$ we have $c^\varepsilon(t) \in \mathbb{Q}$ we get immediately uniform L^∞ -bounds on c^ε . Hence, we have (after extracting a suitable subsequence, which is not relabeled) a weak limit $c^0 \in L^2([0, T], \mathbb{Q})$. We want to improve the convergence to strong convergence. Already in the proof of the convergence of the solutions c^ε in Sec. 2.5 it became clear that there are two different controls, namely (i) the tendency to go to microscopic equilibrium and (ii) the dissipation through the slow reactions. From (i) we will obtain control of the distance of c^ε from $X_{\text{eq}} = PX$ by estimating $(I - P)c^\varepsilon$, but we are not able to control $(I - P)\dot{c}^\varepsilon$. From (ii) we obtain an *a priori* bound for $P\dot{c}^\varepsilon$, and the major task is to show that these two complementary pieces of information are enough to obtain compactness.

Subsequently, we will drop ε in the notations for w^ε , $\kappa_{in}^{\alpha, \varepsilon}$, and $\mathcal{R}_{S, \varepsilon}$, and so on. Of course, we will keep the important factor $1/\varepsilon$ in $\mathcal{R}_\varepsilon^* = \mathcal{R}_S^* + \frac{1}{\varepsilon} \mathcal{R}_F^*$.

The following result shows the convergence of sequences to the subspace $X_{\text{eq}} = PX$ of microscopic equilibria. Recall the decomposition $X = X_{\text{eq}} \oplus X_{\text{fast}}$ from (2.6) and the projection $P = NM$ such that $X_{\text{eq}} = PX$ and $X_{\text{fast}} = (I - P)X$. In particular, the semi-norm $c \mapsto |(I - P)c|$ is equivalent to $c \mapsto \text{dist}(c, X_{\text{eq}})$.

Lemma 5.8. (Convergence in the direction of fast reactions) *Consider a sequence (c^ε) in $L^2([0, T], \mathbb{Q})$ with $\mathfrak{D}_\varepsilon(c^\varepsilon) \leq C_\mathfrak{D} < \infty$ and $c^\varepsilon \rightharpoonup c^0$ in $L^2([0, T], \mathbb{R}^I)$. Then, there is a constant $C > 0$ such that*

$$\int_0^T |(I - P)c^\varepsilon(t)|^2 dt \leq C\varepsilon.$$

In particular, we have $c^0(t) \in \mathbb{Q}_{\text{eq}} = PQ$ for a.a. $t \in [0, T]$.

Proof. The bound on the dissipation functional $\mathfrak{D}_\varepsilon, \mathcal{R}_\varepsilon \geq 0, \mathcal{R}_S^* \geq 0$ and the relation $C^*(\log p - \log q) = 2(\sqrt{p/q} + \sqrt{q/p} - 2)$ imply

$$C_\mathfrak{D} \geq \mathfrak{D}_\varepsilon(c^\varepsilon) \geq \frac{1}{\varepsilon} \int_0^T \sum_{(i,n) \in \mathbb{F}} \frac{4\kappa_{in}^F}{\sqrt{w_i w_n}} \left(\sqrt{\frac{c_i^\varepsilon}{w_i}} - \sqrt{\frac{c_n^\varepsilon}{w_n}} \right)^2 dt,$$

where the set \mathbb{F} is given in term of the equivalence relation \sim_F , viz.

$$\mathbb{F} := \{(i, n) \in \mathcal{I} \times \mathcal{I} \mid i \sim_F n \text{ and } i < n\}.$$

Using the decomposition $X = X_{\text{eq}} \oplus X_{\text{fast}}$ from (2.6), we see that the semi-norm

$$\|c\|_{\mathbb{F}} := \left(\sum_{(i,n) \in \mathbb{F}} \left(\frac{c_i}{w_i} - \frac{c_n}{w_n} \right)^2 \right)^{1/2}$$

defines a norm on X_{fast} and there exists $C_2 > 0$ such that $|(I - P)c| \leq C_2 \|c\|_{\mathbb{F}}$ on \mathbb{Q} .

Denoting by $\underline{w} > 0$ and $\underline{\kappa} > 0$ lower bounds for all w_i^ε and all κ_{in}^F with $i \sim_F n$, respectively, we obtain the estimate

$$\begin{aligned} \int_0^T |(I - P)c^\varepsilon(t)|^2 dt &\leq C_2^2 \int_0^T \|c^\varepsilon(t)\|_{\mathbb{F}}^2 dt \\ &\leq C_2^2 \int_0^T \sum_{(i,n) \in \mathbb{F}} \left(\sqrt{\frac{c_i^\varepsilon}{w_i}} - \sqrt{\frac{c_n^\varepsilon}{w_n}} \right)^2 \left(\sqrt{\frac{c_i^\varepsilon}{w_i}} + \sqrt{\frac{c_n^\varepsilon}{w_n}} \right)^2 dt \\ &\leq \frac{C_2^2}{\underline{w}^2 \underline{\kappa}} \int_0^T \sum_{(i,n) \in \mathbb{F}} \frac{4\kappa_{in}^F}{\sqrt{w_i w_n}} \left(\sqrt{\frac{c_i^\varepsilon}{w_i}} - \sqrt{\frac{c_n^\varepsilon}{w_n}} \right)^2 dt \leq \frac{C_2^2}{\underline{w}^2 \underline{\kappa}} C_\mathfrak{D} \varepsilon. \end{aligned}$$

By weak lower semicontinuity of semi-norms we find $\int_0^T |(I - P)c^0(t)|^2 dt = 0$ and conclude $c^0(t) = Pc^0(t)$ a.e. on $[0, T]$. This proves the result. \square

The next result shows that we are able to control the time derivative of Pc^ε . Using $\text{range}(P) = \text{range}(N)$ and $NM = \text{id}_Y$ it suffices to control $M\dot{c}^\varepsilon$. For this,

we show that $\mathcal{R}_\varepsilon(c, \cdot)$ restricted to PX has a uniform lower superlinear bound in terms of the superlinear function C , see (A.2).

