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

Preprint ISSN 2198-5855

Consistency and convergence for a family of finite volume discretizations of the Fokker–Planck operator

Martin Heida, Markus Kantner, Artur Stephan

submitted: February 5, 2020 (revision: September 30, 2020)

Weierstrass Institute
Mohrenstr. 39
10117 Berlin
Germany
E-Mail: martin.heida@wias-berlin.de
markus.kantner@wias-berlin.de
artur.stephan@wias-berlin.de

No. 2684
Berlin 2020

2010 *Mathematics Subject Classification.* 35Q84,49M25,65N08.

Key words and phrases. Finite Volume, Fokker–Planck, Scharfetter–Gummel, Stolarsky mean, Consistency, Order of Convergence.

M. H. and A. S. are financed by Deutsche Forschungsgemeinschaft (DFG) through Grant CRC 1114 “Scaling Cascades in Complex Systems”, Project C05 *Effective models for materials and interfaces with multiple scales*. The work of M. K. received funding from the DFG under Germany’s Excellence Strategy – EXC2046: MATH+ (Berlin Mathematics Research Center).

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

Fax: +49 30 20372-303
E-Mail: preprint@wias-berlin.de
World Wide Web: http://www.wias-berlin.de/

Consistency and convergence for a family of finite volume discretizations of the Fokker–Planck operator

Martin Heida, Markus Kantner, Artur Stephan

Abstract

We introduce a family of various finite volume discretization schemes for the Fokker–Planck operator, which are characterized by different Stolarsky weight functions on the edges. This family particularly includes the well-established Scharfetter–Gummel discretization as well as the recently developed square-root approximation (SQRA) scheme. We motivate this family of discretizations both from the numerical and the modeling point of view and provide a uniform consistency and error analysis. Our main results state that the convergence order primarily depends on the quality of the mesh and in second place on the choice of the Stolarsky weights. We show that the Scharfetter–Gummel scheme has the analytically best convergence properties but also that there exists a whole branch of Stolarsky means with the same convergence quality. We show by numerical experiments that for small convection the choice of the optimal representative of the discretization family is highly non-trivial while for large gradients the Scharfetter–Gummel scheme stands out compared to the others.

1 Introduction

The Fokker–Planck equation (FPE), also known as *Smoluchowski equation* or *Kolmogorov forward equation*, is one of the most important equations in theoretical physics and applied mathematics. It describes the time evolution of the probability density function of a particle in an external force field (e.g., fluctuating forces as in Brownian motion). The equation can be generalized to other contexts and observables and has been employed in a broad range of applications, including physical chemistry, protein synthesis, plasma physics and semiconductor device simulation. Thus, there is a huge interest in the development of efficient and robust numerical methods. In the context of finite volume (FV) methods, the central objective is a robust and accurate discretization of the (particle or probability) flux implied by the FPE.

A particularly important discretization scheme for the flux was derived by Scharfetter and Gummel [46] in the context of the drift-diffusion model for electronic charge carrier transport in bipolar semiconductor devices [48]. The typically exponentially varying carrier densities at p-n junctions lead to unphysical results (spurious oscillations), if the flux is discretized in a naive way using standard finite difference schemes [45]. The problem was overcome by considering the flux expression as a one-dimensional boundary value problem along each edge between adjacent mesh nodes. The resulting Scharfetter–Gummel (SG) scheme provides a robust discretization of the flux as it asymptotically approaches the numerically stable discretizations in the drift- (upwind scheme) and diffusion-dominated (central finite difference scheme) limits. The SG-scheme and its several generalizations to more complex physical problem settings are nowadays widely used in semiconductor device simulation [37, 19] and have been extensively studied in the literature [4, 15, 18, 30]. The SG-scheme is also known as *exponential fitting scheme* and was independently discovered by Allan and Southwell [1] and Il'in [28] in different contexts.

DOI 10.20347/WIAS.PREPRINT.2684 Berlin, February 5, 2020/rev. September 30, 2020

Recently, an alternative flux discretization method, called *square-root approximation* (SQRA) scheme, has been derived explicitly for high dimensional problems. The original derivation in [32] aims at applications in molecular dynamics and is based on Markov state models. However, it can also be obtained from a maximum entropy path principle [10] and from discretizing the Jordan–Kinderlehrer–Otto variational formulation of the FPE [39]. In Section 3.2, we provide a derivation of SQRA scheme, which is motivated from the theory of gradient flows. In contrast to the SG-scheme, the SQRA is very recent and only sparsely investigated.

The SG and the SQRA schemes both turn out to be special cases of a family of discretization schemes based on weighted Stolarsky means [47], see Section 3.1. This family is very rich and allows for a general convergence and consistency analysis, which we carry out in Sections 4–5. There are also other discretization schemes available in literature. The Chang–Cooper scheme [6] has been derived for computing ion-electron collisions and uses another Stolarsky mean, namely the logarithmic mean. General discretization schemes using different weights are called B-schemes and have been introduced in [5]. We will recall the corresponding results in Section 1.2 below.

1.1 The FPE and the SG and SQRA discretization schemes

In this work, we consider the stationary Fokker–Planck equation

$$-\nabla\cdot(\kappa\nabla u)-\nabla\cdot(\kappa u\nabla V)=f, \tag{1.1}$$

which can be equivalently written as

$$\operatorname{div}\mathbf{J}(u,V)=f$$

using the flux $\mathbf{J}(u,V)=-\kappa\left(\nabla u+u\nabla V\right)$, where $\kappa>0$ is a (possibly space-dependent) diffusion coefficient and $V:\Omega\to\mathbb{R}$ is a given potential. The flux $\mathbf{J}$ consists of a diffusive part $\kappa\nabla u$ and a drift part $\kappa u\nabla V$, which compensate for the stationary density $\pi=\mathrm{e}^{-V}$ (Boltzmann distribution) as $\mathbf{J}(\mathrm{e}^{-V},V)=0$. This reflects the principle of detailed balance in the thermodynamic equilibrium. The right-hand side f describes possible sink or source terms.

Assumption 1.1. *Unless stated otherwise we assume $V\in C^2(\overline{\Omega})$, $\kappa\in C^1(\overline{\Omega})$, $f\in C(\overline{\Omega})$ real valued functions with $\kappa>0$. The standard boundary conditions are the homogeneous Dirichlet boundary conditions.*

Remark 1.2. Some results also hold for lower regularity and for κ a symmetric strictly positive definite matrix.

In Section 3.1 we derive the following discretization of (1.1)

$$-\sum_{j:\,j\sim i}\frac{m_{ij}}{h_{ij}}\kappa_{ij}S_{ij}\left(\frac{u_j}{\pi_j}-\frac{u_i}{\pi_i}\right)=f_i, \tag{1.2}$$

where $\pi_i=\mathrm{e}^{-V_i}$, $f_i=\int_{\Omega_i}f$ is the integral of f over the i-th cell, $S_{ij}=S_{\alpha,\beta}(\pi_i,\pi_j)$ is a Stolarsky mean of π_i and π_j and $\sum_{j:\,j\sim i}$ denotes the sum over all neighbors of cell i. We sometimes refer to the general form (1.2) as *discrete FPE*.

Assumption 1.3. *Under the Assumption 1.1 we additionally assume that for some $\infty>K>\kappa_0>0$ it holds $K>\kappa_{ij}\geq\kappa_0$.*

The weighted Stolarsky means [47]

$$S_{\alpha,\beta}\left(x,y\right)=\left(\frac{\beta\left(x^{\alpha}-y^{\alpha}\right)}{\alpha\left(x^{\beta}-y^{\beta}\right)}\right)^{\frac{1}{\alpha-\beta}},\qquad \alpha\neq 0,\,\beta\neq 0,\,\alpha\neq\beta,\,x\neq y \tag{1.3}$$

generalize the logarithmic mean and other means and can be extended to the critical points $\alpha=0$, $\beta=0$, $\alpha=\beta$, $x=y$ in a continuous way, see Tab. 2. An interesting aspect of the above representation is that all these schemes preserve positivity with the discrete linear operator being an M-matrix. This can be seen introducing the relative density $U=u/\pi$ for which Eq. (1.2) yields

$$-\sum_{j:\,j\sim i}\frac{m_{ij}}{h_{ij}}\kappa_{ij}S_{ij}\left(U_j-U_i\right)=f_i, \tag{1.4}$$

which is a discretization of the elliptic equation

$$-\nabla\cdot\left(\kappa\pi\nabla U\right)=f, \tag{1.5}$$

where the discrete Fokker–Planck operator becomes a purely diffusive second order operator in U. Furthermore, if κ is a symmetric strictly positive definite uniformly elliptic matrix, the operator in Eq. (1.4) is also symmetric strictly positive definite and uniformly elliptic. In the latter setting, we can thus rule out the occurrence of spurious oscillations in our discretization.

The above formulation underpins the diffusive character both of the discrete and the continuous FPE. Using the relation $S_{\alpha,\beta}\left(x,y\right)=x\,S_{\alpha,\beta}\left(1,y/x\right)$ and introducing the weight function

$$B_{\alpha,\beta}\left(x\right)=S_{\alpha,\beta}\left(1,\mathrm{e}^{-x}\right)\quad\text{with}\quad B_{\alpha,\beta}\left(-x\right)=\mathrm{e}^{x}B_{\alpha,\beta}\left(x\right), \tag{1.6}$$

Eq. (1.2) can equally be reformulated as

$$-\sum_{j:\,i\sim j}\frac{m_{ij}}{h_{ij}}\kappa_{ij}\left(B_{\alpha,\beta}\left(V_i-V_j\right)u_j-B_{\alpha,\beta}\left(V_j-V_i\right)u_i\right)=f_i.$$

Two special cases of particular interest are

$$B_{0,-1}\left(V_i-V_j\right)=\frac{V_i-V_j}{\mathrm{e}^{V_i-V_j}-1}=S_{0,-1}\left(\pi_i,\pi_j\right)\pi_j^{-1}, \tag{1.7}$$

$$B_{1,-1}\left(V_i-V_j\right)=\mathrm{e}^{-\frac{1}{2}\left(V_i-V_j\right)}=S_{1,-1}\left(\pi_i,\pi_j\right)\pi_j^{-1}. \tag{1.8}$$

With regard to Tab. 2 below, these coefficients are known as the Bernoulli function $B_{0,-1}$ (for SG) and the SQRA-coefficient $B_{1,-1}$. FV schemes with general weight functions B have been investigated in [5, 34] (B-schemes).

In the purely diffusive regime, i.e., for $V_i-V_j\to 0$, it holds $B_{\alpha,\beta}\left(V_i-V_j\right)\to 1$ for all α, β, such that the Stolarsky scheme approaches a discrete analogue of the diffusive part of the continuous flux $J_{ij}=\kappa_{ij}\left(B_{\alpha,\beta}\left(V_j-V_i\right)u_i-B_{\alpha,\beta}\left(V_i-V_j\right)u_j\right)/h_{ij}$.

In the drift-dominated regime, i.e., for $V_j-V_i\to\pm\infty$, the various $B_{\alpha,\beta}$ behave differently. While $B_{1,-1}\left(V_i-V_j\right)$ cannot be controlled in a reasonable way, asymptotics of $B_{0,-1}$ recover the upwind scheme

$$J_{i,j}\to-\kappa_{i,j}\frac{V_j-V_i}{h_{i,j}}\begin{cases}u_j & \text{if } V_j>V_i\\ u_i & \text{if } V_j<V_i\end{cases}, \tag{1.9}$$

which is a robust discretization of the drift part of the flux, where the density u is evaluated in the donor cell of the flux. Hence, the Bernoulli function $B_{0,-1}$ interpolates between the appropriate discretizations for the drift- and diffusion-dominated limits, which is why the SG scheme is the preferred FV scheme for Fokker–Planck type operators. Mathematically, this is formulated in Section 5.2.

1.2 Major contributions of this work

As main contribution, we investigate the order of convergence for the general Stolarsky scheme. Furthermore, we provide a derivation of the general Stolarsky mean FV discretization in Section 3.1 and discuss the gradient structure of the discretization schemes in view of the natural gradient structure of the FPE in Section 3.2.

In recent years, convergence order has been derived for many different schemes. In [31], quantitative convergence of order $O(h^2)$ for several upwind schemes on rectangular grids has been shown. In [2] the finite volume Scharfetter–Gummel discretization (of steady convection diffusion equations) is connected to a finite element method and convergence of order $O(h)$ is obtained by using results from [51]. Investigating general B-schemes, [5] proved strong convergence in L^2 for the solutions of the FV scheme to the continuous solution. Recently, convergence of order $O(h)$ for general B-schemes including SG, SQRA as well as Stolarsky means has been proved in 1D [34]. Independently, convergence for the SQRA discretization has been investigated in [39] in 1D, [11] (formally, rectangular meshes) and [24] using G-convergence on grids with random weights.

