Modeling of chemical reaction systems with detailed balance using gradient structures

Jan Maas[1], Alexander Mielke[2] [3]

submitted: April 7, 2020

[1] Institute of Science and Technology Austria Am Campus 1 3400 Klosterneuburg Austria E-Mail: jan.maas@ist.ac.at	[2] Weierstrass Institute Mohrenstr. 39 10117 Berlin Germany E-Mail: alexander.mielke@wias-berlin.de

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

Modeling of chemical reaction systems with detailed balance using gradient structures
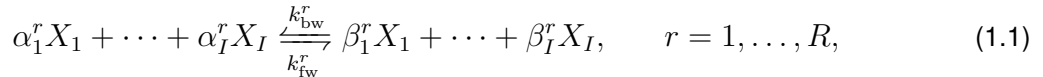
Jan Maas, Alexander Mielke

Abstract

We consider various modeling levels for spatially homogeneous chemical reaction systems, namely the chemical master equation, the chemical Langevin dynamics, and the reaction-rate equation. Throughout we restrict our study to the case where the microscopic system satisfies the detailed-balance condition. The latter allows us to enrich the systems with a gradient structure, i.e. the evolution is given by a gradient-flow equation. We present the arising links between the associated gradient structures that are driven by the relative entropy of the detailed-balance steady state. The limit of large volumes is studied in the sense of evolutionary Γ-convergence of gradient flows. Moreover, we use the gradient structures to derive hybrid models for coupling different modeling levels.

1 Introduction

In this work we discuss different models for chemical reactions taking place in a container of volume V. Throughout we assume that the spatial extent of the container and the position of the chemical species are irrelevant, which means that we are looking at a well-stirred system. We assume that the system is composed of I different species named X_1 to X_I, which may represent different molecules, e.g., $X_1 = \mathrm{H}_2$, $X_2 = \mathrm{O}_2$, and $X_3 = \mathrm{H}_2\mathrm{O}$. We assume that these I species undergo R different reactions of mass-action type:

$$\alpha_1^r X_1 + \cdots + \alpha_I^r X_I \underset{k_{\mathrm{fw}}^r}{\overset{k_{\mathrm{bw}}^r}{\rightleftharpoons}} \beta_1^r X_1 + \cdots + \beta_I^r X_I, \qquad r = 1, \ldots, R, \tag{1.1}$$

where the vectors $\boldsymbol{\alpha}^r, \boldsymbol{\beta}^r \in \mathbb{N}_0^I$ contain the stoichiometric coefficients, and $k_{\mathrm{fw}}^r, k_{\mathrm{bw}}^r > 0$ are the forward and backward reaction rates, see Section 2. The reaction $2\mathrm{H}_2 + \mathrm{O}_2 \rightleftharpoons 2\mathrm{H}_2\mathrm{O}$ would lead to the vectors $\boldsymbol{\alpha} = (2, 1, 0)$ and $\boldsymbol{\beta} = (0, 0, 2)$.

Denoting by $\boldsymbol{c} = (c_1, \ldots, c_I) \in \boldsymbol{C} := [0, \infty[^I$ the vector of nonnegative densities, the simplest model is the macroscopic **reaction-rate equation** (RRE), which is a system of ODEs on the state space $\boldsymbol{C}$:

$$\dot{\boldsymbol{c}} = -\boldsymbol{R}(\boldsymbol{c}) \quad \text{with } \boldsymbol{R}(\boldsymbol{c}) := \sum_{r=1}^R \left(k_{\mathrm{fw}}^r \boldsymbol{c}^{\boldsymbol{\alpha}^r} - k_{\mathrm{bw}}^r \boldsymbol{c}^{\boldsymbol{\beta}^r} \right) \left(\boldsymbol{\alpha}^r - \boldsymbol{\beta}^r \right). \tag{RRE}$$

Here the monomials $\boldsymbol{c}^{\boldsymbol{\alpha}^r} := \Pi_{i=1}^I c_I^{\alpha_i^r}$ indicate that the probability for the right number of particles for the rth reaction to meet is given by a simple product of the corresponding densities, i.e., we assume that the positions of the particles are independent.

A truly microscopic model can be obtained as a stochastic process. Here we count the number of particles $N_i^V(t)$ for each species X_i and consider the random vector $\boldsymbol{N}^V(t) = (N_1^V(t), \ldots, N_I^V(t)) \in$

$\mathcal{N} := \mathbb{N}_0^I$. A forward or backward reaction of type r is modeled as an instantaneous event where the particle numbers jump from $\boldsymbol{N}^V(t) + \boldsymbol{\alpha}^r$ to $\boldsymbol{N}^V(t) + \boldsymbol{\beta}^r$ or vice versa. The corresponding jump rates in a volume of size $V > 0$ are given by $k_{\text{fw}}^r \, \mathbb{B}_V^{\boldsymbol{\alpha}^r}(\boldsymbol{N}^V(t))$ and $k_{\text{bw}}^r \, \mathbb{B}_V^{\boldsymbol{\beta}^r}(\boldsymbol{N}^V(t))$ respectively; see (3.1) for the definition of $\mathbb{B}_V^{\boldsymbol{\alpha}}(\boldsymbol{n})$.

Here we study the vector of probabilities

$$\boldsymbol{u}(t) \in \mathscr{P}(\mathcal{N}) := \big\{ \boldsymbol{v} = (v_{\boldsymbol{n}})_{\boldsymbol{n} \in \mathcal{N}} \,\big|\, v_{\boldsymbol{n}} \geq 0, \, \sum_{\boldsymbol{n} \in \mathcal{N}} v_n = 1 \big\}$$

that describes the probability distribution of the random variable $\boldsymbol{N}^V(t)$. The time evolution of $\boldsymbol{u}(t)$ is given by the **chemical master equation** (CME), i.e., the Kolmogorov forward equation associated with the continuous time Markov chain above. This is a countable linear system of ODEs:

$$\dot{\boldsymbol{u}}(t) = \mathcal{B}_V \boldsymbol{u}(t), \qquad \boldsymbol{u}(0) = \boldsymbol{u}_0, \tag{CME}$$

where $\mathcal{B}_V$ is an (unbounded) linear operator on $\ell^1(\mathcal{N})$, see Section 3, where also existence and uniqueness of solutions is discussed. We refer to [ML*11] for a short introduction to the CME and to [Gil92] for a justification.

The basis of this work is the observation from [Mie11] that (RRE) can be interpreted as a gradient flow if the reaction system satisfies the detailed-balance condition, i.e., there exists a positive equilibrium $\boldsymbol{c}_* = (c_i^*)_{i=1,\ldots,I} \in {]0,\infty[}^I$ such that

$$\kappa_*^r := k_{\text{fw}}^r \boldsymbol{c}_*^{\boldsymbol{\alpha}^r} = k_{\text{bw}}^r \boldsymbol{c}_*^{\boldsymbol{\beta}^r} \text{ for } r = 1, \ldots, R.$$

Defining the Boltzmann entropy E and the Onsager operator $\mathbb{K}$ via

$$E(\boldsymbol{c}) = \sum_{i=1}^I \lambda_{\text{B}}\big(\frac{c_i}{c_i^*}\big) c_i^* \qquad \text{with } \lambda_{\text{B}}(z) = z \log z - z + 1 \text{ and}$$

$$\mathbb{K}(\boldsymbol{c}) = \sum_{r=1}^R \kappa_*^r \, \Lambda\big(\frac{\boldsymbol{c}^{\boldsymbol{\alpha}^r}}{\boldsymbol{c}_*^{\boldsymbol{\alpha}^r}}, \frac{\boldsymbol{c}^{\boldsymbol{\beta}^r}}{\boldsymbol{c}_*^{\boldsymbol{\beta}^r}}\big) \big(\boldsymbol{\alpha}^r - \boldsymbol{\beta}^r\big) \otimes \big(\boldsymbol{\alpha}^r - \boldsymbol{\beta}^r\big) \in \mathbb{R}_{\text{sym},\geq 0}^{I \times I}$$

where $\Lambda(a,b) = \int_0^1 a^s b^{1-s} \, \mathrm{d}s$ is the logarithmic mean, we see that (RRE) is generated by the gradient system $(\boldsymbol{C}, E, \mathbb{K})$, namely $\dot{\boldsymbol{c}} = -\boldsymbol{R}(\boldsymbol{c}) = -\mathbb{K}(\boldsymbol{c})\mathrm{D}E(\boldsymbol{c})$. In Section 2.5 we also discuss further gradient structures, e.g. those used in [MPR14, MP*17, MiS19].

If (RRE) satisfies the detailed-balance condition, then (CME) does so with an equilibrium distribution $\boldsymbol{w}^V \in \mathscr{P}(\mathcal{N})$ that is explicitly given as a product of one-dimensional Poisson distributions with mean $c_i^* V$, namely (cf. Theorem 3.1),

$$w_{\boldsymbol{n}}^V = \prod_{i=1}^I e^{-c_i^* V} \frac{(c_i^* V)^{n_i}}{n_i!} \text{ for all } \boldsymbol{n} = (n_1, \ldots, n_I) \in \mathcal{N}.$$

Consequently, we are also able to interpret (CME) as a gradient flow induced by a gradient system $(\mathscr{P}(\mathcal{N}), \mathcal{E}_V, \mathcal{K}_V)$, see (3.7). Here $\mathcal{E}_V(\boldsymbol{u})$ is again the Boltzmann entropy with respect to $\boldsymbol{w}^V$, but now divided by the volume V:

$$\mathcal{E}_V(\boldsymbol{u}) = \frac{1}{V} \sum_{\boldsymbol{n} \in \mathcal{N}} \lambda_{\text{B}}\big(\frac{u_{\boldsymbol{n}}}{w_{\boldsymbol{n}}^V}\big) w_{\boldsymbol{n}}^V = \frac{1}{V} \sum_{\boldsymbol{n} \in \mathcal{N}} u_{\boldsymbol{n}} \log u_{\boldsymbol{n}} + \sum_{\boldsymbol{n} \in \mathcal{N}} u_{\boldsymbol{n}} \frac{1}{V} \log \frac{1}{w_{\boldsymbol{n}}^V}. \tag{1.2}$$

Large-volume approximations using gradient structures

A major challenge in modeling chemical reactions is the question of understanding the transition from small-volume effects to the macroscopic behavior in large volumes. The first breakthrough was obtained in [Kur67, Kur69, Kur70, Kur72] by connecting the particle numbers $\boldsymbol{N}(t) \in \mathcal{N}$ to the concentrations $\boldsymbol{c} \in \boldsymbol{C}$ and showing that

$$\frac{1}{V}\boldsymbol{N}^V(0) \to \boldsymbol{c}_0 \text{ almost surely implies } \frac{1}{V}\boldsymbol{N}^V(t) \to \boldsymbol{c}(t) \text{ almost surely for all } t > 0, \qquad (1.3)$$

where $t \mapsto \boldsymbol{c}(t)$ is the solution of (RRE) with $\boldsymbol{c}(0) = \boldsymbol{c}_0$. This result may be interpreted as a justification for the RRE in terms of the Markovian model. In [MRP16, MP*17] a dynamic large deviation principle is applied to $\frac{1}{V}\boldsymbol{N}^V(\cdot)$, which leads to a rate functional that generates a gradient structure $(\boldsymbol{C}, \mathcal{E}, \Psi_{\mathrm{cosh}})$; see Section 2.5. Recent large deviation results for chemical reaction networks can be found in [ADE18b, ADE18a].

In this paper we study the limit $V \to \infty$ for the gradient system $(\mathscr{P}(\mathcal{N}), \mathcal{E}_V, \mathcal{K}_V)$, and hence for (CME), in the sense of evolutionary Γ-convergence for gradient systems, as introduced in [SaS04, Ser11] and further developed in [Mie16, DFM19]. For this purpose we use a suitable embedding $\iota_V : \mathscr{P}(\mathcal{N}) \to \mathscr{P}(\boldsymbol{C})$ (Section 4) and obtain the coarse grained gradient system $(\mathscr{P}(\boldsymbol{C}), \boldsymbol{E}, \boldsymbol{K})$ with

$$\boldsymbol{E}(\varrho) = \int_{\boldsymbol{C}} E(\boldsymbol{c})\varrho(\mathrm{d}\boldsymbol{c}) \quad \text{and} \quad \big(\boldsymbol{K}(\varrho)\xi\big)(\boldsymbol{c}) = -\operatorname{div}_{\boldsymbol{c}}\big(\varrho(\boldsymbol{c})\mathbb{K}(\boldsymbol{c})\nabla_{\boldsymbol{c}}\xi(\boldsymbol{c})\big).$$

In particular, the coarse grained gradient flow equation is the **_Liouville equation_**

$$\dot{\varrho}(t, \boldsymbol{c}) = \operatorname{div}_{\boldsymbol{c}}\big(\varrho(t, \boldsymbol{c})\boldsymbol{R}(\boldsymbol{c})\big), \qquad \varrho_{t=0} = \varrho_0, \qquad (\text{Lio})$$

associated with (RRE); here we used that $\xi = \mathrm{D}_\varrho \boldsymbol{E} = E$ and $\boldsymbol{R} = -\mathbb{K}\mathrm{D}_{\boldsymbol{c}}E$. Thus, in this scaling a pure transport equation remains, while all diffusion disappears, as can be seen in the factor $1/V$ before the middle sum in (1.2). In particular, our result is consistent with Kurtz' result (1.3): by assuming $\varrho(0) = \delta_{\boldsymbol{c}_0} \in \mathscr{P}(\boldsymbol{C})$ we obtain $\varrho(t) = \delta_{\boldsymbol{c}(t)}$. While Kurtz works directly on the Markovian random variables, we work at the level of their distributions:

$$
\begin{array}{ccc}
\boldsymbol{u} \in \mathscr{P}(\mathcal{N}) & \xrightarrow[\text{here}]{\iota_V(\boldsymbol{u}^V) \to \varrho} & \varrho \in \mathscr{P}(\boldsymbol{C}) \\
\text{CME: } \dot{\boldsymbol{u}} = -\mathcal{K}_V(\boldsymbol{u})\mathrm{D}\mathcal{E}_V(\boldsymbol{u}) & & \text{Liouville: } \partial_t\varrho = -\boldsymbol{K}(\varrho)\mathrm{D}\boldsymbol{E}(\varrho) \\
\uparrow & & \uparrow \\
\boldsymbol{N} \in \mathcal{N} & \xrightarrow[\text{Kurtz}]{\frac{1}{V}\boldsymbol{N}^V \to \boldsymbol{c}} & \boldsymbol{c} \in \boldsymbol{C} \\
\text{Markovian model} & & \text{RRE: } \dot{\boldsymbol{c}} = -\mathbb{K}(\boldsymbol{c})\mathrm{D}E(\boldsymbol{c}).
\end{array}
$$

Our convergence result for the gradient systems $(\mathscr{P}(\mathcal{N}), \mathcal{E}_V, \mathcal{K}_V)$ to the limiting gradient system $(\mathscr{P}(\boldsymbol{C}), \boldsymbol{E}, \boldsymbol{K})$ can be seen as a concrete example of the *EDP convergence* of gradient systems as discussed in [LM*17, DFM19]. Another example treating the convergence of "Markovian discretizations" towards a Fokker–Planck equation is studied in [DiL15]; see also [FaS16, EF*16, Sch19] for applications to interacting particle systems.

In addition to the extreme cases V finite and $V \to \infty$ it is also important to study the case of intermediate V, where $\frac{1}{V}\boldsymbol{N}^V(t)$ already behaves continuously but still shows some fluctuations of standard deviation $1/\sqrt{V}$, see [WiS17] for a numerical approach to treat the hierarchy via a suitable

hybrid method. In [Kur78] it is shown that the random vector $t \mapsto \boldsymbol{X}^V(t) \in \boldsymbol{C}$ obtained by solving the stochastic differential equation

$$\mathrm{d}\boldsymbol{X}^V(t) = -\boldsymbol{R}(\boldsymbol{X}^V(t))\,\mathrm{d}t + \frac{1}{\sqrt{V}}\Big(\Sigma^{\mathrm{fw}}(\boldsymbol{X}^V(t))\,\mathrm{d}\mathfrak{B}^{\mathrm{fw}}(t) + \Sigma^{\mathrm{bw}}(\boldsymbol{X}^V(t))\,\mathrm{d}\mathfrak{B}^{\mathrm{bw}}(t)\Big)$$

with independent Brownian vectors $\mathfrak{B}^{\mathrm{fw}}(t)$, $\mathfrak{B}^{\mathrm{bw}}(t) \in \mathbb{R}^R$, and (1.4)

$$\Sigma^{\mathrm{fw}}(\boldsymbol{X}) = \Big(\big(\tfrac{\kappa^r \boldsymbol{X}^{\boldsymbol{\alpha}^r}}{\boldsymbol{c}_*^{\boldsymbol{\alpha}^r}}\big)^{1/2}(\boldsymbol{\alpha}^r - \boldsymbol{\beta}^r)\Big)_r, \quad \Sigma^{\mathrm{bw}}(\boldsymbol{X}) = \Big(\big(\tfrac{\kappa^r \boldsymbol{X}^{\boldsymbol{\beta}^r}}{\boldsymbol{c}_*^{\boldsymbol{\beta}^r}}\big)^{1/2}(\boldsymbol{\beta}^r - \boldsymbol{\alpha}^r)\Big)_r \in \mathbb{R}^{I \times R},$$

(see [Kur78, Eqn. (1.7)]) yields an improved approximation as $\frac{1}{V}\boldsymbol{N}^V(t) = \boldsymbol{X}^V(t) + O\big((\log V)/V\big)$, while $\frac{1}{V}\boldsymbol{N}^V(t) = \boldsymbol{c}(t) + O\big(1/\sqrt{V}\big)$. This model is a so-called diffusion approximation, which in the reaction context also is termed 'chemical Langevin dynamics'. In [Gil00, Eqn. (23)] and [WiS17, Eqn. (7)] the stochastic differential equation (1.4) is called **chemical Langevin equation** (CLE).

The associated Kolmogorov forward equation takes the form

$$\dot{\rho} = \frac{1}{V}\sum_{i,j=1}^I \partial^2_{ij}\big(\rho\,\widehat{\mathbb{K}}_{\mathrm{CLE}}(\boldsymbol{c})_{ij}\big) + \mathrm{div}\big(\rho\boldsymbol{R}(\boldsymbol{c})\big) \text{ with } \widehat{\mathbb{K}}_{\mathrm{CLE}} = \frac{1}{2}\big(\Sigma^{\mathrm{fw}}(\Sigma^{\mathrm{fw}})^\mathsf{T} + \Sigma^{\mathrm{bw}}(\Sigma^{\mathrm{bw}})^\mathsf{T}\big). \quad (1.5)$$

Here the diffusion matrix $\widehat{\mathbb{K}}_{\mathrm{CLE}}$ can be written in the explicit form

$$\widehat{\mathbb{K}}_{\mathrm{CLE}}(\boldsymbol{c}) = \sum_{r=1}^R \kappa^r \frac{1}{2}\Big(\frac{\boldsymbol{c}^{\boldsymbol{\alpha}^r}}{\boldsymbol{c}_*^{\boldsymbol{\alpha}^r}} + \frac{\boldsymbol{c}^{\boldsymbol{\beta}^r}}{\boldsymbol{c}_*^{\boldsymbol{\beta}^r}}\Big)\,(\boldsymbol{\alpha}^r - \boldsymbol{\beta}^r) \otimes (\boldsymbol{\alpha}^r - \boldsymbol{\beta}^r) \quad (1.6)$$

that is different from $\mathbb{K}(\boldsymbol{c})$, because in the former the arithmetic mean while in the latter the logarithmic mean is taken.

One drawback of the chemical Langevin equation (1.5) is that it cannot be written as gradient flow of the relative entropy, as the Einstein relation for the drift flux and the diffusion flux is not satisfied. Therefore we propose other approximations that stay inside the theory of gradient flows and seem to work sufficiently well if the concentrations are not too large or small. Our simplest approximation is given by the gradient system $\big(\mathscr{P}(\boldsymbol{C}), \widetilde{\boldsymbol{E}}_V, \boldsymbol{K}\big)$ with

$$\widetilde{\boldsymbol{E}}_V(\varrho) = \int_{\boldsymbol{C}}\Big(\frac{1}{V}\rho\log\rho + \rho E\Big)\,\mathrm{d}\boldsymbol{c}, \text{ where } \varrho = \rho\,\mathrm{d}\boldsymbol{c},$$

which leads to the linear **Fokker–Planck equation**

$$\dot{\rho} = \mathrm{div}\Big(\frac{1}{V}\mathbb{K}(\boldsymbol{c})\nabla\rho + \rho\boldsymbol{R}\Big). \quad (\mathrm{FPE})$$

In Section 5 we show that by systematically deriving higher-order corrections to $\widetilde{\boldsymbol{E}}_V$ and $\boldsymbol{K}$ we can recover the asymptotically correct diffusion matrix $\widehat{\mathbb{K}}_{\mathrm{CLE}}$ while keeping the gradient structure, but have to accept several additional terms, or switch over to the notion of *asymptotic gradient flow structures* in the sense of [BB*17].

Hybrid modeling using gradient structures

A major advantage of the gradient flow description is that the different structures can be combined to obtain hybrid models, in which the set of chemical species is divided into subclasses which may

be treated differently depending on the desired or needed accuracy. Our approach is based on the idea of model reduction for gradient structures. The idea is to approximate a complicated gradient structure $(\mathsf{X}, \mathsf{E}_X, \mathsf{K}_X)$ by a simpler one $(\mathsf{Y}, \mathsf{E}_Y, \mathsf{K}_Y)$ via an embedding mapping $\mathsf{x} = \Phi(\mathsf{y})$. Staying within the class of gradient systems has the advantage that the most important features of the original system can be preserved. In particular, decay of the driving functional along the approximate flow holds automatically. By contrast, such crucial features could get lost in a direct approach based on the evolution equation itself.

In Section 6 we shall deal with three examples for hybrid models where it is essential to keep V as a large but finite parameter. First, we shall consider a hybrid model in which an RRE is coupled to a Fokker–Planck equation. Here the set of species is divided into two classes: $\boldsymbol{C} = \boldsymbol{C}_\mathrm{s} \times \boldsymbol{C}_\mathrm{m}$. Some of them will be described *stochastically* (s), while others are described *macroscopically* (m). This leads to a gradient flow structure on the hybrid state space $\mathsf{Y} = \mathscr{P}(\boldsymbol{C}_\mathrm{s}) \times \boldsymbol{C}_\mathrm{m}$. The resulting gradient flow equation turns out to be a mean-field equation, in which the density of the component $\boldsymbol{c}_\mathrm{s}$ satisfies a linear equation which is nonlinearly coupled to an ODE for the component $\boldsymbol{c}_\mathrm{m}$.

We also study the coupling of an RRE for macroscopic variables to a CME for n microscopic variables. This leads to a hybrid system on $\mathscr{P}(\mathbb{N}_0^n) \times \boldsymbol{C}_\mathrm{m}$. Finally we analyze a mixed CME / Fokker–Planck model with state space $\mathscr{P}(\mathfrak{N})$, in which the underlying space $\mathfrak{N} := \{0, 1, \ldots, N-1\} \cup [N/V, \infty[$ contains a mixture of discrete and continuous components.

The present work concentrates solely on the analytical underpinnings of hybrid modeling for CME; for numerical approaches to CME and to spatio-temporal CME we refer to [AC*05, HeL07, MuK07, Hig08, Eng09, Jah11, DoK14, WiS17].

Notational conventions. Throughout the paper we will consistently use the following notation to distinguish the different modeling levels.

Reaction-rate equation: The RRE is denoted by $(\boldsymbol{C}, E, \mathbb{K})$:
state and state space $\boldsymbol{c} \in \boldsymbol{C} := [0, \infty[^I$, steady state $\boldsymbol{c}^* = \boldsymbol{c}_*$, dual variable $\boldsymbol{\zeta}$
energy functional $E(\boldsymbol{c})$, Onsager operator $\mathbb{K}(\boldsymbol{c})$
conserved quantities $\mathbb{Q}\boldsymbol{c} = \boldsymbol{q}$, stoichiometric subsets $\boldsymbol{I}(\boldsymbol{q}) = \{\, \boldsymbol{c} \in \boldsymbol{C} \mid \mathbb{Q}\boldsymbol{c} = \boldsymbol{q} \,\}$.

Chemical master equation: The CME is denoted by $(\mathscr{P}(\mathcal{N}), \mathcal{E}_V, \mathcal{K}_V)$:
state and state space $\boldsymbol{u} = (u_{\boldsymbol{n}})_{\boldsymbol{n} \in \mathcal{N}} \in \mathscr{P}(\mathcal{N}) \subset$, steady state $\boldsymbol{w}^V$, dual variable $\boldsymbol{\mu}$
energy functional $\mathcal{E}_V(\boldsymbol{u})$, Onsager operator $\mathcal{K}_V(\boldsymbol{u})$
invariant subsets $\mathcal{I}(\overline{\boldsymbol{n}}) = \{\, \boldsymbol{n} \in \mathcal{N} \mid \mathbb{Q}\boldsymbol{n} = \mathbb{Q}\overline{\boldsymbol{n}} \,\}$.

Liouville equation: The LE is denoted by $(\mathscr{P}(\boldsymbol{C}), \boldsymbol{E}, \boldsymbol{K})$:
state and state space $\varrho = \rho\,\mathrm{d}\boldsymbol{c} \in \mathscr{P}(\boldsymbol{C})$, steady state $\delta_{\boldsymbol{c}_*}$, dual variable ξ
energy functional $\boldsymbol{E}(\varrho) = \int_{\boldsymbol{C}} E(\boldsymbol{c})\,\mathrm{d}\varrho(\boldsymbol{c})$, Onsager operator $\boldsymbol{K}(\varrho) = -\operatorname{div}\big(\varrho(\cdot)\mathbb{K}(\cdot)\nabla\square\big)$

Fokker–Planck equation: The FPE is denoted by $(\mathscr{P}(\boldsymbol{C}), \widetilde{\boldsymbol{E}}_V, \boldsymbol{K})$:
state and state space $\varrho \in \mathcal{P}(\boldsymbol{C}) := \mathscr{P}(\boldsymbol{C})$, steady state W_V, dual variable ξ
energy functional $\widetilde{\boldsymbol{E}}_V(\varrho) = \int_{\boldsymbol{C}} \big(\frac{1}{V}\rho(\boldsymbol{c})\log\rho(\boldsymbol{c}) + \rho(\boldsymbol{c})E(\boldsymbol{c})\big)\,\mathrm{d}\boldsymbol{c}$, Onsager operator $\boldsymbol{K}$.

Hybrid systems are denoted by "mathfrak" letters:
$(\mathscr{P}(\boldsymbol{C}_\mathrm{s}) \times \boldsymbol{C}_\mathrm{m}, \mathfrak{E}_V^{\text{FP-RR}}, \mathfrak{K}_V^{\text{FP-RR}})$ for coupling FPE and RRE
$(\mathscr{P}(\mathbb{N}_0^J) \times \boldsymbol{C}_\mathrm{m}, , \mathfrak{E}_V^{\text{CM-RR}}, \mathfrak{K}_V^{\text{CM-RR}})$ for coupling CME and RRE
$(\mathscr{P}(\mathfrak{N}_{V,N}), \mathfrak{E}_{V,N}, \mathfrak{K}_{V,N})$ for merging discrete and continuous modeling for one species.

The space of all signed Borel measures of bounded variation on $\boldsymbol{C}$ is denoted by $\mathscr{M}(\boldsymbol{C})$.

2 Reaction rate equations

We denote by $\boldsymbol{c} = (c_1, \ldots, c_I) \in \boldsymbol{C} := [0, \infty[^I$ the concentrations of I different chemical species $X_1, \ldots, X_I$ reacting according to the mass action law, i.e., the reactions

$$\alpha_1^r X_1 + \cdots + \alpha_I^r X_I \xrightleftharpoons[k_{\mathrm{fw}}^r]{k_{\mathrm{bw}}^r} \beta_1^r X_1 + \cdots + \beta_I^r X_I \qquad (2.1)$$
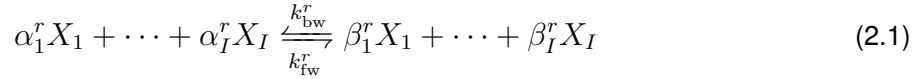
for $r = 1, \ldots, R$, where R is the number of possible reactions, $\boldsymbol{\alpha}^r, \boldsymbol{\beta}^r \in \mathbb{N}_0^I$ are the vectors of the stoichiometric coefficients, and $k_{\mathrm{fw}}^r, k_{\mathrm{bw}}^r > 0$ are the forward and backward reaction-rates. In general these rates may depend on $\boldsymbol{c}$, but for simplicity we keep them as constants in this work. A typical example is the splitting of water into hydrogen and oxygen, namely $2\mathrm{H}_2 + \mathrm{O}_2 \leftrightharpoons 2\mathrm{H}_2\mathrm{O}$.

The corresponding *reaction-rate equations* (RRE) are given via the ODE system

$$\dot{\boldsymbol{c}} = -\boldsymbol{R}(\boldsymbol{c}) \quad \text{with } \boldsymbol{R}(\boldsymbol{c}) := \sum_{r=1}^{R} \left(k_{\mathrm{fw}}^r \boldsymbol{c}^{\boldsymbol{\alpha}^r} - k_{\mathrm{bw}}^r \boldsymbol{c}^{\boldsymbol{\beta}^r} \right) \left(\boldsymbol{\alpha}^r - \boldsymbol{\beta}^r \right), \qquad (2.2)$$

where $\boldsymbol{c}^{\boldsymbol{\alpha}} = c_1^{\alpha_1} \cdots c_I^{\alpha_I}$, see [FeH77, Grö83, ÉrT89].

2.1 Stoichiometry, conservation, and decomposition of the state space

The stoichiometric subspace $\mathbb{S} \subset \mathbb{R}^I$ and its orthogonal complement $\mathbb{S}^\perp$ are defined via

$$\mathbb{S} := \mathrm{span}\{\, \boldsymbol{\alpha}^r - \boldsymbol{\beta}^r \mid r = 1, \ldots, R \,\}, \quad \mathbb{S}^\perp := \{\, \boldsymbol{\xi} \in \mathbb{R}^I \mid \boldsymbol{\xi} \cdot \boldsymbol{\mu} = 0 \text{ for all } \boldsymbol{\mu} \in \mathbb{S} \,\}. \qquad (2.3)$$

For each $\boldsymbol{\xi} \in \mathbb{S}^\perp$ the function $C_{\boldsymbol{\xi}}(\boldsymbol{c}) = \boldsymbol{\xi} \cdot \boldsymbol{c}$ defines a first integral, which easily follows from $\boldsymbol{\xi} \cdot \boldsymbol{R}(\boldsymbol{c}) \equiv 0$. These conservation laws often go under the name *conservation of atomic species*, see [ÉrT89]. Suppose now that $\mathbb{S}^\perp$ is a non-trivial subspace of $\mathbb{R}^I$. We shall argue that the RRE induces a decomposition of the state space $\boldsymbol{C} = [0, \infty[^I$ into affine invariant subsets. (If $\mathbb{S}^\perp = \{\boldsymbol{0}\}$, the only invariant set is $\boldsymbol{C}$ itself.)

Choosing a basis $\{\, \boldsymbol{m}_k \in \mathbb{R}^I \mid k = 1, \ldots, m_{\mathbb{W}} \,\}$ of $\mathbb{S}^\perp$ we define the matrix $\mathbb{Q} \in \mathbb{R}^{m_{\mathbb{W}} \times I}$, which has the rows $\boldsymbol{m}_k \in \mathbb{R}^I$. By construction we have $\mathbb{Q}[\mathbb{S}] = \{\boldsymbol{0}\}$, and we conclude that the solutions $\boldsymbol{c}$ of (2.2) conserve $\mathbb{Q}\boldsymbol{c}$ as follows:

$$\dot{\boldsymbol{c}} = -\boldsymbol{R}(\boldsymbol{c}) \quad \Longrightarrow \quad \mathbb{Q}\boldsymbol{c}(t) = \mathbb{Q}\boldsymbol{c}(0) \text{ for } t > 0. \qquad (2.4)$$

By construction every affine conserved quantity is of the form $\boldsymbol{\xi} \cdot \boldsymbol{c} + q$ for some $\boldsymbol{\xi} \in \mathbb{S}^\perp$ and $q \in \mathbb{R}$. This allows us to decompose the full state space $\boldsymbol{C} = [0, \infty[^I$ into the invariant, affine subsets $(\boldsymbol{c}_0 + \mathbb{S}) \cap \boldsymbol{C}$ for $\boldsymbol{c}_0 \in \boldsymbol{C}$. Using the notation

$$\mathfrak{Q} := \{\, \mathbb{Q}\boldsymbol{c} \in \mathbb{R}^{m_{\mathbb{W}}} \mid \boldsymbol{c} \in \boldsymbol{C} \,\}$$

we define, for all $\boldsymbol{q} \in \mathfrak{Q}$, the sets

$$\boldsymbol{I}(\boldsymbol{q}) := \{\, \boldsymbol{c} \in \boldsymbol{C} \mid \mathbb{Q}\boldsymbol{c} = \boldsymbol{q} \,\}. \qquad (2.5)$$

Then, $\boldsymbol{q}_1 \neq \boldsymbol{q}_2$ implies $\boldsymbol{I}(\boldsymbol{q}_1) \cap \boldsymbol{I}(\boldsymbol{q}_2) = \emptyset$, and we have $\boldsymbol{C} = \bigcup_{\boldsymbol{q} \in \mathfrak{Q}} \boldsymbol{I}(\boldsymbol{q})$. Let us note that this decomposition does not depend on the choice of the orthonormal basis which determines the matrix $\mathbb{Q}$, although the set $\boldsymbol{I}(\boldsymbol{q})$ does depend on $\mathbb{Q}$. Note also that we can always write $\boldsymbol{I}(\boldsymbol{q}) = (\boldsymbol{c} + \mathbb{S}) \cap \boldsymbol{C}$ for some arbitrary $\boldsymbol{c} \in \boldsymbol{C}$ satisfying $\mathbb{Q}\boldsymbol{c} = \boldsymbol{q}$.

2.2 Detailed balance and the Wegscheider matrix

We say that the above reaction system fulfills the *condition of detailed balance* if there exists a positive equilibrium density vector $\boldsymbol{c}_* \in \,]0, \infty[^I$ such that all reactions are simultaneously in equilibrium, i.e.,

$$\kappa_*^r := k_{\mathrm{fw}}^r \boldsymbol{c}_*^{\boldsymbol{\alpha}^r} = k_{\mathrm{bw}}^r \boldsymbol{c}_*^{\boldsymbol{\beta}^r} \quad \text{for } r = 1, \ldots, R. \tag{2.6}$$

This condition implies that $\boldsymbol{R}(\boldsymbol{c}_*) = 0$, but we emphasize that this condition is stronger in general cases. The condition of detailed balance is also called the condition of *microscopic reversibility*, see [ÉrT89, p. 45] or [DeM84] for a general discussion of these concepts.