Proposition 5.9. (Convergence in the direction of slow reactions) *Consider a sequence (c^ε) in $L^2([0, T], \mathbb{Q})$ with $\mathfrak{D}_\varepsilon(c^\varepsilon) \leq C_{\mathfrak{D}} < \infty$ and $c^\varepsilon \rightharpoonup c^0$ in $L^2([0, T]; X)$. Then, there is a constant $C_W > 0$ such that*

$$\int_0^T C \left(\frac{1}{C_W} |P\dot{c}^\varepsilon(t)| \right) dt \leq C_W. \tag{5.7}$$

Moreover, $Pc^\varepsilon \rightharpoonup Pc^0 \in W^{1,1}([0, T]; \mathbb{Q})$ and $Pc^\varepsilon \rightarrow Pc^0$ in $C^0([0, T]; P\mathbb{Q})$.

With Lemma 5.8 we have $c^\varepsilon \rightarrow c^0$ strongly in $L^2([0, T], \mathbb{Q})$ and $c^0 = Pc^0 \in W^{1,1}([0, T]; \mathbb{Q})$.

Proof. To show a lower bound for $\mathcal{R}_\varepsilon(c, Pv)$ we first derive an upper bound for $\mathcal{R}_\varepsilon^*(c, \tilde{\xi})$ for $\tilde{\xi} \in P^*X^*$. Use $\mathcal{R}_{F,\varepsilon}^*(c, \tilde{\xi}) = 0$ and set $\bar{\kappa} := \sup \{ \kappa_{in}^{S,\varepsilon} \mid 1 \leq i < n \leq I, \varepsilon \in]0, 1[\}$ to obtain

$$\mathcal{R}_\varepsilon^*(c, \tilde{\xi}) = \sum_{i < j} \kappa_{in}^{S,\varepsilon} \sqrt{c_i c_j} C^*(\tilde{\xi}_i - \tilde{\xi}_j) \leq \sum_{i < j} \bar{\kappa} \frac{1}{2} C^*(\sqrt{2}|\tilde{\xi}|) \leq aC^*(\sqrt{2}|\tilde{\xi}|)$$

with $a = I^2 \bar{\kappa} / 4$. Next, Legendre transform, $\mathcal{R}_{F,\varepsilon}^*(c, \tilde{\xi}) = 0$ by (5.1) and the bound $|P^*\xi| \leq C_P |\xi| / \sqrt{2}$ yield the lower bound

$$\begin{aligned} \mathcal{R}_\varepsilon(c, v) &\geq \sup \{ \langle \tilde{\xi}, v \rangle - \mathcal{R}_\varepsilon^*(c, \tilde{\xi}) \mid \tilde{\xi} \in P^*X^* \} \\ &= \sup \{ \langle P^*\xi, v \rangle - \mathcal{R}_{S,\varepsilon}^*(c, P^*\xi) \mid \xi \in X^* \} \\ &\geq \sup \{ \langle \xi, Pv \rangle - aC^*(\sqrt{2}|P^*\xi|) \mid \xi \in X^* \} \\ &\geq \sup \{ \langle \xi, Pv \rangle - aC^*(C_P|\xi|) \mid \xi \in X^* \} \\ &= aC(|Pv| / (aC_P)). \end{aligned}$$

Applying this to $v = \dot{c}^\varepsilon$ we find

$$\int_0^T aC \left(\frac{|P\dot{c}^\varepsilon(t)|}{aC_P} \right) dt \leq \int_0^T \mathcal{R}_\varepsilon(c^\varepsilon(t), \dot{c}^\varepsilon(t)) dt \leq \mathfrak{D}_\varepsilon(c^\varepsilon) \leq C_{\mathfrak{D}},$$

which gives (5.7) with $C_W = \max\{aC_P, C_{\mathfrak{D}}/a\}$.

With the superlinearity of C , we obtain $Pc^\varepsilon \rightharpoonup Pc^0$ in $W^{1,1}([0, T]; PX)$. Moreover, the sequence Pc^ε is also equicontinuous, which is seen as follows. By (5.7) and (A.2) we have $\int_0^T |P\dot{c}^\varepsilon(t)| \log(2 + |P\dot{c}^\varepsilon(t)|) dt \leq C_1$. For $R > 0$ we set $\Sigma(R, \varepsilon) = \{t \in [0, T] \mid |P\dot{c}^\varepsilon(t)| \geq R\}$. Thus, for $t_1 < t_2$ we obtain the estimate

$$\begin{aligned} &|Pc^\varepsilon(t_2) - Pc^\varepsilon(t_1)| \\ &\leq \int_{t_1}^{t_2} |P\dot{c}^\varepsilon(t)| dt \end{aligned}$$

$$\begin{aligned} &\leq \int_{[t_1, t_2] \setminus \Sigma(R, \varepsilon)} |P\dot{c}^\varepsilon(t)| dt + \int_{\Sigma(R, \varepsilon)} |P\dot{c}^\varepsilon(t)| \frac{\log(2 + |P\dot{c}^\varepsilon(t)|)}{\log(2 + R)} dt \\ &\leq (t_2 - t_1)R + \frac{C_1}{\log(2 + R)}. \end{aligned}$$

The last sum can be made smaller than any $\varepsilon > 0$ by choosing first $R = R(\varepsilon) := \exp(2C_1/\varepsilon)$ and then assuming $t_2 - t_1 < \delta(\varepsilon) := \varepsilon/(2R(\varepsilon))$. This shows $|Pc^\varepsilon(t_2) - Pc^\varepsilon(t_1)| < \varepsilon$ whenever $|t_2 - t_1| < \delta(\varepsilon)$, which is the desired equicontinuity. By the Arzelà–Ascoli theorem we obtain uniform convergence.