Here, we are going to derive estimates for the order of convergence in the energy norm for general Stolarsky schemes. We benefit from analytical properties of Stolarsky means and uses the general theory of consistent meshes in the sense of the recent work [8]. We will see that the error naturally splits into the consistency error for the discretization of the Laplace operator (the consistency of the elliptic operator) plus an error which is due to the convective part. Here we have the possibility to study the error in terms of U and of u, in both cases in the energy norm. While the error in terms of U can be directly inferred from the diffusive estimate in Lemma 2.12, one can also apply a splitting into diffusion- and convection-part of the error, both in terms of U and u. The order of convergence is in general limited by the consistency of the mesh but can be improved up to order $O(h)$ in u (on all Voronoï grids), resp. $O(h^2)$ in U (on cubic grids). It is interesting to observe that the optimal Stolarsky mean can be different in the variables u and U for the same problem on the same mesh. This is indicated by the numerical experiment of Example 7.1.

Despite the latter discrepancy for u and U, the Stolarsky scheme $S_{0,-1}$ (SG scheme) turns out to be special among all schemes as it yields order $O(h^2)$ in u on cubic grids (Theorem 6.3). Due to a perturbation result (Corollary 4.4), the good convergence properties of the SG scheme carry over to every Stolarsky scheme where $\alpha + \beta = -1$.

Using the notations of Section 2, we formulate the above in the following theorems, where $\mathcal{R}_{\mathcal{T}_h} u$ is the pointwise evaluation of u in the centers of the Voronoï cells. Hence the constraint $d \le 4$ in this work stems from the condition $H^2(\Omega) \hookrightarrow C(\overline{\Omega})$.

Theorem 1.4. *Let $d \le 4$ and $\mathcal{T}_h = (\mathcal{V}_h, \mathcal{E}_h, \mathcal{P}_h)$ and κ, V satisfying Assumptions 1.1 and 1.3 such that $\mathcal{T}_h = (\mathcal{V}_h, \mathcal{E}_h, \mathcal{P}_h)$ is a family of φ-consistent meshes (Def. 2.14) with $\mathrm{diam}\mathcal{T}_h \to 0$ as $h \to 0$ and let the assumptions of Lemma 5.1 hold. If $u \in H^2(\Omega)$ is the solution of* (1.1) *and $u_{\mathcal{T}_h}$ the solution of* (1.2) *then*

$$\left\| u_{\mathcal{T}_h} - \mathcal{R}_{\mathcal{T}_h} u \right\|_{H_{\mathcal{T},\kappa}}^2 \le C_1 \left(\|u\|_{H^2} + \|u\|_\infty \|V\|_{H^2} \right) \varphi(h)^2 + C_2 h^2 \,,$$

where C_1 depends on $\mathcal{T}_h$ and κ and C_2 additionally depends on $\|V\|_{C^2}$ and $\|u\|_{H^2}$. In case $S_* = S_{0,-1}$ or $S_* = S_{\alpha,\beta}$ with $\alpha + \beta = -1$ and $u \in C^1(\overline{\Omega})$ the above can be improved to

$$\left\| u_{\mathcal{T}_h} - \mathcal{R}_{\mathcal{T}_h} u \right\|_{H_{\mathcal{T},\kappa}}^2 \le C_1 \left(\|u\|_{H^2} + \|u\|_\infty \|V\|_{H^2} \right) \varphi(h)^2 + C_2 h^4 \,.$$

Proof. This is a consequence of Definition 2.14 together with Lemma 2.12, Theorem 5.6 and Corollary 5.7. □

Remark 1.5. As a consequence of former works (see Propositions 2.15 and 2.16) it holds $\varphi(h) = O(h)$ on Voronoï grids and $\varphi(h) = O(h^2)$ on cubic grids. This explains the next result.

Theorem 1.6. *Let $d \le 4$ and $\mathcal{T}_h = (\mathcal{V}_h, \mathcal{E}_h, \mathcal{P}_h)$ and κ, V satisfying Assumptions 1.1 and 1.3 such that $\mathcal{T}_h = (\mathcal{V}_h, \mathcal{E}_h, \mathcal{P}_h)$ is a family of* cubic *φ-consistent meshes (Def. 2.14) with $\mathrm{diam}\mathcal{T}_h \to 0$ as $h \to 0$ and let the assumptions of Lemma 5.1 hold. If $u \in H^2(\Omega)$ is the solution of* (1.1) *and $u_{\mathcal{T}_h}$ the solution of* (1.2) *then*

$$\|u_{\mathcal{T}_h} - \mathcal{R}_{\mathcal{T}_h} u\|^2_{H_{\mathcal{T},\kappa}} \le Ch^2\,,$$

where C depends on $\mathcal{T}_h$, κ, $\|V\|_{C^2}$ and $\|u\|_{H^2}$. In case $S_* = S_{0,-1}$ or $S_* = S_{\alpha,\beta}$ with $\alpha + \beta = -1$ the above can be improved to

$$\|u_{\mathcal{T}_h} - \mathcal{R}_{\mathcal{T}_h} u\|^2_{H_{\mathcal{T},\kappa}} \le Ch^4\,.$$

Proof. This is a consequence of Lemma 2.12, Propositions 2.16 (resp. Theorem 6.2) and Theorem 6.3. □

We note at this point, that these estimates are only “worst case” estimates, while the true rate of convergence could also be better. In Section 4 we will see that the rate of convergence is close for different Stolarsky means which share the same value of $\alpha + \beta$. I.e. the difference in the error due to switching $S_{\alpha,\beta}$ with $S_{\tilde{\alpha},\tilde{\beta}}$ is of order h^3 if $\tilde{\alpha} + \tilde{\beta} = \alpha + \beta$, see Corollary 4.3. This explains the shape of the error graphs in Figs. 2 (a, c) and 3 (a, c).

Although we treat the Stolarsky means as an explicit example, note that some of the main results also hold for other smooth means.

1.3 Outlook

The results of this work suggest to search for “optimal” parameters α and β in the choice of the Stolarsky mean in order to reduce the error of the approximation as much as possible. However, from an analytical point of view, the quest for such optimal α and β is quite challenging. Moreover, since the optimal choice might vary locally, depending on the local properties of the potential V, we suggest to implement a learning algorithm that provides suitable parameters α and β depending on the local structure of V and the mesh.

1.4 Outline of this work

After some preliminaries regarding notation and a priori estimates in Section 2, we present a mathematical derivation of the SG scheme in Section 3.1 and discuss its formal relation to SQRA. We will then provide a derivation of SQRA from physical principles in Section 3.2, based on the Jordan–Kinderlehrer–Otto [29] formulation of the FPE. In Section 3.1, we show that SG and SQRA are elements of a huge family of discretization schemes (1.2).

Section 5 provides the error analysis and estimates for the consistency and the order of convergence. We distinguish the cases of small and large gradients and have a particular look at cubic meshes. Section 6 specifies the results to cubic grids.

Finally, we show hat the optimal choice of S_* depends on V and f, but is not unique. If $S_{\alpha,\beta}$ denotes one of the Stolarsky means, we will prove in Section 4 that the Stolarsky means satisfying $\alpha + \beta = \mathrm{const.}$ show similar quantitative convergence behavior as suggested in Corollary 4.3. Finally, this result is illustrated in Section 7 by numerical simulations.

2 Preliminaries and notation

We collect some concepts and notation, which will frequently be used in this work.

2.1 The Mesh

For a subset $A \subset \mathbb{R}^d$, $\overline{A}$ is the topological closure of A.

Definition 2.1. Let $\Omega \subset \mathbb{R}^d$ be a polygonal domain. A finite volume mesh of Ω is a triangulation $\mathcal{T} = (\mathcal{V}, \mathcal{E}, \mathcal{P})$ consisting of a family of control volumes $\mathcal{V} := \{\Omega_i, i = 1, \dots, N\}$ which are convex polytope cells, a family of $(d-1)$-dimensional interfaces

$$\begin{aligned} \mathcal{E} &:= \mathcal{E}_\Omega \cup \mathcal{E}_\partial \\ \mathcal{E}_\Omega &:= \left\{\sigma_{ij} \subset \mathbb{R}^d \,:\, \sigma_{ij} = \partial\Omega_i \cap \partial\Omega_j\right\} \\ \mathcal{E}_\partial &:= \left\{\sigma \subset \mathbb{R}^d \,:\, \sigma = \partial\Omega_i \cap \partial\Omega \text{ is flat}\right\} \end{aligned}$$

and points $\mathcal{P} = \{x_i, i = 1, \dots, N\}$ with $x_i \in \overline{\Omega_i}$ satisfying

(i) $\bigcup_i \overline{\Omega}_i = \overline{\Omega}$

(ii) For every i there exists $\mathcal{E}_i \subset \mathcal{E}$ such that $\overline{\Omega}_i \backslash \Omega_i = \bigcup_{\sigma \in \mathcal{E}_i} \sigma$. Furthermore, $\mathcal{E} = \bigcup_i \mathcal{E}_i$.

(iii) For every i, j either $\overline{\Omega}_i \cap \overline{\Omega}_j = \emptyset$ or $\overline{\Omega}_i \cap \overline{\Omega}_j = \overline{\sigma}$ for $\sigma \in \mathcal{E}_i \cap \mathcal{E}_j$ which will be denoted σ_{ij}.

The mesh is called h-consistent if

(iv) The Family $(x_i)_{i=1\dots N}$ is such that $x_i \neq x_j$ if $i \neq j$ and the straight line D_{ij} going through x_i and x_j is orthogonal to σ_{ij}.

and admissible if

(v) For any boundary interface $\sigma \in \mathcal{E}_\partial \cap \mathcal{E}_i$ it holds $x_i \notin \sigma$ and for $D_{i,\sigma}$ the line through x_i orthogonal to σ it holds that $D_{i,\sigma} \cap \sigma \neq \emptyset$ and let $y_\sigma := D_{i,\sigma} \cap \sigma$.

Property (iv) is assumed in [22] in order to prove a strong form of consistency in the sense of Definition 2.14 below. It is satisfied for example for Voronoï discretizations.

We write m_i for the volume of Ω_i and for $\sigma \in \mathcal{E}$ we denote m_σ its $(d-1)$-dimensional mass. In case $\sigma_{ij} \in \mathcal{E}_i \cap \mathcal{E}_j$ we write $m_{ij} := m_{\sigma_{ij}}$. For the sake of simplicity, we consider $\tilde{\mathcal{P}} := (x_i)_{i=1,\dots,N}$ and $\mathcal{P} := \tilde{\mathcal{P}} \cup \{y_\sigma \,:\, \sigma \in \mathcal{E}_\partial,$ according to (v)$\}$. We extend the enumeration of $\tilde{\mathcal{P}}$ to $\mathcal{P} = (x_j)_{j=1,\dots,\tilde{N}}$ and write $i \sim j$ if $x_i, x_j \in \tilde{\mathcal{P}}$ with $\mathcal{E}_i \cap \mathcal{E}_j \neq \emptyset$. Similarly, if $x_i \in \tilde{\mathcal{P}}$ and $x_j = y_\sigma$ for $\sigma \in \mathcal{E}_i$ we write $\sigma_{ij} := \sigma$ and $i \sim j$. Finally, we write $h_{ij} = |x_i - x_j|$.

We further call

$$\mathcal{P}^* := \{u : \mathcal{P} \to \mathbb{R}\}\,, \quad \tilde{\mathcal{P}}^* := \left\{u : \tilde{\mathcal{P}} \to \mathbb{R}\right\}\,, \quad \text{and} \quad \mathcal{E}^* := \{w : \mathcal{E} \to \mathbb{R}\}$$

the discrete functions from $\mathcal{P}$ resp. $\tilde{\mathcal{P}}$ resp. $\mathcal{E}$ to $\mathbb{R}$.

symbol	meaning	symbol	meaning
κ	diffusion coefficient	U	u/π
κ_{ij}	$\frac{\bar\kappa_i\bar\kappa_j}{\bar\kappa_i\frac{d_{i,ij}}{h_{ij}}+\bar\kappa_j\frac{d_{j,ij}}{h_{ij}}}$	u	density
κ^*,κ_*	$0<\kappa_*\le\kappa\le\kappa^*<\infty$	m_i	$\mathrm{vol}(\Omega_i)$
V	real potential on $\Omega\subset\mathbb{R}^d$	h_i	$\mathrm{diam}(\Omega_i)$
V^*,V_*	$-\infty<V_*\le V\le V^*<\infty$	σ_{ij}	$\partial\Omega_i\cap\partial\Omega_j$
π	stat. measure $\mathrm{e}^{-V(x)}$ on Ω	m_{ij}	area of σ_{ij}
π_i	stat. measure $\mathrm{e}^{-V(x_i)}$ on Ω_i	$\mathbf{h}_{ij}$	x_i-x_j
u_i	$u(x_i)$	h_{ij}	$\lvert\mathbf{h}_{ij}\rvert$
$\bar f_i$	$\frac{1}{\lvert\Omega_i\rvert}\int_{\Omega_i}f\mathrm{d}x$	$d_{i,ij}$	$\mathrm{dist}\left(x_i,\sigma_{ij}\right)$
f_i	$m_i\bar f_i$	$\mathrm{diam}\mathcal{T}$	diameter, i.e. $\sup_{i\sim j}\lvert x_i-x_j\rvert$
$\mathbf{J}$	$-\kappa\left(\nabla u+u\nabla V\right)$	$J^S_{ij}U$	$-\frac{\kappa_{ij}}{h_{ij}}S_{ij}\left(U_j-U_i\right)$

Table 1: Commonly used notations.