We are looking for a characterization of detailed balance. Let $\mathbb{W} \in \mathbb{Z}^{R \times I}$ be the matrix which has the row vectors $\boldsymbol{\gamma}^r := \boldsymbol{\alpha}^r - \boldsymbol{\beta}^r \in \mathbb{Z}^I$, $r = 1, \ldots, R$. We call $\mathbb{W}$ the *Wegscheider matrix* because of the pioneering work in [Weg02]. We then have

$$\mathbb{S} = \operatorname{Ran} \mathbb{W}^{\mathsf{T}} \quad \text{and} \quad \mathbb{S}^{\perp} = \operatorname{Ker} \mathbb{W},$$

which explains the abbreviation $m_{\mathbb{W}} := \dim \mathbb{S}^{\perp} = \dim \operatorname{Ker} \mathbb{W}$. Since $\boldsymbol{c}_*$ is strictly positive, we can take the logarithm of the polynomial conditions (2.6) and find the equivalent linear system

$$\mathbb{W} \log \boldsymbol{c}_* = \big(\log(k_{\mathrm{bw}}^r / k_{\mathrm{fw}}^r) \big)_{r=1,\ldots,R}, \quad \text{where } \log \boldsymbol{c} = \big(\log c_i \big)_{i=1,\ldots,I}. \tag{2.7}$$

By Fredholm's alternative, (2.7) is solvable if and only if

$$\boldsymbol{y} \cdot \big(\log(k_{\mathrm{bw}}^r / k_{\mathrm{fw}}^r) \big)_{r=1,\ldots,R} = 0 \quad \text{for all } \boldsymbol{y} \in \operatorname{Ker} \mathbb{W}^{\mathsf{T}}. \tag{2.8}$$

These conditions on the reaction coefficients k_{fw}^r and k_{bw}^r are called *Wegscheider conditions* (see, e.g., [Weg02, ScS89, VlR09, GlM13]). By choosing a basis of $\operatorname{Ker} \mathbb{W}^{\mathsf{T}}$ and exponentiation they can be rewritten as polynomial conditions without referring to the equilibrium state $\boldsymbol{c}_*$.

Let $n_{\mathbb{W}} := \dim(\operatorname{Ker} \mathbb{W}^{\mathsf{T}}) \in \mathbb{N}_0$ denote the number of Wegscheider conditions. Then the following assertions hold:

(i) If the stoichiometric vectors $\boldsymbol{\alpha}^r - \boldsymbol{\beta}^r$, $r = 1, \ldots, R$, are linearly independent, then $\operatorname{Ker} \mathbb{W}^{\mathsf{T}} = \{\boldsymbol{0}\}$, hence there is no Wegscheider condition.

(ii) If $\boldsymbol{\alpha}^r - \boldsymbol{\beta}^r$, $r = 1, \ldots, R$, are linearly dependent, then $\dim(\operatorname{Ker} \mathbb{W}^{\mathsf{T}}) > 0$ and non-trivial Wegscheider conditions appear.

Since $\dim(\operatorname{Ran} \mathbb{W}) = \dim(\operatorname{Ran} \mathbb{W}^{\mathsf{T}}) = \dim \mathbb{S}$ by standard linear algebra, the number of Wegscheider conditions can be expressed as

$$n_{\mathbb{W}} = R - \dim \mathbb{S} = R - I + \dim(\operatorname{Ker} \mathbb{W}) = R - I + m_{\mathbb{W}}.$$

Hence, if the number R of reactions is smaller than the number I of species, the Wegscheider conditions can usually be satisfied easily.

Remark 2.1 (Wellposedness of RRE) *We conclude this subsection with a statement concerning the well-posedness of the RRE given as in Theorem 2.2 below. For all $\boldsymbol{c}(0) \in \boldsymbol{C} = [0, \infty[^I$ there exists a unique global solution $\boldsymbol{c} : [0, \infty[\to \boldsymbol{C}$. Local existence for solutions starting in the interior of $\boldsymbol{C}$ is trivial, as $\boldsymbol{R}$ is a polynomial vector field. Since the relative entropy E is a coercive Liapunov functional, the solutions cannot blow up and stay inside a region $B_R(0) \cap \boldsymbol{C}$ for some $R > 0$.*

Moreover, solutions cannot leave this region via the boundary $\partial \boldsymbol{C}$, since the vector field is either tangential to $\partial \boldsymbol{C}$ or points inwards. Indeed, if $c_j(t_0) = 0$ for some j, then

$$\dot{c}_j(t_0) = -R_j(\boldsymbol{c}(t_0)) = -\sum_{r=1}^{R} \kappa_*^r \left(\frac{\boldsymbol{c}^{\boldsymbol{\alpha}^r}(t_0)}{\boldsymbol{c}_*^{\boldsymbol{\alpha}^r}} - \frac{\boldsymbol{c}^{\boldsymbol{\beta}^r}(t_0)}{\boldsymbol{c}_*^{\boldsymbol{\beta}^r}} \right) \left(\alpha_j^r - \beta_j^r \right) \geq 0,$$

because each term in the sum is nonpositive: If $\alpha_j^r = \beta_j^r$ or $\min\{\alpha_j^r, \beta_j^r\} > 0$, then the term is 0. Thus, we are left with the cases $(\alpha_j^r, \beta_j^r) \in \{(n, 0), (0, n)\}$ for some positive n. In the first case $c_j(t_0) = 0$ implies $\boldsymbol{c}^{\boldsymbol{\alpha}^r}(t_0) = 0$ and the result follows, and the second case is similar.

2.3 The reaction-rate equations as a gradient system

We show that a RRE satisfying the detailed-balance condition can be generated by a gradient system $(\boldsymbol{C}, E, \mathbb{K})$. Here, the state space $\boldsymbol{C} := [0, \infty[^I$ contains all possible concentration vectors $\boldsymbol{c}$. The driving functional is the relative entropy $E : \boldsymbol{C} \to [0, \infty[$ and the Onsager matrix $\mathbb{K}$ is chosen suitably (recall that $\lambda_{\mathrm{B}}(z) = z \log z - z + 1 \geq 0$):

$$E(\boldsymbol{c}) := \sum_{i=1}^{I} \lambda_{\mathrm{B}}\left(\frac{c_i}{c_i^*}\right) c_i^* \ \text{ and } \ \mathbb{K}(\boldsymbol{c}) = \sum_{r=1}^{R} \kappa_*^r \Lambda\left(\frac{\boldsymbol{c}^{\boldsymbol{\alpha}^r}}{\boldsymbol{c}_*^{\boldsymbol{\alpha}^r}}, \frac{\boldsymbol{c}^{\boldsymbol{\beta}^r}}{\boldsymbol{c}_*^{\boldsymbol{\beta}^r}}\right) \left(\boldsymbol{\alpha}^r - \boldsymbol{\beta}^r\right) \otimes \left(\boldsymbol{\alpha}^r - \boldsymbol{\beta}^r\right), \tag{2.9}$$

where the logarithmic-mean function Λ is given via

$$\Lambda(a, b) = \int_0^1 a^s b^{1-s} \, \mathrm{d}s = \frac{a - b}{\log a - \log b}. \tag{2.10}$$

The following result shows that a RRE (2.2) satisfying the detailed-balance condition (2.6) is indeed generated by the gradient system $(\boldsymbol{C}, E, \mathbb{K})$. This was first established in [Yon08, Sect. VII] to derive entropy bounds for hyperbolic conservation laws in reactive flows and was rederived in [Mie11] in the context of reaction diffusion systems including electric charge-interactions. It is interesting to note that for continuous time Markov chains (CTMC), which form a special subclass of RRE with linear reactions, there are several distinct gradient structures, see [Maa11, Prop. 4.2] and [Mie13, Thm. 3.1] and Section 2.4. However, in the case of nonlinear reactions according to the mass-action law, only the gradient structure with the Boltzmann entropy remains. The key fact is the logarithm identity $(\boldsymbol{\alpha} - \boldsymbol{\beta}) \cdot \log \boldsymbol{c} = \log(\boldsymbol{c}^{\boldsymbol{\alpha} - \boldsymbol{\beta}})$.

Theorem 2.2 (Gradient structure for RRE) *If the RRE* (2.2) *satisfies the detailed-balance condition* (2.6) *for a positive steady state $\boldsymbol{c}_* = (c_i^*)_{i=1,\dots,I}$, then it has the gradient structure $(\boldsymbol{C}, E, \mathbb{K})$ defined in* (2.9)*, namely $\dot{\boldsymbol{c}} = -\boldsymbol{R}(\boldsymbol{c}) = -\mathbb{K}(\boldsymbol{c})\mathrm{D}E(\boldsymbol{c})$.*

Proof. Multiplying $\mathrm{D}E(\boldsymbol{c}) = (\log(c_i/c_i^*))_{i=1,\dots,I}$ by $\boldsymbol{\alpha}^r - \boldsymbol{\beta}^r \in \mathbb{R}^I$ we obtain

$$\begin{aligned} (\log(c_i/c_i^*))_{i=1,\dots,I} \cdot \left(\boldsymbol{\alpha}^r - \boldsymbol{\beta}^r\right) &= \sum_{i=1}^{I} \left(\alpha_i^r \log(c_i/c_i^*) - \beta_i^r \log(c_i/c_i^*) \right) \\ &= \log\left(\frac{\boldsymbol{c}^{\boldsymbol{\alpha}^r}}{\boldsymbol{c}_*^{\boldsymbol{\alpha}^r}}\right) - \log\left(\frac{\boldsymbol{c}^{\boldsymbol{\beta}^r}}{\boldsymbol{c}_*^{\boldsymbol{\beta}^r}}\right), \end{aligned} \tag{2.11}$$

which is the denominator of $\Lambda\left(\frac{\boldsymbol{c}^{\boldsymbol{\alpha}^r}}{\boldsymbol{c}_*^{\boldsymbol{\alpha}^r}}, \frac{\boldsymbol{c}^{\boldsymbol{\beta}^r}}{\boldsymbol{c}_*^{\boldsymbol{\beta}^r}}\right)$. Hence, using $\Lambda(a, b)(\log a - \log b) = a - b$ gives

$$\mathbb{K}(\boldsymbol{c})\mathrm{D}E(\boldsymbol{c}) = \sum_{r=1}^{R} \kappa_*^r \left(\frac{\boldsymbol{c}^{\boldsymbol{\alpha}^r}}{\boldsymbol{c}_*^{\boldsymbol{\alpha}^r}} - \frac{\boldsymbol{c}^{\boldsymbol{\beta}^r}}{\boldsymbol{c}_*^{\boldsymbol{\beta}^r}}\right) \left(\boldsymbol{\alpha}^r - \boldsymbol{\beta}^r\right) \overset{\mathrm{DB}}{=} \sum_{r=1}^{R} \left(k_{\mathrm{fw}}^r \boldsymbol{c}^{\boldsymbol{\alpha}^r} - k_{\mathrm{bw}}^r \boldsymbol{c}^{\boldsymbol{\beta}^r}\right) \left(\boldsymbol{\alpha}^r - \boldsymbol{\beta}^r\right) = \boldsymbol{R}(\boldsymbol{c}),$$

where we used the detailed-balance condition (2.6) in $\overset{\mathrm{DB}}{=}$. Thus, the assertion is established. ∎

Summarizing the above derivations, we have rewritten the RRE in thermodynamic form

$$\dot{\boldsymbol{c}} = -\boldsymbol{R}(\boldsymbol{c}) = -\mathbb{K}(\boldsymbol{c})\boldsymbol{\mu} \quad \text{with} \ \ \boldsymbol{\mu} = \mathrm{D}E(\boldsymbol{c}), \tag{2.12}$$

which is also called the Onsager principle [Ons31, OnM53]. The latter states that the rate (flux) of a macroscopic variable is given as the product of a symmetric positive definite matrix $\mathbb{K}$ and the thermodynamic driving force $-\boldsymbol{\mu}$, see e.g. [DeM84, Ch. X, § 4]. The symmetry $\mathbb{K} = \mathbb{K}^\top$ is related to microscopic reversibility, i.e., detailed balance, see also [MPR14, MRP16]. Subsequently, we refer to $\mathbb{K}$ as the Onsager operator or matrix.

Here we clearly see the advantage of using the Onsager operator $\mathbb{K}$ to write the RRE as a gradient system,as opposed to working with the Riemannian tensor: we do not have to take care of the fact that $\mathbb{K}$ is not invertible except if $\mathbb{S} = \mathbb{R}^I$.

2.4 Continuous time Markov chains as a gradient system

The forward equation for a reversible CTMC on a discrete space $\{1, 2, \ldots, I\}$ is a special case of the RRE considered above. In this case all reactions are of the form

$$X_i \overset{k_{\mathrm{bw}}^{ij}}{\underset{k_{\mathrm{fw}}^{ij}}{\rightleftharpoons}} X_j \ \ \text{for} \ \ 1 \le i < j \le I,$$
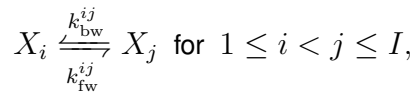
and the reaction rates k_{fw}^{ij} (resp. k_{bw}^{ij}) are interpreted as the transition rates from i to j (resp. from j to i). The reaction-rate equation is the linear system of ODEs

$$\dot{\boldsymbol{c}} = -\boldsymbol{R}(\boldsymbol{c}) = \mathcal{A}\boldsymbol{c} \quad \text{with} \ \mathcal{A}\boldsymbol{c} = -\sum_{i<j} \left(k_{\mathrm{fw}}^{ij} c_i - k_{\mathrm{bw}}^{ij} c_j \right)(\boldsymbol{e}_i - \boldsymbol{e}_j), \tag{2.13}$$

and the detailed-balance condition for the equilibrium state $\boldsymbol{c}_*$ takes the form

$$\kappa_*^{ij} := c_i^* k_{\mathrm{fw}}^{ij} = c_j^* k_{\mathrm{bw}}^{ij} \ \ \text{for} \ \ 1 \le i < j \le I. \tag{2.14}$$

Using this condition, the RRE can be written coordinate-wise as

$$\frac{\dot{c}_i}{c_i^*} = \sum_{j<i} k_{\mathrm{bw}}^{ji} \left(\frac{c_j}{c_j^*} - \frac{c_i}{c_i^*} \right) + \sum_{j>i} k_{\mathrm{fw}}^{ij} \left(\frac{c_j}{c_j^*} - \frac{c_i}{c_i^*} \right), \ \ \text{or equiv.,} \ \dot{c}_i = \sum_{j \ne i} \kappa_*^{ij} \left(\frac{c_j}{c_j^*} - \frac{c_i}{c_i^*} \right).$$

Here we used the notational convention that $\kappa_*^{ij} := \kappa_*^{ji}$ for $j < i$. The relative entropy E is as above and the Onsager matrix takes the form

$$\mathbb{K}_{\mathrm{M}}(\boldsymbol{c}) = \sum_{i<j} \kappa_*^{ij} \Lambda\left(\frac{c_i}{c_i^*}, \frac{c_j}{c_j^*} \right)(\boldsymbol{e}_i - \boldsymbol{e}_j) \otimes (\boldsymbol{e}_i - \boldsymbol{e}_j), \tag{2.15}$$

where $\boldsymbol{e}_i \in \mathbb{R}^I$ denotes the i-th unit vector. We then have the gradient structure $(\boldsymbol{C}, E, \mathbb{K}_{\mathrm{M}})$, namely

$$\dot{\boldsymbol{c}} = \mathcal{A}\boldsymbol{c} = -\mathbb{K}_{\mathrm{M}}(\boldsymbol{c})\mathrm{D}E(\boldsymbol{c}).$$

This gradient flow structure has been found in the independent works [Maa11] (which deals with Markov chains exclusively) and [Mie11] (in the setting of reaction-diffusion systems, in which Markov

chains are implicitly contained). The related work [CH*12] deals with discretizations of Fokker–Plank equations.

In fact, for the construction of gradient structures for Markov chains $\dot{\boldsymbol{c}} = \mathcal{A}\boldsymbol{c}$ we do not need the summation rule for logarithms. Hence, following [Maa11, Mie13] there are more general gradient structures. Choosing a strictly convex function $\phi : [0, \infty[\to \mathbb{R}$ that is smooth on $]0, \infty[$ we set

$$E^{\phi}(\boldsymbol{c}) := \sum_{i=1}^{n} c_i^* \phi\left(\frac{c_i}{c_i^*}\right), \quad \mathbb{K}_{\mathrm{M}}^{\phi}(\boldsymbol{c}) = \sum_{j=2}^{n} \sum_{i=1}^{j-1} \kappa_*^{ij} \Phi\left(\frac{c_i}{c_i^*}, \frac{c_j}{c_j^*}\right) (\boldsymbol{e}_i - \boldsymbol{e}_j) \otimes (\boldsymbol{e}_i - \boldsymbol{e}_j), \quad (2.16)$$

where $\Phi(a, b) = (a - b)/(\phi'(a) - \phi'(b))$ for $0 < a \neq b$ and $\Phi(a, a) = 1/\phi''(a)$. The gradient flow structure $(\boldsymbol{C}, E, \mathbb{K}_{\mathrm{M}})$ corresponds to the case where $\phi = \lambda_{\mathrm{B}} : z \mapsto z \log z - z + 1$.

Proposition 2.3 (Gradient structure for CTMC) *If the CTMC* (2.13) *satisfies the detailed-balance condition* (2.14) *for a positive steady state* $\boldsymbol{c}_* = (c_i^*)_{i=1,\dots,I}$*, then it has the gradient structures* $(\boldsymbol{C}, E^{\phi}, K_{\mathrm{M}}^{\phi})$*, namely* $\dot{\boldsymbol{c}} = \mathcal{A}\boldsymbol{c} = -K_{\mathrm{M}}^{\phi}(\boldsymbol{c})\mathrm{D}E^{\phi}(\boldsymbol{c})$.

Remark 2.4 *The construction in Proposition 2.3 does not extend to general RRE. There one would need to replace the quantity* $\Lambda\left(\frac{\boldsymbol{c}^{\boldsymbol{\alpha}^r}}{\boldsymbol{c}_*^{\boldsymbol{\alpha}^r}}, \frac{\boldsymbol{c}^{\boldsymbol{\beta}^r}}{\boldsymbol{c}_*^{\boldsymbol{\beta}^r}}\right)$ *in* (2.9) *by* $\left(\frac{\boldsymbol{c}^{\boldsymbol{\alpha}^r}}{\boldsymbol{c}_*^{\boldsymbol{\alpha}^r}} - \frac{\boldsymbol{c}^{\boldsymbol{\beta}^r}}{\boldsymbol{c}_*^{\boldsymbol{\beta}^r}}\right)/\left((\boldsymbol{\alpha}^r - \boldsymbol{\beta}^r) \cdot \phi'(\frac{\boldsymbol{c}}{\boldsymbol{c}_*})\right)$*, but this quantity can be negative in general. As a consequence, the corresponding Onsager matrix would not be positive definite. This cannot happen for Markov chains (i.e., when* $\boldsymbol{\alpha} = \boldsymbol{e}_i$ *and* $\boldsymbol{\beta} = \boldsymbol{e}_j$*), by virtue of the convexity of* ϕ.

In the following we will mainly concentrate on the gradient structure $(\boldsymbol{C}, E, \mathbb{K}_{\mathrm{M}})$ with the logarithmic entropy, as it is the only one that connects with the RRE.

2.5 Generalized gradient structures

For Markov chains and RRE there are several families of *generalized gradient structures* $(\boldsymbol{C}, E, \Psi^*)$ where the quadratic function $\Psi^*(\boldsymbol{c}, \boldsymbol{\zeta}) = \frac{1}{2}\langle \boldsymbol{\zeta}, \mathbb{K}(\boldsymbol{c})\boldsymbol{\zeta}\rangle$ is replaced by a general *dual dissipation potential* $\Psi^*(\boldsymbol{c}, \cdot) : \mathbb{R}^I \to [0, \infty[$ that is continuous and convex and satisfies $\Psi^*(\boldsymbol{c}, 0) = 0$.

In the case of RRE, the monomial terms $\boldsymbol{c}^{\boldsymbol{\alpha}}$ can only be generated by the logarithmic summation rule $\sum_{i=1}^{I} \log(b_i) = \log\left(\Pi_{i=1}^{I} b_i\right)$. Hence, we stick to the relative entropy E defined in (2.9), i.e., $\phi(z) = \lambda_{\mathrm{B}}(z)$. However, we may replace the linear Onsager principle $\dot{\boldsymbol{c}} = -\mathbb{K}(\boldsymbol{c})\mathrm{D}E(\boldsymbol{c})$ by the more general nonlinear form $\dot{\boldsymbol{c}} = \partial_{\boldsymbol{\zeta}}\Psi^*(\boldsymbol{c}, -\mathrm{D}E(\boldsymbol{c}))$.

To define Ψ^* we choose an arbitrary family of smooth dissipation functionals $\psi_r : \mathbb{R} \to [0, \infty[$, i.e., $\psi_r(0) = \psi_r'(0) = 0$ and $\psi_r'' > 0$ and define the dissipation potential

$$\Psi^*(\boldsymbol{c}, \boldsymbol{\zeta}) = \sum_{r=1}^{R} L_r(\boldsymbol{c})\, \psi_r\left((\boldsymbol{\alpha}^r - \boldsymbol{\beta}^r) \cdot \boldsymbol{\zeta}\right) \text{ with } L_r(\boldsymbol{c}) = \kappa_*^r \frac{\frac{\boldsymbol{c}^{\boldsymbol{\beta}^r}}{\boldsymbol{c}_*^{\boldsymbol{\beta}^r}} - \frac{\boldsymbol{c}^{\boldsymbol{\alpha}^r}}{\boldsymbol{c}_*^{\boldsymbol{\alpha}^r}}}{\psi_r'\left(\log\frac{\boldsymbol{c}^{\boldsymbol{\beta}^r}}{\boldsymbol{c}_*^{\boldsymbol{\beta}^r}} - \log\frac{\boldsymbol{c}^{\boldsymbol{\alpha}^r}}{\boldsymbol{c}_*^{\boldsymbol{\alpha}^r}}\right)}. \quad (2.17)$$

Using (2.11) we easily obtain $-\boldsymbol{R}(\boldsymbol{c}) = \partial_{\boldsymbol{\zeta}}\Psi^*(\boldsymbol{c}, -\mathrm{D}E(\boldsymbol{c}))$, i.e., $\dot{\boldsymbol{c}} = -\boldsymbol{R}(\boldsymbol{c})$ is generated by the generalized gradient system $(\boldsymbol{C}, E, \Psi^*)$.

The case $\psi_r(\zeta) = \frac{1}{2}\zeta^2$ leads to the quadratic dissipation potential in (2.9), i.e., the functions L_r are given in terms of the logarithmic mean. In [AGH02] the choices $\psi_r(\pm\zeta) = \mathrm{e}^{\zeta} - 1 - \zeta$ is used. Based

on a derivation via the large deviation principle (see [MPR14, MRP16, MP*17]) a special role is played by the choice of a "cosh-type" function ψ_r:

$$\psi_r(\zeta) = \mathsf{C}^*(\zeta) := 4\cosh\left(\frac{1}{2}\zeta\right) - 4 \quad \text{giving} \quad L_r(\boldsymbol{c}) = \kappa_*^r \left(\frac{\boldsymbol{c}^{\boldsymbol{\alpha}^r}}{\boldsymbol{c}_*^{\boldsymbol{\beta}^r}} \, \frac{\boldsymbol{c}^{\boldsymbol{\beta}^r}}{\boldsymbol{c}_*^{\boldsymbol{\beta}^r}}\right)^{1/2}. \tag{2.18}$$

Here C^* is normalized such that $\mathsf{C}^*(\zeta) = \frac{1}{2}\zeta^2 + O(\zeta^4)$. Hence, the dual dissipation potential takes the form

$$\Psi_{\cosh}^*(\boldsymbol{c}, \boldsymbol{\zeta}) := \sum_{r=1}^{R} \kappa_*^r \left(\frac{\boldsymbol{c}^{\boldsymbol{\alpha}^r}}{\boldsymbol{c}_*^{\boldsymbol{\alpha}^r}} \, \frac{\boldsymbol{c}^{\boldsymbol{\beta}^r}}{\boldsymbol{c}_*^{\boldsymbol{\beta}^r}}\right)^{1/2} \mathsf{C}^*\big((\boldsymbol{\alpha}^r - \boldsymbol{\beta}^r) \cdot \boldsymbol{\zeta}\big). \tag{2.19}$$

It is shown in [MiS19, Prop. 4.1] that this generalized gradient structure is distinguished as the only tilt-invariant gradient structure for CTMCs.

3 The chemical master equation

3.1 Modeling discrete particle numbers via CME

The chemical master equation (CME) is a CTMC that is defined on the set $\mathcal{N} = \mathbb{N}_0^I$ where $\boldsymbol{n} = (n_1, \ldots, n_I) \in \mathcal{N}$ is the vector of particle numbers, see [ML*11] for an introduction. This means that $n_i \in \mathbb{N}_0$ denotes the number of particles of species X_i in a sufficiently big volume, whose size is denoted by $V > 0$. The modeling assumes that all particles move randomly in this big volume (well-stirred tank reactor) so that they can meet independently. The dynamics is formulated in terms of the probabilities

$$u_{\boldsymbol{n}}(t) = \text{probability that at time } t \text{ there are } n_i \text{ particles of species } X_i \text{ for } i = 1, \ldots, I.$$

All the R reaction pairs may happen independently of each other according to the number of the available atoms needed for the reactions and the reaction coefficients $k_{\text{fw}}^r \geq 0$ and $k_{\text{bw}}^r \geq 0$, respectively. Moreover, the jump intensities

$$k_{\text{fw}}^r \mathbb{B}_V^{\boldsymbol{\alpha}^r}(\boldsymbol{n}) \text{ from } \boldsymbol{n} + \boldsymbol{\alpha}^r \text{ to } \boldsymbol{n} + \boldsymbol{\beta}^r \quad \text{and} \quad k_{\text{bw}}^r \mathbb{B}_V^{\boldsymbol{\beta}^r}(\boldsymbol{n}) \text{ from } \boldsymbol{n} + \boldsymbol{\beta}^r \text{ to } \boldsymbol{n} + \boldsymbol{\alpha}^r$$

also depend on the volume V, as n_i denotes the absolute particle number, while for the reaction the densities $c_i = n_i/V$ matter. The specific form of $\mathbb{B}_V^{\boldsymbol{\alpha}}(\boldsymbol{n})$ (cf. [Kur70, ML*11]) reads

$$\mathbb{B}_V^{\boldsymbol{\alpha}}(\boldsymbol{n}) = \begin{cases} \dfrac{V(\boldsymbol{n}+\boldsymbol{\alpha})!}{V^{|\boldsymbol{\alpha}|}\boldsymbol{n}!} & \text{for } \boldsymbol{n} \in \mathcal{N}, \\ 0 & \text{for } \boldsymbol{n} \notin \mathcal{N}, \end{cases} \quad \text{where } \boldsymbol{n}! = \prod_{i=1}^{I} n_i! \,. \tag{3.1}$$

To avoid clumsy notation we defined $\mathbb{B}_V^{\boldsymbol{\alpha}}(\boldsymbol{n})$ for all $\boldsymbol{n} \in \mathbb{Z}^I$, but $\mathbb{B}_V^{\boldsymbol{\alpha}}(\boldsymbol{n}) = 0$ if $\boldsymbol{n} \notin \mathcal{N}$. We also see that $\mathbb{B}_V^{\boldsymbol{\alpha}}(\boldsymbol{n}) \approx V\boldsymbol{c}^{\boldsymbol{\alpha}}$ for $\boldsymbol{c} = \frac{1}{V}\boldsymbol{n}$, where the factor V indicates that the number of reactions is proportional to the volume of the container, if the densities are kept constant.

The CME associated with the RRE (2.2) is the Kolmogorov forward equation for the probability distributions $\boldsymbol{u} = (u_{\boldsymbol{n}})_{\boldsymbol{n}\in\mathcal{N}} \in \mathscr{P}(\mathcal{N})$, namely

$$\dot{\boldsymbol{u}} = \sum_{r=1}^{R} \overline{\mathcal{B}}_V^r \boldsymbol{u} \text{ with } \big(\overline{\mathcal{B}}_V^r \boldsymbol{u}\big)_{\boldsymbol{n}} = k_{\text{fw}}^r\big(\mathbb{B}_V^{\boldsymbol{\alpha}^r}(\boldsymbol{n}-\boldsymbol{\beta}^r)\, u_{\boldsymbol{n}+\boldsymbol{\alpha}^r-\boldsymbol{\beta}^r} - \mathbb{B}_V^{\boldsymbol{\alpha}^r}(\boldsymbol{n}-\boldsymbol{\alpha}^r)\, u_{\boldsymbol{n}}\big)$$

$$+ k_{\text{bw}}^r\big(\mathbb{B}_V^{\boldsymbol{\beta}^r}(\boldsymbol{n}-\boldsymbol{\alpha}^r)\, u_{\boldsymbol{n}-\boldsymbol{\alpha}^r+\boldsymbol{\beta}^r} - \mathbb{B}_V^{\boldsymbol{\beta}^r}(\boldsymbol{n}-\boldsymbol{\beta}^r)\, u_{\boldsymbol{n}}\big). \tag{3.2}$$

The rth forward reaction from $\boldsymbol{n}+\boldsymbol{\alpha}^r$ to $\boldsymbol{n}+\boldsymbol{\beta}^r$ can only happen (i.e., $\mathbb{B}_V^{\boldsymbol{\alpha}^r}(\boldsymbol{n}) > 0$) if $\boldsymbol{n} \geq \boldsymbol{0}$. Hence any occurring $u_{\boldsymbol{m}}$ with $\boldsymbol{m} \notin \mathcal{N}$ is multiplied by intensity 0, so in (3.2) we may set $u_{\boldsymbol{m}} \equiv 0$ for all $\boldsymbol{m} \notin \mathcal{N}$. The operators $\overline{\mathcal{B}}_V^r$ are the adjoints of the Markov generators $\mathcal{Q}_V^r$ given by

$$(\mathcal{Q}_V^r \boldsymbol{\mu})_{\boldsymbol{n}} = k_{\mathrm{fw}}^r \mathbb{B}_V^{\boldsymbol{\alpha}^r}(\boldsymbol{n}-\boldsymbol{\alpha}^r)(\mu_{\boldsymbol{n}-\boldsymbol{\alpha}^r+\boldsymbol{\beta}^r} - \mu_{\boldsymbol{n}}) + k_{\mathrm{bw}}^r \mathbb{B}_V^{\boldsymbol{\beta}^r}(\boldsymbol{n}-\boldsymbol{\beta}^r)(\mu_{\boldsymbol{n}+\boldsymbol{\alpha}^r-\boldsymbol{\beta}^r} - \mu_{\boldsymbol{n}}) \quad (3.3)$$

for $\boldsymbol{\mu} = (\mu_{\boldsymbol{n}})_{\boldsymbol{n} \in \mathcal{N}}$.

We emphasize that the RRE as well as the CME are uniquely specified if the reaction network (2.1), the reaction rates k_{fw}^r and k_{bw}^r, and the volume $V > 0$ are given. Hence, there are obviously close relations between both models, in particular for $V \gg 1$, see [Kur70, Gil92, AC*05, ACK10, WiS17].

So far, we have not used the detailed-balance condition, i.e., we can even allow for $k_{\mathrm{bw}}^r = 0$ in the above considerations. In all cases, the Kolmogorov forward equation is an infinite-dimensional linear ODE as in Section 2.4. The following result shows that the detailed-balance condition is inherited from the RRE to the CME, and moreover a simple equilibrium $\boldsymbol{w}^V$ can be given explicitly as a product distribution of individual Poisson distributions, namely $m \mapsto \mathrm{e}^{-Vc_i^*}(Vc_i^*)^m/m!$. This result can also be retrieved from [ACK10] by combining Theorems 4.1 and 4.5 there, where it is shown that the weaker "complex-balance condition" is sufficient to guarantee that the Poisson distribution $\boldsymbol{w}^V$ is an equilibrium for CME.

For completeness we give a short and independent proof of the fundamental result that for RRE with detailed balance the associated CME satisfies detailed balance again.