The final convergence follows from $c^\varepsilon = Pc^\varepsilon + (I - P)c^\varepsilon$ via Lemma 5.8, and the last statement from $Pc^0(t) = c^0(t)$ a.e. in $[0, T]$. □

5.4. The liminf estimate

For the limit passage $\varepsilon \rightarrow 0$ we use a technique, which was introduced formally in Ref. 20 and exploited in Ref. 23 for the study of the large-volume limit in chemical master equations. It relies on the idea that the velocity part $\mathfrak{D}_\varepsilon^{\text{vel}} = \int \mathcal{R}_\varepsilon dt$ of the dissipation functional \mathfrak{D}_ε can be characterized by Legendre transform using a classical result of Rockafellar:

Theorem 5.10. (Theorem 2 in Ref. 36) *Let $f : [0, T] \times \mathbb{R}^n \rightarrow \mathbb{R}_\infty$ be a normal, convex integrand and with conjugate f^* . Assume there exist $u_\circ \in L^1([0, T]; \mathbb{R}^n)$ and $\xi_\circ \in L^\infty([0, T]; \mathbb{R}^n)$ such that $t \mapsto f(t, u_\circ(t))$ and $t \mapsto f^*(t, \xi_\circ(t))$ are integrable, then the functionals*

$$\begin{aligned} I_f : &\begin{cases} L^1([0, T]; \mathbb{R}^n) \rightarrow \mathbb{R}_\infty, \\ u \mapsto \int_0^T f(t, u(t)) dt \end{cases} \quad \text{and} \\ I_{f^*} : &\begin{cases} L^\infty([0, T]; \mathbb{R}^n) \rightarrow \mathbb{R}_\infty, \\ \eta \mapsto \int_0^T f^*(t, \eta(t)) dt \end{cases} \end{aligned}$$

are proper convex functionals that are conjugate to each other with respect to the dual pairing $(u, \eta) \mapsto \int_0^T \langle \xi(t), u(t) \rangle dt$, viz. for all $u \in L^1([0, T]; \mathbb{R}^n)$ we have

$$\begin{aligned} &\int_0^T f(t, u(t)) dt \\ &= \sup \left\{ \int_0^T (\langle \eta(t), u(t) \rangle - f^*(t, \eta(t))) dt \mid \eta \in L^\infty([0, T]; \mathbb{R}^n) \right\}. \end{aligned} \tag{5.8}$$

We apply this result with $f(t, u) = \mathcal{R}_\varepsilon(c(t), u)$ and obtain, for $\varepsilon \in [0, 1]$,

$$\begin{aligned} \mathfrak{D}_\varepsilon(c) &= \sup \{ \mathfrak{B}_\varepsilon(c, \dot{c}, \xi) \mid \xi \in L^\infty([0, T]; X^*) \}, \quad \text{where} \\ \mathfrak{B}_\varepsilon(c, u, \xi) &:= \mathfrak{B}_\varepsilon^{\text{vel}}(c, u, \xi) + \mathfrak{D}_\varepsilon^{\text{slope}}(c) \quad \text{with} \\ \mathfrak{B}_\varepsilon^{\text{vel}}(c, u, \xi) &:= \int_0^T (\langle \xi(t), u(t) \rangle - \mathcal{R}_\varepsilon^*(c(t), \xi(t))) \, dt \quad \text{and} \\ \mathfrak{D}_\varepsilon^{\text{slope}}(c) &= \int_0^T \mathcal{R}_\varepsilon^*(c, -D\mathcal{E}_\varepsilon(c(t))) \, dt. \end{aligned} \tag{5.9}$$

The assumptions are easily satisfied as we may choose $u_o \equiv 0$ and $\eta_o \equiv 0$.

With these preparations we obtain the liminf estimate in a straightforwardly.

Theorem 5.11. (Liminf estimate) *The weak convergence $c^\varepsilon \rightharpoonup c^0$ in $L^2([0, T]; Q)$ implies $\liminf_{\varepsilon \rightarrow 0} \mathfrak{D}_\varepsilon(c^\varepsilon) \geq \mathfrak{D}_0(c^0)$, where \mathfrak{D}_0 is defined via \mathcal{E}_0 and \mathcal{R}_{eff} in (5.3).*

Proof. We may assume that $\alpha_* := \liminf_{\varepsilon \rightarrow 0} \mathfrak{D}_\varepsilon(c^\varepsilon) < \infty$, since otherwise the desired estimate is trivially satisfied.

Step 1. Strong convergence and limit characterization: Using Proposition 5.9 gives

$$c^\varepsilon \rightarrow c^0 \quad \text{strongly in } L^2([0, T]; Q) \quad \text{and} \quad c^0 = Pc^0 \in W^{1,1}([0, T]; \mathbb{R}^I).$$

Step 2. Slope part: Because of $Pc^0(t) = c^0(t)$ we know $\xi_0(t) = D\mathcal{E}_0(c^0(t)) \in M^*X^*$ which implies $\chi_{\Xi}(-D\mathcal{E}_0(c^0(t))) = 0$ on $[0, T]$. Hence, dropping the nonnegative term $\mathcal{R}_{F,\varepsilon}^*(c^\varepsilon, -D\mathcal{E}_\varepsilon(c^\varepsilon(t)))$ and setting $\mathcal{S}_\varepsilon(c) := \mathcal{R}_{S,\varepsilon}^*(c, -D\mathcal{E}_\varepsilon(c))$ we obtain

$$\liminf_{\varepsilon \rightarrow 0} \mathfrak{D}_\varepsilon^{\text{slope}}(c^\varepsilon) \geq \liminf_{\varepsilon \rightarrow 0} \int_0^T \mathcal{S}_\varepsilon(c^\varepsilon(t)) \, dt \stackrel{*}{=} \int_0^T \mathcal{S}_0(c^0(t)) \, dt = \mathfrak{D}_0^{\text{slope}}(c^0).$$

In the passage $\stackrel{*}{=}$ we use the strong convergence $c^\varepsilon \rightarrow c^0$ and the continuity of

$$\begin{aligned} [0, 1] \times Q \ni (\varepsilon, c) &\mapsto \mathcal{S}_\varepsilon(c) = \mathcal{R}_{S,\varepsilon}^*(c, -D\mathcal{E}_\varepsilon(c)) \\ &= \sum_{i < n} \frac{4\kappa_{in}^{S,\varepsilon}}{w_i^\varepsilon w_n^\varepsilon} \left(\sqrt{\frac{c_i}{w_i^\varepsilon}} - \sqrt{\frac{c_n}{w_n^\varepsilon}} \right)^2. \end{aligned} \tag{5.10}$$