In this work, we consider function with

$$\text{discrete homogeneous Dirichlet boundary conditions:}\qquad \forall\sigma\in\mathcal{E}_\partial:\quad u(y_\sigma)=0\,. \tag{2.1}$$

We write the latter also as $u_i=0$ if $x_i\in\mathcal{P}\backslash\tilde{\mathcal{P}}$. Hence, in what follows we write

$$\forall x_i\in\tilde{\mathcal{P}}:\ a_i:=a(x_i)\,,\qquad\text{with}\quad\sum_i a_i:=\sum_{x_i\in\tilde{\mathcal{P}}}a(x_i)\,.$$

For $w\in\mathcal{E}^*$ we write $w_{ij}:=w(\sigma)$ if $\sigma_{ij}=\sigma$. Then for fixed i the expression

$$\sum_{j:\,i\sim j}w_{ij}:=\sum_{\sigma_{ij}\in\mathcal{E}_i}w_{ij}$$

is the sum over all w_{ij} such that $\mathcal{E}_i\cap\mathcal{E}_j\neq\emptyset$ and

$$\sum_i\sum_{j:\,i\sim j}w_{ij}=\sum_{j\sim i}w_{ij}:=\sum_{\sigma_{ij}\in\mathcal{E}}w_{ij}:=\sum_{\sigma\in\mathcal{E}}w(\sigma)$$

is the sum over all edges.

Moreover, we define the diameter of a triangulation $\mathcal{T}$ as

$$\mathrm{diam}\mathcal{T}=\sup_{i\sim j}\lvert x_i-x_j\rvert.$$

The identity

$$\sum_i\sum_{j:j\sim i}A_{ij}=\sum_{j\sim i}\left(A_{ij}+A_{ji}\right) \tag{2.2}$$

will frequently be used throughout this paper, where we often encounter the case $A_{ij}=\alpha_{ij}U_i$ with $\alpha_{ij}=-\alpha_{ji}$:

$$\sum_i\sum_{j:j\sim i}\alpha_{ij}U_i=\sum_{j\sim i}\left(\alpha_{ij}U_i+\alpha_{ji}U_j\right)=\sum_{j\sim i}\alpha_{ij}\left(U_i-U_j\right)\ . \tag{2.3}$$

Formula (2.2) in particular allows for a discrete integration by parts for functions satisfying (2.1):

$$\sum_i \sum_{j:j\sim i} (U_j - U_i) U_i = \sum_{j\sim i} ((U_j - U_i) U_i + (U_i - U_j) U_j) = -\sum_{j\sim i} (U_j - U_i)^2 . \tag{2.4}$$

On a given mesh $\mathcal{T} = (\mathcal{V}, \mathcal{E}, \mathcal{P})$, we consider the linear discrete operator $\mathcal{L}_\kappa^\mathcal{T} : \mathcal{P}^* \to \mathcal{P}^*$, which is defined by a family of non-negative weights $\kappa : \mathcal{E} \to \mathbb{R}$ and acts on functions $u \in \mathcal{P}^*$ via

$$\forall x_i \in \mathcal{P} : \left(\mathcal{L}_\kappa^\mathcal{T} u\right)_i := \sum_{i\sim j} \kappa_{ij} \frac{m_{ij}}{h_{ij}} (u_j - u_i) . \tag{2.5}$$

While (2.5) is very general, it is shown in [22], Lemma 3.3, that the property (iv) of Definition 2.1 comes up with some special consistency properties for the choice of

$$\kappa_{ij} := \frac{\bar\kappa_i \bar\kappa_j}{\bar\kappa_i \frac{d_{i,ij}}{h_{ij}} + \bar\kappa_j \frac{d_{j,ij}}{h_{ij}}} , \tag{2.6}$$

where $d_{i,ij}$ and $d_{j,ij}$ are the distances between σ_{ij} and x_i and x_j respectively and averaged diffusion coefficient is defined by $\overline{\kappa}_i = m_i^{-1} \int_{\Omega_i} \kappa(x) \mathrm{d}x$.

Lemma 2.2 (A consistency lemma, [22]). *Let the $\mathcal{T} = (\mathcal{V}, \mathcal{E}, \mathcal{P})$ satisfy Definition 2.1 (i)–(v) and let $d \in \{2,3\}$ and let h_{ij} be uniformly bounded from above and from below. Then for every $u \in H^2(\Omega)$ it holds*

$$\left| \int_{\sigma_{ij}} \kappa \nabla u \cdot \boldsymbol{\nu}_{ij} - \kappa_{ij} \frac{m_{ij}}{h_{ij}} (u(x_j) - u(x_i)) \right| \le C m_{ij}^{\frac{1}{2}} h_{ij}^{\frac{1}{2}} \|u\|_{H^2(\Omega_i \cup \Omega_j)} .$$

Lemma 2.2 was one of the motivations to provide a more general and powerful concept of consistency in [8], as we will discuss in Section 2.5

2.2 Poincaré inequalities

In order to derive the a priori estimates in Section 2.3 we need to exploit (discrete) Poincaré inequalities to estimate $\|u\|_{L^2(\Omega)}$ by $\|\nabla u\|_{L^2(\Omega)}$ or $\|u^\mathcal{T}\|_{L^2(\mathcal{P})}$ by $\|Du^\mathcal{T}\|_{L^2(\mathcal{E})}$, where $(Du^\mathcal{T})_{ij} = u_j - u$. In particular, we use the following theorem which can be found e.g. in [16] or can be proved using Lemma A.1 applied to piecewise constant functions on the cells with $C_\# \le \frac{\mathrm{diam}\Omega}{h_0}$ and the choice $|\boldsymbol{\eta}| > \mathrm{diam}\Omega$.

Theorem 2.3. *Given a mesh $\mathcal{T} = (\mathcal{V}, \mathcal{E}, \mathcal{P})$ let $h_{\inf} := \inf\{|x-y| : (x,y) \in \mathcal{P}^2\} > 0$ and $h_{\sup} := \sup\{|x-y| : (x,y) \in \mathcal{P}^2\} > 0$ correspondingly. Then for every $u \in L^2(\mathcal{P})$ satisfying (2.1) and for every $\boldsymbol{\eta} \in \mathbb{R}^d$ it holds*

$$\int_\Omega \left| \sum_i u_i \chi_{\Omega_i}(x) - \sum_i u_i \chi_{\Omega_i}(x + \boldsymbol{\eta}) \right|^2 dx \le |\boldsymbol{\eta}| \left(\mathrm{diam}\Omega \frac{h_{\sup}}{h_{\inf}} \sum_{i\sim j} \frac{m_{ij}}{h_{ij}} (u_j - u_i)^2 \right) , \tag{2.7}$$

and particularly

$$\|u\|_{L^2(\mathcal{P})}^2 \le (\mathrm{diam}\Omega)^2 \frac{h_{\sup}}{h_{\inf}} \sum_{i\sim j} m_{ij} (u_j - u_i)^2 . \tag{2.8}$$

2.3 Existence and a priori estimates

In what follows, we study the properties of (1.4)–(1.5). Putting $U \equiv 1$ in both of these equations, we immediately see that the Boltzmann distribution $u_i := \pi_i = \exp(-V(x_i)) = \exp(-V_i)$, resp. the continuous version $u = \pi$ is the stationary solution for $f = 0$. Hence, from the standard theory of elliptic systems ([14] Chapter 6), we have the following theorem.

Theorem 2.4. *Let Ω be as above and $f \in \mathrm{L}^2(\Omega)$, $\kappa \in C^1\left(\overline{\Omega} : \mathbb{R}^{d\times d}\right)$ such that κ is uniformly bounded, symmetric and elliptic and $V \in \mathrm{C}^2(\overline{\Omega})$. Then there is a unique $u \in \mathrm{H}^2(\Omega) \cap \mathrm{H}^1_0(\Omega)$ solving $-\nabla\cdot(\kappa\nabla u) - \nabla\cdot(\kappa u\nabla V) = f$ in the weak sense.*

Furthermore, we find the following.

Theorem 2.5. *Let $\mathcal{T} = (\mathcal{V}, \mathcal{E}, \mathcal{P})$ be an admissible mesh in the sense of Definition 2.1 and let $c > 0$ such that $\kappa_{ij} > c$ for every i, j. Furthermore, let $\pi > 0$. Then there exists a unique solution $U^{\mathcal{T}} \in L^2(\tilde{\mathcal{P}})$ to* (1.4) *satisfying discrete homogeneous Dirichlet boundary conditions* (2.1)*.*

Proof. Multiplying (1.4) with $\phi \in L^2(\tilde{\mathcal{P}})$ and applying (2.4) we find

$$\begin{aligned}\sum_i f_i\phi_i &= \sum_i -\phi_i \sum_{j:j\sim i} \frac{m_{ij}}{h_{ij}}\kappa_{ij}S_{ij}\left(U_j^{\mathcal{T}} - U_i^{\mathcal{T}}\right)\\ &= \sum_{i\sim j} \frac{m_{ij}}{h_{ij}}\kappa_{ij}S_{ij}\left(U_j^{\mathcal{T}} - U_i^{\mathcal{T}}\right)(\phi_j - \phi_i)\ .\end{aligned}$$

The right hand side is a strictly positive symmetric bilinear form in $L^2(\tilde{\mathcal{P}})$ due to the Poincaré inequality (2.8). Hence there exists a unique solution $U^{\mathcal{T}}$ by the Lax–Milgram theorem. □

Having shown the existence of solutions to (1.5) and (1.4), we recall the derivation of some natural a priori estimates for both the continuous Fokker–Planck equation and the discretization.

Continuous FPE Let u, resp. $U = u/\pi$, be a solution of the stationary Fokker–Planck equation (1.5) with homogeneous Dirichlet boundary conditions. Testing with U, we get from a standard calculation that

$$\int_\Omega \tfrac{1}{\kappa\pi}|\kappa\pi\nabla U|^2 \le C\int_\Omega f^2\,. \tag{2.9}$$

Furthermore, the standard theory of elliptic equations (e.g., [14]) yields that $\|U\|_{H^2(\Omega)} \le C\,\|f\|_{L^2}$, where C depends on the C^1-norm of $\kappa\pi$ and the Poincaré-constant.