Theorem 3.1 (Detailed balance for CME) *Let $\mathbb{B}_V^{\boldsymbol{\alpha}}(\boldsymbol{n})$ be given in the form* (3.1). *Assume that* (2.2) *has the equilibrium $\boldsymbol{c}_* \in]0,\infty[^I$ satisfying the detailed-balance condition* (2.6). *Then the equilibrium $\boldsymbol{w}^V := (w_{\boldsymbol{n}}^V)_{\boldsymbol{n} \in \mathcal{N}} \in \mathscr{P}(\mathcal{N})$ given by*

$$w_{\boldsymbol{n}}^V = \frac{1}{Z_V^*} \frac{(V\boldsymbol{c}_*)^{\boldsymbol{n}}}{\boldsymbol{n}!} \quad \text{with } Z_V^* := \Pi_{i=1}^I \mathrm{e}^{Vc_i^*}$$

satisfies the detailed-balance condition for the CME (3.2), *namely*

$$\forall\, r = 1,\dots,R \;\forall\, \boldsymbol{n} \in \mathcal{N}: \quad k_{\mathrm{fw}}^r \mathbb{B}_V^{\boldsymbol{\alpha}^r}(\boldsymbol{n}) w_{\boldsymbol{n}+\boldsymbol{\alpha}^r}^V = k_{\mathrm{bw}}^r \mathbb{B}_V^{\boldsymbol{\beta}^r}(\boldsymbol{n}) w_{\boldsymbol{n}+\boldsymbol{\beta}^r}^V = \kappa_*^r V w_{\boldsymbol{n}}^V =: \widehat{\nu}_V^{\boldsymbol{n},r}.$$

Proof. For each reaction we obtain the relation

$$k_{\mathrm{fw}}^r \mathbb{B}_V^{\boldsymbol{\alpha}^r}(\boldsymbol{n}) w_{\boldsymbol{n}+\boldsymbol{\alpha}^r}^V = k_{\mathrm{fw}}^r \frac{V(\boldsymbol{n}+\boldsymbol{\alpha}^r)!}{V^{|\boldsymbol{\alpha}^r|}\boldsymbol{n}!} \frac{(V\boldsymbol{c}_*)^{\boldsymbol{n}+\boldsymbol{\alpha}^r}}{Z_V^*(\boldsymbol{n}+\boldsymbol{\alpha}^r)!} = k_{\mathrm{fw}}^r V \boldsymbol{c}_*^{\boldsymbol{\alpha}^r} \frac{(V\boldsymbol{c}_*)^{\boldsymbol{n}}}{Z_V^* \boldsymbol{n}!} = V\kappa_*^r w_{\boldsymbol{n}}^V.$$

Analogously we obtain the same result for $k_{\mathrm{bw}} \mathbb{B}_V^{\boldsymbol{\beta}^r}(\boldsymbol{n}) w_{\boldsymbol{n}+\boldsymbol{\beta}^r}^V$, where the detailed-balance condition (2.6) is used in the definition of κ_*^r. ∎

Using the detailed-balance coefficients $\widehat{\nu}_V^{\boldsymbol{n},r}$ we can rewrite the operator $\overline{\mathcal{B}}_V^r$ from (3.2) in a symmetrically balanced form as

$$\overline{\mathcal{B}}_V^r \boldsymbol{u} = \sum_{\boldsymbol{n} \in \mathcal{N}} \widehat{\nu}_V^{\boldsymbol{n},r} \left(\frac{u_{\boldsymbol{n}+\boldsymbol{\alpha}^r}}{w_{\boldsymbol{n}+\boldsymbol{\alpha}^r}^V} - \frac{u_{\boldsymbol{n}+\boldsymbol{\beta}^r}}{w_{\boldsymbol{n}+\boldsymbol{\beta}^r}^V} \right) \left(\boldsymbol{e}^{(\boldsymbol{n}+\boldsymbol{\beta}^r)} - \boldsymbol{e}^{(\boldsymbol{n}+\boldsymbol{\alpha}^r)} \right), \quad (3.4)$$

where $\boldsymbol{e}^{(\boldsymbol{m})}$ is the unit vector, i.e., $\boldsymbol{e}_{\boldsymbol{n}}^{(\boldsymbol{m})} = \delta_{\boldsymbol{n}-\boldsymbol{m}}$.

It is important to realize that in general the steady state for the detailed-balance condition is highly non-unique, because of the discrete versions

$$\mathcal{I}(\overline{\boldsymbol{n}}) := \{\, \boldsymbol{n} \in \mathcal{N} \mid \mathbb{Q}\boldsymbol{n} = \mathbb{Q}\overline{\boldsymbol{n}} \,\} \subset \mathcal{N}$$

of the invariant stoichiometric subspaces $I(q) = \{\, c \in C \mid \mathbb{Q}c = q \,\} \subset C$. Indeed, choosing $\overline{n}$ arbitrary and defining $\overline{w} = (\overline{w}_n) \in \mathscr{P}(\mathcal{N})$ via $\overline{w}_n = \frac{1}{Z}w_n^V$ for $n \in \mathcal{I}(\overline{n})$ and $\overline{w}_n = 0$ elsewhere, we obtain another equilibrium for the CME (3.2). Defining convex combination we obtain a rich family of steady states.

The following counterexamples show that the above result, which is central to our work, cannot be expected for systems not satisfying the detailed-balance condition.

Example 3.2 (Equation without detailed balance) *For $a, b \in \mathbb{N}$ we consider the RRE*

$$\dot{c} = 2a - 4b\,c + 2\,(1-c^2), \tag{3.5}$$

which consists of two individual reaction pairs, namely $X \overset{2a}{\underset{4b}{\rightleftharpoons}} \emptyset$ and $2X \overset{1}{\underset{1}{\rightleftharpoons}} \emptyset$ with the individual steady states $c_{(1)} = a/(2b)$ and $c_{(2)} = 1$. The joint steady state of (3.5) is $c_* = (1+a+b^2)^{1/2} - b$, and we have detailed balance if and only if $a = 2b$.

Building the CME according to (3.2) based on the two reaction pairs we obtain

$$\dot{u}_n = 2aVu_{n-1} - (2aV+4bn)u_n + 4b(n+1)u_{n+1} + Vu_{n-2} - \left(V + \tfrac{n(n-1)_+}{V}\right)u_n + \tfrac{(n+2)(n+1)}{V}u_{n+2}.$$

For the case $a = 2$ and $b = 1$, where the detailed-balance condition holds with $c_* = 1 = c_{(1)} = c_{(2)}$, we obtain

$$\dot{u}_n = Vu_{n-2} + 4Vu_{n-1} - \left(5V + 4n + \tfrac{n(n-1)_+}{V}\right)u_n + 4(n+1)u_{n+1} + \tfrac{(n+2)(n+1)}{V}u_{n+2},$$

and it is easy to check that $\widetilde{w}^V = (\mathrm{e}^{-V}V^n/n!)_{n\in\mathbb{N}_0}$ is a steady state.

However, for $a = 7$ and $b = 1$ the detailed-balance condition fails with $c_{(1)} = 7/2 > c_* = 2 > c_{(2)} = 1$. The CME reads

$$\dot{u}_n = Vu_{n-2} + 14Vu_{n-1} - \left(15V + 4n + \tfrac{n(n-1)_+}{V}\right)u_n + 4(n+1)u_{n+1} + \tfrac{(n+2)(n+1)}{V}u_{n+2}.$$

An explicit calculation shows that the Poisson distribution based on $c_* = 2$, i.e., $\widetilde{w}_n^V = \mathrm{e}^{-2V}(2V)^n/n!$, is not a steady state. Indeed, inserting $\widetilde{w}^V$ into the right-hand side of the last equation we find (for $n \geq 1$)

$$\dot{u}_n|_{u=\widetilde{w}^V} = \frac{\mathrm{e}^{-2V}(2V)^{n-2}}{n!}\left(-12V^3 + 12nV^2 - 3n(n-1)V\right) \neq 0 \text{ for general } n \in \mathbb{N}.$$

Example 3.3 (Microscopic versus macroscopic detailed balance) *We may also consider a RRE that looks macroscopically as being in detailed balance, but is generated by a microscopic model that is not in detailed balance. The two reactions $\emptyset \overset{2}{\rightarrow} X$ and $2X \overset{1}{\rightarrow} \emptyset$ produce the RRE $\dot{c} = 2(1-c^2)$ that has the equilibrium $c_* = 1$. However the CME reads*

$$\dot{u}_n = 2Vu_{n-1} - \left(2V + \frac{n(n-1)}{V}\right)u_n + \frac{(n+2)(n+1)}{V}u_{n+2}.$$

Again, the Poisson distribution $\widetilde{w}^V$ with $w_n^V = \mathrm{e}^{-V}V^n/n!$ is not the equilibrium:

$$\dot{u}_n|_{u=\widetilde{w}^V} = \frac{\mathrm{e}^{-V}V^{n-1}}{n!}\left(2Vn - V^2 - n(n-1)\right) \neq 0.$$

Note that the reversible reaction pair $2X \overset{1}{\underset{1}{\rightleftharpoons}} \emptyset$ yields the same RRE, and its associated CME satisfies the detailed-balance condition.

3.2 Existence and uniqueness of solutions of CME

In this part we establish well-posedness for the CME. We do this by combining classical results from the theory of Markov chains with abstract semigroup theory.

For fixed $\boldsymbol{n}_0 \in \mathcal{N}$ we construct a special Green's function $p_t(\boldsymbol{n}_0, \cdot)$. General Markov chain theory (e.g., [Lig10, Ch. 2]) implies that there exist a unique minimal solution $[0, \infty[\times \mathcal{N} \ni (t, \boldsymbol{n}) \mapsto p_t(\boldsymbol{n}_0, \boldsymbol{n})$ to the backward equation

$$\dot{p}_t(\boldsymbol{n}_0, \boldsymbol{n}) = \sum_{r=1}^R \Big(k_{\text{fw}}^r \, \mathbb{B}_V^{\boldsymbol{\alpha}^r}(\boldsymbol{n}_0 - \boldsymbol{\alpha}^r) \big(p_t(\boldsymbol{n}_0 - \boldsymbol{\alpha}^r + \boldsymbol{\beta}^r, \boldsymbol{n}) - p_t(\boldsymbol{n}_0, \boldsymbol{n}) \big)$$

$$+ k_{\text{bw}}^r \, \mathbb{B}_V^{\boldsymbol{\beta}^r}(\boldsymbol{n}_0 - \boldsymbol{\beta}^r) \big(p_t(\boldsymbol{n}_0 + \boldsymbol{\alpha}^r - \boldsymbol{\beta}^r, \boldsymbol{n}) - p_t(\boldsymbol{n}_0, \boldsymbol{n}) \big) \Big)$$

associated with the CME with initial condition $p_0(\boldsymbol{n}_0, \boldsymbol{n}) = \delta_{\boldsymbol{n}_0}(\boldsymbol{n})$. This minimal solution is non-negative and satisfies $p_t(\boldsymbol{n}_0, \boldsymbol{n}) \geq 0$ and $\sum_{\boldsymbol{n} \in \mathcal{N}} p_t(\boldsymbol{n}_0, \boldsymbol{n}) \leq 1$, but for general CTMC it can happen that the latter inequality is strict, which means that the corresponding Markov chain explodes in finite time. We will show that explosion does not happen for CME with detailed balance.

For the functional analytic existence and uniqueness result we use the sequence spaces

$$\ell^p(\mathcal{N}) := \Big\{ \boldsymbol{u} = (u_{\boldsymbol{n}})_{\boldsymbol{n} \in \mathcal{N}} \,\Big|\, \sum_{\boldsymbol{n} \in \mathcal{N}} |u_{\boldsymbol{n}}|^p < \infty \Big\}$$

as well as the weighted spaces

$$\mathrm{L}^p(\mathcal{N}, \boldsymbol{w}^V) := \Big\{ \boldsymbol{v} = (v_{\boldsymbol{n}})_{\boldsymbol{n} \in \mathcal{N}} \,\Big|\, \sum_{\boldsymbol{n} \in \mathcal{N}} \big| \frac{v_{\boldsymbol{n}}}{w_{\boldsymbol{n}}^V} \big|^p < \infty \Big\}$$

with the corresponding norms and the usual modification for $p = \infty$. Now, we consider the transition semigroup $(\mathcal{P}_t)_{t \geq 0}$ defined by

$$(\mathcal{P}_t \boldsymbol{v})_{\boldsymbol{n}} := \sum_{\boldsymbol{m} \in \mathcal{N}} p_t(\boldsymbol{n}, \boldsymbol{m}) v_{\boldsymbol{m}}, \quad \boldsymbol{v} = (v_{\boldsymbol{m}}) \in \ell^\infty(\mathcal{N}),$$

which we shall study by induction over the number R of reactions using the Trotter-Kato formula, where the detailed-balance condition guarantees that each subsystem is a contraction semigroup on $\mathrm{L}^2(\mathcal{N}, \boldsymbol{w}^V)$.

Theorem 3.4 *Assume that the detailed balance condition* (2.6) *holds. Then, the semigroup* $(\mathcal{P}_t)_{t \geq 0}$ *extends to a* C_0*-semigroup of contractions on* $\mathrm{L}^p(\mathcal{N}, \boldsymbol{w}^V)$ *for all* $1 \leq p < \infty$. *Moreover, the semigroup is selfadjoint on* $\mathrm{L}^2(\mathcal{N}, \boldsymbol{w}^V)$ *and Markovian, i.e.,* $\mathcal{P}_t \mathbf{1} = \mathbf{1}$ *for all* $t \geq 0$.

A related existence result for the Markov semigroup of the CME was established in [GaY14], which however does not apply to the case of reversible RRE, because of the restrictions on the growth of the transition rates.

Proof. All of the above statements follow from the general theory of continuous time Markov chains, except for the Markovianity. To show the latter, we first consider the case of a single reaction, thus $R = 1$. Each of the irreducible components of the state space $\mathcal{N}$ is then one-dimensional (see also [MaM20]), and the Markov chain is a birth-death chain on a countable (possibly finite) set.

If there exist two components of $\boldsymbol{\alpha} - \boldsymbol{\beta}$ with opposite sign, then each of the irreducible components of the state space $\mathcal{N}$ is finite. Therefore it is clear that the Markov chain does not explode in finite time.

Suppose now that all components of $\boldsymbol{\alpha} - \boldsymbol{\beta}$ have equal sign, say $\alpha_i - \beta_i \geq 0$ for all $i = 1, \ldots, I$, and at least one component is strictly positive. Then each of the infinite irreducible components of $\mathcal{N}$ is of the form

$$\{ \, \boldsymbol{n}^{(k)} := \boldsymbol{n}^{(0)} + k(\boldsymbol{\alpha} - \boldsymbol{\beta}) \mid k \in \mathbb{N}_0 \, \}$$

for some $\boldsymbol{n}^{(0)} \in \mathcal{N}$, and the restricted Markov process is a birth-death process with birth rate b_k and death rate d_k given by

$$\mathsf{b}_k := k_{\mathrm{bw}} \mathbb{B}_V^{\boldsymbol{\beta}}(\boldsymbol{n}^{(k)} - \boldsymbol{\beta}) \ \text{ from } \boldsymbol{n}^{(k)} \text{ to } \boldsymbol{n}^{(k+1)} \quad \text{and}$$
$$\mathsf{d}_k := k_{\mathrm{fw}} \mathbb{B}_V^{\boldsymbol{\alpha}}(\boldsymbol{n}^{(k)} - \boldsymbol{\alpha}) \ \text{ from } \boldsymbol{n}^{(k)} \text{ to } \boldsymbol{n}^{(k-1)}.$$

Reuter's criterion ([Reu57, Thm. 11]) gives a characterization of non-explosion for birth-death chains; it asserts that the chain is non-explosive if and only if

$$\sum_{k \geq j \geq 0} \mathsf{r}_{j,k} = \infty, \quad \text{where } \mathsf{r}_{j,k} := \frac{\mathsf{d}_k \cdot \ldots \cdot \mathsf{d}_{j+1}}{\mathsf{b}_k \cdot \ldots \cdot \mathsf{b}_j}.$$

In our setting we have $\frac{\mathsf{d}_{k+1}}{\mathsf{b}_k} = (V\boldsymbol{c}^*)^{\boldsymbol{\beta}-\boldsymbol{\alpha}} \frac{\boldsymbol{n}^{(k+1)}!}{\boldsymbol{n}^{(k)}!}$, so that $\mathsf{r}_{0,k} = \frac{1}{\mathsf{b}_k}(V\boldsymbol{c}^*)^{k(\boldsymbol{\beta}-\boldsymbol{\alpha})} \frac{\boldsymbol{n}^{(k)}!}{\boldsymbol{n}^{(0)}!}$, and therefore

$$\sum_{k \geq j \geq 0} \mathsf{r}_{j,k} \geq \sum_{k \geq 0} \mathsf{r}_{0,k} \geq \frac{V^{|\boldsymbol{\beta}|-1}}{k_{\mathrm{bw}} \boldsymbol{n}^{(0)}!} \sum_{k \geq 0} \frac{(\boldsymbol{n}^{(k)} - \boldsymbol{\beta})!}{(V\boldsymbol{c}^*)^{k(\boldsymbol{\alpha}-\boldsymbol{\beta})}}.$$

Since the summands tend to ∞ as $k \to \infty$, we infer that the latter sum is infinite; hence the Markov chain is non-explosive, or equivalently $\mathcal{P}_t \mathbf{1} = \mathbf{1}$ (see [Lig10, Thm. 2.33]).

The case of multiple reactions follows by induction on the number of reactions R. Indeed, for $\mathcal{R} \subseteq \{1, \ldots, R\}$, let $(\mathcal{P}_t^{\mathcal{R}})_{t \geq 0}$ denote the semigroup corresponding to the reactions $r \in \mathcal{R}$. Then the Trotter product formula for contraction semigroups on $\mathrm{L}^2(\mathcal{N}, \boldsymbol{w}^V)$ (see e.g., [Dav80]) asserts that

$$\mathcal{P}_t^{\{1,\ldots,R+1\}} = \lim_{n \to \infty} \left(\mathcal{P}_{t/n}^{\{1,\ldots,R\}} \mathcal{P}_{t/n}^{\{R+1\}} \right)^n$$

strongly in $\mathrm{L}^2(\mathcal{N}, \boldsymbol{w}^V)$. Note that we can apply this formula, since the detailed-balance conditions hold for all reactions simultaneously, hence all of the semigroups are contractive on the same space $\mathrm{L}^2(\mathcal{N}, \boldsymbol{w}^V)$. We also observe that the class of finitely supported functions is a core for each of the generators. The Markovianity of $\mathcal{P}^{\{1,\ldots,R+1\}}$ thus follows from the Trotter formula and the Markovianity of $\mathcal{P}^{\{1,\ldots,R\}}$ and $\mathcal{P}^{\{R+1\}}$. ∎

Remark 3.5 *The mere existence of a probability distribution satisfying the detailed-balance equations is not sufficient to guarantee non-explosion of a continuous time Markov chain. It might happen that the chain jumps infinitely often in a finite time interval, see [Nor97, Sec. 3.5] for an example. The previous result shows that this phenomenon does not occur in CME satisfying the detailed-balance condition.*

It remains to transfer the results from $\mathrm{L}^1(\mathcal{N}, \boldsymbol{w}^V)$ to $\ell^1(\mathcal{N})$. Denoting by $\mathcal{Q}$ the generator of the C_0-semigroup $(\mathcal{P}_t)_{t \geq 0}$ on $\mathrm{L}^1(\mathcal{N}, \boldsymbol{w}^V)$, we define the operator $\mathcal{B} : \mathrm{Dom}(\mathcal{B}) \subseteq \ell^1(\mathcal{N}) \to \ell^1(\mathcal{N})$ by

$$\mathcal{B}\boldsymbol{u} = \boldsymbol{w}^V \mathcal{Q}(\boldsymbol{u}/\boldsymbol{w}^V), \quad \mathrm{Dom}(\mathcal{B}) = \{ \, \boldsymbol{u} \in \ell^1(\mathcal{N}) \mid \boldsymbol{u}/\boldsymbol{w}^V \in \mathrm{Dom}(\mathcal{Q}) \, \}.$$

This definition of $\mathcal{B}$ is consistent with the explicit formula for $\mathcal{B}$ given above. Since $\mathcal{Q}$ generates a C_0-semigroup of contractions on $\mathrm{L}^1(\mathcal{N}, \boldsymbol{w}^V)$, it follows that $\mathcal{B}$ generates a C_0-semigroup $(\mathcal{P}_t)_{t \geq 0}$ of contractions on $\ell^1(\mathcal{N})$. Furthermore, since $\mathcal{P}_t$ preserves positivity and $\mathcal{P}_t \mathbf{1} = \mathbf{1}$, it follows that $\mathscr{P}(\mathcal{N})$ is invariant under the semigroup generated by $\mathcal{B}$.

As an immediate consequence we obtain global well-posedness for the CME in $\mathscr{P}(\mathcal{N})$.

Theorem 3.6 (Global well-posedness of the CME) *Let the detailed-balance condition* (2.6) *hold. Then, for all* $\boldsymbol{u}_0 \in \mathscr{P}(\mathcal{N})$ *there exists a unique mild solution* $\boldsymbol{u} : [0, \infty) \to \mathscr{P}(\mathcal{N})$ *to the CME* (3.2) *satisfying* $\boldsymbol{u}(0) = \boldsymbol{u}_0$.

3.3 Gradient structures for CME

Since the CME is the forward equation associated with a reversible CTMC, we can formulate it as a gradient flow in view of Proposition 2.3. Indeed, for a strictly convex function $\phi : [0, \infty[\to \mathbb{R}$ that is smooth on $]0, \infty[$, let us write

$$\mathcal{E}_V^\phi(\boldsymbol{u}) := \sum_{\boldsymbol{n} \in \mathcal{N}} w_{\boldsymbol{n}}^V \, \phi\Big(\frac{u_{\boldsymbol{n}}}{w_{\boldsymbol{n}}^V}\Big), \tag{3.6}$$

$$\mathcal{K}_V^\phi(\boldsymbol{u}) := \sum_{\boldsymbol{n} \in \mathcal{N}} \sum_{r=1}^{R} \widehat{\nu}_V^{\boldsymbol{n},r} \, \Phi\Big(\frac{u_{\boldsymbol{n}+\boldsymbol{\alpha}^r}}{w_{\boldsymbol{n}+\boldsymbol{\alpha}^r}^V}, \frac{u_{\boldsymbol{n}+\boldsymbol{\beta}^r}}{w_{\boldsymbol{n}+\boldsymbol{\beta}^r}^V}\Big) \, (\boldsymbol{e}^{(\boldsymbol{n}+\boldsymbol{\alpha}^r)} - \boldsymbol{e}^{(\boldsymbol{n}+\boldsymbol{\beta}^r)}) \otimes (\boldsymbol{e}^{(\boldsymbol{n}+\boldsymbol{\alpha}^r)} - \boldsymbol{e}^{(\boldsymbol{n}+\boldsymbol{\beta}^r)}),$$

where Φ is defined after (2.16), $\widehat{\nu}_V^{\boldsymbol{n},r}$ is given in Theorem 3.1, and $\boldsymbol{e}^{(\boldsymbol{m})}$ denotes the $\boldsymbol{m}$-th unit vector in $\ell^1(\mathcal{N})$. The following result is then a special case of Proposition 2.3.

Proposition 3.7 (Quadratic gradient structures for CME) *If the RRE* (2.2) *satisfies the detailed-balance condition* (2.6) *for a positive steady state* $\boldsymbol{c}_* = (c_i^*)_{i=1,\dots,I}$, *then the associated CME has the gradient structure* $(\mathscr{P}(\mathcal{N}), \mathcal{E}_V^\phi, \mathcal{K}_V^\phi)$ *defined in* (3.6), *namely*

$$\dot{\boldsymbol{u}} = \mathcal{B}_V \boldsymbol{u} = -\mathcal{K}_V^\phi(\boldsymbol{u}) \mathrm{D}\mathcal{E}_V^\phi(\boldsymbol{u}).$$

In the following we will mainly be concerned with the case that $\mathcal{E}_V^\phi$ is the logarithmic entropy, where ϕ is the Boltzmann function $\lambda_{\mathrm{B}}(z) = z \log z - z + 1$. In that case we obtain

$$\mathcal{E}_V(\boldsymbol{u}) := \frac{1}{V} \sum_{\boldsymbol{n} \in \mathcal{N}} w_{\boldsymbol{n}}^V \lambda_{\mathrm{B}}\Big(\frac{u_{\boldsymbol{n}}}{w_{\boldsymbol{n}}^V}\Big) = \frac{1}{V} \sum_{\boldsymbol{n} \in \mathcal{N}} \big(u_{\boldsymbol{n}} \log u_{\boldsymbol{n}} - u_{\boldsymbol{n}} \log w_{\boldsymbol{n}}^V\big), \tag{3.7}$$

$$\mathcal{K}_V(\boldsymbol{u}) := V \sum_{r=1}^{R} \sum_{\boldsymbol{n} \in \mathcal{N}} \widehat{\nu}_V^{\boldsymbol{n},r} \, \Lambda\Big(\frac{u_{\boldsymbol{n}+\boldsymbol{\alpha}^r}}{w_{\boldsymbol{n}+\boldsymbol{\alpha}^r}^V}, \frac{u_{\boldsymbol{n}+\boldsymbol{\beta}^r}}{w_{\boldsymbol{n}+\boldsymbol{\beta}^r}^V}\Big) \, (\boldsymbol{e}^{(\boldsymbol{n}+\boldsymbol{\alpha}^r)} - \boldsymbol{e}^{(\boldsymbol{n}+\boldsymbol{\beta}^r)}) \otimes (\boldsymbol{e}^{(\boldsymbol{n}+\boldsymbol{\alpha}^r)} - \boldsymbol{e}^{(\boldsymbol{n}+\boldsymbol{\beta}^r)}),$$

where the logarithmic mean $\Lambda(a, b)$ is defined in (2.10). The above definitions do not only restrict to the entropy function $\phi = \lambda_{\mathrm{B}}$, but also introduce a normalization with respect to the volume V. Hence, $\mathcal{E}_V$ can be seen as an entropy per unit volume. The corresponding scaling of $\mathcal{K}_V$ was chosen such that the evolution equation $\dot{\boldsymbol{u}} = -\mathcal{K}_V(\boldsymbol{u}) \mathrm{D}\mathcal{E}_V(\boldsymbol{u})$ is the same as $\dot{\boldsymbol{u}} = -\mathcal{K}_V^\phi(\boldsymbol{u}) \mathrm{D}\mathcal{E}_V^\phi(\boldsymbol{u})$.

For later purposes we also provide the cosh-type gradient structure for CME, whose relevance and usefulness is discussed in [MPR14, MP*17, FrL19, MiS19]. Recall the definition of C^* in (2.18) and note the special scaling via the volume V in (3.8) below, which is needed because $\boldsymbol{\Psi}_{\mathrm{cosh},V}^*(\boldsymbol{u}, \cdot)$ is not scaling invariant.

Proposition 3.8 (cosh-type gradient structure for CME) *If the RRE* (2.2) *satisfies the detailed-balance condition* (2.6) *for a positive steady state* $\boldsymbol{c}_* = (c_i^*)_{i=1,\dots,I}$, *then the associated CME has the gradient structure* $(\mathscr{P}(\mathcal{N}), \mathcal{E}_V, \boldsymbol{\Psi}_{\mathrm{cosh},V}^*)$ *with* $\mathcal{E}_V$ *from* (3.7) *and*

$$\boldsymbol{\Psi}_{\mathrm{cosh},V}^*(\boldsymbol{u}, \boldsymbol{\mu}) := \frac{1}{V} \sum_{r=1}^{R} \sum_{\boldsymbol{n} \in \mathcal{N}} \widehat{\nu}_V^{\boldsymbol{n},r} \Big(\frac{u_{\boldsymbol{n}+\boldsymbol{\alpha}^r}}{w_{\boldsymbol{n}+\boldsymbol{\alpha}^r}^V} \frac{u_{\boldsymbol{n}+\boldsymbol{\beta}^r}}{w_{\boldsymbol{n}+\boldsymbol{\beta}^r}^V}\Big)^{1/2} \mathsf{C}^*\Big(V(\mu_{\boldsymbol{n}+\boldsymbol{\beta}^r} - \mu_{\boldsymbol{n}+\boldsymbol{\alpha}^r})\Big). \tag{3.8}$$

Proof. The desired formula $\sum_{r=1}^{R} \overline{\mathcal{B}}_V^r \boldsymbol{u} = \mathrm{D}_\mu \Psi_{\mathrm{cosh},V}^* \big(\boldsymbol{u}, -\mathrm{D}\mathcal{E}_V(\boldsymbol{u}) \big)$ follows easily by recalling $\overline{\mathcal{B}}_V^r$ from (3.4) and by using $\sqrt{ab}\,(\mathrm{C}^*)'\big(\log a - \log b\big) = a - b$ and $\mathrm{D}\mathcal{E}_V(\boldsymbol{u}) = \frac{1}{V}\big(\log(u_{\boldsymbol{n}}/w_{\boldsymbol{n}}^V)\big)_{\boldsymbol{n}\in\mathcal{N}}$. $\blacksquare$

4 Liouville and Fokker–Planck equations

For general evolutionary equations one can define a measure-valued flow in the phase space that is given by transporting the measures according to the semiflow of the original equation. The evolution equation describing this measure-valued flow is the Liouville equation. For our RRE $\dot{\boldsymbol{c}} = -\boldsymbol{R}(\boldsymbol{c})$ in $\boldsymbol{C} := [0,\infty[^I$ we assume that we have a global semiflow $\boldsymbol{c}(t) = \Phi_t(\boldsymbol{c}(0))$ and consider probability measures $\varrho(t,\cdot) \in \mathscr{P}(\boldsymbol{C})$ that are obtained by transporting ϱ_0 with Φ_t, namely

$$\varrho(t,\cdot) = \Phi_t^{\#}\varrho_0, \quad \text{i.e., } \forall\,\psi \in \mathrm{C}_b(\boldsymbol{C}) : \int_{\boldsymbol{C}} \psi(\boldsymbol{c})\varrho(t,\mathrm{d}\boldsymbol{c}) = \int_{\boldsymbol{C}} \psi(\Phi_t(\boldsymbol{c}))\varrho_0(\mathrm{d}\boldsymbol{c}).$$

In particular, if $\varrho_0 = \sum_{k=1}^m a_k \delta_{\boldsymbol{c}_0^k}$, then $\varrho(t,\cdot) = \sum_{k=1}^m a_k \delta_{\Phi_t(\boldsymbol{c}_0^k)}(\cdot)$.

It is now easy to see that $t \mapsto \varrho_t \in \mathscr{P}(\boldsymbol{C})$ satisfies the Liouville equation

$$\partial_t \varrho(t,\boldsymbol{c}) = \mathrm{div}\big(\varrho(t,\boldsymbol{c})\boldsymbol{R}(\boldsymbol{c})\big), \tag{4.1}$$

in the sense of distributions. We will regard (4.1) as an evolution equation in the space $\mathscr{P}(\boldsymbol{C})$. We will not always notationally distinguish between an absolutely continuous probability measure and its density, but if we want to distinguish them we will write $\varrho(\mathrm{d}\boldsymbol{c}) = \rho(\boldsymbol{c})\,\mathrm{d}\boldsymbol{c}$ with $\rho \in \mathrm{L}^1(\boldsymbol{C})$.

The goal of this section is to give a rigorous connection between the CME for $V \to \infty$ and the Liouville equation in terms of the associated gradient structures.

4.1 The Liouville equation as a gradient system

We show that the gradient structure $\dot{\boldsymbol{c}} = -\boldsymbol{R}(\boldsymbol{c}) = -\mathbb{K}(\boldsymbol{c})\mathrm{D}E(\boldsymbol{c})$ for the RRE, which was discussed in Section 2.3, induces a natural gradient structure for the Liouville equation. Consider the "Otto-Wasserstein-type" Onsager operator $\boldsymbol{K}(\varrho)$ that acts on functions $\xi : \boldsymbol{C} \to \mathbb{R}$ via

$$\boldsymbol{K}(\varrho)\xi = -\,\mathrm{div}\big(\varrho\,\mathbb{K}\,\nabla\xi\big),$$

where div and ∇ are taken with respect to $\boldsymbol{c} \in \mathbb{R}^I$. We also consider the affine potential energy functional $\boldsymbol{E} : \mathscr{P}(\boldsymbol{C}) \to [0,+\infty]$ defined by

$$\boldsymbol{E}(\varrho) = \int_{\boldsymbol{C}} E(\boldsymbol{c})\,\mathrm{d}\varrho(\boldsymbol{c}). \tag{4.2}$$

In the next result we identify the formal gradient structure for the Liouville equation.

Proposition 4.1 (Gradient structure for the Liouville equation) *If the RRE* (2.2) *satisfies the detailed-balance condition* (2.6) *for a positive steady state* $\boldsymbol{c}_* = (c_i^*)_{i=1,\ldots,I}$, *then the associated Liouville equation has the gradient structure* $(\mathscr{P}(\boldsymbol{C}), \boldsymbol{E}, \boldsymbol{K})$, *namely*

$$\dot{\varrho} = -\boldsymbol{K}(\varrho)\mathrm{D}\boldsymbol{E}(\varrho) = \mathrm{div}\big(\varrho\mathbb{K}\nabla E\big) = \mathrm{div}(\varrho\boldsymbol{R}). \tag{4.3}$$

Proof. Let $\varrho \in \mathscr{P}(\boldsymbol{C})$ and let $\sigma \in \mathscr{M}(\boldsymbol{C})$ be a signed measure of finite total variation such that $\sigma(\boldsymbol{C}) = 0$ and $\varrho + h\sigma \in \mathscr{P}(\boldsymbol{C})$ for $|h|$ sufficiently small. Then we have

$$\frac{\boldsymbol{E}(\varrho + h\sigma) - \boldsymbol{E}(\varrho)}{h} = \int_{\boldsymbol{C}} E(\boldsymbol{c}) \, \mathrm{d}\sigma(\boldsymbol{c}),$$

hence $\mathrm{D}\boldsymbol{E}(\varrho) = E$ for all ϱ. Therefore, $-\boldsymbol{K}(\varrho)\mathrm{D}\boldsymbol{E}(\varrho) = \mathrm{div}\left(\varrho\,\mathbb{K}\,\nabla E\right) = \mathrm{div}(\varrho\boldsymbol{R})$. The gradient flow equation $\dot{\varrho} = -\boldsymbol{K}(\varrho)\mathrm{D}\boldsymbol{E}(\varrho)$ is thus given by the Liouville equation (4.1). ∎

4.2 Passing to the limit from CME to Liouville

In this section we shall demonstrate that the gradient flow structure for the CME converges in a suitable sense to the gradient structure for the Liouville equation if $V \to \infty$.

More precisely, we will show that after a suitable V-dependent embedding of $\mathscr{P}(\mathcal{N})$ into $\mathscr{P}(\boldsymbol{C})$ the proper scalings of the functionals $\mathcal{E}_V$ and $\Psi_V^* : (\boldsymbol{u}, \boldsymbol{\mu}) \mapsto \frac{1}{2}\boldsymbol{\mu} \cdot \mathcal{K}_V(\boldsymbol{u})\boldsymbol{\mu}$ converge in the sense of Γ-convergence to the corresponding structures for the Liouville equation given by the gradient system $(\mathscr{P}(\boldsymbol{C}), \boldsymbol{E}, \boldsymbol{K})$, see Section 4.3 to 4.5. Following the approach in [SaS04, Ser11, Mie16], and in particular [LM*17], we are then able to establish the convergence for $V \to \infty$ of solutions $\boldsymbol{u}^V : [0, \infty[\to \mathscr{P}(\mathcal{N})$ of the CME $\dot{\boldsymbol{u}}^V = -\mathcal{K}_V(\boldsymbol{u}^V)\mathrm{D}\mathcal{E}_V(\boldsymbol{u}^V)$ to the solution $\varrho : [0, \infty[\to \mathscr{P}(\boldsymbol{C})$ of the Liouville equation $\dot{\varrho} = -\boldsymbol{K}(\varrho)\mathrm{D}\boldsymbol{E}(\varrho)$, thereby recovering Kurtz' result (1.3), see Section 4.6.