Step 3. Velocity part: We exploit the Rockafellar representation (5.9) together with the fact that $\dot{c}^0(t) = P\dot{c}^0(t)$ a.e. in $[0, T]$. The latter condition allows us to test only by functions $\xi = P^*\xi \in L^\infty([0, T]; X^*)$, which leads to the estimate

$$\begin{aligned} &\liminf_{\varepsilon \rightarrow 0} \mathfrak{D}_\varepsilon^{\text{vel}}(c^\varepsilon) \\ &\geq \liminf_{\varepsilon \rightarrow 0} \mathfrak{B}_\varepsilon^{\text{vel}}(c^\varepsilon, \dot{c}^\varepsilon, P^*\xi) \stackrel{\text{a}}{=} \liminf_{\varepsilon \rightarrow 0} \int_0^T (\langle \xi, P\dot{c}^\varepsilon \rangle - \mathcal{R}_{S,\varepsilon}^*(c^\varepsilon, \xi)) \, dt \\ &\stackrel{\text{b}}{=} \int_0^T (\langle \xi, P\dot{c}^0 \rangle - \mathcal{R}_S^*(c^0, \xi)) \, dt = \mathfrak{B}_0^{\text{vel}}(c^0, \dot{c}^0, \xi), \end{aligned}$$

where in $\stackrel{a}{=}$ we used $\mathcal{R}_\varepsilon^*(c, \xi) = \mathcal{R}_{S, \varepsilon}^*(c, \xi)$ whenever $\xi = P^*\xi$, see (5.1). In $\stackrel{b}{=}$ we exploited the weak convergence $Pc^\varepsilon \rightharpoonup Pc^0$ established in Proposition 5.9 as well as the strong convergence $c^\varepsilon \rightarrow c^0$ together with the continuity of $(\varepsilon, c) \mapsto \mathcal{R}_{S, \varepsilon}^*(c, \xi)$.

Now we exploit Rockafellar’s characterization (5.9) to return to $\mathfrak{D}_0^{\text{vel}}(c^0)$, namely

$$\begin{aligned} \mathfrak{D}_0^{\text{vel}}(c^0) &= \sup \{ \mathfrak{B}_0^{\text{vel}}(c^0, \dot{c}^0, \xi) \mid \xi \in L^\infty([0, T]; X^*) \} \\ &= \sup \left\{ \int_0^T (\langle \xi, \dot{c}^0 \rangle - \mathcal{R}_S^*(c^0, \xi) - \chi_\Xi(\xi)) \, dt \mid \xi \in L^\infty([0, T]; X^*) \right\} \\ &= \sup \{ \mathfrak{B}_0^{\text{vel}}(c^0, \dot{c}^0, \xi) \mid \xi = P^*\xi \in L^\infty([0, T]; X^*) \}. \end{aligned}$$

With the above estimate we conclude $\liminf_{\varepsilon \rightarrow 0} \mathfrak{D}_\varepsilon^{\text{vel}}(c^\varepsilon) \geq \mathfrak{D}_0^{\text{vel}}(c^0)$.

Adding this to the estimate in Step 2 we obtain the full liminf estimate. □

5.5. Construction of the recovery sequence

Now we construct the recovery sequence for the Mosco-convergence of the dissipation functionals \mathfrak{D}_ε . This provides the required limsup estimate $\limsup_{\varepsilon \rightarrow 0} \mathfrak{D}_\varepsilon(c^\varepsilon) \leq \mathfrak{D}_0(c^0)$ along at least one sequence with the strong convergence $c^\varepsilon \rightarrow c^0$ in $L^2([0, T]; \mathbb{Q})$. For this we use in Step 2(b) an approximation result by piecewise affine functions \widehat{c}_N introduced in Theorem 2.6, Step 3 of Ref. 21 and adapted to state-dependent dissipation potentials in Corollary 3.3 in Ref. 2.

Theorem 5.12. (Recovery sequences) *For every $c^0 \in L^2([0, T]; \mathbb{Q})$ there exists a sequence $(c^\varepsilon)_{\varepsilon \in]0, 1[}$ with $c^\varepsilon \rightarrow c^0$ in $L^2([0, T]; \mathbb{Q})$ such that $\lim_{\varepsilon \rightarrow 0} \mathfrak{D}_\varepsilon(c^\varepsilon) = \mathfrak{D}_0(c^0)$.*

Proof. Step 1. The case $\mathfrak{D}_0(c^0) = \infty$. We choose the constant sequence $c^\varepsilon = c^0$ and claim $\mathfrak{D}_\varepsilon(c^\varepsilon) = \mathfrak{D}_\varepsilon(c^0) \rightarrow \infty$. Because of $\mathfrak{D}_0(c^0) = \infty$ one of the following conditions is false:

$$(i) \ c^0(t) \in \mathbb{Q}_{\text{eq}} \text{ a.e. in } [0, T] \quad \text{or} \quad (ii) \ C(|Pc^0(\cdot)|) \in L^1([0, T]).$$

If (i) is false, then $c^0(t) \notin \mathbb{Q}_{\text{eq}}$ for $t \in \mathcal{T} \subset [0, T]$, where $|\mathcal{T}| = \int_{\mathcal{T}} 1 \, dt > 0$. Setting $\mathcal{F}_\varepsilon(c) := \mathcal{R}_{F, \varepsilon}^*(c, -D\mathcal{E}_\varepsilon(c))$ we have

$$\begin{aligned} \mathfrak{D}_\varepsilon^{\text{slope}}(c^0) &= \int_0^T \left(\mathcal{R}_{S, \varepsilon}^*(c^0, -D\mathcal{E}_\varepsilon(c^0)) + \frac{1}{\varepsilon} \mathcal{R}_{F, \varepsilon}^*(c^0, -D\mathcal{E}_\varepsilon(c^0)) \right) dt \\ &\geq \frac{1}{\varepsilon} \int_0^T \mathcal{F}_\varepsilon(c^0(t)) \, dt. \end{aligned}$$

However, for $t \in \mathcal{T}$ we have $\mathcal{F}_\varepsilon(c^0(t)) \rightarrow \mathcal{F}_0(c^0(t)) > 0$. Thus, $\mathfrak{D}_\varepsilon^{\text{slope}}(c^0) \rightarrow \infty$ follows which implies $\mathfrak{D}_\varepsilon(c^0) \rightarrow \infty$.

If (ii) is false, then $\mathfrak{D}_\varepsilon^{\text{vel}}(c^0) = \infty$ for all $\varepsilon > 0$ and we are done.

Step 2. Preliminary recovery sequences for the case $\mathfrak{D}_0(c^0) < \infty$. In the sub-steps (a)–(c) we discuss three approximations for general c^0 .