Discrete FPE Let $U_i^{\mathcal{T}}$ be a solution of (1.4) with $f_i = m_i\bar{f}_i = \int_{\Omega_i} f\,\mathrm{d}x$ (as specified in the Tab. 1), i.e.,

$$\forall i: \qquad -\sum_{j:j\sim i}\frac{m_{ij}}{h_{ij}}\kappa_{ij}S_{ij}\left(U_j^{\mathcal{T}} - U_i^{\mathcal{T}}\right) = m_i\bar{f}_i\,.$$

Then, multiplying with $U_i^{\mathcal{T}}$, summing over all $x_i \in \mathcal{P}$ and using (2.4), we conclude with help of the discrete Poincaré inequality (see Theorem 2.3 below)

$$\begin{aligned}\sum_{j\sim i}\frac{m_{ij}}{h_{ij}}\kappa_{ij}S_{ij}\left(U_j^{\mathcal{T}} - U_i^{\mathcal{T}}\right)^2 &= \sum_i m_i\bar{f}_iU_i^{\mathcal{T}} \le \sum_i\left((U_i^{\mathcal{T}})^2 m_i + \tfrac{1}{\pi_i}\bar{f}_i^2 m_i\right)\\ \Rightarrow \sum_{j\sim i}\frac{m_{ij}}{h_{ij}}\kappa_{ij}S_{ij}\left(U_j^{\mathcal{T}} - U_i^{\mathcal{T}}\right)^2 &\le C\sum_i m_i\bar{f}_i^2.\end{aligned}$$

The last estimate can be rewritten as

$$\sum_{j\sim i}\frac{m_{ij}}{h_{ij}}\kappa_{ij}\frac{1}{S_{ij}\kappa_{ij}^2}\left(\kappa_{ij}S_{ij}(U_j^{\mathcal{T}}-U_i^{\mathcal{T}})\right)^2\le C\sum_i \bar{f}_i^2 m_i. \tag{2.10}$$

2.4 Gradients, Fluxes and L^2-spaces

In order to derive and formulate variational consistence errors for the discrete FPE (1.4), we introduce the discrete fluxes

$$\begin{aligned} J_{ij}^S U^{\mathcal{T}} &:= -\frac{\kappa_{ij}}{h_{ij}}S_{ij}\left(U_j^{\mathcal{T}}-U_i^{\mathcal{T}}\right),\\ \overline{J}_{ij}U &:= -\frac{1}{m_{ij}}\int_{\sigma_{ij}}\kappa\pi\nabla U\cdot\boldsymbol{\nu}_{ij}\,. \end{aligned} \tag{2.11}$$

In particular, if $S_{ij}=\sqrt{\pi_i\pi_j}$ we get the flux of the SQRA $J_{ij}^{\text{SQRA}}U^{\mathcal{T}} := -\kappa_{ij}\sqrt{\pi_i\pi_j}\left(U_j^{\mathcal{T}}-U_i^{\mathcal{T}}\right)/h_{ij}$. The quantity $J_{ij}^S U^{\mathcal{T}}$ can indeed be considered as a flux in the sense that it will be shown to approximate $\overline{J}_{ij}$, S_{ij} is a discrete approximation of $\pi|_{\sigma_{ij}}$, κ_{ij} is a discrete approximation of $\kappa|_{\sigma_{ij}}$. The differences $\left(U_j^{\mathcal{T}}-U_i^{\mathcal{T}}\right)/h_{ij}$ take the role of gradients ∇U in the continuous problem and hence we refer to them as *discrete gradients* even though they are 1-dimensional objects.

While former approaches focus on the rate of convergence of $\left(u_j^{\mathcal{T}}-u_i^{\mathcal{T}}\right)/h_{ij}\to\nabla u$, we additionally follow the approach of [8] applied to U and are interested in the rate of convergence of $J_{ij}^S U^{\mathcal{T}}\to\mathbf{J}(U)$, which is an indirect approach to the original problem as this rate of convergence is directly related to $\left(U_j^{\mathcal{T}}-U_i^{\mathcal{T}}\right)/h_{ij}\to\nabla U$.

In view of the natural norms for the variational consistency (see (2.16) f.f.), we introduce the following

$$\begin{aligned} &\forall U\in L^2(\Omega): &&\|U\|_{L^2(\Omega)}^2 := \int_\Omega U^2\mathrm{d}x &&\|U\|_{L^2_\pi(\Omega)}^2 := \int_\Omega \tfrac{1}{\pi}U^2\mathrm{d}x\\ &\forall U\in\mathcal{P}^*: &&\|U\|_{L^2(\mathcal{P})}^2 := \sum_{i\in\mathcal{P}} m_i U_i^2 &&\|U\|_{L^2_\pi(\mathcal{P})}^2 := \sum_{i\in\mathcal{P}} m_i\tfrac{1}{\pi_i}U_i^2\\ &\forall J\in\mathcal{E}^*: &&\|J\|_{L^2(\mathcal{E})}^2 := \sum_{i\sim j} m_{ij}h_{ij}J_{ij}^2 &&\|J\|_{L^2_S(\mathcal{E})}^2 := \sum_{i\sim j} m_{ij}h_{ij}\frac{1}{S_{ij}}J_{ij}^2 \end{aligned} \tag{2.12}$$

Let us introduce the discrete flux $J^S U^{\mathcal{T}}\in\mathcal{E}^*$ via $J^S U^{\mathcal{T}}(\sigma_{ij}) := J_{ij}^S U^{\mathcal{T}}$ and similarly also $\frac{1}{\kappa}J^S U^{\mathcal{T}}\in\mathcal{E}^*$ via $J^S U^{\mathcal{T}}(\sigma_{ij}) := \frac{1}{\kappa_{ij}}J_{ij}^S U^{\mathcal{T}}$. With all the above notations, our a priori estimates (2.9) and (2.10) now read

$$\begin{aligned} \|\frac{1}{\sqrt{\kappa}}\mathbf{J}(U)\|_{L^2_\pi(\Omega)}^2 &\le C\|f\|_{L^2_\pi(\Omega)}^2\\ \left\|\frac{1}{\sqrt{\kappa}}J^S U^{\mathcal{T}}\right\|_{L^2_S(\mathcal{E})}^2 &\le C\|\bar{f}\|_{L^2_\pi(\mathcal{P})}^2\,. \end{aligned}$$

Assuming that the diffusion coefficient is bounded, i.e. $\kappa^*\ge\kappa\ge\kappa_*$, we further get

$$\begin{aligned} \tfrac{1}{\kappa_*}\|\mathbf{J}(U)\|_{L^2_\pi(\Omega)}^2 &\le C\|f\|_{L^2_\pi(\Omega)}^2\\ \frac{1}{\kappa^*}\|J^S U^{\mathcal{T}}\|_{L^2_S(\mathcal{E})}^2 &\le C\|\bar{f}\|_{L^2_\pi(\mathcal{P})}^2\,. \end{aligned}$$

Remark 2.6 (Naturalness of norms)**.** Let us discuss why these norms are natural to consider. The left norms in (2.12) can be interpreted as the Euclidean L^2-norms on Ω, $\mathcal{P}$ and $\mathcal{E}$, while the right norms are the natural norms for the study of the Fokker–Planck equation as they are weighted with the inverse of the Boltzmann distribution π, resp. π_i. Note that assuming V is bounded from above and below, the L^2-norms $\|\cdot\|_{L^2_\pi(\Omega)}$ and $\|\cdot\|_{L^2(\Omega)}$ are equivalent and the same holds true for the two norms in the discrete setting.

Given a discretization $\mathcal{T}$, the linear map

$$C_c\left(\mathbb{R}^d\right) \to \mathbb{R}\,, \qquad f \mapsto \sum_{i\in\mathcal{P}} m_i f(x_i)$$

defines an integral on Ω w.r.t. a discrete measure $\mu_\mathcal{T}$ having the property that $\mu_\mathcal{T} \to \mathcal{L}^d$ vaguely, where $\mathcal{L}^d$ is the d-dimensional Lebesgue measure. In particular $\mu_\mathcal{T}(A) \to \mathcal{L}^d\left(A\right)$ for every bounded measurable set with $\mathcal{L}^d\left(\partial A\right) = 0$. The norm $\|U\|^2_{L^2(\mathcal{P})}$ is simply the L^2-norm based on the measure $\mu_\mathcal{T}$.

Similar considerations work also for the norm on $\mathcal{E}^*$. The norm $\|\cdot\|^2_{L^2(\mathcal{E})}$ is given via a measure $\tilde{\mu}_\mathcal{T}$ having the property

$$\tilde{\mu}_\mathcal{T} : C_c\left(\mathbb{R}^d\right) \to \mathbb{R}\,, \qquad f \mapsto \sum_{i\sim j} m_{ij} h_{ij} f(x_{ij})\,,$$

with the property that $\tilde{\mu}_\mathcal{T} \to d\cdot\mathcal{L}^d$ vaguely: every Voronoï cell Ω_i consists of disjoint cones with mass $\frac{1}{d} m_{ij} h_{ij}$, where one has to account for all cones with $j \sim i$. In particular, we obtain $\tilde{\mu}_\mathcal{T}(A) \approx d\cdot\mathcal{L}(A)$ for Lipschitz domains – an estimate which then becomes precise in the limit. Without going into details, let us mention that heuristically the prefactor d balances the fact that $J_{ij} \approx \frac{(\mathbf{x}_i - \mathbf{x}_j)}{|x_i - x_j|}\cdot\nabla U$ which yields for functions $U \in C^1_c\left(\mathbb{R}^d\right)$:

$$\sum_{i\sim j} m_{ij} h_{ij} \left|\frac{(\mathbf{x}_i - \mathbf{x}_j)}{|x_i - x_j|}\cdot\nabla U\right|^2 \to \int_{\mathbb{R}^d} |\nabla U|^2\,.$$

For the particular case of a rectangular mesh, this is straight forward to verify.

2.5 Consistency and inf-sup stability

Results such as Lemma 2.2 motivated the authors of the recent paper [8] to define the concepts of consistency and inf-sup stability as discussed in the following. For readability, we will restrict the general framework of [8] to cell-centered finite volume schemes and refer to general concepts only as far as needed.

Definition 2.7 (inf-sup stability)**.** A bilinear form $a_\mathcal{T}$ on $L^2\left(\mathcal{P}\right)$ for a given mesh $\mathcal{T} = \left(\mathcal{V}, \mathcal{E}, \mathcal{P}\right)$ is called *(uniformly) inf-sup stable* with respect to a norm $\|\cdot\|_{H_\mathcal{T}}$ on a subspace $H_\mathcal{T} \subset L^2\left(\mathcal{P}\right)$ if there exists $\gamma > 0$ (independent from $\mathcal{T}$) such that

$$\forall u \in H_\mathcal{T} : \quad \gamma\, \|u\|_{H_\mathcal{T}} \le \sup_{v\in H_\mathcal{T}} \frac{a_\mathcal{T}\left(u,v\right)}{\|v\|_{L^2(\mathcal{P})}}\,.$$

Usually, and particularly in our setting, $a_\mathcal{T}$ is the discretization of a continuous bilinear form, say e.g. $a\left(u,v\right) = \int_\Omega \nabla u\cdot\left(\kappa\nabla v\right)$. We are interested in discretizing the problem

$$\forall v \in H^1_0\left(\Omega\right) : \quad a\left(u,v\right) = l\left(v\right)\,, \tag{2.13}$$

where $l : H^1_0\left(\Omega\right) \to \mathbb{R}$ is a continuous linear map, and in the convergence of the solutions $u_\mathcal{T}$ of the discrete problems

$$\forall v \in L^2\left(\mathcal{T}\right) : \quad a_\mathcal{T}\left(u_\mathcal{T}, v\right) = l_\mathcal{T}\left(v\right) \tag{2.14}$$

to the solutions u for (2.13).

Definition 2.8 (Consistency)**.** Let $B \subset H_0^1(\Omega)$ be a continuously embedded Banach subspace and for given $\mathcal{T} = (\mathcal{V}, \mathcal{E}, \mathcal{P})$ consider continuous linear operators $\mathcal{R}_\mathcal{T} : B \to L^2(\mathcal{P})$ with uniform bound. Let u be the solution to the linear equation (2.13) and let $l_\mathcal{T} : L^2(\mathcal{P}) \to \mathbb{R}$ be a family of linear functionals. The *variational consistency error* of $u \in B$ is the linear form $\mathfrak{E}_\mathcal{T}(u; \cdot) : L^2(\mathcal{P}) \to \mathbb{R}$ where

$$\mathfrak{E}_\mathcal{T}(u; \cdot) := l_\mathcal{T}(\cdot) - a_\mathcal{T}(\mathcal{R}_\mathcal{T} u, \cdot) .$$

Let now a family $(\mathcal{T}, a_\mathcal{T}, l_\mathcal{T})$ with $\mathrm{diam}\mathcal{T} \to 0$ be given and consider the corresponding family of linear discrete problems (2.14) and let $u \in B$ be a solution of (2.13). We say that

$$\text{consistency holds if} \quad \|\mathfrak{E}_\mathcal{T}(u; \cdot)\|_{H_\mathcal{T}^*} := \sup_{v \in H_\mathcal{T} \setminus \{0\}} \frac{|\mathfrak{E}_\mathcal{T}(u; v)|}{\|v\|_{H_\mathcal{T}}} \to 0 \quad \text{as} \quad \mathrm{diam}\mathcal{T} \to 0 .$$

Remark 2.9. A typical situation is the case $d \leq 3$, where $H^2(\Omega) \cap H_0^1(\Omega) \hookrightarrow C_0(\Omega)$ continuously. We then might set $B = H^2(\Omega) \cap H_0^1(\Omega)$ and $(\mathcal{R}_\mathcal{T} u)_i := u(x_i)$.

Consistency measures the rate at which $\mathcal{R}_\mathcal{T} u - u_\mathcal{T} \to 0$ and particularly provides a positive answer to the question whether the numerical scheme converges, at least if the solution of (2.13) lies in B. This is formulated in Theorem 10 of [8].