The main tool for proving this evolutionary Γ-convergence for gradient systems is the so-called *energy–dissipation principle*, cf. [Mie16, Sec. 3.3], which states that $\boldsymbol{u}^V$ solves the CME if and only if for all $T > 0$ the following energy-dissipation estimate holds:

$$\mathcal{E}_V(\boldsymbol{u}^V(T)) + \int_0^T \left(\Psi_V(\boldsymbol{u}^V, \dot{\boldsymbol{u}}^V) + \Psi_V^*\left(\boldsymbol{u}^V, -\mathrm{D}\mathcal{E}_V(\boldsymbol{u}^V)\right)\right) \mathrm{d}t \leq \mathcal{E}_V(\boldsymbol{u}^V(0)), \qquad (4.4)$$

where we use the quadratic dissipation potential Ψ_V and its Legendre dual Ψ_V^* defined via $\Psi_V^*(\boldsymbol{u}, \boldsymbol{\mu}) := \frac{1}{2}\langle\boldsymbol{\mu}, \mathcal{K}_V(\boldsymbol{u})\boldsymbol{\mu}\rangle$ with $\mathcal{K}_V$ from (3.7), namely

$$\Psi_V^*(\boldsymbol{u}, \boldsymbol{\mu}) = \frac{V}{2} \sum_{\boldsymbol{n} \in \mathcal{N}} \sum_{r=1}^R \widehat{\nu}_V^{\boldsymbol{n},r} \Lambda\left(\frac{u_{\boldsymbol{n}+\boldsymbol{\alpha}^r}}{w_{\boldsymbol{n}+\boldsymbol{\alpha}^r}^V}, \frac{u_{\boldsymbol{n}+\boldsymbol{\beta}^r}}{w_{\boldsymbol{n}+\boldsymbol{\beta}^r}^V}\right)\left(\mu_{\boldsymbol{n}+\boldsymbol{\alpha}^r} - \mu_{\boldsymbol{n}+\boldsymbol{\beta}^r}\right)^2 \qquad (4.5a)$$

$$= \frac{V}{2} \sum_{\boldsymbol{n} \in \mathcal{N}} \sum_{r=1}^R \Lambda\left(k_{\mathrm{fw}}^r \, \mathbb{B}_V^{\boldsymbol{\alpha}^r}(\boldsymbol{n})\, u_{\boldsymbol{n}+\boldsymbol{\alpha}^r}, k_{\mathrm{bw}}^r \, \mathbb{B}_V^{\boldsymbol{\beta}^r}(\boldsymbol{n})\, u_{\boldsymbol{n}+\boldsymbol{\beta}^r}\right)\left(\mu_{\boldsymbol{n}+\boldsymbol{\alpha}^r} - \mu_{\boldsymbol{n}+\boldsymbol{\beta}^r}\right)^2, \qquad (4.5b)$$

where the second form uses Theorem 3.1 and is especially useful to perform the limit $V \to \infty$, see the proof of Proposition 4.6.

We refer to [DaS14, Mie16] for this equivalence and general methods for proving such results on evolutionary Γ-convergence. In [DiL15] a similar approach was used to establish the convergence of CTMC to a Fokker–Planck equation. However, there the convergence of a parabolic equation is established, where upper and lower bounds of the density can be used. Here, the importance is that our limit measures $\varrho(t)$ may not have densities; indeed, because we want to recover the Kurtz result (1.3) we are interested in the "deterministic case" $\varrho(t) = \delta_{\boldsymbol{c}(t)}$. So our analysis has to be more careful in dealing with general limit measures. For this, we use the dualization approach introduced in [LM*17] where $t \mapsto \Psi_V(\boldsymbol{u}^V, \dot{\boldsymbol{u}}^V)$ is estimated from below by $\langle\dot{\boldsymbol{u}}^V, \boldsymbol{\mu}^V\rangle - \Psi_V^*(\boldsymbol{u}^V, \boldsymbol{\mu}^V)$ for suitably chosen recovery functions $t \mapsto \boldsymbol{\mu}^V(t)$.

In order to compare probability measures on different spaces $\mathcal{N}$ and $\boldsymbol{C}$, we consider a suitable embedding $\iota_V : \mathscr{P}(\mathcal{N}) \to \mathscr{P}(\boldsymbol{C})$. Here $\iota_V(\boldsymbol{u})$ is simply obtained by assigning the mass of $\boldsymbol{u}$ at $\boldsymbol{n} \in \mathcal{N}$ uniformly to the cube

$$A_{\boldsymbol{n}}^V := \left[\tfrac{n_1}{V}, \tfrac{n_1+1}{V} \right[\times \cdots \times \left[\tfrac{n_I}{V}, \tfrac{n_I+1}{V} \right[\subseteq \boldsymbol{C}.$$

More explicitly, $\iota_V(\boldsymbol{u})$ is given by

$$\iota_V : \mathscr{P}(\mathcal{N}) \to \mathscr{P}(\boldsymbol{C}); \quad \boldsymbol{u} \mapsto \iota_V(\boldsymbol{u}) = \varrho = \rho \,\mathrm{d}\boldsymbol{c} \text{ with } \rho(\boldsymbol{c}) := V^I \sum_{\boldsymbol{n} \in \mathcal{N}} u_{\boldsymbol{n}} \mathbb{1}_{A_{\boldsymbol{n}}^V}(\boldsymbol{c}), \quad (4.6)$$

where $\mathbb{1}_A$ denotes the indicator function with $\mathbb{1}_A(b) = 1$ for $b \in A$ and 0 otherwise. The corresponding dual operation acting on functions $\xi \in \mathrm{C_b}(\boldsymbol{C})$ is given by

$$\iota_V^* : \mathrm{C_b}(\boldsymbol{C}) \to \ell^\infty(\mathcal{N}); \quad (\iota_V^* \xi)(\boldsymbol{n}) = V^I \int_{\boldsymbol{c} \in A_{\boldsymbol{n}}^V} \xi(\boldsymbol{c}) \,\mathrm{d}\boldsymbol{c}. \quad (4.7)$$

The final convergence result will be formulated in Theorem 4.7, which will be a direct consequence of the following three estimates

Section 4.3 $\qquad \iota_V(\boldsymbol{u}^V) \overset{*}{\rightharpoonup} \varrho \qquad\qquad \Rightarrow \qquad\qquad \boldsymbol{E}(\varrho) \le \liminf\limits_{V \to \infty} \mathcal{E}_V(\boldsymbol{u}^V);$

Section 4.4 $\qquad \iota_V(\boldsymbol{u}^V) \overset{*}{\rightharpoonup} \varrho \qquad \Rightarrow \Psi_{\mathrm{Lio}}^*(\varrho, \mathrm{D}\boldsymbol{E}(\varrho)) \le \liminf\limits_{V \to \infty} \Psi_V^*(\boldsymbol{u}^V, \mathrm{D}\mathcal{E}_V(\boldsymbol{u}^V));$

Section 4.5 $\quad \iota_V(\boldsymbol{u}^V) \overset{*}{\rightharpoonup} \varrho, \; \xi \in \mathrm{C_c^1}(\boldsymbol{C}) \Rightarrow \qquad \Psi_{\mathrm{Lio}}^*(\varrho, \xi) \ge \limsup\limits_{V \to \infty} \Psi_V^*(\boldsymbol{u}^V, \iota_V^* \xi);$

where the dual dissipation potential Ψ_{Lio}^* is defined via

$$\Psi_{\mathrm{Lio}}^*(\varrho, \xi) = \frac{1}{2} \int_{\boldsymbol{C}} \nabla \xi(\boldsymbol{c}) \cdot \mathbb{K}(\boldsymbol{c}) \nabla \xi(\boldsymbol{c}) \,\mathrm{d}\varrho(\boldsymbol{c}).$$

We will see in Section 4.6 that the limsup estimate for the dual potential Ψ_V^* in Section 4.5 provides a weak form of a liminf estimate for the primal potential Ψ_V.

A fundamental fact of the chosen gradient structures of the underlying Markov processes is that all the three terms in the energy-dissipation principle define convex functionals, which is of considerable help in proving the desired liminf estimates. Note that the convergence $\iota_V(\boldsymbol{u}^V) \overset{*}{\rightharpoonup} \varrho$ is rather weak. However, we can use that the coefficients of the transition rates defining the CME are quite regular, so that the other parts in the integral converge in a much better sense. Moreover, the functionals $\varrho \mapsto \boldsymbol{E}(\varrho)$ and $\varrho \mapsto \Psi_{\mathrm{Lio}}^*(\varrho, \mathrm{D}\boldsymbol{E}(\varrho))$ are in fact linear in ϱ.

4.3 Γ-limit of the relative entropies

We also define $\boldsymbol{X}_V := \iota_V(\mathscr{P}(\mathcal{N})) \subset \mathscr{P}(\boldsymbol{C})$ and $W_V = \iota_V(w^V) \in \boldsymbol{X}_V$ and consider the functionals

$$\widehat{\mathcal{E}}_V : \mathscr{P}(\boldsymbol{C}) \to [0, \infty], \qquad \widehat{\mathcal{E}}_V(\varrho) = \begin{cases} \widetilde{\mathcal{E}}_V(\varrho) & \text{if } \varrho \in \boldsymbol{X}_V, \\ \infty & \text{otherwise,} \end{cases}$$

where $\widetilde{\mathcal{E}}_V : \mathscr{P}(\boldsymbol{C}) \to [0, \infty]$ is defined via

$$\widetilde{\mathcal{E}}_V(\varrho) = \frac{1}{V} \mathrm{Ent}(\varrho | W_V \mathrm{d}\boldsymbol{c}) = \begin{cases} \frac{1}{V} \int_{\boldsymbol{C}} \lambda_{\mathrm{B}}(\rho/W_V) W_V \,\mathrm{d}\boldsymbol{c} & \text{for } \varrho = \rho \,\mathrm{d}c, \\ \infty & \text{otherwise.} \end{cases}$$

These definitions are chosen such that $\mathcal{E}_V(\boldsymbol{u}) = \widetilde{\mathcal{E}}_V(\iota_V(\boldsymbol{u})) = \widehat{\mathcal{E}}_V(\iota_V(\boldsymbol{u}))$ for all $\boldsymbol{u} \in \mathscr{P}(\mathcal{N})$.

Finally we define a natural inverse of ι_V, namely

$$\varkappa_V : \ \mathscr{P}(\boldsymbol{C}) \to \mathscr{P}(\mathcal{N}); \ \varrho \mapsto \Big(\varrho\big(A_{\boldsymbol{n}}^V\big)\Big)_{\boldsymbol{n}\in\mathcal{N}}, \tag{4.8}$$

such that $P_V := \iota_V \circ \varkappa_V$ is a projection from $\mathscr{P}(\boldsymbol{C})$ onto $\boldsymbol{X}_V \subset \mathscr{P}(\boldsymbol{C})$.

To understand the limit of $\widehat{\mathcal{E}}_V$ for $V \to \infty$ we will use the representation

$$\widetilde{\mathcal{E}}_V(\rho\mathrm{d}\boldsymbol{c}) = \frac{1}{V}\int_{\boldsymbol{C}} \lambda_{\mathrm{B}}(\rho/W_V)W_V\,\mathrm{d}\boldsymbol{c} = \int_{\boldsymbol{C}}\Big(\frac{1}{V}\rho\log\rho + \rho E_V(\boldsymbol{c})\Big)\mathrm{d}\boldsymbol{c}$$

$$\text{with} \ \ E_V(\boldsymbol{c}) = \frac{1}{V}\log\Big(\frac{1}{W_V(\boldsymbol{c})}\Big) = -I\frac{\log V}{V} - \frac{1}{V}\log \boldsymbol{w}_{\boldsymbol{n}}^V \ \ \text{for } \boldsymbol{c} \in A_{\boldsymbol{n}}^V. \tag{4.9}$$

In Lemma 4.2 below we will show that E_V converges pointwise to E as defined in (2.9). To quantify the latter convergence, we use the classical lower and upper bounds of [Nan59] for Stirling's formula:

$$\forall\, n \in \mathbb{N}_0: \ \ n! = \sqrt{2\pi k_n}\Big(\frac{n}{\mathrm{e}}\Big)^n \ \ \text{with } k_0 = \frac{1}{2\pi}$$

$$\text{and } k_n = n + \frac{1}{6} + \frac{\gamma_n}{124/5 + 72n} \ \text{with } \gamma_n \in [0.9, 1] \text{ for } n \geq 1. \tag{4.10}$$

Using this estimate and recalling E from (2.9) we obtain the following estimate.

Lemma 4.2 (Pointwise bound for E_V) *For all $\boldsymbol{c}^* > 0$ there exist $K_* > 0$ and $V_* > 0$ such that for all $V \geq V_*$ the following bounds hold:*

$$|E_V(\boldsymbol{c}) - E(\boldsymbol{c})| \leq \frac{K_*}{V}\big(\log V + E(\boldsymbol{c})\big) \ \ \text{for all } \boldsymbol{c} \in \boldsymbol{C}. \tag{4.11}$$

Proof. We decompose the error via

$$E_V(\boldsymbol{c}) - E(\boldsymbol{c}) = \Big(E_V(\boldsymbol{c}) - E\big(\tfrac{1}{V}\boldsymbol{n}\big)\Big) + \Big(E\big(\tfrac{1}{V}\boldsymbol{n}\big) - E(\boldsymbol{c})\Big) \tag{4.12}$$

with $\boldsymbol{n}$ defined by $\boldsymbol{c} \in A_{\boldsymbol{n}}^V$. For the second term we use the convexity of λ_{B} and the estimate $\log z \leq 1 + \lambda_{\mathrm{B}}(z)$. Hence, we have

$$c_i^*\Big[\lambda_{\mathrm{B}}\big(\tfrac{c_i}{c_i^*}\big) - \lambda_{\mathrm{B}}\big(\tfrac{n_i}{Vc_i^*}\big)\Big] \leq \big(c_i - \tfrac{n_i}{V}\big)\log\big(\tfrac{c_i}{c_i^*}\big) \leq \frac{1}{V}\Big(1 + \lambda_{\mathrm{B}}\big(\tfrac{c_i}{c_i^*}\big)\Big)$$

$$\leq \frac{\max\{I, 1/c_i^*\}}{V}\Big(\frac{1}{I} + c_i^*\lambda_{\mathrm{B}}\big(\tfrac{c_i}{c_i^*}\big)\Big).$$

Summing this inequality over $i = 1, \dots, I$ we obtain the upper bound

$$E(\boldsymbol{c}) - E\big(\tfrac{1}{V}\boldsymbol{n}\big) \leq \frac{K_1}{V}\big(1 + E(\boldsymbol{c})\big) \ \ \text{with } K_1 = \max\{I, 1/c_1^*, \dots, 1/c_I^*\}. \tag{4.13}$$

For the opposite direction we use (a) that λ_{B} decreases on $[0, 1]$ and the convexity of λ_{B} which implies (b) $\lambda_{\mathrm{B}}(z_1) - \lambda_{\mathrm{B}}(z_2) \leq \lambda_{\mathrm{B}}(0) - \lambda_{\mathrm{B}}(z_2 - z_1)$ for $0 \leq z_1 \leq z_2$. This yields

$$E\big(\tfrac{1}{V}\boldsymbol{n}\big) - E(\boldsymbol{c}) = \sum_{i=1}^{I} c_i^*\Big[\lambda_{\mathrm{B}}\big(\tfrac{n_i}{Vc_i^*}\big) - \lambda_{\mathrm{B}}\big(\tfrac{c_i}{c_i^*}\big)\Big] \overset{\text{(b)}}{\leq} \sum_{i=1}^{I} c_i^*\Big[\lambda_{\mathrm{B}}(0) - \lambda_{\mathrm{B}}\big(\tfrac{c_i}{c_i^*} - \tfrac{n_i}{Vc_i^*}\big)\Big]$$

$$\overset{\text{(a)}}{\leq} \sum_{i=1}^{I} c_i^*\Big[\lambda_{\mathrm{B}}(0) - \lambda_{\mathrm{B}}\big(\tfrac{1}{Vc_i^*}\big)\Big] = \frac{1}{V}\sum_{i=1}^{I}\big(1 + \log(Vc_i^*)\big) \overset{\text{(c)}}{\leq} 2I\frac{\log V}{V}, \tag{4.14}$$

if $V \geq V_1^* := \max\left\{ \max\{1/c_i^*, \mathrm{e}c_i^*\} \mid i = 1, ..., I \right\}$, where $Vc_i^* \geq 1$ and $V \geq \mathrm{e}c_i^*$ are needed in (a) and (c), respectively. Together with (4.13) this controls the second error term in (4.12), viz.

$$\left| E(\boldsymbol{c}) - E\left(\tfrac{1}{V}\boldsymbol{n}\right) \right| \leq \frac{K_2}{V}\left(\log V + E(\boldsymbol{c}) \right) \text{ for } V \geq V_2^* = \max\{\mathrm{e}, V_1^*\}, \tag{4.15}$$

where $K_2 = \max\{2I, K_1\}$.

For controlling the first error term in (4.12) we use (4.9) and obtain the identity

$$E_V(\boldsymbol{c}) - E\left(\tfrac{1}{V}\boldsymbol{n}\right) = -I\frac{\log V}{V} + \frac{1}{2V}\sum_{i=1}^{I} \log(2\pi k_{n_i}) \quad \text{for all } \boldsymbol{c} \in A_{\boldsymbol{n}}^V, \tag{4.16}$$

with k_n from (4.10). Because of $2\pi k_n \geq 1$ we obtain, for all $V \geq 1$, the lower bound

$$E_V(\boldsymbol{c}) - E\left(\tfrac{1}{V}\boldsymbol{n}\right) \geq -I\frac{\log V}{V} \geq -\frac{I}{V}\left(\log V + E(\boldsymbol{c}) \right).$$

For the upper bound we use $2\pi k_0 = 1$ and $2\pi k_n \leq 8n$ for $n \geq 1$. Hence for $n_i \geq 1$ we obtain, using again the estimate $\log z \leq 1 + \lambda_{\mathrm{B}}(z)$,

$$\log(2\pi k_{n_i}) \leq \log(8n_i) \leq \log(8c_i^*V) + \log\left(\frac{c_i}{c_i^*}\right) \leq \log V + \log(8\mathrm{e}c_i^*) + \frac{1}{c_i^*}c_i^*\lambda_{\mathrm{B}}\left(\frac{c_i}{c_i^*}\right).$$

Summation over $i = 1, \ldots, I$ yields, for all $\boldsymbol{c} \in A_{\boldsymbol{n}}^V$ and $V \geq V_3^* := 8\mathrm{e}\max\{c_1^*, \ldots, c_I^*\}$, the upper bound

$$E_V(\boldsymbol{c}) - E\left(\frac{\boldsymbol{n}}{V}\right) \leq \frac{K_3}{V}E(\boldsymbol{c}) \text{ with } K_3 = \max\left\{ \frac{1}{2c_1^*}, \ldots, \frac{1}{2c_I^*} \right\}.$$

Together with the lower estimate we control the first error term in (4.12) via

$$\left| E_V(\boldsymbol{c}) - E\left(\tfrac{1}{V}\boldsymbol{n}\right) \right| \leq \frac{K_4}{V}\left(\log V + E(\boldsymbol{c}) \right) \text{ for } V \geq V_4^* = \max\{1, V_3^*\},$$

where $K_4 = \max\{I, K_3\}$.

Adding the estimates for first and the second error term (4.12) we obtain the desired estimate (4.11) with the choices $K_* = K_2 + K_4$ and $V_* = \max\{V_2^*, V_4^*\}$. ∎

For consistency of notation we remark that $\widetilde{\mathcal{E}}_V(\varrho)$ can be rewritten as

$$\widetilde{\mathcal{E}}_V(\rho\mathrm{d}\boldsymbol{c}) = \int_{\boldsymbol{C}} \left(\frac{1}{V}\log\rho(\boldsymbol{c}) + E_V(\boldsymbol{c})\right)\rho(\boldsymbol{c})\,\mathrm{d}\boldsymbol{c}\,,$$

provided that this integral exists. The limit functional $\boldsymbol{E}$ is given by

$$\boldsymbol{E} : \mathscr{P}(\boldsymbol{C}) \to [0, \infty]; \quad \varrho \mapsto \int_{\boldsymbol{C}} E(\boldsymbol{c})\,\mathrm{d}\varrho(\boldsymbol{c}), \tag{4.17}$$

where we use that E is a continuous and non-negative function, so that $\boldsymbol{E}$ can be defined everywhere but attains the value $+\infty$ if ϱ does not decay suitably at infinity. We will use the following semi-continuity result.

Lemma 4.3 (Lower semi-continuity of E) *For sequences $(\varrho_k)_k \subset \mathscr{P}(C)$ with $\varrho_k \overset{*}{\rightharpoonup} \varrho_\infty$, we have $E(\varrho_\infty) \leq \liminf_{k\to\infty} E(\varrho_k)$.*

Proof. For cut-off functions $\chi \in C_c(C)$ with $\chi(c) \in [0,1]$ we have $E(\chi\varrho_k) \to E(\chi\varrho_\infty)$ by weak* convergence and continuity of E. Using $\chi \leq 1$ yields $E(\chi\varrho_\infty) \leq \liminf_{k\to\infty} E(1\varrho_k)$. Choosing a non-decreasing sequence χ_n with $\chi_n(c) \to 1$ for all $c \in C$ we have $E(\chi_n\varrho_\infty) \to E(1\varrho_\infty)$ by Beppo Levi's monotone convergence, and the assertion follows. ∎

The following result gives the Γ-convergence of $\mathcal{E}_V$ to E with respect to the sequential weak* convergence as well as the equi-coercivity.

Theorem 4.4 (Γ-convergence of $\mathcal{E}_V$ to E) *Let $\widehat{\mathcal{E}}_V$ and E be defined on $\mathscr{P}(C)$ as above. Then we have the following properties:*

(a) Compactness / equi-coercivity:

$$\exists\, V_*, C, c > 0 \,\forall\, V \geq V_* \,\forall\, \varrho \in \mathscr{P}(C) : \quad \widehat{\mathcal{E}}_V(\varrho) \geq -C + cE(\varrho). \tag{4.18}$$

(b) Weak liminf estimate:*

$$\varrho_V \overset{*}{\rightharpoonup} \varrho \text{ in } \mathscr{P}(C) \quad \Longrightarrow \quad \liminf_{V\to\infty} \widehat{\mathcal{E}}_V(\varrho_V) \geq E(\varrho). \tag{4.19}$$

(c) Limsup estimate / recovery sequence:

$$\forall\, \widehat{\varrho} \in \mathscr{P}(C) \,\exists\, (\widehat{\varrho}_V)_{V\geq 1} : \quad \widehat{\mathcal{E}}_V(\widehat{\varrho}_V) \to E(\widehat{\varrho}) \text{ and } \widehat{\varrho}_V \overset{*}{\rightharpoonup} \widehat{\varrho}, \tag{4.20}$$

where we may take $\widehat{\varrho}_V = P_V\widehat{\varrho} = \iota_V\big(\varkappa_V(\widehat{\varrho})\big)$.

Proof. Obviously it is sufficient to show the lower bound (a) and the liminf estimate (b) for the smaller functional $\widetilde{\mathcal{E}}_V$, and for $\varrho = \rho\,\mathrm{d}c$ with $\rho \in \mathrm{L}^1(C)$ (resp. $\varrho_V = \rho_V\,\mathrm{d}c$ with $\rho_V \in \mathrm{L}^1(C)$). We use the elementary convexity estimate

$$\forall\, r \geq 0,\ a, w > 0 : \quad w\lambda_{\mathrm{B}}(r/w) = r\log(r/w) - r + w \geq r\log(a/w) - a + w.$$

We choose $r(c) = \varrho(c)$, $w(c) = W_V(c)$, and $a(c) = \mathrm{e}^{-|c|_1} = \Pi_{i=1}^I \mathrm{e}^{-c_i} > 0$. Note that $a \in \mathrm{L}^\infty(C) \cap \mathscr{P}(C)$ and W_V/a is bounded from above, for any fixed V. Hence, $c \mapsto \log(a(c)/W_V(c)) = -|c|_1 + VE_V(c)$ is bounded from below, and we can integrate the above estimate to obtain the lower bound

$$\widetilde{\mathcal{E}}_V(\varrho) \geq \frac{1}{V}\int_C \log\big(a(c)/W_V(c)\big)\,\mathrm{d}\varrho(c) = \int_C \Big(E_V(c) - \frac{|c|_1}{V}\Big)\,\mathrm{d}\varrho(c).$$

Since there exists a constant $K_1 > 0$ such that $|c|_1 \leq K_1\big(1 + E(c)\big)$ and since E_V satisfies the lower bound in (4.11), we obtain the lower bound

$$\widetilde{\mathcal{E}}_V(\varrho) \geq \int_C E(c)\,\mathrm{d}\varrho(c) - \frac{K_*+K_1}{V}\int_C \big(\log V + E(c)\big)\,\mathrm{d}\varrho(c)$$

$$= E(\varrho) - \frac{K_*+K_1}{V}\big(\log V + E(\varrho)\big).$$

This immediately implies (4.18) in part (a) with $V_*/\log V_* = 2(K_*+K_1)$. Moreover, if $\varrho_V \overset{*}{\rightharpoonup} \varrho$ then we have the lower bound $\widehat{\mathcal{E}}_V(\varrho_V) \geq E(\varrho_V) - \frac{K_*+K_1}{V}\big(\log V + E(\varrho_V)\big)$ and the liminf estimate (4.19) follows from Lemma 4.3.

To show part (c) we use the indicated recovery sequence and the upper bounds for E_V from (4.11). For a given $\widehat{\varrho} \in \mathscr{P}(\boldsymbol{C})$ we define $\widehat{\varrho}_V = \iota_V(\varkappa_V(\widehat{\varrho}))$. For an arbitrary continuous and bounded test function ψ we define the piecewise constant approximation ψ_V via averaging over $A_{\boldsymbol{n}}^V$. We obtain

$$\int_{\boldsymbol{C}} \psi(\boldsymbol{c}) \, \mathrm{d}\widehat{\varrho}_V(\boldsymbol{c}) = \int_{\boldsymbol{C}} \psi_V(\boldsymbol{c}) \, \mathrm{d}\widehat{\varrho}_V(\boldsymbol{c}) = \int_{\boldsymbol{C}} \psi_V(\boldsymbol{c}) \, \mathrm{d}\widehat{\varrho}(\boldsymbol{c}) \to \int_{\boldsymbol{C}} \psi(\boldsymbol{c}) \, \mathrm{d}\widehat{\varrho}(\boldsymbol{c}),$$

where the convergence follows via Lebesgue's dominated convergence from the pointwise convergence $\psi_V \to \psi$ and the uniform boundedness of ψ_V. Thus, we conclude $\widehat{\varrho}_V \overset{*}{\rightharpoonup} \widehat{\varrho}$.

To show convergence of $\widehat{\mathcal{E}}_V(\widehat{\varrho}_V)$ it suffices to prove the upper bound $\limsup_{V \to \infty} \widehat{\mathcal{E}}_V(\widehat{\varrho}_V) \leq \boldsymbol{E}(\widehat{\varrho})$. For this we use the bound $\widehat{\rho}_V(\boldsymbol{c}) \leq V^I = 1/\mathrm{vol}(A_{\boldsymbol{n}}^V)$ and the fact that $\widehat{\rho}_V$ and E_V are constant on the same cubes to obtain

$$\widehat{\mathcal{E}}_V(\widehat{\varrho}_V) = \int_{\boldsymbol{C}} \big(\frac{\log \widehat{\rho}_V(\boldsymbol{c})}{V} + E_V(\boldsymbol{c}) \big) \, \mathrm{d}\widehat{\rho}_V(\boldsymbol{c}) \leq \frac{I \log V}{V} + \int_{\boldsymbol{C}} E_V(\boldsymbol{c}) \, \mathrm{d}\widehat{\varrho}(\boldsymbol{c}),$$

where now only the measure $\widehat{\varrho}$ is left. The first term tends to 0 for $V \to \infty$, and the second can be estimated from above using the upper estimate in (4.11), which yields

$$\widehat{\mathcal{E}}_V(\widehat{\varrho}_V) \leq \frac{I \log V}{V} + \int_{\boldsymbol{C}} E(\boldsymbol{c}) + \frac{K_*}{V} \big(\log V + E(\boldsymbol{c}) \big) \, \mathrm{d}\widehat{\varrho}(\boldsymbol{c}) = \big(1 + \frac{K_*}{V} \big) \boldsymbol{E}(\widehat{\varrho}) + \frac{I + K_*}{V} \log V.$$

This implies the desired upper bound for $V \to \infty$, and the proof is complete. ∎

4.4 A liminf estimate for the dual dissipation functional

Here we provide the liminf estimate for the dual dissipation potential $\Psi_V^*(\boldsymbol{u}^V, \mathrm{D}\mathcal{E}_V(\boldsymbol{u}^V))$ based on the lower bound

$$\Psi_{\mathrm{Lio}}^*(\varrho, \mathrm{D}\boldsymbol{E}(\varrho)) = \frac{1}{2} \int_{\boldsymbol{C}} \nabla E(\boldsymbol{c}) \cdot \mathbb{K}(\boldsymbol{c}) \nabla E(\boldsymbol{c}) \, \mathrm{d}\varrho(\boldsymbol{c}). \tag{4.21}$$

We observe that the latter term is linear in ϱ while the former term is convex in $\boldsymbol{u}^V$. Indeed, introducing the convex function $G(a, b) = (a - b)(\log a - \log b)$ for $a, b > 0$ and noting the relation $\Lambda(a, b)(\log a - \log b)^2 = G(a, b)$ we have

$$\Psi_V^*(\boldsymbol{u}^V, \mathrm{D}\mathcal{E}_V(\boldsymbol{u}^V)) = \frac{1}{2V} \sum_{r=1}^{R} \sum_{\boldsymbol{n} \in \mathcal{N}} \widehat{\nu}_V^{\boldsymbol{n},r} G\Big(\frac{u_{\boldsymbol{n}+\boldsymbol{\alpha}^r}^V}{w_{\boldsymbol{n}+\boldsymbol{\alpha}^r}^V}, \frac{u_{\boldsymbol{n}+\boldsymbol{\beta}^r}^V}{w_{\boldsymbol{n}+\boldsymbol{\beta}^r}^V} \Big). \tag{4.22}$$

To establish the linear lower bound we use the elementary, affine lower bound

$$\forall\, a, b > 0, \omega \in \mathbb{R}: \quad G(a, b) \geq g(\omega)\, a + g(-\omega)\, b, \quad \text{where } g(\omega) := 1 - \mathrm{e}^{-\omega} + \omega. \tag{4.23}$$

This estimate follows easily by convexity, $G(a, b) \geq G(\mathrm{e}^\omega, 1) + \mathrm{D}G(\mathrm{e}^\omega, 1) \cdot (a - \mathrm{e}^\omega, b - 1)$, and 1-homogeneity giving $G(\mathrm{e}^\omega, 1) = \mathrm{D}G(\mathrm{e}^\omega, 1) \cdot (\mathrm{e}^\omega, 1)$. Note that equality holds in (4.23) if $\omega = \log(a/b)$. Moreover, we have $g(\omega) + g(-\omega) = 2 - \mathrm{e}^\omega - \mathrm{e}^{-\omega} \leq 0$, so a careful choice of ω depending on $\boldsymbol{n}$ will be necessary to obtain a good lower bound with a positive leading term.

Proposition 4.5 *We have the liminf estimate*

$$\iota_V(\boldsymbol{u}^V) \overset{*}{\rightharpoonup} \varrho \text{ in } \mathscr{P}(\boldsymbol{C}) \quad \Longrightarrow \quad \Psi_{\mathrm{Lio}}^*(\varrho, \mathrm{D}\boldsymbol{E}(\varrho)) \leq \liminf_{V \to \infty} \Psi_V^*(\boldsymbol{u}^V, \mathrm{D}\mathcal{E}_V(\boldsymbol{u}^V)).$$

Proof. The special forms of $\mathbb{K}(c)$, $E(\boldsymbol{c})$, and Ψ^*_{Lio} in (4.21) give the formula

$$\Psi^*_{\text{Lio}}(\varrho, \mathrm{D}\boldsymbol{E}(\varrho)) = \frac{1}{2} \int_C \sum_{r=1}^R \kappa^r_* \, G\Big(\frac{\boldsymbol{c}^{\boldsymbol{\alpha}^r}}{\boldsymbol{c}^{\boldsymbol{\alpha}^r}_*}, \frac{\boldsymbol{c}^{\boldsymbol{\beta}^r}}{\boldsymbol{c}^{\boldsymbol{\beta}^r}_*}\Big) \mathrm{d}\varrho(\boldsymbol{c}). \tag{4.24}$$

Since Ψ^*_V and Ψ^*_{Lio} are defined as sums over $r = 1, \dots, R$ of nonnegative terms, it suffices to show the result for each r separately, where we suppress the index r.