Step 2(a). Positivity for the case $\varepsilon = 0$. We set $\tilde{c}_\delta(t) := \delta w^0 + (1 - \delta)c^0(t)$ and claim that $\mathfrak{D}_0(\tilde{c}_\delta) \rightarrow \mathfrak{D}_0(c^0) < \infty$ for $\delta \searrow 0$. As \mathfrak{D}_0 is convex and lower semicontinuous (cf. see Proposition 5.2), we have $\liminf_{\delta \searrow 0} \mathfrak{D}_0(\tilde{c}_\delta) \geq \mathfrak{D}_0(c^0)$.

Obviously, $\tilde{c}_\delta \geq (1 - \delta)c^0$ holds componentwise, and hence the explicit form of \mathcal{R}_0^* gives

$$\mathcal{R}_{\text{eff}}^*(\tilde{c}_\delta, \xi) \geq (1 - \delta)\mathcal{R}_{\text{eff}}^*(c^0, \xi), \quad \text{and thus } \mathcal{R}_{\text{eff}}(\tilde{c}_\delta, v) \leq (1 - \delta)\mathcal{R}_{\text{eff}}\left(c^0, \frac{1}{1 - \delta}v\right).$$

Inserting $v = \dot{\tilde{c}}_\delta = (1 - \delta)\dot{c}^0$ into the latter estimate gives

$$\mathfrak{D}_0^{\text{vel}}(\tilde{c}_\delta) = \int_0^T \mathcal{R}_{\text{eff}}(\tilde{c}_\delta, \dot{\tilde{c}}_\delta) dt \leq \int_0^T (1 - \delta)\mathcal{R}_{\text{eff}}(c^0, \dot{c}^0) dt = (1 - \delta)\mathfrak{D}_0^{\text{vel}}(c^0),$$

which proves the desired claim of Step 2(a), because $\mathfrak{D}_0^{\text{slope}}(\tilde{c}_\delta) \rightarrow \mathfrak{D}_0^{\text{slope}}(c^0)$ is trivial.

Step 2(b). We stay with $\varepsilon = 0$ and, by Step 2(a), may assume for some $c_* > 0$ that

$$c^0(t) \in \mathbf{Q}_{c_*} := \{c \in \mathbf{Q} \mid \forall i \in \mathcal{I}: c_i \geq c_*\} \text{ for all } t \in [0, T].$$

We now approximate c^0 by a function $\hat{c}_N \in W^{1,\infty}([0, T]; PX)$ still satisfying $\hat{c}_N(t) \in \mathbf{Q}_{c_*}$.

For $N \in \mathbb{N}$ we define $\hat{c}_N : [0, T] \rightarrow PX$ as the piecewise affine interpolant of the nodal points $\hat{c}_N(kT/N) = c^0(kT/N)$ for $k = 0, 1, \dots, N$. We also define the piecewise constant interpolant $\bar{c}_N : [0, T] \rightarrow \mathbf{Q}_{c_*}$ via $\bar{c}_N(t) = c^0(kT/N)$ for $t \in](k - 1)T/N, kT/N]$. Then, using $c^0 \in W^{1,1}([0, T]; PX) \subset C^0([0, T]; PX)$ we have

$$\begin{aligned} \hat{c}_N &\rightarrow c^0 \quad \text{in } W^{1,1}([0, T]; PX) \quad \text{and in } C^0([0, T]; PX) \quad \text{and} \\ \bar{c}_N &\rightarrow c^0 \quad \text{in } L^\infty([0, T]; PX). \end{aligned}$$

We now set

$$\alpha_N := \|c^0 - \hat{c}_N\|_{L^\infty} + \|c^0 - \bar{c}_N\|_{L^\infty}$$

and obtain $\alpha_N \rightarrow 0$.

These uniform estimates can be used in conjunction with the uniform continuity of $c \mapsto \mathcal{R}_{\text{eff}}^*(c, \xi)$ when restricted to \mathbf{Q}_{c_*} . Clearly $\mathbf{Q}_{c_*} \ni c \mapsto \sqrt{c_i c_n}$ is Lipschitz continuous, and we call the Lipschitz constant λ^* . The special form of $\mathcal{R}_{\text{eff}}^*$ then implies

$$\begin{aligned} \forall c, \tilde{c} \in \mathbf{Q}_{c_*} \quad \forall \xi \in X^* : & \quad |\mathcal{R}_{\text{eff}}^*(c, \xi) - \mathcal{R}_{\text{eff}}^*(\tilde{c}, \xi)| \\ & \leq \Lambda^* |c - \tilde{c}| \mathcal{R}_{\text{eff}}^*(c, \xi) \quad \text{with } \Lambda^* = \lambda^* \bar{\kappa}. \end{aligned}$$

Assuming $|c - \tilde{c}| \leq \alpha$ and $\Lambda^* \alpha < 1$ and applying the Legendre transform we find

$$(1 - \Lambda^* \alpha) \mathcal{R}_{\text{eff}}\left(c, \frac{1}{1 - \Lambda^* \alpha} v\right) \geq \mathcal{R}_{\text{eff}}(\tilde{c}, v) \geq (1 + \Lambda^* \alpha) \mathcal{R}_{\text{eff}}\left(c, \frac{1}{1 + \Lambda^* \alpha} v\right).$$

Exploiting the scaling property (A.4) we arrive at the estimates

$$\frac{1}{1 - \Lambda^* \alpha} \mathcal{R}_{\text{eff}}(c, v) \geq \mathcal{R}_{\text{eff}}(\tilde{c}, v) \geq \frac{1}{1 + \Lambda^* \alpha} \mathcal{R}_{\text{eff}}(c, v).$$

To estimate the velocity part of the dissipation functional as in Refs. 21 and 2 we introduce

$$\mathfrak{J}(c, v) := \int_0^T \mathcal{R}_{\text{eff}}(c(t), v(t)) dt,$$

which allows us to use different approximations for c^0 and for \dot{c}^0 . We obtain

$$\begin{aligned} \mathfrak{D}_0^{\text{vel}}(\hat{c}_N) &= \mathfrak{J}(\hat{c}_N, \hat{c}_N) \leq (1 + \Lambda^* \alpha_N) \mathfrak{J}(\bar{c}_N, \hat{c}_N) \\ &\stackrel{*}{\leq} (1 + \Lambda^* \alpha_N) \mathfrak{J}(\bar{c}_N, \dot{c}^0) \leq (1 + \Lambda^* \alpha_N)^2 \mathfrak{J}(c^0, \dot{c}^0) \\ &= (1 + \Lambda^* \alpha_N)^2 \mathfrak{D}_0^{\text{vel}}(c^0). \end{aligned}$$

For the estimate $\stackrel{*}{\leq}$ we split $[0, T]$ into the subintervals $S_k^N :=](k - 1)T/N, kT/N[$, where \bar{c}_N and \hat{c}_N are equal to the constants $c^0(kT/N)$ and $\frac{T}{N} \int_{S_N} \dot{c}^0(t) dt$, respectively. Then, Jensen’s inequality for the convex function $\mathcal{R}_{\text{eff}}(\bar{c}_N, \cdot)$ gives the desired estimate.