Theorem 2.10 (Theorem 10, [8])**.** *Using the above notation, it holds*

$$\|u_\mathcal{T} - \mathcal{R}_\mathcal{T} u\|_{H_\mathcal{T}} \leq \gamma^{-1} \|\mathfrak{E}_\mathcal{T}(u; \cdot)\|_{H_\mathcal{T}^*} \tag{2.15}$$

In our setting, $\|\cdot\|_{H_\mathcal{T}} = \|\cdot\|_{H_{\mathcal{T},\kappa}}$ (see (2.16)) is a norm on $L^2(\mathcal{P})$ defined in terms of the discrete gradients. By the discrete Poincaré inequality, (2.15) also implies a convergence estimate for the discrete solutions itself. The theorem can be understood as a requirement on the regularity of u, resp. the right hand side of (2.13) for convergence of the scheme.

We introduce

$$H_\mathcal{T} := \left\{ u \in L^2(\mathcal{P}) : u \text{ satisfies hom. Dir. b.c. (2.1)} \right\}$$

with the $H_\mathcal{T}$-norm through

$$\|u\|_{H_{\mathcal{T},\kappa}} := \sum_{i \sim j} \frac{m_{ij}}{h_{ij}} \kappa_{ij} (u_j - u_i)^2 \tag{2.16}$$

and find by the uniform bound $\kappa_{ij} > \kappa_0$ that $\|\cdot\|_{H_{\mathcal{T},\kappa}}$ and the following norms are equivalent:

$$\|u\|_{L^2, H_{\mathcal{T},\kappa}} := \|u\|_{H_{\mathcal{T},\kappa}} + \|u\|_{L^2(\mathcal{P})} , \qquad \|u\|_{L^2, H_\mathcal{T}} := \|u\|_{L^2, H_{\mathcal{T},1}} = \sum_{i \sim j} \frac{m_{ij}}{h_{ij}} (u_j - u_i)^2 + \|u\|_{L^2(\mathcal{P})} .$$

Due to the discrete Poincaré inequality (2.8), this holds uniformly, i.e. for every κ there exist constants $C_1, C_2, C_3, C_4 > 0$ independent from $\mathcal{T}$ such that for all functions $u \in L^2(\mathcal{P})$ with homogeneous Dirichlet boundary values

$$\|u\|_{H_{\mathcal{T},1}} \leq C_1 \|u\|_{L^2, H_\mathcal{T}} \leq C_2 \|u\|_{L^2, H_{\mathcal{T},\kappa}} \leq C_3 \|u\|_{H_{\mathcal{T},\kappa}} \leq C_4 \|u\|_{H_{\mathcal{T},1}} . \tag{2.17}$$

Using these relations, we can prove the following theorem for the bilinear discrete and continuous forms

$$a(u, v) = \int_\Omega \nabla u \cdot \kappa \nabla v + u \nabla V \cdot \kappa \nabla v ,$$

$$a_\mathcal{T}(u, v) = \sum_{i \sim j} \frac{m_{ij}}{h_{ij}} S_{i,j} \kappa_{ij} \left(\frac{u_j}{\pi_j} - \frac{u_i}{\pi_i} \right) (v_j - v_i) .$$

We furthermore need the following relation

$$a_1b_1 - a_2b_2 = \frac{1}{2}\left(a_1 - a_2\right)\left(b_1 + b_2\right) + \frac{1}{2}\left(a_1 + a_2\right)\left(b_1 - b_2\right) . \tag{2.18}$$

Lemma 2.11. *Under the Assumption 1.1 the following holds: Let $\mathcal{T} = (\mathcal{V}, \mathcal{E}, \mathcal{P})$ be a family of meshes that satisfy Assumption 1.3 and inequalities* (2.8) *and* (2.17) *uniformly for functions $u \in H_\mathcal{T}$. Then $a_\mathcal{T}$ is uniformly inf-sup stable for $\|\cdot\|_{H_{\mathcal{T},\omega}}$, where $\omega = \kappa$ or $\omega = 1$ and where γ_ω in both cases depends on Ω, $\frac{h_{\sup}}{h_{\inf}}$, K, κ_0, $\|\pi\|_\infty$ and $\|\nabla\pi\|_\infty$.*

Proof. We first observe that (2.18) yields

$$U_j - U_i = \frac{u_j}{\pi_j} - \frac{u_i}{\pi_i} = \frac{1}{2}\frac{(\pi_i + \pi_j)}{\pi_i\pi_j}\left(u_j - u_i\right) + \frac{1}{2}\left(u_i + u_j\right)\left(\pi_i^{-1} - \pi_j^{-1}\right) .$$

Introducing $\bar{u}_{ij} := \frac{1}{2}\left(u_i + u_j\right)$ we obtain with the triangle inequality

$$2\sum_{i\sim j}\frac{m_{ij}}{h_{ij}}S_{i,j}\kappa_{ij}\left(\left(U_j - U_i\right)\left(U_j - U_i\right) + \bar{u}_{ij}^2\left(\pi_j^{-1} - \pi_i^{-1}\right)^2\right) \geq \sum_{i\sim j}\frac{m_{ij}}{h_{ij}}S_{i,j}\kappa_{ij}\frac{1}{4}\left(\frac{(\pi_i + \pi_j)}{\pi_i\pi_j}\right)^2\left(u_j - u_i\right)^2 .$$

Observing that

$$\sum_{i\sim j}\frac{m_{ij}}{h_{ij}}S_{i,j}\kappa_{ij}\bar{u}_{ij}^2\left(\pi_j^{-1} - \pi_i^{-1}\right)^2 \leq 2\sum_i U_i^2\sum_{j:j\sim i}\frac{m_{ij}}{h_{ij}}S_{i,j}\kappa_{ij}\pi_i^2\left(\pi_j^{-1} - \pi_i^{-1}\right)^2$$

and exploiting (2.17) we observe that

$$\sum_{i\sim j}\frac{m_{ij}}{h_{ij}}S_{i,j}\kappa_{ij}\left(U_j - U_i\right)\left(U_j - U_i\right) \geq C\left\|u\right\|_{H_{\mathcal{T},1}} ,$$

where C depends on Ω, $\frac{h_{\sup}}{h_{\inf}}$, K, κ_0, $\|\pi\|_\infty$ and $\|\nabla\pi\|_\infty$. On the other hand

$$\sum_{i\sim j}\frac{m_{ij}}{h_{ij}}S_{i,j}\kappa_{ij}\left(U_j - U_i\right)\left(U_j - U_i\right) = a_\mathcal{T}(u, U) \leq \sup_{v\in H_\mathcal{T}}\frac{a_\mathcal{T}(u,v)}{\|v\|_{L^2(\mathcal{P})}} ,$$

which together implies uniform inf-sup stability. □

Next we derive $\mathfrak{E}_\mathcal{T}$ in terms of κ, π and $\mathcal{T}$ and provide an estimate on $\mathfrak{E}_\mathcal{T}$. The main message of Lemma 2.12 is that the consistency error can be estimated by two separate expressions, one estimating the error contributed by the diffusive term and one estimating the error contributed by the convective term in the FPE.

Lemma 2.12. *Let $k \geq 1$ such that $H^k(\Omega)$ embeds into $C(\overline{\Omega})$ let $u \in H^k(\Omega) \cap H_0^1(\Omega)$ be a solution to*

$$\forall v \in H_0^1(\Omega): \quad a(u, v) = l(v) ,$$

where

$$l(v) = \int_\Omega f\, v , \qquad l_\mathcal{T}(v) = \sum_i f_i v_i .$$

Using the notation (2.11) *the consistency of u is given through*

$$\begin{aligned}\mathfrak{E}_{\mathcal{T},\mathrm{FPE},\kappa}(u; v) &= \sum_{i\sim j}\left(v_j - v_i\right)\left(m_{ij}J_{ij}^S U - m_{ij}\overline{J}_{ij}U\right) &(2.19)\\ &= \mathfrak{E}_{\mathcal{T},\kappa}(u; v) + \mathfrak{E}_{\mathcal{T},\kappa,\mathrm{conv}}(u; v) , &(2.20)\end{aligned}$$

where with $u_i = (\mathcal{R}_{\mathcal{T}}u)_i$

$$\mathfrak{E}_{\mathcal{T},\kappa}(u;v) = \sum_{i\sim j}(v_j - v_i)\left(\int_{\sigma_{ij}}\kappa\nabla u\cdot\boldsymbol{\nu}_{ij} - \frac{m_{ij}}{h_{ij}}\kappa_{ij}(u_j - u_i)\right), \tag{2.21}$$

$$\mathfrak{E}_{\mathcal{T},\kappa,\mathrm{conv}}(u;v) = \sum_{i\sim j}(v_j - v_i)\left(\int_{\sigma_{ij}}\kappa u\nabla V\cdot\boldsymbol{\nu}_{ij} - \frac{m_{ij}}{h_{ij}}\kappa_{ij}\left(\frac{S_{ij}-\pi_j}{\pi_j}u_j - \frac{S_{ij}-\pi_i}{\pi_i}u_i\right)\right). \tag{2.22}$$

In particular, for both $\omega = \kappa$ *or* $\omega = 1$ *we obtain*

$$\left\|\mathfrak{E}_{\mathcal{T},\mathrm{FPE},\kappa}(u;v)\right\|^2_{H^*_{\mathcal{T},\omega}} \le \sum_{i\sim j}\frac{h_{ij}}{m_{ij}}\omega_{ij}^{-1}\left(m_{ij}J^S_{ij}U - m_{ij}\overline{J}_{ij}U\right)^2 \tag{2.23}$$

$$\le 2\left\|\mathfrak{E}_{\mathcal{T},\kappa}(u;\cdot)\right\|^2_{H^*_{\mathcal{T},\omega}} + 2\left\|\mathfrak{E}_{\mathcal{T},\kappa,\mathrm{conv}}(u;\cdot)\right\|^2_{H^*_{\mathcal{T},\omega}} \tag{2.24}$$

$$\left\|\mathfrak{E}_{\mathcal{T},\kappa}(u;\cdot)\right\|^2_{H^*_{\mathcal{T},\omega}} \le |\mathfrak{E}|_{\mathcal{T},\kappa,\omega}(u) := \sum_{i\sim j}\frac{h_{ij}}{m_{ij}}\omega_{ij}^{-1}\left(\int_{\sigma_{ij}}\kappa\nabla u\cdot\boldsymbol{\nu}_{ij} - \frac{m_{ij}}{h_{ij}}\kappa_{ij}\left((\mathcal{R}_{\mathcal{T}}u)_j - (\mathcal{R}_{\mathcal{T}}u)_i\right)\right)^2, \tag{2.25}$$

$$\left\|\mathfrak{E}_{\mathcal{T},\kappa,\mathrm{conv}}(u;\cdot)\right\|^2_{H^*_{\mathcal{T},\omega}} \le |\mathfrak{E}|_{\mathcal{T},\kappa,\omega,\mathrm{conv}}(u) \tag{2.26}$$

$$:= \sum_{i\sim j}\frac{h_{ij}}{m_{ij}}\omega_{ij}^{-1}\left(\int_{\sigma_{ij}}\kappa u\nabla V\cdot\boldsymbol{\nu}_{ij} - \frac{m_{ij}}{h_{ij}}\kappa_{ij}\left(\frac{S_{ij}-\pi_j}{\pi_j}u_j - \frac{S_{ij}-\pi_i}{\pi_i}u_i\right)\right)^2. \tag{2.27}$$

Remark. The expression for $|\mathfrak{E}|_{\mathcal{T},\kappa}(u)$ was explicitly provided before in [8].

Proof. In what follows, we combine ideas of the proofs of Theorems 27 and 33 in [8]. However, since our grid and our coefficients have a simple structure, our calculations are much shorter. We first observe that the definition of f_i and (1.1) imply

$$f_i = \int_{\Omega_i} f = -\int_{\Omega_i}\nabla\cdot(\kappa\nabla u + \kappa u\nabla V).$$

Hence, Gauß' theorem yields

$$l_{\mathcal{T}}(v) = -\sum_i v_i\int_{\Omega_i}\nabla\cdot(\kappa\nabla u + \kappa u\nabla V) = \sum_{i\sim j}(v_j - v_i)\int_{\sigma_{ij}}(\kappa\nabla u + \kappa u\nabla V)\cdot\boldsymbol{\nu}_{ij}$$

and hence (2.19). By an abuse of notation we write $u_j := (\mathcal{R}_{\mathcal{T}}u)_j$ and $u_i = (\mathcal{R}_{\mathcal{T}}u)_i$ for simplicity. Then we obtain

$$S_{ij}\left(\frac{u_j}{\pi_j} - \frac{u_i}{\pi_i}\right) = (u_j - u_i) + \left(\frac{S_{ij}-\pi_j}{\pi_j}u_j - \frac{S_{ij}-\pi_i}{\pi_i}u_i\right)$$

and hence

$$\begin{aligned}\mathfrak{E}_{\mathcal{T},\mathrm{FPE},\kappa}(u;v) &= l_{\mathcal{T}}(\cdot) - a_{\mathcal{T}}(\mathcal{R}_{\mathcal{T}}u,\cdot)\\ &= \sum_{i\sim j}(v_j - v_i)\left(\int_{\sigma_{ij}}\kappa\nabla u\cdot\boldsymbol{\nu}_{ij} - \frac{m_{ij}}{h_{ij}}\kappa_{ij}(u_j - u_i)\right).\\ &\quad+\sum_{i\sim j}(v_j - v_i)\left(\int_{\sigma_{ij}}\kappa u\nabla V\cdot\boldsymbol{\nu}_{ij} - \frac{m_{ij}}{h_{ij}}\kappa_{ij}\left(\frac{S_{ij}-\pi_j}{\pi_j}u_j - \frac{S_{ij}-\pi_i}{\pi_i}u_i\right)\right).\end{aligned}$$

From here we conclude by direct calculation and by the definition of the dual norm $\|\cdot\|^2_{H^*_{\mathcal{T},\omega}}$. □

A particular focus of the calculations below will lie on the following structure.