Inserting (4.23) into (4.22) yields, with $\omega_{\boldsymbol{n}} \in \mathbb{R}$ to be fixed afterwards,

$$\Psi^*_V(\boldsymbol{u}^V, \mathrm{D}\mathcal{E}_V(\boldsymbol{u}^V)) \geq \frac{1}{2V} \sum_{\boldsymbol{n} \in \mathcal{N}} \widehat{\nu}^{\boldsymbol{n}}_V \Big(g(\omega_{\boldsymbol{n}}) \frac{u^V_{\boldsymbol{n}+\boldsymbol{\alpha}}}{w^V_{\boldsymbol{n}+\boldsymbol{\alpha}}} + g(-\omega_{\boldsymbol{n}}) \frac{u^V_{\boldsymbol{n}+\boldsymbol{\beta}}}{w^V_{\boldsymbol{n}+\boldsymbol{\beta}}}\Big)$$

$$= \frac{\kappa_*}{2} \sum_{\boldsymbol{n} \in \mathcal{N}} \Big(g(\omega_{\boldsymbol{n}}) \mathbb{A}^{\boldsymbol{\alpha}}_V(\boldsymbol{n}) u^V_{\boldsymbol{n}+\boldsymbol{\alpha}} + g(-\omega_{\boldsymbol{n}}) \mathbb{A}^{\boldsymbol{\beta}}_V(\boldsymbol{n}) u^V_{\boldsymbol{n}+\boldsymbol{\beta}}\Big)$$

$$\text{with } \mathbb{A}^{\boldsymbol{\delta}}_V(\boldsymbol{n}) := \frac{w^V_{\boldsymbol{n}}}{w^V_{\boldsymbol{n}+\boldsymbol{\delta}}} = \frac{(\boldsymbol{n}+\boldsymbol{\delta})!}{(\boldsymbol{c}_* V)^{\boldsymbol{\delta}} \boldsymbol{n}!},$$

where we used the detailed-balance conditions from Theorem 3.1 for the last identity. Rearranging the sum and recalling that $\mathbb{A}^{\boldsymbol{\delta}}(\boldsymbol{n}) = 0$ for $\boldsymbol{n} \notin \mathcal{N}$ we find

$$\Psi^*_V(\boldsymbol{u}^V, \mathrm{D}\mathcal{E}_V(\boldsymbol{u}^V)) \geq \frac{\kappa_*}{2} \sum_{\boldsymbol{n} \in \mathcal{N}} h^V_{\boldsymbol{n}} u^V_{\boldsymbol{n}} \quad \text{with } h^V_{\boldsymbol{n}} := g(\omega_{\boldsymbol{n}-\boldsymbol{\alpha}}) \mathbb{A}^{\boldsymbol{\alpha}}_V(\boldsymbol{n}-\boldsymbol{\alpha}) + g(-\omega_{\boldsymbol{n}-\boldsymbol{\beta}}) \mathbb{A}^{\boldsymbol{\beta}}_V(\boldsymbol{n}-\boldsymbol{\beta}).$$

We now choose $\omega_{\boldsymbol{n}} = \log\big(\mathbb{A}^{\boldsymbol{\alpha}}_V(\boldsymbol{n})/\mathbb{A}^{\boldsymbol{\beta}}_V(\boldsymbol{n})\big)$ for $\boldsymbol{n} \in \mathcal{N}$ and $\omega_{\boldsymbol{n}} = 0$ otherwise and find, for all $\boldsymbol{n}$ with $\boldsymbol{n} \geq \boldsymbol{\alpha}$ or $\boldsymbol{n} \geq \boldsymbol{\beta}$, the relation

$$h^V_{\boldsymbol{n}} = G\big(\mathbb{A}^{\boldsymbol{\alpha}}_V(\boldsymbol{n}-\boldsymbol{\alpha}), \mathbb{A}^{\boldsymbol{\beta}}_V(\boldsymbol{n}-\boldsymbol{\beta})\big) + f^V_{\boldsymbol{n}} \quad \text{with}$$

$$f^V_{\boldsymbol{n}} := \mathbb{A}^{\boldsymbol{\alpha}}_V(\boldsymbol{n}-\boldsymbol{\alpha}) - \mathbb{A}^{\boldsymbol{\alpha}}_V(\boldsymbol{n}-\boldsymbol{\beta}) + \mathbb{A}^{\boldsymbol{\beta}}_V(\boldsymbol{n}-\boldsymbol{\beta}) - \mathbb{A}^{\boldsymbol{\beta}}_V(\boldsymbol{n}-\boldsymbol{\alpha})$$

$$+ \mathbb{A}^{\boldsymbol{\alpha}}_V(\boldsymbol{n}-\boldsymbol{\alpha}) \log\Big(\frac{\mathbb{A}^{\boldsymbol{\beta}}_V(\boldsymbol{n}-\boldsymbol{\beta})}{\mathbb{A}^{\boldsymbol{\beta}}_V(\boldsymbol{n}-\boldsymbol{\alpha})}\Big) + \mathbb{A}^{\boldsymbol{\beta}}_V(\boldsymbol{n}-\boldsymbol{\beta}) \log\Big(\frac{\mathbb{A}^{\boldsymbol{\alpha}}_V(\boldsymbol{n}-\boldsymbol{\alpha})}{\mathbb{A}^{\boldsymbol{\alpha}}_V(\boldsymbol{n}-\boldsymbol{\beta})}\Big)$$

The idea is now that as $\frac{1}{V}\boldsymbol{n} \to \boldsymbol{c} > \boldsymbol{0}$ we have the convergences

$$\mathbb{A}^{\boldsymbol{\delta}}_V(\boldsymbol{n}-\boldsymbol{\alpha}) \to \boldsymbol{c}^{\boldsymbol{\delta}}/\boldsymbol{c}^{\boldsymbol{\delta}}_* \quad \text{and} \quad \mathbb{A}^{\boldsymbol{\delta}}_V(\boldsymbol{n}-\boldsymbol{\beta}) \to \boldsymbol{c}^{\boldsymbol{\delta}}/\boldsymbol{c}^{\boldsymbol{\delta}}_*,$$

which yields $f^V_{\boldsymbol{n}} \to 0$ and $h^V_{\boldsymbol{n}} \to G(\boldsymbol{c}^{\boldsymbol{\alpha}}/\boldsymbol{c}^{\boldsymbol{\alpha}}_*, \boldsymbol{c}^{\boldsymbol{\beta}}/\boldsymbol{c}^{\boldsymbol{\beta}}_*)$ as desired. To be more precise we define, for all $\varepsilon \in \,]0, 1[$, the functions

$$G_\varepsilon(a, b) = -\varepsilon + \min\{(1-\varepsilon)G(a, b), 1/\varepsilon\},$$

which converge monotonely to $G(a, b)$ for $\varepsilon \searrow 0$. A lengthy calculation using the explicit structure of $\mathbb{A}^{\boldsymbol{\delta}}(\boldsymbol{n})$ shows that for all $\varepsilon > 0$ there exists $V_\varepsilon \gg 1$ such that $h^V_{\boldsymbol{n}} \geq G_\varepsilon(\mathbb{A}^{\boldsymbol{\alpha}}_V(\boldsymbol{n}), \mathbb{A}^{\boldsymbol{\beta}}_V(\boldsymbol{n}))$ for all $V \geq V_\varepsilon$ and all $\boldsymbol{n}$. Even more, if we define the functions $H_V : \boldsymbol{C} \to \mathbb{R}$; $\boldsymbol{c} \mapsto \sum_{\boldsymbol{n} \in \mathcal{N}} h^V_{\boldsymbol{n}} \mathbb{1}_{A^V_{\boldsymbol{n}}}(\boldsymbol{c})$, then, for all $\varepsilon > 0$ there exists $\widetilde{V}_\varepsilon \gg 1$ such that

$$\forall V \geq \widetilde{V}_\varepsilon \, \forall \boldsymbol{c} \in \boldsymbol{C} : \; H_V(\boldsymbol{c}) \geq \mathfrak{H}_\varepsilon(\boldsymbol{c}) := G_\varepsilon\Big(\frac{\boldsymbol{c}^{\boldsymbol{\alpha}}}{\boldsymbol{c}^{\boldsymbol{\alpha}}_*}, \frac{\boldsymbol{c}^{\boldsymbol{\beta}}}{\boldsymbol{c}^{\boldsymbol{\beta}}_*}\Big).$$

Hence, using the definition of ι_V we find the lower bound

$$\Psi_V^*(\boldsymbol{u}^V, \mathrm{D}\mathcal{E}_V(\boldsymbol{u}^V)) \geq \frac{\kappa_*}{2} \sum_{\boldsymbol{n}} h_{\boldsymbol{n}}^V u_{\boldsymbol{n}}^V = \frac{\kappa_*}{2} \int_C H_V(\boldsymbol{c}) \,\mathrm{d}\iota_V(\boldsymbol{u}^V)(\boldsymbol{c}) \geq \frac{\kappa_*}{2} \int_C \mathfrak{H}_\varepsilon(\boldsymbol{c}) \,\mathrm{d}\iota_V(\boldsymbol{u}^V)(\boldsymbol{c}).$$

Since $\mathfrak{H}_\varepsilon$ is lower semi-continuous and bounded, this implies the liminf estimate

$$\iota_V(\boldsymbol{u}^V) \overset{*}{\rightharpoonup} \varrho \quad \Longrightarrow \quad \liminf_{V \to \infty} \Psi_V^*(\boldsymbol{u}^V, \mathrm{D}\mathcal{E}_V(\boldsymbol{u}^V)) \geq \frac{\kappa_*}{2} \int_C \mathfrak{H}_\varepsilon(\boldsymbol{c}) \,\mathrm{d}\varrho(\boldsymbol{c}).$$

Because $\varepsilon > 0$ was arbitrary we can use the monotone convergence $\mathfrak{H}_\varepsilon(\boldsymbol{c}) \nearrow G\left(\frac{\boldsymbol{c}^{\boldsymbol{\alpha}}}{\boldsymbol{c}_*^{\boldsymbol{\alpha}}}, \frac{\boldsymbol{c}^{\boldsymbol{\beta}}}{\boldsymbol{c}_*^{\boldsymbol{\beta}}}\right)$ to conclude the desired result for each of the R reactions

$$\liminf_{V \to \infty} \Psi_V^{r,*}(\boldsymbol{u}^V, \mathrm{D}\mathcal{E}_V(\boldsymbol{u}^V)) \geq \frac{\kappa_*}{2} \int_C G\left(\frac{\boldsymbol{c}^{\boldsymbol{\alpha}^r}}{\boldsymbol{c}_*^{\boldsymbol{\alpha}^r}}, \frac{\boldsymbol{c}^{\boldsymbol{\beta}^r}}{\boldsymbol{c}_*^{\boldsymbol{\beta}^r}}\right) \mathrm{d}\varrho(\boldsymbol{c})$$

$$= \frac{\kappa_*}{2} \int_C \Lambda\left(\frac{\boldsymbol{c}^{\boldsymbol{\alpha}^r}}{\boldsymbol{c}_*^{\boldsymbol{\alpha}^r}}, \frac{\boldsymbol{c}^{\boldsymbol{\beta}^r}}{\boldsymbol{c}_*^{\boldsymbol{\beta}^r}}\right) \left(\nabla E(\boldsymbol{c}) \cdot (\boldsymbol{\alpha}^r - \boldsymbol{\beta}^r)\right)^2 \mathrm{d}\varrho(\boldsymbol{c}).$$

Summation over $r = 1, \ldots, R$ yields the full result for Ψ_V^*. ∎

4.5 A liminf estimate for the dissipation functional

In the evolutionary Γ-convergence method of [SaS04, Ser11, Mie16] it is standard to provide a liminf estimate for the *primal dissipation potential* Ψ_V which in our case is defined via the Legendre transform

$$\Psi_V(\boldsymbol{u}, \boldsymbol{v}) = \sup \left\{ \sum_{\boldsymbol{n} \in \mathcal{N}} u_{\boldsymbol{n}} \xi_{\boldsymbol{n}} - \Psi_V^*(\boldsymbol{u}, \boldsymbol{\xi}) \,\Big|\, \boldsymbol{\xi} = (\xi_{\boldsymbol{n}})_{\boldsymbol{n} \in \mathcal{N}} \right\}.$$

However, as our theory relies on the dualization $\Psi_V(\boldsymbol{u}, \boldsymbol{v}) \geq \sum_{\boldsymbol{n} \in \mathcal{N}} u_{\boldsymbol{n}} \xi_{\boldsymbol{n}} - \Psi_V^*(\boldsymbol{u}, \boldsymbol{\xi})$ it will be sufficient to have the following limsup estimate for Ψ_V^*, which crucially relies on the concavity of the map $(a, b) \mapsto \Lambda(a, b)$.

Proposition 4.6 *Consider any pair $(\varrho, \xi) \in \mathscr{P}(C) \times \mathrm{C}_c^1(C)$ and set $\boldsymbol{\xi}^V = \iota_V^* \xi : \mathcal{N} \to \mathbb{R}$ with ι_V^* defined in (4.7). Then, for every family $(\boldsymbol{u}^V)_{V>1}$ we have the limsup estimate*

$$\iota_V(\boldsymbol{u}^V) \overset{*}{\rightharpoonup} \varrho \quad \Longrightarrow \quad \limsup_{V \to \infty} \Psi_V^*(\boldsymbol{u}^V, \boldsymbol{\xi}^V) \leq \Psi_{\mathrm{Lio}}^*(\varrho, \xi) = \frac{1}{2} \int_C \nabla \xi \cdot \mathbb{K} \nabla \xi \,\mathrm{d}\varrho(\boldsymbol{c}). \quad (4.25)$$

Proof. As in the proof of Proposition 4.5 we can exploit that Ψ_V^* is a sum of non-negative terms over $r = 1, \ldots, R$. Hence, it is sufficient to show the desired limsup estimate for each reaction individually. For notational simplicity we drop the reaction index r.

Defining $\varrho^V = \rho^V \mathrm{d}\boldsymbol{c} = \iota_V(\boldsymbol{u}^V)$, relation (4.5b) leads us to the integral representation

$$\Psi_V^*(\boldsymbol{u}^V, \boldsymbol{\xi}^V) = \frac{\kappa_*}{2} \int_{\boldsymbol{c} \in C} \Lambda\left(\rho^{V,a}(\boldsymbol{c}), \rho^{V,b}(\boldsymbol{c})\right) M_V^\xi(\boldsymbol{c}) \,\mathrm{d}\boldsymbol{c},$$

where

$$\rho^{V,a}(\boldsymbol{c}) = a_V(\boldsymbol{c})\rho^V(\boldsymbol{c} + \tfrac{1}{V}\boldsymbol{\alpha}), \quad \rho^{V,b}(\boldsymbol{c}) = b_V(\boldsymbol{c})\rho^V(\boldsymbol{c} + \tfrac{1}{V}\boldsymbol{\beta}),$$

and the functions a_V, b_V, and M_V^ξ are given

$$a_V(\boldsymbol{c}) = \frac{\mathbb{B}_V^{\boldsymbol{\alpha}}(\boldsymbol{n})}{V \boldsymbol{c}_*^{\boldsymbol{\alpha}}}, \quad b_V(\boldsymbol{c}) = \frac{\mathbb{B}_V^{\boldsymbol{\beta}}(\boldsymbol{n})}{V \boldsymbol{c}_*^{\boldsymbol{\beta}}}, \quad M_V^\xi(\boldsymbol{c}) = V^2 \big(\boldsymbol{\xi}_{\boldsymbol{n}+\boldsymbol{\alpha}}^V - \boldsymbol{\xi}_{\boldsymbol{n}+\boldsymbol{\beta}}^V\big)^2 \quad \text{for } \boldsymbol{c} \in A_{\boldsymbol{n}}^V.$$

Using $\xi \in \mathrm{C}_{\mathrm{c}}^1(\boldsymbol{C})$ there exists $R > 0$ such that $\mathrm{sppt}\, M_V^\xi \subset \boldsymbol{C}_R := B_R(0) \cap \boldsymbol{C}$, and we have uniform convergence

$$\|a_V - a_\infty\|_{\mathrm{L}^\infty(\boldsymbol{C}_R)} + \|b_V - b_\infty\|_{\mathrm{L}^\infty(\boldsymbol{C}_R)} + \|M_V^\xi - (\nabla \xi \cdot \boldsymbol{\gamma})^2\|_{\mathrm{L}^\infty(\boldsymbol{C}_R)} \to 0 \text{ as } V \to \infty,$$

where $a_\infty(\boldsymbol{c}) = \boldsymbol{c}^{\boldsymbol{\alpha}}/\boldsymbol{c}_*^{\boldsymbol{\alpha}}$, $b_\infty(\boldsymbol{c}) = \boldsymbol{c}^{\boldsymbol{\beta}}/\boldsymbol{c}_*^{\boldsymbol{\beta}}$, and $\boldsymbol{\gamma} = \boldsymbol{\alpha} - \boldsymbol{\beta}$. Using $\Lambda(r,t) \leq \frac{1}{2}(r+t)$, the uniform boundedness of a_V and b_V on $\boldsymbol{C}_R$, and that ϱ_V is a probability measure, we see that in the limsup of $\Psi_V^*(\boldsymbol{u}^V, \boldsymbol{\xi}^V)$ we can replace M_V^ξ by $\big((\boldsymbol{\alpha} - \boldsymbol{\beta}) \cdot \nabla \xi\big)^2$ without changing the limsup in the left-hand side of (4.25).

Next we consider the functionals $F : \mathscr{M}(\boldsymbol{C}_R) \times \mathscr{M}(\boldsymbol{C}_R) \to [0, +\infty]$ given by

$$F(\varrho_1, \varrho_2) = \int_{\boldsymbol{C}_R} f(\rho_1(\boldsymbol{c}), \rho_2(\boldsymbol{c})) \big(\boldsymbol{\gamma} \cdot \nabla \xi(\boldsymbol{c})\big)^2 \mathrm{d}\boldsymbol{c} \text{ with } f(r,t) = \begin{cases} r + t - \Lambda(r,t), & \text{for } r, t \geq 0, \\ +\infty, & \text{else.} \end{cases}$$

Note that $f(r,t) \geq \frac{1}{2}(r+t)$. Moreover, f is convex and positively homogeneous of degree 1. Thus, F is weak* lower semi-continuous on $\mathscr{M}(\boldsymbol{C}_R) \times \mathscr{M}(\boldsymbol{C}_R)$, cf. [FoL07, Thm. 6.57]. Now using the convergences

$$\varrho^{V,a} \overset{*}{\rightharpoonup} a_\infty \varrho\big|_{\boldsymbol{C}_R} \quad \text{and} \quad \varrho^{V,b} \overset{*}{\rightharpoonup} b_\infty \varrho\big|_{\boldsymbol{C}_R} \quad \text{as } V \to \infty,$$

we obtain the liminf estimate $\liminf_{V \to \infty} F(\varrho^{V,a}, \varrho^{V,b}) \geq F(a_\infty \varrho, b_\infty \varrho)$.

Thus, in the view of the identity

$$\int_{\boldsymbol{C}_R} \Lambda(\rho^{V,a}, \rho^{V,b})(\boldsymbol{\gamma} \cdot \nabla \xi(\boldsymbol{c}))^2 \mathrm{d}\boldsymbol{c} = \int_{\boldsymbol{C}_R} (\rho^{V,a} + \rho^{V,b})\big(\boldsymbol{\gamma} \cdot \nabla \xi(\boldsymbol{c})\big)^2 \mathrm{d}\boldsymbol{c} - F(\varrho^{V,a}, \varrho^{V,b}),$$

and observing that the first term on the right-hand side is weak* continuous, the limsup for $V \to \infty$ gives

$$\begin{aligned} \limsup_{V \to \infty} \Psi_V^*(\boldsymbol{u}^V, \boldsymbol{\xi}^V) &= \frac{\kappa_*}{2} \limsup_{V \to \infty} \int_{\boldsymbol{C}_R} \Lambda(\rho^{V,a}, \rho^{V,b})(\boldsymbol{\gamma} \cdot \nabla \xi(\boldsymbol{c}))^2 \mathrm{d}\boldsymbol{c} \\ &\leq \frac{\kappa_*}{2} \int_{\boldsymbol{C}_R} (a_\infty + b_\infty)\big(\boldsymbol{\gamma} \cdot \nabla \xi\big)^2 \mathrm{d}\varrho - \frac{\kappa_*}{2} F(a_\infty \varrho, b_\infty \varrho) \\ &= \frac{\kappa_*}{2} \int_{\boldsymbol{C}_R} \Lambda(a_\infty, b_\infty)\big(\boldsymbol{\gamma} \cdot \nabla \xi(\boldsymbol{c})\big)^2 \mathrm{d}\varrho(\boldsymbol{c}) = \Psi_{\mathrm{Lio}}^*(\varrho, \xi). \end{aligned}$$

This is the desired result for one reaction, and the full result follows by summation over $r = 1, \ldots, R$ and the definition of $\mathbb{K}$, namely $\nabla \xi \cdot \mathbb{K} \nabla \xi = \sum_{r=1}^R \kappa_*^r \Lambda\big(\frac{\boldsymbol{c}^{\boldsymbol{\alpha}^r}}{\boldsymbol{c}_*^{\boldsymbol{\alpha}^r}}, \frac{\boldsymbol{c}^{\boldsymbol{\beta}^r}}{\boldsymbol{c}_*^{\boldsymbol{\beta}^r}}\big)(\boldsymbol{\gamma}^r \cdot \nabla \xi)^2$. ∎

4.6 Convergence of solutions

Here we provide the general convergence result as $V \to \infty$ for the appropriately embedded solutions $\boldsymbol{u}^V : [0, \infty[\to \mathscr{P}(\mathcal{N})$ of the CME to the solutions $\varrho : [0, \infty[\to \mathscr{P}(\boldsymbol{C})$ of the Liouville equation, which is a simple transport along the solutions of the RRE $\dot{\boldsymbol{c}} = -\boldsymbol{R}(\boldsymbol{c}) = -\mathbb{K}(\boldsymbol{c})\mathrm{D}E(\boldsymbol{c})$. Our approach follows the strategy of evolutionary Γ-convergence as initiated in [SaS04, Ser11] with the new idea of dualization as introduced in [LM*17].

Theorem 4.7 (Evolutionary Γ-convergence of CME to Liouville) *For all $V > 1$ consider a solution $\boldsymbol{u}^V : [0, \infty[\to \mathscr{P}(\mathcal{N})$ of the CME (3.2). Assume that the initial conditions are well-prepared in the sense that*

$$\iota_V(\boldsymbol{u}^V(0)) \overset{*}{\rightharpoonup} \varrho^0 \text{ in } \mathscr{P}(\boldsymbol{C}) \quad \text{and} \quad \mathcal{E}_V(\boldsymbol{u}^V(0)) \to \boldsymbol{E}(\varrho^0).$$

Then, for all $t > 0$, we have the convergence

$$\iota_V(\boldsymbol{u}^V(t)) \overset{*}{\rightharpoonup} \varrho(t) \text{ in } \mathscr{P}(\boldsymbol{C}) \quad \text{and} \quad \mathcal{E}_V(\boldsymbol{u}^V(t)) \to \boldsymbol{E}(\varrho(t)),$$

where $\varrho : [0, \infty[\to \mathscr{P}(\boldsymbol{C})$ is the unique solution of the Liouville equation (4.3) starting at $\varrho(0) = \varrho^0$, i.e., for all $\varphi \in \mathrm{C}_{\mathrm{c}}^1([0, T] \times \boldsymbol{C})$ with $\varphi(T, \cdot) = 0$ we have

$$\int_{\boldsymbol{C}} \varphi(0, \boldsymbol{c}) \varrho^0(\mathrm{d}\boldsymbol{c}) + \int_0^T \int_{\boldsymbol{C}} \Big(\partial_t \varphi(t, \boldsymbol{c}) - \nabla\varphi(t, \boldsymbol{c}) \cdot \mathbb{K}(\boldsymbol{c}) \nabla E(\boldsymbol{c}) \Big) \varrho(t, \mathrm{d}\boldsymbol{c}) \, \mathrm{d}t = 0. \qquad (4.26)$$

Moreover, for all $r, s \in [0, T]$ with $r < s$ we have the energy identity

$$\boldsymbol{E}(\varrho(s)) + 2 \int_r^s \Psi_{\mathrm{Lio}}^* \big(\varrho(t), -\mathrm{D}\boldsymbol{E}(\varrho(t)) \big) \, \mathrm{d}t = \boldsymbol{E}(\varrho(r)). \qquad (4.27)$$

For the proof we use the energy-dissipation principle for $V \geq 1$ and pass to the limit in each of the terms. If $\boldsymbol{u}^V$ is a solution of the CME, then for all $T > 0$ we have

$$\mathcal{E}_V(\boldsymbol{u}^V(T)) + \int_0^T \Psi_V(\boldsymbol{u}^V, \dot{\boldsymbol{u}}^V) + \Psi_V^*\big(\boldsymbol{u}^V, -\mathrm{D}\mathcal{E}_V(\boldsymbol{u}^V)\big) \, \mathrm{d}t = \mathcal{E}_V(\boldsymbol{u}^V(0)). \qquad (4.28)$$

Following the ideas in [DiL15] for the passage from a Markov chain to the Fokker–Planck equation or the general methods in evolutionary Γ-convergence, we want to pass to the limit in each of the four terms. As a general fact, it will be sufficient to obtain liminf estimates on the left-hand side, since by a chain-rule argument an estimate with "$\leq$" instead of equality can be turned back into an equality. Moreover, by the assumptions of the theorem we see that the right-hand side converges to the desired limit.

However, it is rather delicate to pass to the limit in the integral $\int_0^T \Psi_V(\boldsymbol{u}^V, \dot{\boldsymbol{u}}^V) \, \mathrm{d}t$, because the potential Ψ_V is only implicitly defined and we expect the limit to be given in terms of the Benamou-Brenier formula for the Wasserstein distance induced by the metric on $(\boldsymbol{C}, \mathbb{K})$. A major difficulty is even to obtain a suitable equi-continuity for the solutions $\boldsymbol{u}^V$ to be able to extract a subsequence converging at all times. In particular, it is unclear how to pass to the limit in $\iota_V(\dot{\boldsymbol{u}}^V(t))$ by a direct argument.

Hence, following [LM*17], we estimate the primal dissipation potential Ψ_V from below using the definition in terms of the Legendre transform of Ψ_V^*. Using additionally an integration by parts we have

$$\int_0^T \Psi_V(\boldsymbol{u}^V, \dot{\boldsymbol{u}}^V) \, \mathrm{d}t \geq \mathfrak{J}_V(\boldsymbol{u}^V, \boldsymbol{\eta}) \quad \text{for all } \boldsymbol{\eta} \in \mathrm{C}^1([0, T]; \ell^\infty(\mathcal{N})) \text{ with}$$

$$\mathfrak{J}_V(\boldsymbol{u}, \boldsymbol{\eta}) := \langle \boldsymbol{u}(T), \boldsymbol{\eta}(T) \rangle - \langle \boldsymbol{u}(0), \boldsymbol{\eta}(0) \rangle - \int_0^T \langle \boldsymbol{u}(t), \dot{\boldsymbol{\eta}}(t) \rangle + \Psi_V^*(\boldsymbol{u}(t), \boldsymbol{\eta}(t)) \, \mathrm{d}t,$$

where $\langle \boldsymbol{u}, \boldsymbol{\eta} \rangle := \sum_{n \in \mathcal{N}} u_{\boldsymbol{n}} \eta_{\boldsymbol{n}}$. With this argument we can replace the energy-dissipation principle (4.28) by the estimate

$$\mathcal{E}_V(\boldsymbol{u}^V(T)) + \mathfrak{J}_V(\boldsymbol{u}^V, \boldsymbol{\eta}) + \int_0^T \Psi_V^*\big(\boldsymbol{u}^V, -\mathrm{D}\mathcal{E}_V(\boldsymbol{u}^V)\big) \, \mathrm{d}t \leq \mathcal{E}_V(\boldsymbol{u}^V(0)), \qquad (4.29)$$

which holds for all differentiable $\boldsymbol{\eta}$. In this equation we are then able to pass to the limit $V \to \infty$, when choosing $\boldsymbol{\eta} = \boldsymbol{\eta}^V = \iota_V^*(\xi)$ for a smooth function ξ.

At the end we are then able to calculate the supremum over all ξ by using the especially simple quadratic structure in ξ, which mirrors the fact that the Liouville equation is a simple transport equation.

Proof of Theorem 4.7.

Step 1: Embedding and uniform a priori bounds. We now consider the family $\boldsymbol{u}^V : [0, T] \to \mathscr{P}(\mathcal{N})$ and embed it into $\mathscr{P}(\boldsymbol{C})$ via ι_V from (4.6). As in [DiL15] we show an equi-continuity in a 1-Wasserstein distance, but introduce an additional weight accounting for our unbounded domain $\boldsymbol{C}$. We define the maximal order p of all reactions via

$$p := \max\{ |\boldsymbol{\alpha}^r|_1, |\boldsymbol{\beta}^r|_1 \mid r = 1, \ldots, R \}.$$

For $\mu \in \mathscr{M}(\boldsymbol{C})$ and for $\varrho_0, \varrho_1 \in \mathscr{P}(\boldsymbol{C})$ we set

$$\|\mu\|_{1\mathrm{W}} := \sup \left\{ \int_{\boldsymbol{C}} f(\boldsymbol{c}) \, \mathrm{d}\mu(\boldsymbol{c}) \, \middle| \, f \in \mathbb{F} \right\} \quad \text{and} \quad d_{1\mathrm{W}}(\varrho_0, \varrho_1) = \|\varrho_0 - \varrho_1\|_{1\mathrm{W}},$$

where $\mathbb{F} := \{ f \in \mathrm{C}^1(\boldsymbol{C}) \mid \sup_{\boldsymbol{C}}(1+|\boldsymbol{c}|^p)|\nabla f(\boldsymbol{c})| \leq 1 \}$.

Using the definition of the Markov generators $\mathcal{Q}_V^r$ in terms of the coefficients $\mathbb{B}_V^{\delta^r}(\boldsymbol{n})$, see (3.3), it is easy to derive the uniform estimate $\|\iota_V(\dot{\boldsymbol{u}}^V(t))\|_{1\mathrm{W}} \leq C_{1\mathrm{W}}$ independently of the initial conditions and $V \geq 1$ (one simply needs $\sum u_{\boldsymbol{n}}^V \equiv 1$). Hence, we obtain the uniform Lipschitz bound

$$d_{1\mathrm{W}}\big(\iota_V(\boldsymbol{u}^V(t)), \iota_V(\boldsymbol{u}^V(s))\big) \leq C_{1\mathrm{W}}|t-s| \quad \text{for all } s, t \in [0, T] \text{ and all } V \geq 1.$$

Moreover, as $\mathcal{E}_V(\boldsymbol{u}^V(t)) \leq \mathcal{E}_V(\boldsymbol{u}^V(0)) \leq \boldsymbol{E}(\varrho^0) + o(1)_{V\to\infty}$ by well-preparedness, the equi-coercivity of $\mathcal{E}_V$ established in (4.18) yields the uniform bound

$$\exists V_* \geq 1, C_{\mathrm{B}} < \infty \, \forall \, t > 0, \, V \geq V_* : \int_{\boldsymbol{C}}(1+|\boldsymbol{c}|)\iota_V(\boldsymbol{u}^V(t)) \, \mathrm{d}\boldsymbol{c} \leq C_{\mathrm{B}}. \tag{4.30}$$

Step 2: Extraction of a subsequence. The subset of $\mathscr{P}(\boldsymbol{C})$ defined by the boundedness of the above first moment is a compact subset of the metric space $(\mathscr{P}(\boldsymbol{C}), d_{1\mathrm{W}})$. Indeed, using Prokhorov's theorem one finds that this set is weak* sequentially compact. Since $d_{1\mathrm{W}}$ is dominated by the bounded Lipschitz metric (which metrizes weak* convergence), the compactness of $(\mathscr{P}(\boldsymbol{C}), d_{1\mathrm{W}})$ follows.

Hence, we can apply the abstract Arzelà-Ascoli theorem in $(\mathscr{P}(\boldsymbol{C}), d_{1\mathrm{W}})$ to extract a subsequence $V_k \to \infty$ and a limit function $\varrho : [0, T] \to \mathscr{P}(\boldsymbol{C})$ such that

$$\forall \, t \in [0, T]: \qquad \iota_V(\boldsymbol{u}^V(t)) \overset{*}{\rightharpoonup} \varrho(t) \text{ in } \mathscr{P}(\boldsymbol{C}), \tag{4.31a}$$

$$\forall \, s, t \in [0, T]: \qquad d_{1\mathrm{W}}(\varrho(t), \varrho(s)) \leq C_{1\mathrm{W}}|t-s|, \tag{4.31b}$$

$$\forall \, t \in [0, T]: \qquad \boldsymbol{E}(\varrho(t)) \leq \boldsymbol{E}(\varrho^0), \tag{4.31c}$$

$$\text{the mapping } t \mapsto \varrho(t) \text{ is weak* continuous.} \tag{4.31d}$$

At first, in place of (4.31a) one obtains $d_{1\mathrm{W}}\big(\iota_V(\boldsymbol{u}^V(t)), \varrho(t)\big) \to 0$. To derive (4.31a), we use the bound (4.30) together with the fact that any bounded continuous function can be uniformly approximated on compact sets by (multiples of) functions in $\mathbb{F}$. Similarly, (4.31d) follows from (4.31b). In particular, combining (4.31d) and the assumption $\iota_V(\boldsymbol{u}^V(0)) \overset{*}{\rightharpoonup} \varrho^0$ we conclude $\varrho(0) = \varrho^0$. Finally, (4.31c) follows via (4.31a) from Theorem 4.4:

$$\boldsymbol{E}(\varrho(t)) \leq \liminf_{V\to\infty} \mathcal{E}_V\big(\boldsymbol{u}^V(t)\big) \leq \liminf_{V\to\infty} \mathcal{E}_V\big(\boldsymbol{u}^V(0)\big) = \boldsymbol{E}(\varrho^0).$$

Step 3: Limit passage in (4.29). Combining (4.31a) for $t = T$ and Theorem 4.4 (cf. (4.19)), the first term satisfies the liminf estimate $\liminf_{V \to \infty} \mathcal{E}_V(\boldsymbol{u}^V(T)) \geq \boldsymbol{E}(\varrho(T))$. For the last term we use the assumption $\mathcal{E}_V(\boldsymbol{u}^V(0)) \to \boldsymbol{E}(\varrho^0) = \boldsymbol{E}(\varrho(0))$.

For the third term we employ Proposition 4.5 for each $t \in [0, T]$ based on (4.31a). Using Fatou's lemma we conclude the liminf estimate

$$\liminf_{V \to \infty} \int_0^T \Psi_V^*\big(\boldsymbol{u}^V(t), -\mathrm{D}\mathcal{E}_V(\boldsymbol{u}^V(t))\big)\,\mathrm{d}t \geq \int_0^T \liminf_{V \to \infty} \Psi_V^*\big(\boldsymbol{u}^V(t), -\mathrm{D}\mathcal{E}_V(\boldsymbol{u}^V(t))\big)\,\mathrm{d}t$$

$$\geq \int_0^T \Psi_{\mathrm{Lio}}^*\big(\varrho(t), -\mathrm{D}\boldsymbol{E}(\varrho(t))\big)\,\mathrm{d}t.$$

Thus, it remains to pass to the limit in $\mathfrak{J}_V(\boldsymbol{u}^V, \eta)$. For this we choose an arbitrary $\xi \in \mathrm{C}_{\mathrm{c}}^1([0, T] \times \boldsymbol{C})$ and define $\boldsymbol{\xi}^V(t) = \iota_V^*(\xi(t))$, cf. (4.7). With this choice we can apply Proposition 4.6 for all $t \in [0, T]$ based on (4.31a). Now, Fatou's lemma yields

$$\liminf_{V \to \infty} \mathfrak{J}_V(\boldsymbol{u}^V, \boldsymbol{\xi}^V) \geq \mathfrak{J}_{\mathrm{Lio}}(\varrho, \xi) \quad \text{where}$$

$$\mathfrak{J}_{\mathrm{Lio}}(\varrho, \xi) := \int_{\boldsymbol{C}} \xi(T, \boldsymbol{c})\varrho(T, \mathrm{d}\boldsymbol{c}) - \int_{\boldsymbol{C}} \xi(0, \boldsymbol{c})\varrho(0, \mathrm{d}\boldsymbol{c})$$

$$- \int_0^T \int_{\boldsymbol{C}} \Big(\partial_t \xi(t, \boldsymbol{c}) + \frac{1}{2}\nabla\xi(t, \boldsymbol{c}) \cdot \mathbb{K}(\boldsymbol{c})\nabla\xi(t, \boldsymbol{c})\Big)\varrho(t, \mathrm{d}\boldsymbol{c})\,\mathrm{d}t.$$

In summary, we conclude that the limit function $\varrho : [0, T] \to \mathscr{P}(\boldsymbol{C})$ satisfies

$$\boldsymbol{E}(\varrho(T)) + \mathfrak{J}_{\mathrm{Lio}}(\varrho, \xi) + \int_0^T \Psi_{\mathrm{Lio}}^*\big(\varrho(t), -\mathrm{D}\boldsymbol{E}(\varrho(t))\big)\,\mathrm{d}t \leq \boldsymbol{E}(\varrho(0)) \tag{4.32}$$

for all $\xi \in \mathrm{C}_{\mathrm{c}}^1([0, T] \times \boldsymbol{C})$.