Since $\mathfrak{D}_0^{\text{slope}}(\hat{c}_N) \rightarrow \mathfrak{D}_0^{\text{slope}}(c^0)$ by the continuity of the integrand \mathcal{S}_0 (cf. (5.10)), the lower semicontinuity of \mathfrak{D}_0 yields $\mathfrak{D}_0(\hat{c}_N) \rightarrow \mathfrak{D}_0(c^0)$.

Step 2(c). Using the Steps 2(a) and 2(b), we now may assume $c^0 \in W^{1,\infty}([0, T], X)$ with $c^0(t) \in \mathbb{Q}_{c^*}$ on $[0, T]$ and define c^ε via the formula

$$c^\varepsilon(t) = \mathbb{D}_{w^\varepsilon} \mathbb{D}_{w^0}^{-1} c^0(t) \quad \text{for } t \in [0, T].$$

This definition gives $\text{D}\mathcal{E}_\varepsilon(c^\varepsilon(t)) \in \Xi$ and hence $\mathcal{S}_\varepsilon^F(c^\varepsilon(t)) = 0$. Hence, the definition of \mathcal{S}_ε in terms of the ratios c_i/w_i^ε (cf. (5.10)) implies $\mathfrak{D}_\varepsilon^{\text{slope}}(c^\varepsilon) \rightarrow \mathfrak{D}_0^{\text{slope}}(c^0)$.

For the velocity part we again use the Rockafellar characterization, namely

$$\mathfrak{D}_\varepsilon^{\text{vel}}(c^\varepsilon) = \sup \{ \mathfrak{B}_\varepsilon^{\text{vel}}(c^\varepsilon, \dot{c}^\varepsilon, \xi) \mid \xi \in L^\infty([0, T]; X^*) \}.$$

Because of the uniform bound for \dot{c}^ε in $L^\infty([0, T]; X)$ we are able to show that the supremum over $\mathfrak{B}_\varepsilon^{\text{vel}}(c^\varepsilon, \dot{c}^\varepsilon, \cdot)$ is attained by maximizers ξ^ε that are uniformly bounded in $L^\infty([0, T]; X^*)$. To see this, we firstly observe that for all $(c, \dot{c}) \in \mathbb{Q} \times X$ the functional $\mathfrak{B}_\varepsilon(c, \dot{c}, \cdot)$ is invariant under translation $\xi \mapsto \xi + c\mathbb{1}$, since $\langle \mathbb{1}, \dot{c} \rangle = 0$ and the dissipation potential $\mathcal{R}_\varepsilon^*(c, \xi)$ only depends on differences $\xi_i - \xi_n$. Hence, to generate compactness for maximizers ξ^ε we fix the first component by setting $\xi_1^\varepsilon = 0$. Then, by $c^0(t) \in \mathbb{Q}_{c^*}$ and the exponential growth of $\mathcal{R}_\varepsilon^*$ we obtain

$$\mathcal{R}_\varepsilon^*(c^\varepsilon(t), \xi) \geq c_o \sum_{i < n: A_{in}^\varepsilon > 0} |\xi_i - \xi_n|^2,$$

with a positive constant $c_o > 0$. By assumption (2.2a) on the connectivity of the reaction network, all vertices can be reached by a reaction path from vertex 1. Hence, there is another constant \tilde{c}_o , such that the estimate

$$\mathcal{R}_\varepsilon^*(c^\varepsilon(t), \xi) \geq c_o \sum_{i < n: A_{in}^\varepsilon > 0} |\xi_i - \xi_n|^2 \geq \tilde{c}_o \sum_{i > 1} |\xi_1 - \xi_i|^2$$

holds. Hence, the maximizers ξ^ε for $\mathfrak{D}_\varepsilon^{\text{vel}}(c^\varepsilon)$ with $\xi_1^\varepsilon = 0$ satisfy

$$\langle \xi^\varepsilon, \dot{c}^\varepsilon(t) \rangle - \mathcal{R}_\varepsilon^*(c^\varepsilon(t), \xi) \leq |\xi^\varepsilon| C \|\dot{c}^0\|_{L^\infty([0,T],X)} - \tilde{c}_0 |\xi^\varepsilon|^2,$$

which implies the uniform bound $\|\xi^\varepsilon\|_{L^\infty([0,T],X^*)} \leq C \|\dot{c}^0\|_{L^\infty([0,T],X)} / \tilde{c}_0$.

We now first choose a subsequence $(\varepsilon_k)_{k \in \mathbb{N}}$ such that $\varepsilon_k \searrow 0$ and $\mathfrak{D}_{\varepsilon_k}^{\text{vel}}(c^{\varepsilon_k}) \rightarrow \beta = \limsup_{\varepsilon \rightarrow 0} \mathfrak{D}_\varepsilon^{\text{vel}}(c^\varepsilon)$. Thus, after selecting a further subsequence (not relabeled) we may assume $\xi^{\varepsilon_k} \rightharpoonup \xi^0$ in $L^2([0, T]; X^*)$. With the strong convergence of $\dot{c}^\varepsilon \rightarrow \dot{c}^0$ we conclude

$$\limsup_{\varepsilon \rightarrow 0} \mathfrak{D}_\varepsilon^{\text{vel}}(c^\varepsilon) = \lim_{k \rightarrow \infty} \mathfrak{B}_{\varepsilon_k}^{\text{vel}}(c^{\varepsilon_k}, \dot{c}^{\varepsilon_k}, \xi^{\varepsilon_k}) \stackrel{*}{\leq} \mathfrak{B}_0^{\text{vel}}(c^0, \dot{c}^0, \xi^0) \leq \mathfrak{D}_0^{\text{vel}}(c^0),$$

where in \leq^* we used the convergence of the duality pairing $\int_0^T \langle \xi^\varepsilon, \dot{c}^\varepsilon \rangle dt$ and a Ioffe-type argument based on the convexity of $\mathcal{R}_\varepsilon^*(c^\varepsilon, \cdot)$ and the lower semicontinuity of $[0, 1] \times X^* \ni (\varepsilon, \xi) \mapsto \mathcal{R}_\varepsilon^*(c^\varepsilon(t), \xi) \in [0, \infty]$, cf. Theorem 7.5 in Ref. 13. Adding the convergence of the slope part, and taking into account the liminf estimate from Theorem 5.11 we obtain the convergence $\lim_{\varepsilon \rightarrow 0} \mathfrak{D}_\varepsilon(c^\varepsilon) = \mathfrak{D}_0(c^0)$.