Lemma 2.13. *Let $\mathcal{T} = (\mathcal{V},\mathcal{E},\mathcal{P})$ be a mesh and $d \le 4$. Let $g \in C(\overline{\Omega})$ and let $g^{\mathcal{T}} \in \mathcal{E}^*$ with $g^{\mathcal{T}}(\sigma_{ij}) = g_{ij}$. Then for every $v \in H^2(\Omega)$ it holds*

$$\sum_{i\sim j}\frac{h_{ij}}{m_{ij}}\omega_{ij}^{-1}\left(\int_{\sigma_{ij}}\kappa g\nabla v\cdot\boldsymbol{\nu}_{ij}-\frac{m_{ij}}{h_{ij}}g_{ij}\kappa_{ij}\left(v_j-v_i\right)\right)^2$$
$$\le\left(\sup_{i,j}|g_{ij}|\right)|\mathfrak{E}|_{\mathcal{T},\kappa,\omega}\left(v;\cdot\right)+\sum_{i\sim j}\frac{h_{ij}}{m_{ij}}\omega_{ij}^{-1}\left(\int_{\sigma_{ij}}\kappa\left(g-g_{ij}\right)\nabla v\cdot\boldsymbol{\nu}_{ij}\right)^2$$

Proof. We obtain

$$\frac{1}{2}\left|\int_{\sigma_{ij}}\kappa g\nabla v\cdot\boldsymbol{\nu}_{ij}-\frac{m_{ij}}{h_{ij}}g_{ij}\kappa_{ij}\left(v_j-v_i\right)\right|^2\le$$
$$\le\left|\int_{\sigma_{ij}}\kappa g\nabla v\cdot\boldsymbol{\nu}_{ij}-\int_{\sigma_{ij}}\kappa g_{ij}\nabla v\cdot\boldsymbol{\nu}_{ij}\right|^2+\left|\int_{\sigma_{ij}}\kappa g_{ij}\nabla v\cdot\boldsymbol{\nu}_{ij}-\frac{m_{ij}}{h_{ij}}g_{ij}\kappa_{ij}\left(v_j-v_i\right)\right|^2$$
$$\le\left|\int_{\sigma_{ij}}\kappa\left(g-g_{ij}\right)\nabla v\cdot\boldsymbol{\nu}_{ij}\right|^2+|g_{ij}|^2\left|\int_{\sigma_{ij}}\kappa\nabla v\cdot\boldsymbol{\nu}_{ij}-\frac{m_{ij}}{h_{ij}}\kappa_{ij}\left(v_j-v_i\right)\right|^2$$

This implies the claim. □

With regard to (2.15) and Lemma 2.2, the above considerations motivate the following definition.

Definition 2.14 (φ-consistency)**.** Let $\mathcal{T}_h = (\mathcal{V}_h,\mathcal{E}_h,\mathcal{P}_h)$ be a family of meshes with $\mathrm{diam}\mathcal{T}_h \to 0$ as $h \to 0$. We say that $\mathcal{T}_h$ is φ-consistent (satisfies φ-consistency) on the subspace $B \subset H_0^1(\Omega)$ if for every $u \in B$ there exists $C \ge 0$ such that for every $h > 0$

$$|\mathfrak{E}|_{\mathcal{T},\kappa,\omega}(u)\le C\left\|u\right\|_{H^2}\varphi\left(h\right)^2 .$$

Hence, we immediately obtain the following.

Proposition 2.15. *Let $d \le 3$. Under the assumptions of Lemma 2.2 and assuming $h_{ij} \le Ch$ for some constant $C > 0$ the mesh is φ-consistent with $\varphi(h) = h$, i.e.*

$$|\mathfrak{E}|_{\mathcal{T},\kappa,\kappa}(u)\le C\left\|u\right\|_{H^2}h^2 .$$

We say that the mesh is h-consistent.

In case the mesh is cubic, we even obtain the following [16].

Proposition 2.16. *Let $d \le 3$ and let the mesh be cubic with all cubes of equal size and let $\kappa \equiv 1$. Under the assumptions of Lemma 2.2 and assuming $h_{ij} \le Ch$ for some constant $C > 0$ the mesh is φ-consistent with $\varphi(h) = h^2$, i.e.*

$$|\mathfrak{E}|_{\mathcal{T},1,1}(u)\le C\left\|u\right\|_{H^2}h^4 .$$

We say that the mesh is h^2-consistent.

3 Derivation of the methods and formal comparison

In this section, we repeat the original derivation of the Scharfetter–Gummel scheme in a more general way and show that both the SG and the SQRA scheme are members of a huge family of discretization schemes. Then we provide a physically motivated derivation of the SQRA scheme which assigns the SQRA a special place in the family of Stolarsky discretizations.

As mentioned in the introduction, also the SG scheme takes a special role, which is of mathematical nature and will discussed in Section 5.2.

3.1 A family of discretization schemes

We Repeat the derivation of the SG scheme from a different point of view to reveal some additional structure and to put it into a broader context.

In one dimension, the Scharfetter–Gummel scheme for the discrete flux on the interval $[0,h]$ is derived under the assumption of constant flux J and constant diffusion coefficient κ on $[0,h]$. In particular, we consider the two-point boundary value problem

$$J=-\kappa\left(u'(x)+u(x)V'(x)\right)\qquad \text{on } [0,h],\qquad u(0)=u_0,\qquad u(h)=u_h, \tag{3.1}$$

for a general potential $V:[0,h]\to\mathbb{R}$ not necessarily assumed to be affine. The general solution reads

$$u(x)=-\left(\frac{1}{\kappa}J\int_0^x \mathrm{e}^V+u_0\mathrm{e}^{V_0}\right)\mathrm{e}^{-V(x)}.$$

The flux can be computed explicitly from the assumption $J=\text{const.}$ and setting $x=h$ in the above formula. This yields

$$J=-\kappa\frac{u_h\mathrm{e}^{V_h}-u_0\mathrm{e}^{V_0}}{\int_0^h\mathrm{e}^V}=-\kappa\frac{1}{h}\left(\frac{1}{h}\int_0^h\pi^{-1}\right)^{-1}\left(\frac{u_h}{\pi_h}-\frac{u_0}{\pi_0}\right)=-\kappa\pi_{\mathrm{mean}}\frac{1}{h}\left(\frac{u_h}{\pi_h}-\frac{u_0}{\pi_0}\right)$$

for the averaged $\pi_{\mathrm{mean}}=\left(\frac{1}{h}\int_0^h\pi^{-1}\right)^{-1}$, which clearly determines the constant flux along the edge. In particular, assuming that V is affine, i.e. $V(x)=\frac{V_h-V_0}{x_h-x_0}(x-x_0)+V_0$, one easily checks that $\pi_{\mathrm{mean}}=(V_h-V_0)/(\mathrm{e}^{V_h}-\mathrm{e}^{V_0})$, which yields the Scharfetter–Gummel discretization. However, a potential can also be approximated not by piecewise affine interpolation but in other ways, resulting in different means π_{mean}. We provide an example of such an approximation for the SQRA in the Appendix A.4.

Generalizing the later considerations to higher dimensions, we find for the flux on the edge between two neighboring points in the discretization from the one dimensional considerations of (2.11) the expression

$$J_{ij}^S u^{\mathcal{T}}:=-\frac{\kappa_{ij}}{h_{ij}}S_{ij}\left(\frac{u_j^{\mathcal{T}}}{\pi_j}-\frac{u_i^{\mathcal{T}}}{\pi_i}\right),$$

where κ_{ij} relates to κ and S_{ij} relates to π_{mean}.

We aim to express π_{mean} by means of the values π_0 and π_h at the boundaries. The choice of this average is non-trivial and determines the quality of the discretization scheme, as we will see below. In the present work, we focus on the (weighted) Stolarsky mean, putting $\pi_{\mathrm{mean}}=S(\pi_i,\pi_j)$ although there are also other means like general f-means ($M_f(x,y)=f\left(\left[f^{-1}(x)+f^{-1}(y)\right]/2\right)$ for a strictly

mean	α	β	$\alpha+\beta$	$S_{\alpha,\beta}(x,y)$	$B_{\alpha,\beta}(x)$
max	$+\infty$	1	$+\infty$	$\max(x,y)$	$\begin{cases} \mathrm{e}^{-x}, & x\le 0 \\ 1, & x>0 \end{cases}$
quadratic mean	4	2	6	$\sqrt{\frac{1}{2}(x^2+y^2)}$	$\sqrt{\frac{1}{2}(1+\mathrm{e}^{-2x})}$
arithmetic mean	2	1	3	$\frac{1}{2}(x+y)$	$\frac{1}{2}(1+\mathrm{e}^{-x})$
logarithmic mean	1	0	1	$(x-y)/\log(x/y)$	$\frac{1}{x}(1-\mathrm{e}^{-x})$
geometric mean (SQRA)	1	-1	0	$\sqrt{xy}$	$\mathrm{e}^{-x/2}$
Scharfetter–Gummel mean	0	-1	-1	$xy\log(x/y)/(x-y)$	$x/(\mathrm{e}^{x}-1)$
harmonic Mean	-2	-1	-3	$2xy/(x+y)$	$2/(\mathrm{e}^{x}+1)$
min	$-\infty$	1	$-\infty$	$\min(x,y)$	$\begin{cases} \mathrm{e}^{x}, & x\le 0 \\ 1, & x>0 \end{cases}$

Table 2: Several mean values expressed as Stolarsky means $S_{\alpha,\beta}$ with corresponding weight functions $B_{\alpha,\beta}$, see Eq. (1.6). The geometric mean corresponds to the SQRA scheme, the $S_{0,-1}$-mean to the Scharfetter–Gummel discretization.

increasing function f). The Stolarsky mean has the advantage that it is a closed formula for a broad family of popular means and that its derivatives can be computed explicitly.

The weighted Stolarsky mean $S_{\alpha,\beta}$ [47] is given as (1.3) whenever these expressions are well defined and continuously extended otherwise, i.e. $S_{\alpha,\beta}(x,x) = x$. We note the symmetry properties $S_{\alpha,\beta}(x,y) = S_{\alpha,\beta}(y,x) = S_{\beta,\alpha}(x,y)$. Interesting special limit cases are

$$S_{0,1}(x,y) = (x-y)/\log(x/y) = \Lambda(x,y)$$

(logarithmic mean), $S_{-1,1}(x,y) = \sqrt{xy}$ (geometric mean) and $S_{0,-1}(x,y) = xy/\Lambda(x,y)$ (Scharfetter–Gummel mean). A list of further Stolarsky means is given in Table 2.

An explicit calculation shows that $\partial_x^2 S_{0,-1}(x,x) = -(3x)^{-1}$ and $\partial_x^2 S_{-1,1}(x,x) = -(4x)^{-1}$. For the general Stolarsky mean $S_{\alpha,\beta}$ one obtains (see Appendix A.3)

$$\begin{aligned} \partial_x S_{\alpha,\beta}(x,x) = \partial_y S_{\alpha,\beta}(x,x) = \frac{1}{2}, \\ \partial_x^2 S_{\alpha,\beta}(x,x) = \partial_y^2 S_*(x,x) = -\partial_{xy}^2 S_*(x,x) = -\partial_{yx}^2 S_*(x,x) = \frac{1}{12x}(\alpha+\beta-3), \end{aligned} \tag{3.2}$$

particularly reproducing the above findings for $\partial_x^2 S_{0,-1}$ and $\partial_x^2 S_{-1,1}$.