Step 4: Energy balance. By inserting $\xi \equiv 0$ in (4.32) we obtain the upper bound

$$\mathcal{D}(\varrho; 0, T) := \int_0^T \int_{\boldsymbol{C}} \nabla E(\boldsymbol{c}) \cdot \mathbb{K}(\boldsymbol{c})\nabla E(\boldsymbol{c})\varrho(t, \mathrm{d}\boldsymbol{c})\,\mathrm{d}t \leq 2\big(\boldsymbol{E}(\varrho(0)) - \boldsymbol{E}(\varrho(T))\big).$$

We want to show energy balance, i.e., equality when the factor 2 is omitted. For this purpose, we observe that the measures $\varrho(t, \cdot) \in \mathscr{P}(\boldsymbol{C})$ decay at infinity such that (4.31c) holds. Hence, we may also use $\xi(t, \boldsymbol{c}) = \lambda E(\boldsymbol{c})$ as testfunctions in (4.32). Writing shortly $e(t) := \boldsymbol{E}(\varrho(t))$ we find $\mathfrak{J}_{\mathrm{Lio}}(\varrho, \lambda E) = \lambda\big(e(T) - e(0)\big) - \frac{\lambda^2}{2}\mathcal{D}(\varrho; 0, T)$ and obtain

$$-\lambda\big(e(0) - e(T)\big) - \frac{\lambda^2}{2}\mathcal{D}(\varrho; 0, T) = \mathfrak{J}_{\mathrm{Lio}}(\varrho, \lambda E) \leq e(0) - e(T) - \frac{1}{2}\mathcal{D}(\varrho; 0, T) \quad \text{for all } \lambda \in \mathbb{R}.$$

Maximizing with respect to λ leads to $(e(0) - e(T))^2/\mathcal{D} \leq 2(e(0) - e(T)) - \mathcal{D}$ which implies $e(0) - e(T) = \mathcal{D}$, or more explicitly $\mathcal{D}(\varrho; 0, T) = \boldsymbol{E}(\varrho(0)) - \boldsymbol{E}(\varrho(T))$, which is the desired energy balance (4.27) for $r = 0$ and $s = T$.

Moreover, we can repeat the calculation on $[0, s]$ with $0 < s < T$ instead of $[0, T]$. The full result (4.27) follows by subtracting the identity on $[0, r]$ from that on $[0, s]$.

Step 5: Weak form of gradient flow equation.   With Step 4 we rewrite (4.32) as

$$\mathfrak{J}_{\mathrm{Lio}}(\varrho, \xi) \le \boldsymbol{E}(\varrho(0)) - \boldsymbol{E}(\varrho(T)) - \frac{1}{2}\mathcal{D}(\varrho; 0, T) = \frac{1}{2}\mathcal{D}(\varrho; 0, T),$$

and know that the left-hand side is maximized by $\xi : (t, \boldsymbol{c}) \mapsto -E(\boldsymbol{c})$. Inserting the test functions $\xi(t, \boldsymbol{c}) = \delta\varphi(t, \boldsymbol{c}) - E(\boldsymbol{c})$ with small $\delta > 0$ and $\varphi \in \mathrm{C}_{\mathrm{c}}^1([0, T] \times \boldsymbol{C})$ with $\varphi(T, \cdot) = 0$ we arrive, after some cancellations and after dividing by $\delta > 0$, at

$$-\int_{\boldsymbol{C}} \varphi(0, \boldsymbol{c})\varrho(0, \mathrm{d}\boldsymbol{c}) - \int_0^T \int_{\boldsymbol{C}} \Big(\partial_t \varphi - \nabla\varphi \cdot \mathbb{K}\nabla\big(E - \tfrac{\delta}{2}\varphi\big) \Big) \varrho(t, \mathrm{d}\boldsymbol{c}) \, \mathrm{d}t \le 0.$$

Taking the limit $\delta \searrow 0$ and replacing φ by $-\varphi$, we obtain the desired result (4.26).

With this, Theorem 4.7 is established.   ∎

5   Approximation via Fokker–Planck equations

In the above section we have seen that the Liouville equation is the proper limit of the CME for $V \to \infty$. However, for finite but large V it can still be advantageous to replace the discrete CME by a continuous PDE with V as a large parameter. In this range the stochastic modeling is done by the so-called *Langevin dynamics*, see [Kur78, Gil00, WiS17], which is based on a stochastic perturbation of the reaction-rate equation (RRE), see (1.4). At the level of probability distributions the corresponding model is the associated Fokker–Planck equation (FPE). We will discuss two different gradient flow approximations: in the first we simply add a suitable "entropic term" to the driving functional, but keep the dissipation fixed (cf. Section 5.2), while in the second we expand $\mathcal{E}_V$ and $\mathcal{K}_V$ such that all terms of order $1/V$ are correct (cf. Section 5.3).

5.1   Improved approximation of the relative entropy

We interpret the sum in the definition of $\mathcal{E}_V$ as a Riemann sum and replace it by a corresponding integral. The main point of the improvement is that we keep the entropy term $\frac{1}{V}\sum u_{\boldsymbol{n}} \log u_{\boldsymbol{n}}$ in the definition of $\mathcal{E}_V(\boldsymbol{u})$, which is in contrast to the limit $\boldsymbol{E}$ obtained in Theorem 4.4. Working with absolutely continuous probability measures $\varrho(\mathrm{d}\boldsymbol{c}) = \rho(\boldsymbol{c})\,\mathrm{d}\boldsymbol{c}$ with $\rho \in \mathrm{L}^1(\boldsymbol{C})$, we can define the V-dependent entropy by

$$\boldsymbol{E}_V(\varrho) = \frac{1}{V} \int_{\boldsymbol{C}} \rho(\boldsymbol{c}) \log\Big(\frac{\rho(\boldsymbol{c})}{W_V(\boldsymbol{c})}\Big) \mathrm{d}\boldsymbol{c}, \tag{5.1}$$

where the equilibrium density $W_V \in \mathrm{L}^1(\boldsymbol{C})$ has to be chosen suitably. A first simple approximation is $\widetilde{W}_V(\boldsymbol{c}) = \frac{1}{\widetilde{Z}_V} \mathrm{e}^{-VE(\boldsymbol{c})}$ with $E(\boldsymbol{c}) = \sum_{i=1}^I c_i^* \lambda_{\mathrm{B}}(c_i/c_i^*)$ as above and $\widetilde{Z}_V = \int_{\boldsymbol{C}} \mathrm{e}^{-VE(\boldsymbol{c})} \, \mathrm{d}\boldsymbol{c}$. However, a better and more refined W_V is obtained using the next order of expansion in Stirling's formula (4.10) as well. For this we use the approximation $k_n \approx n + 1/6$, i.e., $\log(n!) = n \log n - n + \frac{1}{2}\log\big(2\pi(n+\frac{1}{6})\big) + O(1/n^2)$ for $n \to \infty$. Hence, taking the limits $V, |\boldsymbol{n}| \to \infty$ such that $\frac{\boldsymbol{n}}{V} \to \boldsymbol{c}$, we obtain

$$-\frac{1}{V} \log w_{\boldsymbol{n}}^V \approx E(\boldsymbol{c}) + \frac{1}{V} G_V(\boldsymbol{c})$$

with the V-dependent correction $G_V(\boldsymbol{c}) := \frac{1}{2}\sum_{i=1}^I \log\big(2\pi(Vc_i + \frac{1}{6})\big)$ for E.

We now take a probability measure $\varrho = \rho \, \mathrm{d}\boldsymbol{c} \in \mathscr{P}(\boldsymbol{C})$ and a discrete approximation $\boldsymbol{u} \approx \varkappa_V(\varrho) \in \mathscr{P}(\mathcal{N})$, where $\varkappa_V : \mathscr{P}(\boldsymbol{C}) \to \mathscr{P}(\mathcal{N})$ is the natural projection defined in (4.8). Then the Riemann-sum approximation results in

$$
\begin{aligned}
\mathcal{E}_V(\boldsymbol{u}) &= \frac{1}{V} \sum_{\boldsymbol{n} \in \mathcal{N}} u_{\boldsymbol{n}} \log u_{\boldsymbol{n}} - \frac{1}{V} \sum_{\boldsymbol{n} \in \mathcal{N}} u_{\boldsymbol{n}} \log w_{\boldsymbol{n}}^V \\
&\approx \frac{1}{V} \int_{\boldsymbol{C}} \rho(\boldsymbol{c}) \log \rho(\boldsymbol{c}) \, \mathrm{d}\boldsymbol{c} - I \frac{\log V}{V} + \int_{\boldsymbol{C}} \Big(E(\boldsymbol{c}) + \frac{1}{V} G_V(\boldsymbol{c}) \Big) \rho(\boldsymbol{c}) \, \mathrm{d}\boldsymbol{c} \\
&= \frac{1}{V} \int_{\boldsymbol{C}} \rho(\boldsymbol{c}) \log \Big(\frac{\rho(\boldsymbol{c})}{\widehat{W}(V, \boldsymbol{c}, \boldsymbol{c}^*)} \Big) \, \mathrm{d}\boldsymbol{c},
\end{aligned}
$$

where $\widehat{W}(V, \boldsymbol{c}, \boldsymbol{c}_*) = \prod_{i=1}^{I} \widehat{\mathsf{W}}(V, c_i, c_i^*)$ with $\widehat{\mathsf{W}}(V, c, c^*) = \dfrac{V \mathrm{e}^{-Vc^* \lambda_{\mathrm{B}}(c/c^*)}}{\sqrt{2\pi(Vc + 1/6)}}$.

The probability density W_V is then defined by normalizing $\widehat{W}(V, \cdot, \boldsymbol{c}_*)$. We thus set $Z(V, c^*) := \int_0^\infty \widehat{\mathsf{W}}(V, c, c^*) \, \mathrm{d}c$ and

$$
W_V(\boldsymbol{c}) := \prod_{i=1}^{I} \mathsf{W}(V, c_i, c_i^*) \quad \text{with } \mathsf{W}(V, c, c^*) := \frac{\widehat{\mathsf{W}}(V, c, c^*)}{\mathsf{Z}(V, c^*)}. \tag{5.2}
$$

This yields the expansion

$$
-\frac{1}{V} \log W_V(\boldsymbol{c}) = E(\boldsymbol{c}) + \frac{1}{V} E_1^V(\boldsymbol{c}) \quad \text{where} \quad E_1^V(\boldsymbol{c}) = \widehat{\mathsf{z}}(V, \boldsymbol{c}_*) + \frac{1}{2} \sum_{i=1}^{I} \log \big(V c_i + \tfrac{1}{6} \big)
$$

with $\widehat{\mathsf{z}}(V, \boldsymbol{c}_*) := \sum_{i=1}^{I} \log \big(\sqrt{2\pi} \, \mathsf{Z}(V, c_i^*)/V \big)$. In summary, for $\boldsymbol{E}_V$ defined via (5.1) and (5.2) we have

$$
\boldsymbol{E}_V(\varrho) = \boldsymbol{E}(\varrho) + \frac{1}{V} \int_{\boldsymbol{C}} \big(\rho \log \rho + E_1^V \rho \big) \, \mathrm{d}\boldsymbol{c}, \tag{5.3}
$$

and $\mathrm{D}\boldsymbol{E}_V(\varrho)(\boldsymbol{c}) = \frac{1}{V} \log \rho(\boldsymbol{c}) - \frac{1}{V} \log W_V(\boldsymbol{c})$.

5.2 Simple Fokker–Planck approximation

Here we keep the V-independent Onsager operator $\boldsymbol{K}(\varrho) : \xi \mapsto -\operatorname{div}\big(\varrho \mathbb{K} \nabla \xi \big)$ of the Liouville equation and obtain the V-dependent continuous gradient system $(\mathscr{P}(\boldsymbol{C}), \boldsymbol{E}_V, \boldsymbol{K})$. The associated gradient-flow equation $\dot{\varrho} = -\boldsymbol{K}(\varrho) \mathrm{D} \boldsymbol{E}_V(\varrho)$ is the FPE

$$
\dot{\rho} = \operatorname{div} \Big(\frac{1}{V} \mathbb{K}(\boldsymbol{c}) \nabla \rho + \rho \boldsymbol{R}(\boldsymbol{c}) + \rho \boldsymbol{A}_V(\boldsymbol{c}) \Big), \tag{5.4}
$$

where we used $\mathbb{K}(\boldsymbol{c}) \mathrm{D} E(\boldsymbol{c}) = \boldsymbol{R}(\boldsymbol{c})$ and set $\boldsymbol{A}_V(\boldsymbol{c}) := \frac{1}{2} \mathbb{K}(\boldsymbol{c}) \big(\frac{1}{Vc_i + 1/6} \big)_{i=1,\dots,I}$.

We expect that this FPE is a good approximation to the CME for all sufficiently large V. In particular, (5.4) has the steady state $\rho = W_V$, which is close to the discrete steady state $\boldsymbol{w}^V \in \mathscr{P}(\mathcal{N})$ using the embedding as above. In contrast, the only steady states of the Liouville equation (4.1) are concentrated on the equilibria of $\dot{\boldsymbol{c}} = -\boldsymbol{R}(\boldsymbol{c})$. Of course, the FPE still respects the invariant sets $\boldsymbol{I}(\boldsymbol{q})$, because the mobility $\mathbb{K}$ of the Onsager operator $\boldsymbol{K}$ is the same as for the Liouville equation. In particular, $\rho = W_V$ is the unique equilibrium density if and only if $\mathbb{K}$ has full rank, i.e., $\boldsymbol{I}(\boldsymbol{q}) = \boldsymbol{C}$ for all $\boldsymbol{q} \in \mathfrak{Q}$.

The simpler choice $\widetilde{W}_V(\boldsymbol{c}) = \frac{1}{\widetilde{Z}(V)} \mathrm{e}^{-VE(\boldsymbol{c})}$ for the equilibrium yields the relative entropy

$$\widetilde{\boldsymbol{E}}_V(\varrho) = \frac{1}{V} \int_C \rho(\boldsymbol{c}) \log\Big(\frac{\rho(\boldsymbol{c})}{\widetilde{W}_V(\boldsymbol{c})}\Big) \,\mathrm{d}\boldsymbol{c} = \int_C \Big(\frac{1}{V} \log \rho(\boldsymbol{c}) + E(\boldsymbol{c})\Big) \rho(\boldsymbol{c}) \,\mathrm{d}\boldsymbol{c} + \frac{\log(\widetilde{Z}(V))}{V}.$$

The flow equation $\dot{\varrho} = -\boldsymbol{K}(\varrho)\mathrm{D}\widetilde{\boldsymbol{E}}_V(\varrho)$ induced by the gradient system $(\mathscr{P}(\boldsymbol{C}), \widetilde{\boldsymbol{E}}_V, \boldsymbol{K})$ is the simplified FPE

$$\dot{\rho} = \mathrm{div}\,\Big(\frac{1}{V}\mathbb{K}(\boldsymbol{c})\nabla\rho + \rho\boldsymbol{R}(\boldsymbol{c})\Big), \tag{5.5}$$

which is the same as (5.4) but with $\boldsymbol{A}_V \equiv 0$. The simplified equation will be used below as well, since $\widetilde{W}_V$ has a simpler explicit form.

We believe that this approximation is suitable for many purposes. However, it does not produce the correct diffusion as derived in [Kur78, Eqn. (1.7)]. This diffusion correction is used to improve the RRE $\dot{\boldsymbol{c}} = -\boldsymbol{R}(\boldsymbol{c})$ by replacing it by a stochastic differential equation called the chemical Langevin equations (CLE) in [Gil00, WiS17], see (1.4). The associated Fokker–Planck equation takes the form

$$\dot{\rho} = \frac{1}{V}\sum_{i,j=1}^{I} \partial_{ij}^2\big(\rho\widehat{\mathbb{K}}_{\mathrm{CLE}}(\boldsymbol{c})_{ij}\big) + \mathrm{div}\,\big(\rho\boldsymbol{R}(\boldsymbol{c})\big), \tag{5.6}$$

where $\widehat{\mathbb{K}}_{\mathrm{CLE}}(\boldsymbol{c}) \in \mathbb{R}^{I \times I}$ is given in (1.6) and differs from $\mathbb{K}$ as the logarithmic mean $\Lambda(a, b)$ between $a = \boldsymbol{c}^{\boldsymbol{\alpha}^r}/\boldsymbol{c}_*^{\boldsymbol{\alpha}^r}$ and $b = \boldsymbol{c}^{\boldsymbol{\beta}^r}/\boldsymbol{c}_*^{\boldsymbol{\beta}^r}$ is replaced by the arithmetic mean $\frac{1}{2}(a+b)$. Obviously, (5.6) does not have a gradient structure with respect to $\widehat{\mathbb{K}}_{\mathrm{CLE}}$, because there is no function $\boldsymbol{c} \mapsto \widehat{E}(\boldsymbol{c})$ such that $\boldsymbol{R}(\boldsymbol{c}) = \widehat{\mathbb{K}}_{\mathrm{CLE}}(\boldsymbol{c})\nabla\widehat{E}(\boldsymbol{c})$.

5.3 Fokker–Planck equation with higher-order terms

To derive a proper expansion for the term of order $1/V$ in the evolution equation, we work with the V-dependent entropy $\boldsymbol{E}_V$ defined in Section 5.1. Up to an irrelevant V-dependent constant, this functional approximates $\mathcal{E}_V$ from (3.7) up to order $1/V^2$.

Similarly, we need to derive a suitable expansion for the dissipation potential, which can be done for each reaction independently. The discrete dual dissipation potential is given by (4.5b), namely

$$\Psi_V^*(\boldsymbol{u}, \boldsymbol{\xi}) = \frac{V}{2}\sum_{\boldsymbol{n} \in \mathcal{N}} \Lambda\big(k_{\mathrm{fw}}\mathbb{B}_V^{\boldsymbol{\alpha}}(\boldsymbol{n})\,u_{\boldsymbol{n}+\boldsymbol{\alpha}}, k_{\mathrm{bw}}\mathbb{B}_V^{\boldsymbol{\beta}}(\boldsymbol{n})\,u_{\boldsymbol{n}+\boldsymbol{\beta}}\big)\big(\mu_{\boldsymbol{n}+\boldsymbol{\alpha}}-\mu_{\boldsymbol{n}+\boldsymbol{\beta}}\big)^2.$$

For a smooth function $\xi : \boldsymbol{C} \to \mathbb{R}$ we use the second-order accurate midpoint approximation $\boldsymbol{\mu} = \widehat{\boldsymbol{\mu}}_V^\xi : \boldsymbol{n} \mapsto \xi\big(\frac{1}{V}(\boldsymbol{n}+\boldsymbol{\delta})\big)$ with $\boldsymbol{\delta} = \frac{1}{2}(1, \dots, 1)$ to obtain the expansion

$$V\big(\mu_{\boldsymbol{n}+\boldsymbol{\alpha}} - \mu_{\boldsymbol{n}+\boldsymbol{\beta}}\big) = \nabla\xi(\mathsf{c}_{\boldsymbol{n}}^V) \cdot \big(\boldsymbol{\alpha}-\boldsymbol{\beta}\big) + O(1/V^2)_{V \to \infty} \quad \text{with } \mathsf{c}_{\boldsymbol{n}}^V := \frac{1}{V}\Big(\boldsymbol{n}+\frac{\boldsymbol{\alpha}+\boldsymbol{\beta}}{2}+\boldsymbol{\delta}\Big),$$

where we used symmetric difference quotients to obtain second order accuracy. Moreover, for a smooth and sufficiently fast decaying $\varrho = \rho\mathrm{d}\boldsymbol{c} \in \mathscr{P}(\boldsymbol{C})$ we define the associated discrete $\boldsymbol{u} \in \mathscr{P}(\mathcal{N})$ via $\boldsymbol{u} = \varkappa_V(\varrho) = \iota_V^*\varrho$, which yields $V^I u_{\boldsymbol{n}} = \rho\big(\frac{1}{V}(\boldsymbol{n}+\boldsymbol{\delta})\big) + O(1/V^2)$,

$$V^I u_{\boldsymbol{n}+\boldsymbol{\alpha}} = \rho(\mathsf{c}_{\boldsymbol{n}}^V) + \frac{1}{2V}\nabla\rho(\mathsf{c}_{\boldsymbol{n}}^V) \cdot (\boldsymbol{\alpha}-\boldsymbol{\beta}) + O(1/V^2),$$

and similarly for $V^I u_{n+\beta}$. Hence, for the arguments of Λ we find the expansion

$$\frac{1}{V}\mathbb{B}_V^{\alpha}(n)\,V^I u_{n+\alpha} = (\mathsf{c}_n^V)^{\alpha}\rho(\mathsf{c}_n^V) + \frac{1}{V}F_n^V + O(1/V^2)$$

$$\text{with } F_n^V = -(\mathsf{c}_n^V)^{\alpha}\rho(\mathsf{c}_n^V)\sum_{i=1}^{I}\frac{\alpha_i\beta_i}{2(\mathsf{c}_n^V)_i} + \frac{1}{2}(\mathsf{c}_n^V)^{\alpha}\nabla\rho(\mathsf{c}_n^V)\cdot(\alpha-\beta).$$

For all smooth functions $f, g : C \to \mathbb{R}$ with compact support in $\mathrm{int}(C)$, the trapezoidal rule for Riemann integrals gives

$$\sum_{n\in\mathcal{N}}\left(f(\mathsf{c}_n^V) + \frac{1}{V}g(\mathsf{c}_n^V)\right)\frac{1}{V^I} = \int_C\left(f(c) + \frac{1}{V}g(c)\right)\mathrm{d}c + O(1/V^2).$$

Hence, for smooth ρ and ξ we find the expansion

$$\Psi_V^*\big(\varkappa_V(\varrho),\widehat{\boldsymbol{\mu}}_V^{\xi}\big) = \Phi_V^*(\varrho,\xi) + O(1/V^2) \text{ for } V\to\infty \text{ with}$$

$$\Phi_V^*(\varrho,\xi) = \frac{1}{2}\int_C\left(\Lambda(k_{\mathrm{fw}}c^{\alpha}, k_{\mathrm{bw}}c^{\beta})\rho(c) + \frac{1}{V}\Upsilon\big(c,\rho(c),\nabla\rho(c)\big)\right)\big(\nabla\xi(c)\cdot(\alpha-\beta)\big)^2\mathrm{d}c,$$

where the correction term Υ takes the explicit form

$$\Upsilon(c,\rho,p) = \Upsilon_0(c)\rho + \Upsilon_1(c)\,p\cdot(\alpha-\beta) \quad\text{with}$$

$$\Upsilon_0(c) = -\frac{1}{2}\Lambda(k_{\mathrm{fw}}c^{\alpha}, k_{\mathrm{bw}}c^{\beta})\,\alpha\cdot\check{C}\beta \text{ with } \check{C} = \mathrm{diag}(c_i^{-1})_{i=1,\dots,I},$$

$$\Upsilon_1(c) = \Lambda(k_{\mathrm{fw}}c^{\alpha}, k_{\mathrm{bw}}c^{\beta})\frac{k_{\mathrm{fw}}c^{\alpha}+k_{\mathrm{bw}}c^{\beta}-2\Lambda(k_{\mathrm{fw}}c^{\alpha}, k_{\mathrm{bw}}c^{\beta})}{2(k_{\mathrm{fw}}c^{\alpha}-k_{\mathrm{bw}}c^{\beta})}.$$

Here we used the relation $\partial_a\Lambda(a,b) = \frac{\Lambda(a,b)}{a}\frac{a-\Lambda(a,b)}{a-b}$, giving

$$a\partial_a\Lambda(a,b)+b\partial_b\Lambda(a,b) = \Lambda(a,b) \text{ and } a\partial_a\Lambda(a,b)-b\partial_b\Lambda(a,b) = \Lambda(a,b)\frac{a+b-2\Lambda(a,b)}{a-b}.$$

Now we are in the position to calculate the first-order correction to the Liouville equation from the approximate entropy E_V (cf. (5.3)) and the dual dissipation potential Φ_V^* via $\dot{\varrho} = \mathrm{D}_\xi\Phi_V^*\big(\varrho, -\mathrm{D}E_V(\varrho)\big)$, which yields

$$\dot\rho = \mathrm{div}\left\{\left(\rho\,\widehat{a}(c) + \frac{1}{V}\big[\rho\,\widehat{b}_0(c) + \widehat{b}_1(c)\nabla\rho\cdot(\alpha-\beta)\big] + O(1/V^2)\right)(\alpha-\beta)\right\},$$

where the coefficients are given by

$$\widehat{a}(c) = \Lambda(k_{\mathrm{fw}}c^{\alpha}, k_{\mathrm{bw}}c^{\beta})\,(\alpha-\beta)\cdot\nabla E(c) = k_{\mathrm{fw}}c^{\alpha}-k_{\mathrm{bw}}c^{\beta},$$

$$\widehat{b}_0(c) = \Lambda(k_{\mathrm{fw}}c^{\alpha}, k_{\mathrm{bw}}c^{\beta})\,(\alpha-\beta)\cdot\check{C}\delta - \frac{1}{2}(k_{\mathrm{fw}}c^{\alpha}-k_{\mathrm{bw}}c^{\beta})\,\alpha\cdot\check{C}\beta,$$

$$\widehat{b}_1(c) = \Lambda(k_{\mathrm{fw}}c^{\alpha}, k_{\mathrm{bw}}c^{\beta}) + \Upsilon_1(c)(\alpha-\beta)\cdot\nabla E(c) = \frac{1}{2}(k_{\mathrm{fw}}c^{\alpha}+k_{\mathrm{bw}}c^{\beta}).$$

It is interesting to see the cancellation in the term $\widehat{b}_1$, where Υ_1 did not have a sign, but after multiplication with $(\alpha-\beta)\cdot\nabla E(c)$ it becomes positive and increases the logarithmic mean $\Lambda(k_{\mathrm{fw}}c^{\alpha}, k_{\mathrm{bw}}c^{\beta})$ to the arithmetic mean $\frac{1}{2}(k_{\mathrm{fw}}c^{\alpha}+k_{\mathrm{bw}}c^{\beta})$. Moreover, the coefficient b_0 consists of two terms, the first

of which corresponds (up to order $1/V^2$) to the correction $\boldsymbol{A}_V$ in (5.4) arising from the improvement of $\boldsymbol{E}_V$, while the second term arises from improving the dissipation potential Φ_V^*, namely via Υ_0.

Putting these derivations together, summing over $r = 1, \ldots, R$ different reactions, and dropping all terms of order $1/V^2$, we find the following approximative Fokker–Planck equation:

$$\dot{\rho}(t, \boldsymbol{c}) = \mathrm{div}_{\boldsymbol{c}} \left(\frac{1}{V} \widehat{\mathbb{K}}_{\mathrm{CLE}}(\boldsymbol{c}) \nabla \rho(t, \boldsymbol{c}) + \rho(t, \boldsymbol{c}) \boldsymbol{R}(\boldsymbol{c}) + \frac{1}{V} \rho(t, \boldsymbol{c}) \boldsymbol{B}(\boldsymbol{c}) \right) \tag{5.7}$$

where $\boldsymbol{R}(\boldsymbol{c}) = \mathbb{K}(\boldsymbol{c}) \mathrm{D} E(\boldsymbol{c})$, $\boldsymbol{B}(\boldsymbol{c}) = \sum_{r=1}^R \widehat{b}_0^r(\boldsymbol{c})(\boldsymbol{\alpha}^r - \boldsymbol{\beta}^r)$, and $\widehat{\mathbb{K}}_{\mathrm{CLE}}$ is given in (1.6).

The big disadvantage of equation (5.7) is that it is generally no longer a gradient system. However, it may be considered as an equation with an asymptotic gradient flow structure in the sense of [BB*17]. To find the simplest true gradient system that is compatible with the Fokker–Planck equation (5.7), we have to find a true dual dissipation potential $\widehat{\Phi}_V^*$ that is non-negative and coincides with Φ_V^* from above to lowest order. To keep the notation light, we again explain the construction for the case of one reaction only and set $\Lambda_0(\boldsymbol{c}) = \Lambda(k_{\mathrm{fw}} \boldsymbol{c}^{\boldsymbol{\alpha}}, k_{\mathrm{bw}} \boldsymbol{c}^{\boldsymbol{\beta}})$. Our simplest choice is

$$\widehat{\Phi}_V^*(\rho, \xi) = \int_C \left(\Lambda_0(\boldsymbol{c}) \rho(\boldsymbol{c}) + \frac{1}{V} \Upsilon_0(\boldsymbol{c}) \rho(\boldsymbol{c}) + \frac{1}{V} \Upsilon_1(\boldsymbol{c}) \nabla \rho(\boldsymbol{c}) \cdot (\boldsymbol{\alpha} - \boldsymbol{\beta}) \right.$$
$$\left. + \frac{\Upsilon_2(\boldsymbol{c})}{V^2} \rho(\boldsymbol{c}) + \frac{\Upsilon_3(\boldsymbol{c})}{V^2} \frac{\left(\nabla \rho(\boldsymbol{c}) \cdot (\boldsymbol{\alpha} - \boldsymbol{\beta}) \right)^2}{\rho(\boldsymbol{c})} \right) \left(\nabla \xi(\boldsymbol{c}) \cdot (\boldsymbol{\alpha} - \boldsymbol{\beta}) \right)^2 \mathrm{d}\boldsymbol{c},$$

where the higher-order corrections $\Upsilon_2(\boldsymbol{c})$ and $\Upsilon_3(\boldsymbol{c})$ need to be chosen such that $\widehat{\Phi}_V^*(\rho, \xi)$ is still coercive. Choosing $\theta_1, \theta_2 \in \,]0, 1[$ with $\theta_1 < \theta_2$, we may require

$$\Lambda_0(\boldsymbol{c}) + \frac{\Upsilon_0(\boldsymbol{c})}{V} + \frac{\Upsilon_2(\boldsymbol{c})}{V^2} \geq \theta_2 \Lambda_0(\boldsymbol{c}) \quad \text{and} \quad 4\theta_1 \Lambda_0(\boldsymbol{c}) \Upsilon_3(\boldsymbol{c}) \geq \Upsilon_1(\boldsymbol{c})^2$$

for all $V > 1$, so that $\widehat{\Phi}_V^*(\rho, \xi) \geq (\theta_2 - \theta_1) \int_C \Lambda_0(\boldsymbol{c}) \rho(\boldsymbol{c}) \left(\nabla \xi(\boldsymbol{c}) \cdot (\boldsymbol{\alpha} - \boldsymbol{\beta}) \right)^2 \mathrm{d}\boldsymbol{c}$. The bounds for $\Upsilon_2(\boldsymbol{c})$ and $\Upsilon_3(\boldsymbol{c})$ hold for the following choices (or any bigger ones)

$$\Upsilon_2(\boldsymbol{c}) = \frac{\Lambda_0(\boldsymbol{c})}{16(1 - \theta_2)} \left(\boldsymbol{\alpha} \cdot \check{C} \boldsymbol{\beta} \right)^2 \quad \text{and} \quad \Upsilon_3(\boldsymbol{c}) = \frac{1}{4\theta_1 \Lambda_0(\boldsymbol{c})} \Upsilon_1(\boldsymbol{c})^2.$$

Of course, we fix the energy functional to be the improved entropy functional $\boldsymbol{E}_V$ from (5.3), and the gradient system $(\mathscr{P}(\boldsymbol{C}), \boldsymbol{E}_V, \widehat{\Phi}_V^*)$ has the associated gradient-flow equation $\dot{\varrho} = \mathrm{D}_\xi \widehat{\Phi}_V^*(\varrho, -\mathrm{D}\boldsymbol{E}_V(\varrho))$. With $\mathrm{D}\boldsymbol{E}_V(\varrho) = \frac{1}{V}(1 + \log \rho) + E + \frac{1}{V} E_1^V$ we find

$$\dot{\rho} = \mathrm{div} \left(\left[\widehat{a}_0^V(\boldsymbol{c}) \rho + \frac{\widehat{a}_1^V(\boldsymbol{c})}{V} \nabla_\gamma \rho + \frac{\widehat{a}_2^V(\boldsymbol{c})}{V^2} \frac{(\nabla_\gamma \rho)^2}{\rho} + \frac{\widehat{a}_3^V(\boldsymbol{c})}{V^3} \frac{(\nabla_\gamma \rho)^3}{\rho^2} \right] \gamma \right)$$
$$\text{with } \widehat{a}_0^V = \Lambda_\Upsilon^V \left(\nabla_\gamma E + \frac{1}{V} \nabla_\gamma E_V^1 \right), \quad \widehat{a}_1^V = \Lambda_\Upsilon^V + \Upsilon_1 \left(\nabla_\gamma E + \frac{1}{V} \nabla_\gamma E_V^1 \right),$$
$$\widehat{a}_2^V = \Upsilon_1 + \Upsilon_3 \left(\nabla_\gamma E + \frac{1}{V} \nabla_\gamma E_V^1 \right), \quad \text{and } \widehat{a}_3^V(\boldsymbol{c}) = \Upsilon_3, \tag{5.8}$$
$$\text{where } \Lambda_\Upsilon^V(\boldsymbol{c}) = \Lambda_0(\boldsymbol{c}) + \frac{\Upsilon_0(\boldsymbol{c})}{V} + \frac{\Upsilon_2(\boldsymbol{c})}{V^2}, \quad \gamma = \boldsymbol{\alpha} - \boldsymbol{\beta}, \quad \text{and } \nabla_\gamma f = \nabla f \cdot \gamma.$$

Because $\nabla_\gamma E_1^V$ is of order $1/V$, we see that this equation involves terms up to order $1/V^4$, namely through $\widehat{a}_0^V$ and through $\widehat{a}_2^V / V^2$.

Clearly, our gradient-flow equation (5.8) is much more complicated than those generated by the asymptotic gradient-flow structures in the sense of [BB*17], where higher order terms are simply dropped.