Step 3. Construction of recovery sequences for the case $\mathfrak{D}_0(c^0)$. We now apply the approximation steps discussed in Step 2 and show that it is possible to choose an suitable diagonal sequence for getting the desired recovery sequence.

For a general c^0 we apply the approximation as indicated in the sub-steps 2(a)–2(c) and set

$$c^{\delta, N, \varepsilon} = \mathbb{D}_{w^\varepsilon}^{-1} \mathbb{D}_{w_0} (\delta w^0 + (1 - \delta) \hat{c}_N^0).$$

We easily obtain $\|c^0 - c^{\delta, N, \varepsilon}\|_{L^2([0,T],X)} \leq C(\delta + \alpha_N + \varepsilon) \rightarrow 0$ if $\delta \rightarrow 0$, $N \rightarrow \infty$, and $\varepsilon \rightarrow 0$. Moreover, the difference in the dissipation functionals \mathfrak{D}_ε can be estimated via

$$|\mathfrak{D}_\varepsilon(c^{\delta, N, \varepsilon}) - \mathfrak{D}_0(c^0)| \leq A(\delta) + B_\delta(N) + C_{\delta, N}(\varepsilon), \quad \text{where}$$

$$A(\delta) = |\mathfrak{D}_0(\tilde{c}^\delta) - \mathfrak{D}_0(c^0)| \quad \text{with } \tilde{c}^\delta(t) = \delta w^0 + (1 - \delta) c^0(t),$$

$$B_\delta(N) = |\mathfrak{D}_0(c^{\delta, N}) - \mathfrak{D}_0(\tilde{c}^\delta)| \quad \text{with } c^{\delta, N}(t) = \delta w^0 + (1 - \delta) \hat{c}_N^0(t),$$

$$C_{\delta, N}(\varepsilon) = |\mathfrak{D}_\varepsilon(c^{\delta, N, \varepsilon}) - \mathfrak{D}_0(c^{\delta, N})|.$$

A recovery sequence $c^\varepsilon \rightarrow c^0$ with $\mathfrak{D}_\varepsilon(c^\varepsilon) \rightarrow \mathfrak{D}_0(c^0)$ is now obtained by a standard diagonal argument. □

Appendix A. Special Properties of Cosh Gradient Structures

Here we discuss a few special properties that are characterizing for the function C and C^* and this lead to corresponding properties of $\mathcal{R}_{\text{cosh}}^*$.

We consider the special non-quadratic dissipation functional C and its Legendre dual C^* given by

$$C(v) := 2v \operatorname{arsinh}(v/2) - 2\sqrt{4 + v^2} + 4 \quad \text{and} \quad C^*(\xi) := 4 \cosh(\xi/2) - 4.$$

Then, we have $C(v) = \frac{1}{2}v^2 + O(v^4)$ and $C^*(\xi) = \frac{1}{2}\xi^2 + O(\xi^4)$. The function C^* has the following properties:

$$C^*(\log p - \log q) = 2 \left(\sqrt{\frac{p}{q}} + \sqrt{\frac{q}{p}} - 2 \right), \quad C^{*'}(\log p - \log q) = \frac{p - q}{\sqrt{pq}}. \quad (\text{A.1})$$

In addition, we have superlinear growth of C :

$$\frac{1}{2}|s| \log(1 + |s|) \leq C(s) \leq |s| \log(1 + |s|) \quad \text{for all } s \in \mathbb{R}. \quad (\text{A.2})$$

The first of the following scaling properties follows easily by considering the power series expansion of C^* , the second by Legendre transform:

$$\forall \lambda \geq 1 \forall s, \zeta \in \mathbb{R} : \quad C^*(\lambda \zeta) \geq \lambda^2 C^*(\zeta) \quad \text{and} \quad C(\lambda s) \leq \lambda^2 C(s). \quad (\text{A.3})$$

This implies the corresponding scaling property for \mathcal{R}_{\cosh} , namely

$$\forall \lambda \geq 1 \forall c \in \mathbb{Q} \forall v, \xi \in \mathbb{R}^I : \quad \mathcal{R}_{\cosh}^*(c, \lambda \xi) \geq \lambda^2 \mathcal{R}_{\cosh}^*(c, \xi) \quad \text{and} \quad \mathcal{R}_{\cosh}(c, \lambda v) \leq \lambda^2 \mathcal{R}_{\cosh}(c, v). \quad (\text{A.4})$$

Acknowledgments

This research has been partially supported by Deutsche Forschungsgemeinschaft (DFG) through the Collaborative Research Center SFB 1114 ‘‘Scaling Cascades in Complex Systems’’ (Project Number 235221301), subproject C05 ‘‘Effective Models for Materials and Interfaces with Multiple Scales’’. The authors are grateful to Mark Peletier for stimulating discussions and to two anonymous referees for helpful comments.