Interestingly, the derivation of the SQRA in Section 2.2 of [32] relies on the assumption that the flux through a FV-interface has to be proportional to $\left(u_j^{\mathcal{T}}/\pi_j - u_i^{\mathcal{T}}/\pi_i\right)$ with the proportionality factor given by a suitable mean of π_i and π_j. The choice of $S_{-1,1}$ in [32] seems arbitrary, yet it yields very good results [49, 17, 11].

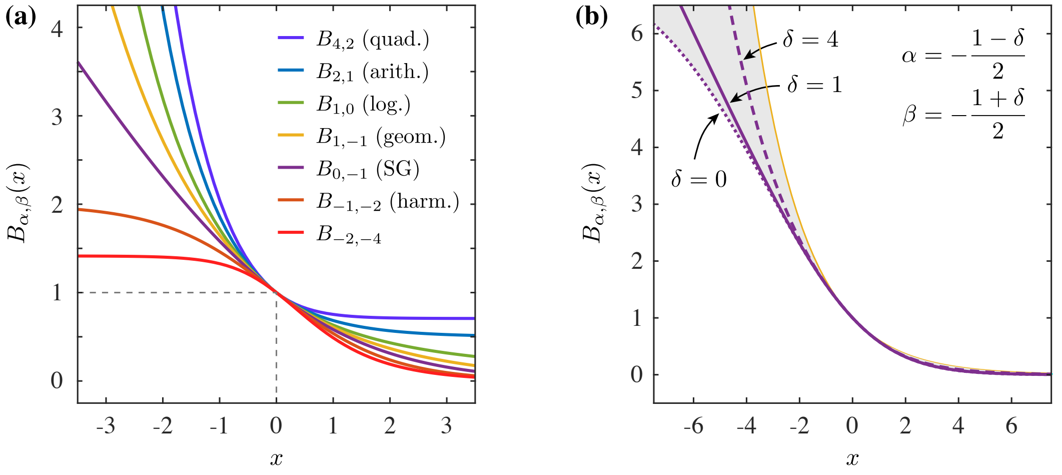

Fig. 1. (a) Weight functions $B_{\alpha,\beta}$ of the discrete flux scheme for different Stolarsky means $S_{\alpha,\beta}$ according to Eq. (1.6), cf. Tab. 2. **(b)** Weight functions for $\alpha + \beta = -1$ using the parametrization $\alpha = -(1-\delta)/2$, $\beta = -(1+\delta)/2$ for $\delta \geq 0$. The SG-mean $(\alpha, \beta) = (0, -1)$ is obtained for $\delta = 1$. The grey shaded region indicates the full range $\delta = [0, \infty)$, where the limit $\delta \to \infty$ is given by the weight function $\mathrm{e}^{-x/2}$ of the SQRA scheme.

3.2 The Wasserstein gradient structure of the Fokker–Planck operator and the SQRA method

The choice of S_* turns out to be crucial for the convergence properties. In this section, we look at physical structures which are desirable to be preserved in the discretization procedure. Our considerations are based on the variational structure of the Fokker–Planck equation. Let us note at this point that a physically reasonable discretization is not necessarily the best from the rate of convergence point of view. Indeed, this last point will be underlined by numerical simulations in Section 7. However, the physical consideration is helpful to understand the family of Stolarsky discretizations from a further, different point of view.

In [29] it was proved that the Fokker–Planck equation

$$\dot{u} = \nabla \cdot (\kappa \nabla u + \kappa u \nabla V) \tag{3.3}$$

has the gradient flow formulation $\dot{u} = \partial_\xi \Psi^*(u, -\mathrm{D}E(u))$ where

$$E(u) = \int_\Omega u \log u + Vu - u + 1 = \int_\Omega u \log\left(\frac{u}{\pi}\right) - u + 1\,, \qquad \Psi^*(u,\xi) = \frac{1}{2}\int_\Omega \kappa u \left|\nabla\xi\right|^2\,, \tag{3.4}$$

and $\pi = \mathrm{e}^{-V}$ is the stationary solution of (3.3). Indeed, one easily checks that $\mathrm{D}E(u) = \log u + V = \log(u/\pi)$ and $\partial_\xi \Psi^*(u,\xi) = -\nabla \cdot (\kappa u \nabla \xi)$ such that it formally holds

$$\partial_\xi \Psi^*(u,\xi)|_{\xi=-\mathrm{D}E(u)} = -\nabla \cdot (\kappa u \nabla\xi)|_{\xi=-\mathrm{D}E(u)} = \nabla \cdot \left(\kappa u \left(\frac{\nabla u}{u} + \nabla V\right)\right) = \nabla \cdot (\kappa \nabla u + \kappa u \nabla V) = \dot{u}.$$

Given a particular partial differential equation, the gradient structure might not be unique. For example, the simple parabolic equation $\partial_t u = \Delta u$ can be described by (3.4) with $V = 0$. But at the same time one might choose $E(u) = \int u^2$ with $\Psi^*(\xi) = \int |\nabla\xi|^2$, which plays a role in phase field modeling (see [26] and references therein) or $E(u) = -\int \log u$ with $\Psi^*(\xi) = \int u^2 |\nabla\xi|^2$.

In view of this observation, one might pose the question about “natural” gradient structures of the discretization schemes. This is reasonable if one believes that discretization schemes should incorporate the underlying physical principles. The energy functional is clearly prescribed by (3.4) with the natural discrete equivalent

$$E_{\mathcal{T}}(u) = \sum_i m_i \left(u_i \log\left(\frac{u_i}{\pi_i}\right) - u_i + 1 \right) . \tag{3.5}$$

The discrete linear evolution equation can be expected to be linear. Since we identified the continuous flux to be $\mathbf{J} = -\kappa\pi\nabla U$ with $U = u/\pi$, we expect the form

$$\dot{u}_i m_i = \partial_\xi \Psi^*_{\mathcal{T}}(u, -\mathrm{D}E_{\mathcal{T}}(u)) = \sum_{j:i\sim j} \frac{m_{ij}}{h_{ij}} \kappa_{i,j} \pi_{ij} \left(\frac{u_j}{\pi_j} - \frac{u_i}{\pi_i} \right) \tag{3.6}$$

for some suitably averaged π_{ij}. Equation (3.6) can be understood as a time-reversible (or detailed balanced) Markov process on the finite state space $\mathcal{P}$. Recently, various different gradient structures have been suggested for (3.6): [38, 35, 13, 7, 40] for a quadratic dissipation as a generalization of the Jordan–Kinderlehrer–Otto approach; and [43, 42], where a dissipation of cosh-type was appeared in the Large deviation rate functional for a hydrodynamic limit of an interacting particle system. All of them can be written in the abstract form

$$\Psi^*_{\mathcal{T}}(u, \xi) = \frac{1}{2} \sum_i \frac{1}{m_i} \sum_{j:i\sim j} \frac{m_{ij}}{h_{ij}} S_{ij} a_{ij}(u, \pi) \psi^* \left(\xi_i - \xi_j \right) , \tag{3.7}$$

where

$$a_{ij}(u, \pi) = \left(\frac{u_i}{\pi_i} - \frac{u_j}{\pi_j} \right) \partial_\xi \psi^* \left(\log\left(\frac{u_i}{\pi_i}\right) - \log\left(\frac{u_j}{\pi_j}\right) \right)^{-1} . \tag{3.8}$$

In fact, any positive and convex function ψ^* defines a reasonable dissipation functional Ψ^* by (3.7) and (3.8). A special case is when choosing for ψ^* and exponentially fast growing function $\psi^*(r) := \mathsf{C}^*(r) := 2\left(\cosh(r/2) - 1\right)$. Then a_{ij} simplifies to

$$a_{ij}(u, \pi) = \sqrt{\frac{u_i u_j}{\pi_i \pi_j}},$$

and hence, the square root appears. Choosing $S_{ij} = \sqrt{\pi_i \pi_j}$, we end up with a dissipation functional of the form

$$\Psi^*_{\mathcal{T}}(u, \xi) = \sum_i \sum_{j:i\sim j} m_{ij} h_{ij} \sqrt{u_i u_j} \, \frac{1}{h_{ij}^2} \mathsf{C}^* \left(\xi_i - \xi_j \right) . \tag{3.9}$$

There are (at least) three good reasons why choosing this gradient structure, i.e., modeling fluxes in exponential terms: a historical, a mathematical and a physical:

1 Already in Marcelin's PhD thesis from 1915 ([36]) exponential reaction kinetics have been derived, which are still common in chemistry literature.

2 Recently, convergence for families of gradient systems has been derived based on the energy-dissipation principle (the so-called EDP-convergence [41, 33, 12]). Vice versa, the above cosh-gradient structure appears as an effective gradient structure applying EDP-convergence to Wasserstein gradient flow problems [33, 21].

3 Recalling the gradient structure for the continuous Fokker–Planck equation (3.4), we observe that the dissipation mechanism Ψ^* is totally independent of the particular form of the energy

$\mathcal{E}$, which is determined by the potential V. This is physically understandable, since a change of the energy resulting, e.g., from external fields should not influence the dissipation structure. The same holds for the discretized version (3.9). In fact it was shown in [44], that the only discrete gradient structure, where the dissipation does not depend on V resp. $\pi = \mathrm{e}^{-V}$, is the cosh-gradient structure with the SQRA discretization $S_{ij} = S_{-1,1}(\pi_i, \pi_j)$. In particular, this characterizes the SQRA. For convenience, we add a proof for that to the Appendix A.2.

We think that these properties distinguish the SQRA, although in the following the convergence proofs do not really rely on the particular discretization weight S_{ij}.

Remark 3.1 (Convergence of energy and dissipation functional)**.** Let us finally make some comments on the convergence of $E_{\mathcal{T}}$ and $\Psi^*_{\mathcal{T}}$ given in (3.5) and (3.9) to the continuous analogies E and Ψ^*. Γ-convergence can be shown if the fineness of $\mathcal{T}$ tends to 0. For the energies it is clear, since $u \mapsto u\log(u/\pi) - u$ is convex. For the dissipation potentials $\Psi^*_{\mathcal{T}}(u,\xi)$ we observe the following: For smooth functions u and ξ, we have $\frac{1}{h_{ij}^2}\mathsf{C}^*(\xi_i - \xi_j) \approx \frac{1}{2}\left(\frac{\mathbf{x}_i - \mathbf{x}_j}{|x_i - x_j|}\cdot\nabla\xi\right)^2 + O(h_{ij}^2)$ and $\sqrt{u_i u_j} \approx u\left(\frac{1}{2}(x_i + x_j)\right)$. The considerations from Section 2.4 then yield $\Psi^*_{\mathcal{T}}(u,\xi) \approx \frac{1}{2}\int_{\boldsymbol{Q}} u\,|\nabla\xi|^2$.

For quadratic dissipation, qualitative convergence results using the underlying gradient structure and the energy-dissipation principle are obtained in [9] in 1 D, and in [20] for multiple dimensions. In [23] convergence of the associated metric is proved.

4 Comparison of discretization schemes

We mutually compare any two discretization schemes of the form (1.2) in case of Dirichlet boundary conditions. In this case, even though the problem is only defined on $\tilde{\mathcal{P}}$, we can simply sum over all $\mathcal{P}$ once we multiplied with a test function that assumes the value 0 at all $\mathcal{P}\backslash\tilde{\mathcal{P}}$.