There is also the question of well-posedness for equation (5.8). To have parabolicity of the leading terms we need that the mapping $p \mapsto \frac{1}{V}\widehat{a}_1^V p + \frac{1}{V^2}\widehat{a}_2^V p^2 + \frac{1}{V^3}\widehat{a}_3^V p^3$ is monotone, which amounts to asking that $\widehat{a}_1^V + 2\widehat{a}_2^V q + 3\widehat{a}_3^V q^2 \geq 0$ for all $q \in \mathbb{R}$. This can be always be achieved by making Υ_2 very big while keeping Υ_3 constant, since Υ_2 only enters once via $\widehat{a}_1^V$.

5.4 Comparison of models

To appreciate the positive and negative aspects of the different approximations of the CME, we treat the simplest example, namely the linear RRE on $\boldsymbol{C} = [0, \infty[$:

$$\dot{c} = 1 - c \qquad \text{corresponding to the reaction pair } X \overset{1}{\underset{1}{\rightleftharpoons}} \emptyset. \tag{5.9}$$

Obviously, we have the explicit solution $c(t) = 1 + (c(0)-1)\mathrm{e}^{-t}$.

The associated CME for $\boldsymbol{u} = \mathscr{P}(\mathbb{N}_0)$ is given by

$$\dot{u}_n = V u_{n-1} - (V+n)u_n + (n+1)u_{n+1} \quad \text{for } n \in \mathbb{N}_0, \tag{5.10}$$

where $u_{-1} = 0$. Using the linearity in (5.9), which leads to the linearity in n of the coefficients in (5.10), we obtain explicit closed form relations of the evolution of the rescaled expectation $\widehat{e}(t) := \frac{1}{V}\sum_{n\in\mathbb{N}_0} n u_n(t)$ and variance $\widehat{v}(t) := \frac{1}{V^2}\sum_{n\in\mathbb{N}_0} n^2 u_n - \widehat{e}(t)^2$, namely

$$\dot{\widehat{e}}(t) = 1 - \widehat{e}(t) \quad \text{and} \quad \dot{\widehat{v}}(t) = -2\widehat{v}(t) + \frac{1 + \widehat{e}(t)}{V}. \tag{5.11}$$

Moreover, it can be easily checked that for any solution $t \mapsto c(t)$ of the RRE (5.9) the following formula provides the explicit solution of the CME (5.10):

$$u_n(t) = \frac{\mathrm{e}^{-c(t)V}}{n!}\big(c(t)V\big)^n \quad \text{for } n \in \mathbb{N}_0. \tag{5.12}$$

Note that this is expression is compatible with the ODEs (5.11) for the moments, since for these Poisson distributions we have $\widehat{e}(t) = c(t)$ and $\widehat{v}(t) = c(t)/V$.

The Liouville equation and the simple Fokker–Planck equation read

$$\text{(Lio)} \quad \dot{\varrho} = \partial_c\big((c-1)\varrho\big) \quad \text{and} \quad \text{(FP)} \quad \dot{\rho} = \partial_c\Big(\Lambda(1,c)\frac{\partial_c\rho}{V} + (c-1)\rho\Big).$$

The Fokker–Planck equation for the chemical Langevin equation (cf. (5.6)) takes the form

$$\text{(FP}_{\text{CLE}}) \quad \dot{\rho} = \partial_c^2\Big(\frac{1+c}{2V}\rho\Big) + \partial_c\big((c-1)\rho\big).$$

To compare the solutions of (FP) and (FP$_{\text{CLE}}$) with the true solutions of the CME (5.10), we assume that the solutions can be approximated by Gaußians. In general, for multidimensional Fokker–Planck equations of the form $\dot{\rho} = \frac{1}{V}\sum_{ij}\partial_{ij}^2\big(\rho\mathbb{M}_{ij}(\boldsymbol{c})\big) + \text{div}\big(\rho\boldsymbol{R}_V\big)$ the ansatz $\rho(t,\cdot) \sim \mathrm{N}(\boldsymbol{a}(t), \frac{1}{V}\mathbb{A}(t))$ with $\boldsymbol{a}(t) \in \mathbb{R}^d$ and $\mathbb{A}(t) \in \mathbb{R}^{d\times d}_{\text{spd}}$ leads to the necessary conditions

$$\dot{\boldsymbol{a}}(t) = -\boldsymbol{R}(\boldsymbol{a}(t)) \quad \text{and} \quad \dot{\mathbb{A}}(t) = -\mathrm{D}\boldsymbol{R}(\boldsymbol{a}(t))\mathbb{A}(t) - \mathbb{A}(t)\mathrm{D}\boldsymbol{R}(\boldsymbol{a}(t))^{\mathsf{T}} + 2\mathbb{M}(\boldsymbol{a}(t)),$$

see [SaS17] for rigorous results of this type. Applying these formulas to (FP$_{\text{CLE}}$) we obtain

$$\dot{a} = 1 - a \quad \text{and} \quad \dot{A} = -2A + 1 + a,\tag{5.13}$$

hence the ODEs for a and A/V coincide with those for $\widehat{e}$ and $\widehat{v}$ in (5.11).

A similar argument indicates that solutions to (FP) are well approximated by Gaußians with mean a_V and variance A_V satisfying

$$\dot{a}_V = 1 - a_V + \frac{1}{V}\partial_2\Lambda(1, a_V) \quad \text{and} \quad \dot{A}_V = -2A_V + 2\Lambda(1, a_V).\tag{5.14}$$

On the one hand, this clearly indicates that (FP$_{\text{CLE}}$) provides a better approximation to the CME for $t \in [0, T]$. By formally passing to the limit $V \to \infty$ in (5.14), we see that the ODE for a_V is asymptotically correct. This is not the case for the ODE for A_V, since the arithmetic mean in (5.13) is replaced by the logarithmic mean in (5.14). However, the error of $\Lambda(1, c)$ compared to $\frac{1}{2}(1+c)$ is less than 10 % for $c \in [1/3, 3]$ and it converges to 0 for $c \to 1$, i.e., in the limit $t \to \infty$. Equations (5.13) are consistent with Kurtz' central limit theorem, which asserts that the normalized process $\frac{1}{V}\boldsymbol{N}^V(t)$ has fluctuations around $\boldsymbol{c}(t)$ of order $1/\sqrt{V}$, and the rescaled process $\sqrt{V}\left(\frac{1}{V}\boldsymbol{N}^V(t) - \boldsymbol{c}(t)\right)$ converges to a Gaußian process $t \mapsto \boldsymbol{V}(t)$ with covariance matrix $\mathbb{A}$ satisfying $\dot{\mathbb{A}}(t) = -\mathrm{D}\boldsymbol{R}(\boldsymbol{c}(t))\mathbb{A}(t) - \mathbb{A}(t)\mathrm{D}\boldsymbol{R}(\boldsymbol{c}(t))^{\mathsf{T}} + 2\widehat{\mathbb{K}}_{\text{CLE}}(\boldsymbol{c}(t))$, see, e.g., [Kur78, Eqn. (1.9)].

On the other hand, (FP) makes a better prediction for the equilibrium distribution that is attained for $t \to \infty$. For (FP$_{\text{CLE}}$) we have the unique steady state

$$\rho_V^{\text{eq,CLE}}(c) = \frac{1}{Z_V^{\text{CLE}}}\,\mathrm{e}^{-V\widetilde{E}(c)} \quad \text{with} \quad \widetilde{E}(c) = \int_1^c \frac{2b-2+1/V}{b+1}\,\mathrm{d}b = 2c-2-\left(4-\frac{1}{V}\right)\log\frac{1+c}{2}.$$

Thus, $\widetilde{E}$ grows only like c, such that $\rho_V^{\text{eq,CLE}}$ decays exponentially only. In contrast, the equilibrium $\rho_V^{\text{eq,FP}} = Z_V^{-1}\mathrm{e}^{-VE(c)}$ of (FP) produces the correct super-exponential decay of the stationary Poisson distribution equation for the CME (5.10).

5.5 Approximation via cosh-type gradient structure

The derivation of a gradient structure (4.3) for the Liouville equation (4.1) can be repeated very similarly by starting from the cosh-type gradient structure introduced in [MPR14], see Proposition 3.8. We do not give the details here but provide the result only.

Starting from the cosh-type dual dissipation potential $\boldsymbol{\Psi}^*_{\cosh,V}$ defined in (3.8) instead of the quadratic dual potential Ψ^*_V defined in (4.5) we obtain the counterparts to Propositions 4.5 and 4.6 but now with

$$\boldsymbol{\Psi}^*_{\cosh,\text{Lio}}(\varrho, \xi) := \int_{\boldsymbol{c}\in\boldsymbol{C}}\sum_{r=1}^R \kappa^r\left(\frac{\boldsymbol{c}^{\boldsymbol{\alpha}^r}}{\boldsymbol{c}^{\boldsymbol{\alpha}^r}_*}\frac{\boldsymbol{c}^{\boldsymbol{\beta}^r}}{\boldsymbol{c}^{\boldsymbol{\beta}^r}_*}\right)^{1/2}\mathsf{C}^*\big((\boldsymbol{\beta}^r - \boldsymbol{\alpha}^r)\cdot\nabla_{\boldsymbol{c}}\xi(\boldsymbol{c})\big)\mathrm{d}\varrho(\boldsymbol{c}).$$

Without any need to justify the approximation procedure in the sense of Section 4.2 we easily obtain the following result.

Proposition 5.1 (cosh-type gradient structure for the Liouville equation) *The Liouville equation* (4.1) *has the gradient structure* $(\mathscr{P}(\boldsymbol{C}), \boldsymbol{E}, \boldsymbol{\Psi}^*_{\cosh,\text{Lio}})$ *with* $\boldsymbol{E}$ *from* (4.2) *and* $\boldsymbol{\Psi}^*_{\cosh,\text{Lio}}$ *from above.*

Proof. The result follows by using $\mathrm{D}\mathcal{E}(\varrho)(\cdot) = E(\cdot)$, $\nabla_{\boldsymbol{c}}E(\boldsymbol{c}) = \big(\log(c_i/c_i^*)\big)_{i=1,\dots,I}$, and

$$\mathrm{D}_\xi\boldsymbol{\Psi}^*_{\mathrm{cosh,Lio}}\big(\varrho, -\mathrm{D}\mathcal{E}(\varrho)\big)[\eta] = \int_{\boldsymbol{C}}\sum_{r=1}^{R}\kappa^r\Big(\frac{\boldsymbol{c}^{\boldsymbol{\alpha}^r}}{\boldsymbol{c}^{\boldsymbol{\alpha}^r}_*}\frac{\boldsymbol{c}^{\boldsymbol{\beta}^r}}{\boldsymbol{c}^{\boldsymbol{\beta}^r}_*}\Big)^{1/2}(\mathsf{C}^*)'\big(-\boldsymbol{\gamma}^r\cdot\big(\log\frac{c_i}{c_i^*}\big)_i\big)\big[\boldsymbol{\gamma}^r\cdot\nabla\eta\big]\,\mathrm{d}\varrho(\boldsymbol{c}),$$

where $\boldsymbol{\gamma}^r = \boldsymbol{\alpha}^r - \boldsymbol{\beta}^r$. Using $\sqrt{ab}\,(\mathsf{C}^*)'\big(\log(a/b)\big) = a - b$ and the definition of $\boldsymbol{R}$ gives $\mathrm{D}_\xi\boldsymbol{\Psi}^*_{\mathrm{cosh,Lio}}\big(\varrho, -\mathrm{D}\mathcal{E}(\varrho)\big)[\eta] = -\int_{\boldsymbol{C}}\boldsymbol{R}(\boldsymbol{c})\cdot\nabla\eta\,\mathrm{d}\varrho(\boldsymbol{c})$ which is the desired right-hand side of (4.1) when testing with η and integrating by parts. ∎

As in the case of quadratic gradient structure for the Liouville equation we may consider the first-order correction to obtain a Fokker–Planck equation. For this we insert the improved energy $\boldsymbol{E}_V$ defined in (5.3) into the dissipation potential $\boldsymbol{\Psi}^*_{\mathrm{cosh},V}$ (cf. (3.8)) to obtain a quasilinear Fokker–Planck-type equation, namely $\dot\varrho = \mathrm{D}_\xi\boldsymbol{\Psi}^*_{\mathrm{cosh},V}\big(\varrho, -\mathrm{D}\boldsymbol{E}_V(\varrho)\big)$. Using the abbreviations $a_r := \frac{\boldsymbol{c}^{\boldsymbol{\alpha}^r}}{\boldsymbol{c}^{\boldsymbol{\alpha}^r}_*}$ and $b_r := \frac{\boldsymbol{c}^{\boldsymbol{\beta}^r}}{\boldsymbol{c}^{\boldsymbol{\beta}^r}_*}$ we find (note $\nabla_{\boldsymbol{c}}E_1^V(\boldsymbol{c}) = O(1/V)$)

$$\begin{aligned}
\mathrm{D}_\xi\boldsymbol{\Psi}^*_{\mathrm{cosh},V}\big(\varrho, -\mathrm{D}\boldsymbol{E}_V(\varrho)\big) &= \mathrm{D}_\xi\boldsymbol{\Psi}^*_{\mathrm{cosh},V}\Big(\varrho, -\frac{1}{V}\log\rho - E - \frac{1}{V}E_1^V\Big) \\
&= \mathrm{div}\Big(\rho\sum_{r=1}^{R}\kappa^r\sqrt{a_rb_r}\Big[(\mathsf{C}^*)'\big(\log\frac{b_r}{a_r}\big)\boldsymbol{\gamma}^r + (\mathsf{C}^*)''\big(\log\frac{b_r}{a_r}\big)\frac{\boldsymbol{\gamma}^r\cdot\nabla\rho}{V\rho}\boldsymbol{\gamma}^r\Big] + O(1/V^2)\Big).
\end{aligned}$$

Using the identities $\sqrt{ab}\,(\mathsf{C}^*)'\big(\log(b/a)\big) = b - a$ and $\sqrt{ab}\,(\mathsf{C}^*)''\big(\log(b/a)\big) = (a+b)/2$ the FP equation has the expansion

$$\dot\rho(t,\boldsymbol{c}) = \mathrm{div}_{\boldsymbol{c}}\Big(\rho(t,\boldsymbol{c})\boldsymbol{R}(\boldsymbol{c}) + \frac{1}{V}\widehat{\mathbb{K}}_{\mathrm{CLE}}(\boldsymbol{c})\nabla_{\boldsymbol{c}}\rho(t,\boldsymbol{c}) + O(1/V^2)_{V\to\infty}\Big)$$

where $\widehat{\mathbb{K}}_{\mathrm{CLE}}$ is exactly the same as obtained in (1.6) by a completely different approach.

6 Hybrid models

We show in this section how the different gradient structures for RRE, for CME, and for the FPE can be combined to obtain hybrid models, which are combinations of several models depending on the desired accuracy. The importance here is to use the proper rescalings in terms of the volume V to make the different descriptions compatible. We do not consider a full theory, but highlight first the general strategy of model reduction for gradient systems in Section 6.1 and then illustrate this by a simple example in Section 6.2. A nontrivial case of a rigorous coarse graining in this spirit is given in [MiS19], where a linear RRE with a small parameter ε is considered. The elimination of the fast relaxations in the time scale ε leads to a coarse-grained gradient system.

In Section 6.3 we discuss the general coupling of the FPE to a RRE and the similar coupling of the CME to a RRE, both leading to so-called mean-field equations, where a linear equation for a probability density is nonlinearly coupled to an ODE. Finally, we discuss the mixed discrete and continuous description, where the CME is used for small numbers of particles and the FPE is used for larger numbers.

6.1 Coarse graining for gradient systems

If a gradient system (X, E_X, Ψ_X) is more complicated than what is needed, one is interested in approximating the system by a simpler model that still contains the most important features. We explain how this can be done while keeping the gradient structure.

We assume that the relevant states $x \in X$ can be described by states $y \in Y$ and that there is a reconstruction mapping $x = \Phi(y)$, i.e., $\Phi(Y)$ is a subset (or submanifold) of X. We now pull back the gradient structure (X, E_X, Ψ_X) to an approximative gradient structure (Y, E_Y, Ψ_Y). The natural approach is to restrict the energy functional and the (primal) dissipation potential as follows:

$$E_Y(y) = E_X(\Phi(y)) \quad \text{and} \quad \Psi_Y(y, \dot{y}) := \Psi_X(\Phi(y), D\Phi(y)\dot{y}). \tag{6.1}$$

The solutions $y : [0, T] \to Y$ of the coarse-grained gradient system (Y, E_Y, Ψ_Y) will provide good approximations $\widehat{x} : t \mapsto \Phi(y(t)) \in X$ of the true solutions of the full GS (X, E_X, Ψ_X), if the set $\Phi(Y)$ approximates a flow-invariant subset of X.

In reaction systems, the primal dissipation potential Ψ_X is usually not known explicitly. Hence, it is desirable to have a method for reducing the dual dissipation potential Ψ_X^* directly to Ψ_Y^*, in the case where $A = D\Phi(y) : Y \to X$ is injective but its adjoint mapping $A^* : X^* \to Y^*$ has a large kernel. The following exact result will be the motivation for our modeling approximations in the subsequent subsections.

Proposition 6.1 *Consider reflexive Banach spaces* X *and* Y *and a real-valued dissipation potential* $\Psi : X \to [0, \infty[$ *(i.e. lower semicontinuous, convex, and* $\Psi(0) = 0$*) that is superlinear, i.e.* $\Psi(v)/\|v\|_X \to \infty$ *for* $\|v\|_X \to \infty$*. Assume that the bounded linear operator* $A : Y \to X$ *has closed range. Then the dissipation potential* $\widetilde{\Psi} : Y \to [0, \infty[; y \mapsto \Psi(Ay)$ *satisfies*

$$\widetilde{\Psi}^*(\eta) = \inf \big\{ \Psi^*(\xi) \,\big|\, A^*\xi = \eta \big\} \quad \text{for all } \eta \in Y^*, \tag{6.2}$$

where we use the convention $\inf \emptyset = \infty$*.*

Proof. For the proof we use the saddle-point theory in [EkT76, Ch. VI.2].

Fix $\eta \in Y^*$ and assume first that $\eta \notin \mathrm{Ran}(A^*)$. Since $\mathrm{Ran}(A) \subset X$ is closed, the Closed Range Theorem yields that $\mathrm{Ran}(A^*) \subset Y^*$ is closed as well, and $\mathrm{Ran}(A^*) = \mathrm{Ker}(A)^\perp$. Consequently, there exists $\tilde{y} \in \mathrm{Ker}(A)$ such that $\langle \eta, \tilde{y} \rangle \neq 0$, and we obtain

$$\widetilde{\Psi}^*(\eta) = \sup_{y \in Y} \big(\langle \eta, y \rangle - \Psi(Ay) \big) \geq \sup_{\lambda \in \mathbb{R}} \big(\lambda \langle \eta, \tilde{y} \rangle - \Psi(\lambda A\tilde{y}) \big) = \infty.$$

This yields (6.2), since the right-hand side is clearly infinite as well.

Fix now $\eta \in \mathrm{Ran}(A^*)$ and define the Lagrangian function $L : X \times X^* \to [-\infty, \infty[$ via

$$L(x, \xi) = -\langle \xi, x \rangle + \Psi(x) - \chi^*(\xi) \quad \text{with} \quad \chi^*(\xi) = \begin{cases} 0 & \text{for } A^*\xi = \eta, \\ \infty & \text{otherwise.} \end{cases}$$

For notational convenience we set

$$h(x) = \sup_{\xi \in X^*} L(x, \xi), \quad g(\xi) = \inf_{x \in X} L(x, \xi), \quad P := \inf_X h, \quad D := \sup_{X^*} g.$$

Classical duality theory yields the trivial inequality $P \geq D$. Clearly, $L(\cdot, \xi)$ is convex and lower semicontinuous, whereas $L(\mathsf{x}, \cdot)$ is concave and upper semicontinuous, since the boundedness of A^* implies that $\{\, \xi \in \mathsf{X}^* \mid A^*\xi = \eta \,\}$ is closed.

Using $\eta \in \mathrm{Ran}(A^*)$, we find $\xi_\eta \in \mathsf{X}^*$ with $A^*\xi_\eta = \eta$, so that our assumptions guarantee the coercivity of $\mathsf{x} \mapsto L(\mathsf{x}, \xi_\eta) \in \mathbb{R}$. Hence, we can apply [EkT76, Chap. VI, Prop. 2.3], which shows that there is no duality gap:

$$P = \inf_{\mathsf{x} \in \mathsf{X}} h(\mathsf{x}) = \min_{\mathsf{x} \in \mathsf{X}} \left(\sup_{\xi \in \mathsf{X}^*} L(\mathsf{x}, \xi) \right) = \sup_{\xi \in \mathsf{X}^*} \left(\inf_{\mathsf{x} \in \mathsf{X}} L(\mathsf{x}, \xi) \right) = \sup_{\xi \in \mathsf{X}^*} g(\xi) = D. \qquad (6.3)$$

We relate P and D with the two sides in our desired formula (6.2). On the one hand,

$$h(\mathsf{x}) = \sup_{\xi \in \mathsf{X}^*} L(\mathsf{x}, \xi) = \Psi(\mathsf{x}) + \sup_{\xi \in \mathsf{X}^*} \left(\langle \xi, -\mathsf{x} \rangle - \chi^*(\xi) \right)$$

$$= \Psi(\mathsf{x}) - \langle \xi_\eta, \mathsf{x} \rangle + \mu(-\mathsf{x}) \text{ with } \mu(\mathsf{x}) := \sup_{\zeta \in \mathsf{X}^*} \left(\langle \zeta, \mathsf{x} \rangle - \delta_0(A^*\zeta) \right),$$

where in the last step we have substituted $\xi = \xi_\eta + \zeta$ with $A^*\xi_\eta = \eta$ and introduced $\delta_0(\widetilde{\eta}) = 0$ for $\widetilde{\eta} = 0$ and ∞ otherwise. Thus, we conclude

$$h(\mathsf{x}) = \Psi(A\mathsf{y}) - \langle \eta, \mathsf{y} \rangle \text{ for } \mathsf{x} = A\mathsf{y} \quad \text{and} \quad h(\mathsf{x}) = \infty \text{ for } \mathsf{x} \notin \mathrm{Ran}(A).$$

Thus, taking the minimum over all of X is the same as taking it over $\mathrm{Ran}(A)$, namely

$$P = \inf_{\mathsf{x} \in \mathsf{X}} h(\mathsf{x}) = \inf_{\mathsf{y} \in \mathsf{Y}} \left(\Psi(A\mathsf{y}) - \langle \eta, \mathsf{y} \rangle \right) = -\widetilde{\Psi}^*(\eta).$$

On the other hand, the definition of $g(\xi) = \inf_{\mathsf{x} \in \mathsf{X}} L(\mathsf{x}, \xi)$ immediately gives $g(\xi) = -\Psi^*(\xi) - \chi^*(\xi)$. Hence, we arrive at

$$D = \sup_{\xi \in \mathsf{X}^*} g(\xi) = - \inf_{\xi \in \mathsf{X}^*} \left(\Psi^*(\xi) + \chi^*(\xi) \right) = - \inf \left\{\, \Psi^*(\xi) \mid A^*\xi = \eta \,\right\}.$$

As a result, formula (6.2) follows from $P = D$. ∎

In our applications below (as well as in most others) the explicit minimization in (6.2) is too complicated to be executed. However, as the coarse-graining mapping through Φ is usually only an approximation, it may suffice to approximate the minimizers suitably. In general, one has to find an approximation $\xi = \mathsf{M}(\mathsf{y}, \eta) \in \mathsf{X}^*$ and sets

$$\Psi_\mathsf{Y}^*(\mathsf{y}, \eta) = \Psi_\mathsf{X}^*(\Phi(\mathsf{y}), \mathsf{M}(\mathsf{y}, \eta)) \quad \text{or} \quad \tfrac{1}{2}\langle \eta, \widetilde{\mathsf{K}}_\mathsf{Y}(\mathsf{y})\eta \rangle = \langle \mathsf{L}(\mathsf{y})\eta, \mathsf{K}_\mathsf{X}(\Phi(\mathsf{y}))\mathsf{L}(\mathsf{y})\eta \rangle, \qquad (6.4)$$

where $\mathsf{L}(\mathsf{y}) : \mathsf{Y}^* \to \mathsf{X}^*$ is the linear version of M. Of course, when constructing M or L one should keep (6.2) in mind to preserve all interesting properties inherited by the coarse-graining process.

6.2 A simple example: from CME to RRE

We apply the above idea with $(\mathsf{X}, \mathsf{E}_\mathsf{X}, \Psi_\mathsf{X}^*)$ being $(\mathscr{P}(\mathcal{N}), \mathcal{E}_V, \mathcal{K}_V)$ and with $(\mathsf{Y}, \mathsf{E}_\mathsf{Y}, \Psi_\mathsf{Y})$ being $(\boldsymbol{C}, E, \mathbb{K})$. The embedding mapping $\Phi_V : \boldsymbol{C} \to \mathscr{P}(\mathcal{N})$ is given by the Poisson distributions

$$\Phi_V(\boldsymbol{c}) := \left(e^{-V|\boldsymbol{c}|_1} \frac{(V\boldsymbol{c})^{\boldsymbol{n}}}{\boldsymbol{n}!} \right)_{\boldsymbol{n} \in \mathcal{N}}, \quad \text{where } |\boldsymbol{c}|_1 = \sum_1^I c_i.$$

In the simple example $\dot{c} = 1 - c$ treated in Section 5.4 the image of Φ_V defines an exactly invariant submanifold, but this is no longer true for nonlinear equations or systems. Nevertheless our construction provides the surprising identity

$$\mathsf{E}_\mathsf{Y}(\boldsymbol{c}) = \mathcal{E}_V(\Phi_V(\boldsymbol{c})) = E(\boldsymbol{c}),$$

with the old E defined in (2.9) which is independent of V.

To reduce the dual dissipation potential Ψ_X defined via $\mathcal{K}_V$ we use the derivative

$$\mathrm{D}\Phi_V(\boldsymbol{c})\boldsymbol{w} = \left(\mathrm{e}^{-V|\boldsymbol{c}|_1} \frac{(V\boldsymbol{c})^{\boldsymbol{n}}}{\boldsymbol{n}!} \sum_{i=1}^{I} \left(\frac{n_i}{c_i} - V\right) w_i\right)_{\boldsymbol{n}\in\mathcal{N}}.$$

Thus, the adjoint operator $\mathrm{D}\Phi_V(\boldsymbol{c})^*$ maps $\boldsymbol{\mu} = (\mu_{\boldsymbol{n}})$ to $\boldsymbol{\zeta} = (\zeta_i)_{i=1,\dots,I}$ via

$$\boldsymbol{\mu} \mapsto \boldsymbol{\zeta} = \mathrm{D}\Phi_V(\boldsymbol{c})^*\boldsymbol{\mu} = \left(\sum_{\boldsymbol{n}\in\mathcal{N}} \mathrm{e}^{-V|\boldsymbol{c}|_1} \frac{(V\boldsymbol{c})^{\boldsymbol{n}}}{\boldsymbol{n}!} \left(\frac{n_i}{c_i} - V\right) \mu_{\boldsymbol{n}}\right)_{i=1,\dots,I}.$$

In general, one is not able to solve the minimization problem (6.2) that produces Ψ_Y^* from Ψ_X^*, so instead we construct a linear mapping $\boldsymbol{\zeta} \mapsto \widetilde{\boldsymbol{\mu}} = \mathsf{M}_V(\boldsymbol{c})\boldsymbol{\zeta}$ that approximates the minimizer for $V \to \infty$ and satisfies $\boldsymbol{\zeta} = \mathrm{D}\Phi_V(\boldsymbol{c})^*\mathsf{M}_V(\boldsymbol{c})\boldsymbol{\zeta}$. Indeed, we search for $\widetilde{\boldsymbol{\mu}}$ in the linear form $\widetilde{\boldsymbol{\mu}}_{\boldsymbol{n}}^{\boldsymbol{a}} = \boldsymbol{a}\cdot\boldsymbol{n}$ for $\boldsymbol{n} \in \mathcal{N}$ and obtain

$$\mathrm{D}\Phi_V(\boldsymbol{c})^*\widetilde{\boldsymbol{\mu}}^{\boldsymbol{a}} = \mathrm{D}\Phi_V(\boldsymbol{c})^*(\boldsymbol{a}\cdot\boldsymbol{n})_{\boldsymbol{n}\in\mathcal{N}} = \left(\sum_{\boldsymbol{n}\in\mathcal{N}} \mathrm{e}^{-V|\boldsymbol{c}|_1} \frac{(V\boldsymbol{c})^{\boldsymbol{n}}}{\boldsymbol{n}!} \left(\frac{n_i}{c_i} - V\right) \sum_{j=1}^{I} a_j n_j\right)_{i=1,\dots,I}$$

$$= \left(\sum_{j=1}^{I} \sum_{\boldsymbol{n}\in\mathcal{N}} \mathrm{e}^{-V|\boldsymbol{c}|_1} \frac{(V\boldsymbol{c})^{\boldsymbol{n}}}{\boldsymbol{n}!} \left(\frac{n_i a_j n_j}{c_i} - V a_j n_j\right)\right)_{i=1,\dots,I}$$

$$= \left(\sum_{j=1}^{I} \left(V^2 \frac{c_i a_j c_j}{c_i} + \delta_{ij} V a_i - V^2 a_j c_j\right)\right)_{i=1,\dots,I} = V\boldsymbol{a},$$

where we used the identities $\sum_{\boldsymbol{n}\in\mathcal{N}} \mathrm{e}^{-V|\boldsymbol{c}|_1} \frac{(V\boldsymbol{c})^{\boldsymbol{n}}}{\boldsymbol{n}!} n_i = V c_i$ and $\sum_{\boldsymbol{n}\in\mathcal{N}} \mathrm{e}^{-V|\boldsymbol{c}|_1} \frac{(V\boldsymbol{c})^{\boldsymbol{n}}}{\boldsymbol{n}!} n_i n_j = V^2 c_i c_j + \delta_{ij} V c_i$. Thus, we choose the simple operator M_V of the form

$$\boldsymbol{\zeta} \mapsto \boldsymbol{\mu} = \mathsf{M}_V(\boldsymbol{c})\boldsymbol{\zeta} = \left(\frac{1}{V}\boldsymbol{\zeta}\cdot\boldsymbol{n}\right)_{\boldsymbol{n}\in\mathcal{N}}.$$

For inserting $\mu = \mathsf{M}_V(\boldsymbol{c})\boldsymbol{\zeta}$ and $\boldsymbol{u} = \Phi_V(\boldsymbol{c})$ into the full dual dissipation potential Ψ_X^*, we use the form (4.5b) and the relations $(\mathsf{M}_V(\boldsymbol{c})\boldsymbol{\zeta})_{\boldsymbol{n}+\boldsymbol{\alpha}} - (\mathsf{M}_V(\boldsymbol{c})\boldsymbol{\zeta})_{\boldsymbol{n}+\boldsymbol{\beta}} = \frac{1}{V}\boldsymbol{\zeta}\cdot(\boldsymbol{\alpha}-\boldsymbol{\beta})$ and

$$\mathbb{B}_V^{\boldsymbol{\beta}}(\boldsymbol{n}) \left(\Phi_V(\boldsymbol{c})\right)_{\boldsymbol{n}+\boldsymbol{\beta}} = \frac{V(\boldsymbol{n}+\boldsymbol{\beta})!}{V^{|\boldsymbol{\beta}|}\boldsymbol{n}!} \mathrm{e}^{-V|\boldsymbol{c}|_1} \frac{(V\boldsymbol{c})^{\boldsymbol{n}+\boldsymbol{\beta}}}{(\boldsymbol{n}+\boldsymbol{\beta})!} = V\mathrm{e}^{-V|\boldsymbol{c}|_1} \frac{(V\boldsymbol{c})^{\boldsymbol{n}}}{\boldsymbol{n}!} \boldsymbol{c}^{\boldsymbol{\beta}}.$$

With this, we find an approximation of the reduced dual dissipation potential Ψ_Y^*, namely

$$\Psi_\mathsf{Y}^*(\boldsymbol{c}, \boldsymbol{\zeta}) := \Psi_\mathsf{X}^*\left(\Phi_V(\boldsymbol{c}), \mathsf{M}_V(\boldsymbol{c})\boldsymbol{\zeta}\right)$$

$$= \frac{V}{2} \sum_{r=1}^{R} \sum_{\boldsymbol{n}\in\mathcal{N}} V\mathrm{e}^{-V|\boldsymbol{c}|_1} \frac{(V\boldsymbol{c})^{\boldsymbol{n}}}{\boldsymbol{n}!} \Lambda\left(k_{\mathrm{fw}}^r \boldsymbol{c}^{\boldsymbol{\alpha}^r}, k_{\mathrm{bw}}^r \boldsymbol{c}^{\boldsymbol{\beta}^r}\right) \left(\frac{(\boldsymbol{\beta}^r - \boldsymbol{\alpha}^r)\cdot\boldsymbol{\zeta}}{V}\right)^2$$

$$= \frac{1}{2} \sum_{r=1}^{R} \Lambda\left(k_{\mathrm{fw}}^r \boldsymbol{c}^{\boldsymbol{\alpha}^r}, k_{\mathrm{bw}}^r \boldsymbol{c}^{\boldsymbol{\beta}^r}\right) \left((\boldsymbol{\beta}^r - \boldsymbol{\alpha}^r)\cdot\boldsymbol{\zeta}\right)^2 = \frac{1}{2}\boldsymbol{\zeta}\cdot\mathbb{K}(\boldsymbol{c})\boldsymbol{\zeta}.$$

Thus, the gradient system $(\mathsf{Y}, \mathsf{E}_\mathsf{Y}, \Psi_\mathsf{Y}^*)$ obtained by the abstract reduction procedure is exactly given by $(\boldsymbol{C}, E, \mathbb{K})$, which is the gradient system for the RRE (2.2) studied in Theorem 2.2.

6.3 Coupling a RRE to a Fokker–Planck equation

In many applications one is interested in the microscopic description of some variables c_j, while other variables c_i can be described more macroscopically. We first start from the simplified FPE (5.5) as the gradient system $(\mathscr{P}(C), \widetilde{E}_V, K)$ and partition the components of c into *stochastic* and *macroscopic* parts, c_{s} and c_{m} respectively, via

$$c = (c_1, \ldots, c_J, c_{J+1}, \ldots, c_I) = (c_{\mathrm{s}}, c_{\mathrm{m}}) \quad \text{with}$$

$$c_{\mathrm{s}} := (c_1, \ldots, c_J) \in C_{\mathrm{s}} := [0, \infty[^J \text{ and } c_{\mathrm{m}} := (c_{J+1}, \ldots, c_I) \in C_{\mathrm{m}} := [0, \infty[^{I-J},$$

In the notation of Section 6.1 we let $\mathsf{X} = \mathscr{P}(C_{\mathrm{s}} \times C_{\mathrm{m}})$ and $\mathsf{Y} = \mathscr{P}(C_{\mathrm{s}}) \times C_{\mathrm{m}}$.