References

1. H. Attouch, *Variational Convergence of Functions and Operators*, Pitman Advanced Publishing Program (Pitman, 1984).
2. A. Bacho, E. Emmrich and A. Mielke, An existence result and evolutionary Γ -convergence for perturbed gradient systems, *J. Evol. Equ.* **19** (2019) 479–522.
3. A. Bobrowski, *Functional Analysis for Probability and Stochastic Processes* (Cambridge Univ. Press, 2005).
4. D. Bothe, Instantaneous limits of reversible chemical reactions in presence of macroscopic convection, *J. Diff. Eqn.* **193** (2003) 27–48.
5. S.-N. Chow, W. Huang, Y. Li and H. Zhou, Fokker–Planck equations for a free energy functional or Markov process on a graph, *Arch. Ration. Mech. Anal.* **203** (2012) 969–1008.
6. K. Disser, M. Liero and J. Zinsl, Evolutionary Γ -convergence of gradient systems modeling slow and fast chemical reactions, *Nonlinearity* **31** (2018) 3689–3706.
7. P. Dondl, T. Frenzel and A. Mielke, A gradient system with a wiggly energy and relaxed EDP-convergence, *ESAIM Control Optim. Calc. Var.* **25** (2019) 45 pp.
8. R. Durrett, *Probability: Theory and Examples* (Cambridge Univ. Press, 2010).
9. E. B. Dynkin, *Markov Processes* (Springer, 1965).
10. M. Erbar and J. Maas, Ricci curvature of finite Markov chains via convexity of the entropy, *Arch. Ration. Mech. Anal.* **206** (2012) 997–1038.

11. W. Fenchel, On conjugate convex functions, *Canadian J. Math.* **1** (1949) 73–77.
12. N. Fenichel, Geometric singular perturbation theory for ordinary differential equations, *J. Differential Equations* **31** (1979) 53–98.
13. I. Fonseca and G. Leoni, *Modern Methods in the Calculus of Variations: L^p spaces* (Springer, 2007).
14. T. Frenzel and M. Liero, Effective diffusion in thin structures via generalized gradient systems and EDP-convergence, WIAS Preprint 2601.
15. T. Frenzel and A. Mielke, Deriving the kinetic relation for the flux through a membrane via edp-convergence, in preparation.
16. M. Grmela, Multiscale equilibrium and nonequilibrium thermodynamics in chemical engineering, *Adv. Chem. Eng.* **39** (2010) 75–128.
17. M. Heida, M. Kantner and A. Stephan, Consistency and convergence for a family of finite volume discretizations of the Fokker–Planck operator, WIAS preprint 2684, arXiv:2002.09385 (2020).
18. R. Jordan, D. Kinderlehrer and F. Otto, The variational formulation of the Fokker–Planck equation, *SIAM J. Math. Anal.* **29** (1998) 1–17.
19. H.-W. Kang and T. G. Kurtz, Separation of time-scales and model reduction for stochastic reaction networks, *Ann. Appl. Probab.* **23** (2013) 529–583.
20. M. Liero, A. Mielke, M. A. Peletier and D. R. M. Renger, On microscopic origins of generalized gradient structures, *Discr. Cont. Dynam. Syst. Ser. S* **10** (2017) 1–35.
21. M. Liero and S. Reichelt, Homogenization of Cahn–Hilliard-type equations via evolutionary Γ -convergence, *Non. Diff. Eqn. Appl.* **25** (2018) Article ID: 6, 31 pp.
22. J. Maas, Gradient flows of the entropy for finite Markov chains, *J. Funct. Anal.* **261** (2011) 2250–2292.
23. J. Maas and A. Mielke, Modeling of chemical reaction systems with detailed balance using gradient structures, *J. Stat. Phys.*, submitted, WIAS preprint 2712, arXiv:2004.02831.
24. R. Marcellin, Contribution a l’étude de la cinétique physico-chimique, *Ann. Phys.* **III** (1915) 120–231.
25. A. Mielke, A gradient structure for reaction-diffusion systems and for energy-drift-diffusion systems, *Nonlinearity* **24** (2011) 1329–1346.
26. A. Mielke, Geodesic convexity of the relative entropy in reversible Markov chains, *Calc. Var. Part. Diff. Eqn.* **48** (2013) 1–31.
27. A. Mielke, On evolutionary Γ -convergence for gradient systems, in eds. A. Muntean, J. Rademacher and A. Zagaris, *Macroscopic and Large Scale Phenomena: Coarse Graining, Mean Field Limits and Ergodicity* (Springer, 2016), Lecture Notes in Applied Math. Mechanics Vol. 3, pp. 187–249.
28. A. Mielke, A. Montefusco and M. A. Peletier, Exploring families of energy-dissipation landscapes via tilting — three types of EDP convergence, *Contin. Mech. Thermodyn.* Submitted. WIAS preprint 2668, arXiv:2001.01455.
29. A. Mielke, R. I. A. Patterson, M. A. Peletier and D. R. M. Renger, Non-equilibrium thermodynamical principles for chemical reactions with mass-action kinetics, *SIAM J. Appl. Math.* **77** (2017) 1562–1585.
30. A. Mielke, M. A. Peletier and D. R. M. Renger, On the relation between gradient flows and the large-deviation principle, with applications to Markov chains and diffusion, *Potent. Anal.* **41** (2014) 1293–1327.
31. A. Mielke, M. A. Peletier and A. Stephan, EDP-convergence for nonlinear fast-slow reaction systems with detailed balance, in preparation.
32. H. C. Öttinger and M. Grmela, Dynamics and thermodynamics of complex fluids. II. Illustrations of a general formalism, *Phys. Rev. E* **56**(3) (1997) 6633–6655.

33. F. Otto, The geometry of dissipative evolution equations: The porous medium equation, *Comm. Partial Differential Equations* **26** (2001) 101–174.
34. M. A. Peletier, Variational modelling: Energies, gradient flows, and large deviations, preprint (2014), arXiv:1402.1990.
35. M. A. Peletier, F. Redig and K. Vafayi, Large deviations in stochastic heat-conduction processes provide a gradient-flow structure for heat conduction, *J. Math. Phys.* **55** (2014) 093301/19.
36. R. T. Rockafellar, Integrals which are convex functionals, *Pacific J. Math.* **24** (1968) 525–539.
37. A. Stephan, Combinatorial considerations on the invariant measure of a stochastic matrix, WIAS preprint 2627, arXiv:1910.02856 (2019).
38. A. Stephan, On EDP-convergence for gradient systems with different time scales, *Proc. Appl. Math. Mech.* **19** (2019) e201900366.
39. A. Stephan, Coarse-graining via EDP-convergence for linear fast-slow reaction diffusion systems, in preparation.
40. H. Stephan, Inequalities for Markov operators, majorization and the direction of time, WIAS preprint 1896 (2013).
41. A. N. Tikhonov, Systems of differential equations containing small parameters in the derivatives, *Mat. Sb.* **31** (1952) 575–586.
42. W.-A. Yong, An interesting class of partial differential equations, *J. Math. Phys.* **49** (2008) 033503, 21.