For the mapping $\Phi : \mathsf{Y} \to \mathsf{X}$ we choose the product ansatz

$$\Phi_V(\varrho_{\mathrm{s}}, \widehat{c}_{\mathrm{m}})(\mathrm{d}c_1, \ldots, \mathrm{d}c_I) := \varrho_{\mathrm{s}}(\mathrm{d}c_{\mathrm{s}}) \prod_{j=J+1}^{I} \mathsf{W}(c_j; \widehat{c}_j, V) \, \mathrm{d}c_j,$$

where the probability densities $\mathsf{W}(\cdot; \widehat{a}, V)$ are given as follows:

$$\mathsf{W}(a; \widehat{a}, V) := \frac{1}{\widehat{a} \, Z(V\widehat{a})} \exp\left(-V\widehat{a} \, \lambda_{\mathrm{B}}(a/\widehat{a})\right) \text{ with } Z(v) := \int_0^\infty \exp\left(-v \, \lambda_{\mathrm{B}}(z)\right) \mathrm{d}z.$$

According to Section 6.1 the functional $\widetilde{\mathsf{E}}_{\mathsf{Y}}^V = \widetilde{E}_V \circ \Phi_V$ on Y is then given by

$$\widetilde{\mathsf{E}}_{\mathsf{Y}}^V(\varrho_s, \widehat{c}_{\mathrm{m}}) = \int_{C_{\mathrm{s}}} \left(\frac{1}{V}\rho_s(c_s)\log\rho_s(c_s) + \rho_s(c_s)E_{\mathrm{s}}(c_s)\right)\mathrm{d}c_s + \frac{\widetilde{Z}(V)}{V} + \sum_{j=J+1}^{I} \widehat{e}_V(\widehat{c}_j, c_j^*)$$

$$\text{where } \widehat{e}_V(\widehat{a}, a^*) := A(V\widehat{a})\widehat{a}\log\left(\frac{\widehat{a}}{a^*}\right) - \widehat{a} + a^* - \frac{\log\left(\widehat{a}Z(V\widehat{a})\right)}{V}$$

with $E_{\mathrm{s}}(c_{\mathrm{s}}) = \sum_{i=1}^{J} c_i^*\lambda_{\mathrm{B}}(c_i/c_i^*)$ and $A(v) = \int_0^\infty z\exp\left(-v\lambda_{\mathrm{B}}(z)\right)\mathrm{d}z/Z(v)$.

It can be shown that $A(v) \geq 1$ and $e_V(\widehat{a}, a^*) \geq a^*\lambda_{\mathrm{B}}(\widehat{a}/a^*)$ for all V, and for $V \to \infty$ we obtain $e_V(\widehat{a}, a^*) \to a^*\lambda_{\mathrm{B}}(\widehat{a}/a^*)$. To simplify the model we are therefore allowed to replace the last term in $\widetilde{\mathsf{E}}_{\mathsf{Y}}^V$ by the relative entropy $E_{\mathrm{m}}(\widehat{c}_{\mathrm{m}}) = \sum_{j=J+1}^{I} c_j^*\lambda_{\mathrm{B}}(\widehat{c}_j/c_j^*)$ for the RRE. Neglecting the irrelevant constant term $\widetilde{Z}(V)/V$, we obtain the hybrid energy again as a relative entropy, namely

$$\mathfrak{E}_V^{\mathsf{FP\text{-}RR}}(\varrho_s, \widehat{c}_{\mathrm{m}}) = \int_{C_{\mathrm{s}}} \left(\frac{1}{V}\rho_s(c_{\mathrm{s}})\log\rho_s(c_{\mathrm{s}}) + \rho_s(c_{\mathrm{s}})E_{\mathrm{s}}(c_{\mathrm{s}})\right)\mathrm{d}c_{\mathrm{s}} + E_{\mathrm{m}}(\widehat{c}_{\mathrm{m}}).$$

For the Onsager operator we also use a cruder reduction than the minimization advocated in Section 6.1. We simply postulate the Onsager operator K_V via the dual dissipation potential

$$\Psi_{V,\mathsf{FP\text{-}RR}}^*(\varrho_s, \widehat{c}_{\mathrm{m}}; \xi, \zeta) = \frac{1}{2}\int_{C_{\mathrm{s}}} \rho_s(c_{\mathrm{s}}) \begin{pmatrix} \nabla_s\xi(c_{\mathrm{s}}) \\ \zeta \end{pmatrix} \cdot \mathbb{K}(c_{\mathrm{s}}, \widehat{c}_{\mathrm{m}}) \begin{pmatrix} \nabla_s\xi(c_{\mathrm{s}}) \\ \zeta \end{pmatrix} \mathrm{d}c_{\mathrm{s}},$$

where $\xi \in \mathrm{C}^1(C_{\mathrm{s}})$ and $\zeta \in \mathbb{R}^{I-J}$. Indeed, in the sense of the general reduction method explained in Section 6.1 we see that $\mathfrak{K}_V^{\mathsf{FP\text{-}RR}}$ is obtained from K by inserting $\varrho = \varrho_s(\mathrm{d}c_{\mathrm{s}}) \otimes \delta_{\widehat{c}_{\mathrm{m}}}$ and $\Xi = \mathsf{M}(\xi, \zeta) : (c_{\mathrm{s}}, c_{\mathrm{m}}) \mapsto \xi(c_{\mathrm{s}}) + \zeta \cdot c_{\mathrm{m}}$.

Thus, the hybrid model induced by the gradient system $(\mathscr{P}(C_{\mathrm{s}}){\times}C_{\mathrm{m}}, \mathfrak{E}_V^{\text{FP-RR}}, \mathfrak{K}_V^{\text{FP-RR}})$ is given by the coupled system for $\rho \in \mathscr{P}(C_{\mathrm{s}})$ and $\widehat{c}_{\mathrm{m}} \in C_{\mathrm{m}}$:

$$\dot{\rho}(c_{\mathrm{s}}) = \operatorname{div}_{\mathrm{s}}\Big(\mathbb{K}_{\mathrm{ss}}(c_{\mathrm{s}},\widehat{c}_{\mathrm{m}})\big(\tfrac{1}{V}\nabla_{\mathrm{s}}\rho(c_{\mathrm{s}}){+}\rho(c_{\mathrm{s}})\nabla_{\mathrm{s}}E_{\mathrm{s}}(c_{\mathrm{s}})\big) + \rho(c_{\mathrm{s}})\mathbb{K}_{\mathrm{sm}}(c_{\mathrm{s}},\widehat{c}_{\mathrm{m}})\nabla_{\mathrm{m}}E_{\mathrm{m}}(\widehat{c}_{\mathrm{m}})\Big),$$

$$\dot{\widehat{c}}_{\mathrm{m}} = -\int_{C_{\mathrm{s}}}\Big(\mathbb{K}_{\mathrm{sm}}^{\mathsf{T}}(c_{\mathrm{s}},\widehat{c}_{\mathrm{m}})\big(\tfrac{1}{V}\nabla_{\mathrm{s}}\rho(c_{\mathrm{s}}){+}\rho(c_{\mathrm{s}})\nabla_{\mathrm{s}}E_{\mathrm{s}}(c_{\mathrm{s}})\big) + \rho(c_{\mathrm{s}})\mathbb{K}_{\mathrm{mm}}(c_{\mathrm{s}},\widehat{c}_{\mathrm{m}})\nabla_{\mathrm{m}}E_{\mathrm{m}}(\widehat{c}_{\mathrm{m}})\Big)\mathrm{d}c_{\mathrm{s}}.$$

It is interesting to see that the last terms can be rewritten in terms of the RRE $\dot{c} = -\mathbb{K}(c)\mathrm{D}E(c) = -R(c) = -(R_{\mathrm{s}}(c_{\mathrm{s}},\widehat{c}_{\mathrm{m}}), R_{\mathrm{m}}(c_{\mathrm{s}},\widehat{c}_{\mathrm{m}}))$, viz.

$$\dot{\rho}(c_{\mathrm{s}}) = \operatorname{div}_{\mathrm{s}}\Big(\tfrac{1}{V}\mathbb{K}_{\mathrm{ss}}(c_{\mathrm{s}},\widehat{c}_{\mathrm{m}})\nabla_{\mathrm{s}}\rho(c_{\mathrm{s}}) + \rho(c_{\mathrm{s}})R_{\mathrm{s}}(c_{\mathrm{s}},\widehat{c}_{\mathrm{m}})\Big),$$

$$\dot{\widehat{c}}_{\mathrm{m}} = -\int_{C_{\mathrm{s}}}\Big(\tfrac{1}{V}\mathbb{K}_{\mathrm{sm}}^{\mathsf{T}}(c_{\mathrm{s}},\widehat{c}_{\mathrm{m}})\nabla_{\mathrm{s}}\rho(c_{\mathrm{s}}) + \rho(c_{\mathrm{s}})R_{\mathrm{m}}(c_{\mathrm{s}},\widehat{c}_{\mathrm{m}})\Big)\mathrm{d}c_{\mathrm{s}}.$$

This reveals that the system is a classical mean-field model, which is linear in the density ρ for the component c_{s} while it is nonlinearly coupled to the ODE for the component $\widehat{c}_{\mathrm{m}}$.

6.4 Coupling a RRE to a CME

In analogy to the coupling of an RRE for some macroscopic c_{m} to a Fokker–Planck equation we can directly couple the CME to an RRE, which leads to hybrid system defined on $\mathscr{P}(\mathbb{N}_0^J){\times}[0,\infty[^{I-J}$. Instead of given the general derivation as in Section 6.3, we just give an explicit example.

For $\beta \in \mathbb{N}_0$ we consider the simple reaction $X_1 \leftrightarrows \beta X_2$ with stoichiometric vectors $\alpha = (1,0)$, $\beta = (0,\beta)$, and $\gamma = (1,-\beta)$. The associated system of RREs is given by

$$\dot{c}_1 = c_2^\beta - c_1, \quad \dot{c}_2 = \beta\,(c_1 - c_2^\beta). \tag{6.5}$$

We have the conservation relation $\mathbb{Q}c = \beta c_1 + c_2 = q$ and the detailed-balance steady state $c_* = (1,1)^\top$. The associated CME on $\mathcal{N} = \mathbb{N}_0^2$ takes the form

$$\dot{u}_n = (n_1{+}1)u_{n+(1,-\beta)} - \Big(n_1 + \frac{n_2!}{V^{\beta-1}(n_2-\beta)!}c\Big)u_n + \frac{(n_2{+}\beta)!}{V^{\beta-1}n_2!}u_{n+(-1,\beta)} \text{ for } n \in \mathcal{N}. \tag{6.6}$$

The detailed-balance steady state by $w_n^V = (w_{n_1}^V w_{n_2}^V)_{n\in\mathcal{N}}$ with $w_n^V = \mathrm{e}^{-V}V^n/n!$. As in (4.5b) the full Onsager operator $\mathcal{K}_V$ is defined via

$$\langle \mu, \mathcal{K}_V(u)\mu\rangle = \sum_{n\in\mathcal{N}} \Lambda\big(\tfrac{n_1+1}{V}u_{n+(1,0)}, \tfrac{(n_2+\beta)!}{V^\beta n_2!}u_{n+(0,\beta)}\big)\big(V\big(\mu_{n+(1,0)}-\mu_{n+(0,\beta)}\big)\big)^2.$$

We partition $c = (c_1,c_2) = (c_{\mathrm{s}},c_{\mathrm{m}})$, i.e., we keep $c_1 \in [0,\infty[$ in stochastic description via the distribution $v = (v_m)_{m\in\mathbb{N}_0} \in \mathscr{P}(\mathbb{N}_0)$, while $c_2 \in [0,\infty[$ will be treated macroscopically. Thus, we define the gradient system $(\mathscr{P}(\mathbb{N}_0){\times}[0,\infty[, \mathfrak{E}_V^{\text{CM-RR}}, \mathfrak{K}_V^{\text{CM-RR}})$ with relative entropy and Onsager operator defined via

$$\mathfrak{E}_V^{\text{CM-RR}}(v,c_2) = \mathsf{E}(c_2) + \frac{1}{V}\sum_{m\in\mathbb{N}_0} v_m\log(v_m/w_m^V), \quad \text{where } \mathsf{E}(z) = \lambda_{\mathrm{B}}(z),$$

$$\Big\langle \big(\begin{smallmatrix}\xi\\\zeta\end{smallmatrix}\big), \mathfrak{K}_V^{\text{CM-RR}}(v,c_2)\big(\begin{smallmatrix}\xi\\\zeta\end{smallmatrix}\big)\Big\rangle = V\sum_{m\in\mathbb{N}_0} \Lambda\big(\tfrac{m+1}{V}v_{m+1}, v_m c_2^\beta\big)\big(\xi_m + \tfrac{\beta}{V}\zeta - \xi_{m+1}\big)^2,$$

for $\boldsymbol{\xi} : \mathbb{N}_0 \to \mathbb{R}$ and $\zeta \in \mathbb{R}$. Again, $\mathfrak{K}_V^{\text{CM-RR}}$ is obtained from $\mathcal{K}_V$ by inserting $\boldsymbol{u}_{m,n_2} = v_m \delta_{\lfloor V c_2 \rfloor}(n_2)$ and $\Xi = \mathsf{M}(\boldsymbol{\xi}, \zeta) : (m, n_2) \mapsto \xi_m + \frac{1}{V} n_2 \zeta$ and performing an approximation for large V. The associated evolution equation is the hybrid system

$$\dot{v}_m = V c_2^\beta v_{m-1} - \left(m + V c_2^\beta\right) v_m + (m{+}1) v_{m+1} \quad \text{for } m \in \mathbb{N}_0 \text{ (with } v_{-1} = 0),$$

$$\dot{c}_2 = \beta\Big(\frac{1}{V} \sum_{m \in \mathbb{N}} m v_m - c_2^\beta\Big).$$

Clearly, this system is consistent with the conservation law $\mathbb{Q}\boldsymbol{c} = \beta c_1 + c_2 = \text{const.}$, in the sense that $c_1 := \frac{1}{V} \sum_{m \in \mathbb{N}} m v_m$ satisfies $\dot{c}_1 = c_2^\beta - c_1 = -\dot{c}_2/\beta$.

6.5 Combining CME and Fokker–Planck descriptions

We consider the simplest nontrivial model, namely the scalar RRE $\dot{c} = a - bc$ with $a, b > 0$, which is induced by the reaction $\emptyset \overset{b}{\underset{a}{\rightleftharpoons}} X$. This corresponds to $\alpha = 0$, $\beta = 1$, $k_{\text{fw}} = a$, and $k_{\text{bw}} = b$. We have the following three derived gradient systems:

(1) The RRE $\dot{c} = a - bc$ is generated by the gradient system $(\mathbb{R}_+, \mathbb{K}, E)$ with steady state $c_* = a/b$, $\mathbb{K}(c) = \Lambda(a, bc)$, and $E(c) = \frac{a}{b} \lambda_{\text{B}}(bc/a)$.

(2) The associated chemical master equation $\dot{\boldsymbol{u}} = \mathcal{B}_V \boldsymbol{u}$ is generated by the gradient system $(\mathscr{P}(\mathbb{N}_0), \mathcal{E}_V, \mathcal{K}_V)$ and reads

$$\dot{u}_n = V a u_{n-1} - \left(V a + bn\right) u_n + b(n{+}1) u_{n+1} \quad \text{for } n \in \mathbb{N}_0 \quad \text{(with } u_{-1} = 0) \tag{6.7}$$

and has the steady state $\boldsymbol{w}_V = \big(\mathrm{e}^{-Va/b}(Va/b)^n/n!\big)_{n \in \mathbb{N}_0}$. The entropy and Onsager operator are

$$\mathcal{E}_V(\boldsymbol{u}) = \frac{1}{V} \sum_{n \in \mathbb{N}_0} u_n \log(u_n/w_n^V) \quad \text{and}$$

$$\mathcal{K}_V(\boldsymbol{u}) = V^2 a \sum_{n \in \mathbb{N}_0} w_n^V \Lambda\big(\tfrac{u_n}{w_n^V}, \tfrac{u_{n+1}}{w_{n+1}^V}\big)(\boldsymbol{e}_n {-} \boldsymbol{e}_{n+1}) \otimes (\boldsymbol{e}_n {-} \boldsymbol{e}_{n+1}).$$

(3) The associated Fokker–Planck equation (5.4) takes the form

$$\dot{\rho} = \partial_c\Big(\frac{\Lambda(a, bc)}{V} \partial_c \rho + \big(bc{-}a + \frac{\Lambda(a, bc)}{2Vc + 1/3}\big)\rho\Big) \quad \text{for } t, c > 0 \text{ and } \rho(t, 0) = 0. \tag{6.8}$$

This equation has the equilibrium solution $W_V : c \mapsto \mathsf{W}(c; a/b, V)$ (cf. (5.2)) and is generated by the gradient system $(\mathscr{P}(]0, \infty[), \boldsymbol{E}_V, \boldsymbol{K})$ with

$$\boldsymbol{E}_V(\rho) = \frac{1}{V} \int_0^\infty \rho(c) \log\Big(\frac{\rho(c)}{W_V(c)}\Big) \, \mathrm{d}c \quad \text{and} \quad \boldsymbol{K}(\rho)\xi = -\big(\rho \Lambda(a, bc)\xi'\big)'.$$

To combine the description via the CME and the Fokker–Planck equation we consider the mixed state space $\mathfrak{N} := \{0, 1, \ldots, N{-}1\} \cup [N/V, \infty[$. Hence, $n \in \{0, \ldots, N{-}1\}$ counts the number of atoms, while for $n \geq N$ we use the concentration $c = n/V \geq N/V$ as a continuous variable to describe the state. A typical choice could be $1 \ll V \ll N$ to be sure to capture all small discrete effects.

The hybrid gradient system $\left(\mathscr{P}(\mathfrak{N}), \mathfrak{E}_{V,N}, \mathfrak{K}_{V,N}\right)$ is described by measures

$$\mathfrak{u} = \sum_{n=0}^{N-1} u_n \delta_n + U(c)\, \mathrm{d}c|_{[N/V,\infty[} \in \mathscr{P}(\mathfrak{N}).$$

The idea is now to choose $\mathfrak{E}_{V,N}$ and $\mathfrak{K}_{V,N}$ rather than to model the evolution equation.

We first choose the equilibrium state in the form

$$\mathfrak{w}^{V,N} = \sum_{n=0}^{N-1} w_n^V \delta_n + W^V(c)\, \mathrm{d}c := \sum_{n=0}^{N-1} \mathrm{e}^{-Va/b} \frac{(Va/b)^n}{n!} \delta_n + \frac{1}{Z_{V,N}} W_V(c)\, \mathrm{d}c,$$

where $Z_{V,N}$ is uniquely determined by asking $\int_{\mathfrak{N}} \mathrm{d}\mathfrak{w}^{V,N} = 1$. The entropy functional is defined via the obvious relative entropy per volume, namely

$$\begin{aligned}
\mathfrak{E}_{V,N}(\mathfrak{u}) &= \frac{1}{V} \int_{\mathfrak{N}} \log\left(\frac{\mathrm{d}\mathfrak{u}}{\mathrm{d}\mathfrak{w}^{V,N}}\right) \mathrm{d}\mathfrak{u} \\
&= \frac{1}{V} \sum_{n=0}^{N-1} \lambda_{\mathrm{B}}\left(\frac{u_n}{w_n^V}\right) w_n^V + \frac{1}{V} \int_{N/V}^{\infty} \lambda_{\mathrm{B}}\left(\frac{U(c)}{W_V(c)}\right) W_V(c)\, \mathrm{d}c,
\end{aligned}$$

where $\frac{\mathrm{d}\mathfrak{u}}{\mathrm{d}\mathfrak{w}}$ denotes the Radon-Nikodym derivative.

The difficult part is the modeling of the Onsager operator $\mathfrak{K}_{V,N}(\boldsymbol{u})$ as it includes the crucial transfer between the discrete and the continuous parts of the hybrid model. We define $\mathfrak{K}$ in terms of its associated quadratic form acting on smooth functions $\xi : \mathfrak{N} \to \mathbb{R}$, where we write ξ_n for $\xi(n)$ and $W(c)$ for $W_V(c)$:

$$\begin{aligned}
\langle \xi, \mathfrak{K}_{V,N}(\boldsymbol{u})\xi \rangle = {}& V^2 a \sum_{n=1}^{N-1} w_{n-1} \Lambda\left(\frac{u_{n-1}}{w_{n-1}}, \frac{u_n}{w_n}\right)(\xi_{n-1} - \xi_n)^2 \\
&+ V^2 \widehat{a} w_{N-1} \Lambda\left(\frac{u_{N-1}}{w_{N-1}}, \frac{U(N/V)}{W(N/V)}\right)\left(\xi_{N-1} - \xi(N/V)\right)^2 \\
&+ \int_{N/V}^{\infty} \Lambda(a, bc) \xi'(c)^2 U(c)\, \mathrm{d}c.
\end{aligned}$$

While the first and the third terms on the right-hand side give the purely discrete and the continuous parts of the state space, respectively, we see that the second term is the new term that couples the discrete and the continuous parts. The parameter $\widehat{a}$ is still to be chosen, the natural parameter being a.

The evolution equation for $\mathfrak{u}$ is again a linear equation of the form $\dot{\mathfrak{u}} = \mathfrak{B}_{V,N}\mathfrak{u}$, i.e., it corresponds to a continuous-time Markov process. It consists of a discrete part, as in (6.7) but only for $n = 0, \ldots, N-2$, and a continuous part, as in (6.8) but only for $c > N/V$. The new structure is the coupling between the two subsystems which gives rise to the following conditions:

$$\dot{u}_{N-1} = V a\, u_{N-2} - \left(V\widehat{a} + b\,(N-1)\right) u_{N-1} + V\widehat{a}\, \frac{w_{N-1}}{W(N/V)}\, U(N/V),$$

$$0 = V\widehat{a}\left(\frac{w_{N-1}}{W(N/V)}\, U(N/V) - u_{N-1}\right) + \frac{1}{V} W(N/V)\Lambda(a, bN/V)\left(\frac{U}{W}\right)'(N/V).$$

By our definition of $\mathfrak{w}^{V,N}$ we have $\frac{w_{N-1}}{W(N/V)} \approx Nb/(aV^2)$ and see that for $\widehat{a} = a$ these conditions take the approximate form

$$\dot{u}_{N-1} \approx V a\, u_{N-2} - \left(V a + b\,(N-1)\right) u_{N-1} + b\frac{N}{V}\, U\left(\frac{N}{V}\right),$$

$$0 \approx \tfrac{1}{V}\Lambda(a, b\tfrac{N}{V})U'(\tfrac{N}{V}) + b\tfrac{N}{V}U(\tfrac{N}{V}) - aV\, u_{N-1},$$

where the second relation clearly shows the corresponding Robin boundary condition connecting the parabolic Fokker–Planck equation to the discrete system on $\{0, \ldots, N-1\}$. Note that u_n and U are scaled such that $V\, u_{N-1}$ is comparable to $U(N/V)$.

References

[AC*05] A. ALFONSI, E. CANCÈS, G. TURINICI, B. DI VENTURA, and W. HUISINGA. Adaptive simulation of hybrid stochastic and deterministic models for biochemical systems. *ESAIM: Proceedings*, 14, 1–13, 2005.

[ACK10] D. F. ANDERSON, G. CRACIUN, and T. G. KURTZ. Product-form stationary distributions for deficiency zero chemical reaction networks. *Bull. Math. Biology*, 72, 1947–1970, 2010.

[ADE18a] A. AGAZZI, A. DEMBO, and J.-P. ECKMANN. Large deviations theory for Markov jump models of chemical reaction networks. *Ann. Probab.*, 28, 1821–1855, 2018.

[ADE18b] A. AGAZZI, A. DEMBO, and J.-P. ECKMANN. On the geometry of chemical reaction networks: Lyapunov function and large deviations. *J. Stat. Phys.*, 172, 321—-352, 2018.

[AGH02] G. ALBINUS, H. GAJEWSKI, and R. HÜNLICH. Thermodynamic design of energy models of semiconductor devices. *Nonlinearity*, 15(2), 367–383, 2002.

[BB*17] M. BRUNA, M. BURGER, H. RANETBAUER, and M.-T. WOLFRAM. Cross-diffusion systems with excluded volume effects and asymptotic gradient flow structures. *J. Nonlinear Sci.*, 27(2), 687–719, 2017.

[CH*12] 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. Rational Mech. Anal.*, 203(3), 969–1008, 2012.

[DaS14] S. DANERI and G. SAVARÉ. Lecture notes on gradient flows and optimal transport. In Y. Ollivier, H. Pajot, and C. Villani, editors, *Optimal Transportation. Theory and Applications*, chapter 6, pages 100–144. Cambridge Univ. Press, 2014.

[Dav80] E. B. DAVIES. *One-parameter semigroups*, volume 15 of *London Mathematical Society Monographs*. Academic Press Inc., London, 1980.

[DeM84] S. DE GROOT and P. MAZUR. *Non-Equilibrium Thermodynamics*. Dover Publ., New York, 1984.

[DFM19] P. DONDL, T. FRENZEL, and A. MIELKE. A gradient system with a wiggly energy and relaxed EDP-convergence. *ESAIM Control Optim. Calc. Var.*, 25(68), 45 pp, 2019.

[DiL15] K. DISSER and M. LIERO. On gradient structures for Markov chains and the passage to Wasserstein gradient flows. *Networks Heterg. Media*, 10(2), 233–253, 2015.

[DoK14] S. DOLGOV and B. KHOROMSKIJ. Simultaneous state-time approximation of the chemical master equation using tensor product formats. *Numer. Lin. Algebra Appl.*, 22(2), 197–219, 2014.

[EF*16] M. ERBAR, M. FATHI, V. LASCHOS, and A. SCHLICHTING. Gradient flow structure for McKean-Vlasov equations on discrete spaces. *Discr. Cont. Dynam. Systems*, 36(12), 6799–6833, 2016.

[EkT76] I. EKELAND and R. TEMAM. *Convex Analysis and Variational Problems.* North Holland, 1976.

[Eng09] S. ENGBLOM. Spectral approximation of solutions to the chemical master equation. *J. Comp. Appl. Math.*, 229, 208–221, 2009.

[ÉrT89] P. ÉRDI and J. TÓTH. *Mathematical Models of Chemical Reactions. Theory and Applications of Deterministic and Stochastic Models.* Princeton University Press, 1989.

[FaS16] M. FATHI and M. SIMON. The gradient flow approach to hydrodynamic limits for the simple exclusion process. In *From particle systems to partial differential equations. III*, volume 162 of *Springer Proc. Math. Stat.*, pages 167–184. Springer, 2016.

[FeH77] M. FEINBERG and F. J. M. HORN. Chemical mechanism structure and the coincidence of the stoichiometric and kinetic subspaces. *Arch. Rational Mech. Anal.*, 66(1), 83–97, 1977.

[FoL07] I. FONSECA and G. LEONI. *Modern Methods in the Calculus of Variations: L^p spaces.* Springer, 2007.

[FrL19] T. FRENZEL and M. LIERO. Effective diffusion in thin structures via generalized gradient systems and EDP-convergence. *WIAS Preprint 2601*, 2019.

[GaY14] L. GAUCKLER and H. YSERENTANT. Regularity and approximability of the solutions to the chemical master equation. *ESAIM Math. Model. Numer. Anal.*, 48(6), 1757–1775, 2014.

[Gil92] D. T. GILLESPIE. A rigorous derivation of the chemical master equation. *Physica A*, 188, 404–425, 1992.

[Gil00] D. T. GILLESPIE. The chemical Langevin equation. *J. Chem. Phys.*, 113(1), 297–306, 2000.

[GlM13] A. GLITZKY and A. MIELKE. A gradient structure for systems coupling reaction-diffusion effects in bulk and interfaces. *Z. angew. Math. Phys. (ZAMP)*, 64, 29–52, 2013.

[Grö83] K. GRÖGER. Asymptotic behavior of solutions to a class of diffusion-reaction equations. *Math. Nachr.*, 112, 19–33, 1983.

[HeL07] A. HELLANDER and P. LÖTSTEDT. Hybrid method for the chemical master equation. *J. Comput. Physics*, 227(1), 100–122, 2007. CME literatureCME/Hell...

[Hig08] D. J. HIGHAM. Modeling and simulating chemical reactions. *SIAM Rev.*, 50(2), 347–368, 2008.

[Jah11] T. JAHNKE. On reduced models for the chemical master equation. *Multiscale Model. Simul.*, 9(4), 1646–1676, 2011.

[Kur67] T. G. KURTZ. *Convergence of operator semigroups with applications to Markov processes.* PhD thesis, University of Stanford, 1967.

[Kur69] T. G. KURTZ. Extension of Trotter's operator semigroup approximation theorem. *J. Funct. Anal.*, 3, 354–375, 1969.

[Kur70] T. G. KURTZ. Solutions of ordinary differential equations as limits of pure jump processes. *J. Appl. Prob.*, 7, 49–58, 1970.

[Kur72] T. G. KURTZ. The relationship between stochastic and deterministic models of chemical reactions. *J. Chem. Phys.*, 57, 2976–2978, 1972.

[Kur78] T. G. KURTZ. Strong approximation theorems for density dependent Markov chains. *Stochastic Processes Appl.*, 6(3), 223–240, 1977/78.

[Lig10] T. M. LIGGETT. *Continuous time Markov processes*, volume 113 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2010.

[LM*17] M. LIERO, A. MIELKE, M. A. PELETIER, and D. R. M. RENGER. On microscopic origins of generalized gradient structures. *Discr. Cont. Dynam. Systems Ser. S*, 10(1), 1–35, 2017.

[Maa11] J. MAAS. Gradient flows of the entropy for finite Markov chains. *J. Funct. Anal.*, 261, 2250–2292, 2011.

[MaM20] J. MAAS and A. MIELKE. On gradient structures for chemical reactions with detailed balance: II. dissipation distances and geodesic convexity. *In preparation*, 2020.

[Mie11] A. MIELKE. A gradient structure for reaction-diffusion systems and for energy-drift-diffusion systems. *Nonlinearity*, 24, 1329–1346, 2011.

[Mie13] A. MIELKE. Geodesic convexity of the relative entropy in reversible Markov chains. *Calc. Var. Part. Diff. Eqns.*, 48(1), 1–31, 2013.

[Mie16] A. MIELKE. On evolutionary Γ-convergence for gradient systems (Ch. 3). In A. Muntean, J. Rademacher, and A. Zagaris, editors, *Macroscopic and Large Scale Phenomena: Coarse Graining, Mean Field Limits and Ergodicity*, Lecture Notes in Applied Math. Mechanics Vol. 3, pages 187–249. Springer, 2016. Proc. of Summer School in Twente University, June 2012.

[MiS19] A. MIELKE and A. STEPHAN. Coarse graining via EDP-convergence for linear fast-slow reaction systems. *Math. Models Meth. Appl. Sci. (M³AS)*, 2019. Submitted. WIAS preprint 2643.

[ML*11] S. MENZ, J. C. LATORRE, C. SCHÜTTE, and W. HUISINGA. Hybrid stochastic-deterministic solution of the chemical master equation. *Multiscale Model. Simul.*, 2011. Submitted.

[MP*17] 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(4), 1562–1585, 2017.

[MPR14] 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. *Potential Analysis*, 41(4), 1293–1327, 2014.

[MRP16] A. MIELKE, D. R. M. RENGER, and M. A. PELETIER. A generalization of Onsager's reciprocity relations to gradient flows with nonlinear mobility. *J. Non-Equil. Thermodyn.*, 41(2), 141–149, 2016.

[MuK07] B. MUNSKY and M. KHAMMASH. A multiple time interval finite state projection algorithm for the solution to the chemical master equation. *J. Comput. Physics*, 226, 818–835, 2007.

[Nan59] T. S. NANJUNDIAH. Note on Stirling's formula. *Amer. Math. Monthly*, 66, 701–703, 1959.

[Nor97] J. R. NORRIS. *Markov chains*, volume 2. Cambridge University Press, 1997.

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

[Ons31] L. ONSAGER. Reciprocal relations in irreversible processes, I+II. *Physical Review*, 37, 405–426, 1931. (part II, 38:2265–2279).

[Reu57] G. E. H. REUTER. Denumerable Markov processes and the associated contraction semi-groups on l. *Acta Math.*, 97, 1–46, 1957.

[SaS04] E. SANDIER and S. SERFATY. Gamma-convergence of gradient flows with applications to Ginzburg-Landau. *Comm. Pure Appl. Math.*, LVII, 1627–1672, 2004.

[SaS17] D. SANZ-ALONSO and A. M. STUART. Gaussian approximations of small noise diffusions in Kullback–Leibler divergence. *Commun. Math. Sci.*, 15(7), 2087–2097, 2017.

[Sch19] A. SCHLICHTING. Macroscopic limit of the Becker–Döring equation via gradient flows. *ESAIM Control Optim. Calc. Var.*, 25(22), 36 pp., 2019.

[ScS89] S. SCHUSTER and R. SCHUSTER. A generalization of Wegscheider's condition. Implications for properties of steady states and for quasi-steady-state approximation. *J. Math. Chem.*, 3(1), 25–42, 1989.

[Ser11] S. SERFATY. Gamma-convergence of gradient flows on Hilbert spaces and metric spaces and applications. *Discr. Cont. Dynam. Systems Ser. A*, 31(4), 1427–1451, 2011.

[VIR09] M. O. VLAD and J. ROSS. Thermodynamically based constraints for rate coefficients of large biochemical networks. *WIREs Syst. Biol. Med.*, 1, 348–358, 2009.

[Weg02] R. WEGSCHEIDER. Über simultane Gleichgewichte und die Beziehungen zwischen Thermodynamik und Reaktionskinetik homogener Systeme. *Z. Phys. Chemie*, 39, 257–303, 1902.

[WiS17] S. WINKELMANN and C. SCHÜTTE. Hybrid models for chemical reaction networks: Multiscale theory and application to gene regulatory systems. *J. Chem. Physics*, 147(11), 114115/1–18, 2017.

[Yon08] W.-A. YONG. An interesting class of partial differential equations. *J. Math. Phys.*, 49, 033503, 21, 2008.