Let us recall the formula (2.11) for the fluxes

$$J_{ij}^S U = -\frac{\kappa_{ij}}{h_{ij}} S_{ij}(U_j - U_i).$$

Moreover, let $u_i = U_i\pi_i$ and $\tilde{u}_i = \tilde{U}_i\pi_i$ be the solution of the discrete FPE (1.2) for two different smooth mean coefficients $S_{ij} = S(\pi_i, \pi_j)$ and $\tilde{S}_{ij} = \tilde{S}(\pi_i, \pi_j)$ (e.g. once for Scharfetter–Gummel and once for SQRA) such that

$$\sum_{k:k\sim i} m_{ik}h_{ik}J_{ik}^S U = m_i\bar{f}_i \tag{4.1}$$

$$\sum_{k:k\sim i} m_{ik}h_{ik}J_{ik}^{\tilde{S}}\tilde{U} = m_i\bar{f}_i. \tag{4.2}$$

In order to compare the solutions of (4.1) and (4.2) we take the difference of these two equations and multiply with $E_i = U_i - \tilde{U}_i$. We obtain

$$\begin{aligned} 0 &= \sum_i \sum_{k:k\sim i} m_{ik}h_{ik}\left(J_{ik}^S U - J_{ik}^{\tilde{S}}\tilde{U}\right)E_i \\ &= \sum_i \sum_{k:\,k\sim i} \frac{m_{ik}}{h_{ki}}\kappa_{ij}(S_{ik}(U_i - U_k) - \tilde{S}_{ik}(\tilde{U}_i - \tilde{U}_k))E_i \end{aligned}$$

Introducing the notation $\alpha_{ik} = \kappa_{ik}\frac{m_{ik}}{h_{ik}}$ and using (2.3) we get

$$\begin{aligned} 0 &= \sum_{k\sim i}\alpha_{ik}\left(S_{ik}(U_i - U_k) - S_{ik}(\tilde{U}_i - \tilde{U}_k) + \left(S_{ik} - \tilde{S}_{ik}\right)(\tilde{U}_i - \tilde{U}_k)\right)(E_i - E_k) \\ &= \sum_{k\sim i}\alpha_{ik}\left(S_{ik}\left(E_i - E_k\right) + (S_{ik} - \tilde{S}_{ik})\left(\tilde{U}_i - \tilde{U}_k\right)\right)\left(E_i - E_k\right). \end{aligned}$$

Using the notation $\mathrm{D}_{ik}A = A_k - A_i$ for discrete gradients

$$(\tilde{S}_{ik} - S_{ik})\left(\tilde{U}_i - \tilde{U}_k\right)(E_i - E_k) \le \frac{1}{2}\left[S_{ik}\left(\mathrm{D}_{ik}E\right)^2 + \frac{(S_{ik} - \tilde{S}_{ik})^2}{S_{ik}}\left(\mathrm{D}_{ik}\tilde{U}\right)^2\right]$$

we get

$$\frac{1}{2}\sum_{k\sim i}\alpha_{ik}S_{ik}\left(\mathrm{D}_{ik}E\right)^2 \le \frac{1}{2}\sum_{k\sim i}\frac{(\tilde{S}_{ik} - S_{ik})^2}{S_{ik}\tilde{S}_{ik}}\alpha_{ik}\tilde{S}_{ik}\left(\mathrm{D}_{ik}\tilde{U}\right)^2. \tag{4.3}$$

In the case of Stolarsky means the constants are more explicit. We have the following expansion of S_{ij}: writing $\pi_{ij} = \frac{1}{2}\left(\pi_i + \pi_j\right)$, $\pi_+ = \pi_- = \frac{1}{2}\left(\pi_i - \pi_j\right)$ and $\pi_i = \pi_0 + \pi_+$ and $\pi_j = \pi_0 - \pi_-$

$$\begin{aligned} S_{ij} &= S_{\alpha,\beta}\left(\pi_{ij},\pi_{ij}\right) + \frac{1}{2}\left(\pi_+ - \pi_-\right) + \frac{1}{2}\partial_x^2 S_{\alpha,\beta}\left(\pi_{ij},\pi_{ij}\right)\left(\pi_+ + \pi_-\right)^2 + O\left(\pi_\pm^3\right) \\ &= \pi_{ij} + \frac{\frac{1}{3}\left(\alpha+\beta\right) - 1}{8\,\pi_{ij}}\left(\pi_i - \pi_j\right)^2 + O\left(\pi_i - \pi_j\right)^3. \end{aligned} \tag{4.4}$$

In case $(\alpha+\beta) = \left(\tilde{\alpha}+\tilde{\beta}\right)$, we obtain $S_{ij} - \tilde{S}_{ij} = O\left(\pi_i - \pi_j\right)^3$ and hence this yields the following first comparison result:

Proposition 4.1. *Let $\mathcal{T}$ be a mesh with right hand side $f \in L^2(\mathcal{P})$ and let u and $\tilde{u}$ be a two solution of the discrete FPE for different Stolarsky mean coefficients $S_{ij} = S_{\alpha,\beta}\left(\pi_i,\pi_j\right)$ and $\tilde{S}_{ij} = S_{\tilde{\alpha},\tilde{\beta}}\left(\pi_i,\pi_j\right)$ respectively. Then*

$$\begin{aligned} &\frac{1}{2}\sum_{k\sim i}\kappa_{ik}\frac{m_{ik}}{h_{ik}}S_{ik}\left(U_k - \tilde{U}_k - \left(U_i - \tilde{U}_i\right)\right)^2 \\ &\qquad \le \frac{1}{2}\sum_{k\sim i}\left(\frac{\left((\alpha+\beta) - \left(\tilde{\alpha}+\tilde{\beta}\right)\right)^2}{24^2\,\pi_{ij}^2\tilde{S}_{ik}S_{ik}}\left(\pi_i - \pi_k\right)^4 + O\left(\pi_i - \pi_k\right)^5\right)\kappa_{ik}\frac{m_{ik}}{h_{ik}}\left(\tilde{U}_k - \tilde{U}_i\right)^2 \end{aligned}$$

In case $(\alpha+\beta) = \left(\tilde{\alpha}+\tilde{\beta}\right)$ we furthermore find

$$\frac{1}{2}\sum_{k\sim i}\kappa_{ik}\frac{m_{ik}}{h_{ik}}S_{ik}\left(U_k - \tilde{U}_k - \left(U_i - \tilde{U}_i\right)\right)^2 = \frac{1}{2}\sum_{k\sim i}O\left(\pi_i - \pi_k\right)^6\kappa_{ik}\frac{m_{ik}}{h_{ik}}\left(\tilde{U}_k - \tilde{U}_i\right)^2.$$

We aim to refine the above result to an order of convergence result for $J^S U - J^{\tilde{S}}\tilde{U}$.. We introduce a third Stolarsky mean $\hat{S}_{ik} = \hat{S}(\pi_i,\pi_k)$ and find

$$\begin{aligned} \hat{S}_{ik}\left(E_i - E_k\right) &= \hat{S}_{ik}\left(U_i - \tilde{U}_i - \left(U_k - \tilde{U}_k\right)\right) \\ &= S_{ik}(U_i - U_k) - S_{ik}(U_i - U_k) + \tilde{S}_{ik}(\tilde{U}_i - \tilde{U}_k) - \tilde{S}_{ik}(\tilde{U}_i - \tilde{U}_k) + \hat{S}_{ik}\left(U_i - \tilde{U}_i - \left(U_k - \tilde{U}_k\right)\right) \\ &= m_{ik}\alpha_{ik}^{-1}\left(J_{ik}^S U - J_{ik}^{\tilde{S}}\tilde{U}\right) + \left(\hat{S}_{ik} - S_{ik}\right)(U_i - U_k) - \left(\hat{S}_{ik} - \tilde{S}_{ik}\right)\left(\tilde{U}_i - \tilde{U}_k\right). \end{aligned}$$

Hence, we have

$$\sum_{k\sim i}\alpha_{ik}\left(S_{ik}(U_i-U_k)-\tilde{S}_{ik}\left(\tilde{U}_i-\tilde{U}_k\right)\right)(E_i-E_k)$$
$$=\sum_{k\sim i}\frac{h_{ik}}{\kappa_{ik}}m_{ik}\frac{1}{\hat{S}_{ik}}\left(J^S_{ik}U-J^{\tilde{S}}_{ik}\tilde{U}\right)^2$$
$$+\sum_{k\sim i}m_{ik}\frac{1}{\hat{S}_{ik}}\left(J^S_{ik}U-J^{\tilde{S}}_{ik}\tilde{U}\right)\left[\left(\hat{S}_{ik}-S_{ik}\right)(U_i-U_k)+\left(\hat{S}_{ik}-\tilde{S}_{ik}\right)\left(\tilde{U}_i-\tilde{U}_k\right)\right],$$

and using Cauchy–Schwartz inequality, we get

$$\sum_{k\sim i}\alpha_{ik}\left(S_{ik}(U_i-U_k)-\tilde{S}_{ik}\left(\tilde{U}_i-\tilde{U}_k\right)\right)(E_i-E_k)\le-\frac{1}{2}\sum_{k\sim i}\frac{h_{ik}m_{ik}}{\kappa_{ik}}\frac{1}{\hat{S}_{ik}}\left(J^S_{ik}U-J^{\tilde{S}}_{ik}\tilde{U}\right)^2$$
$$+\sum_{k\sim i}\frac{m_{ik}\kappa_{ik}}{h_{ik}\hat{S}_{ik}}\left(\left(\hat{S}_{ik}-S_{ik}\right)^2(U_i-U_k)^2+\left(\hat{S}_{ik}-\tilde{S}_{ik}\right)^2\left(\tilde{U}_i-\tilde{U}_k\right)^2\right).$$

Altogether we obtain

$$\frac{1}{2}\sum_{k\sim i}\frac{h_{ik}m_{ik}}{\kappa_{ik}}\frac{1}{\hat{S}_{ik}}\left(J^S_{ik}U-J^{\tilde{S}}_{ik}\tilde{U}\right)^2\le\sum_{k\sim i}\frac{m_{ik}h_{ik}}{\kappa_{ik}\hat{S}_{ik}S^2_{ik}}\left(\hat{S}_{ik}-S_{ik}\right)^2\left(\frac{\kappa_{ik}}{h_{ik}}S_{ik}\left(U_i-U_k\right)\right)^2$$
$$+\sum_{k\sim i}\frac{m_{ik}h_{ik}}{\kappa_{ik}\hat{S}_{ik}\tilde{S}^2_{ik}}\left(\hat{S}_{ik}-\tilde{S}_{ik}\right)^2\left(\frac{\kappa_{ik}}{h_{ik}}\tilde{S}_{ik}\left(\tilde{U}_i-\tilde{U}_k\right)\right)^2.$$

We make once more use of (4.4) writing $C_{\alpha,\beta}:=\frac{1}{24}(\alpha+\beta)$ and exploiting $\pi_i=\pi_{ij}+\pi_{ij}\left(V_i-V_{ij}\right)+O\left(V_i-V_{ij}\right)^2$ with

$$\pi_i-\pi_j=\pi_{ij}\left(V_i-V_j\right)+O\left(V_i-V_{ij}\right)^2+O\left(V_j-V_{ij}\right)^2$$
$$S_{ij}=\pi_{ij}+O\left(\pi_i-\pi_j\right).$$

Hence, we conclude the following result.

Theorem 4.2. *Let $\mathcal{T}$ be a mesh with right hand side $f\in L^2(\mathcal{P})$ and let u and $\tilde{u}$ be two solutions of the discrete FPE for different Stolarsky means S and $\tilde{S}$. Moreover, let $\hat{S}$ be any Stolarsky mean and assume that either $\alpha+\beta\neq\hat{\alpha}+\hat{\beta}$ or $\tilde{\alpha}+\tilde{\beta}\neq\hat{\alpha}+\hat{\beta}$. Then the solutions u and $\tilde{u}$ of the discretized FPE satisfy the symmetrized error estimate up to higher order*

$$\frac{1}{2}\sum_{k\sim i}\frac{h_{ik}m_{ik}}{\kappa_{ik}}\frac{1}{\hat{S}_{ik}}\left(J^S_{ik}U-J^{\tilde{S}}_{ik}\tilde{U}\right)^2\le\sum_{k\sim i}\frac{m_{ik}h_{ik}}{\kappa_{ik}S_{ik}}\left(C_{\alpha,\beta}-C_{\hat{\alpha},\hat{\beta}}\right)(V_i-V_j)^2\left(J^S_{ik}U\right)^2$$
$$+\sum_{k\sim i}\frac{m_{ik}h_{ik}}{\kappa_{ik}\tilde{S}_{ik}}\left(C_{\tilde{\alpha},\tilde{\beta}}-C_{\hat{\alpha},\hat{\beta}}\right)(V_i-V_j)^2\left(J^{\tilde{S}}_{ik}\tilde{U}\right)^2.$$

More general, for any mean we have

$$\frac{1}{2\kappa^*}\|J^SU-J^{\tilde{S}}\tilde{U}\|^2_{L^2_{\hat{S}}(\mathcal{E})}$$
$$\le\frac{1}{\kappa_*}\left\{\sup_{i,k}\frac{\left(\hat{S}_{ik}-S_{ik}\right)^2}{\hat{S}_{ik}S_{ik}}\left\|J^SU\right\|^2_{L^2_S(\mathcal{E})}+\sup_{i,k}\frac{\left(\hat{S}_{ik}-\tilde{S}_{ik}\right)^2}{\hat{S}_{ik}\tilde{S}_{ik}}\left\|J^{\tilde{S}}\tilde{U}\right\|^2_{L^2_{\tilde{S}}(\mathcal{E})}\right\},\quad(4.5)$$

and in particular for Stolarsky means with $\alpha+\beta=\tilde{\alpha}+\tilde{\beta}=\hat{\alpha}+\hat{\beta}$ we find the following result:

Corollary 4.3. *Let $\mathcal{T}$ be a mesh with right hand side $f \in L^2(\mathcal{P})$ and let u and $\tilde{u}$ be two solutions of the discrete FPE for different Stolarsky mean coefficients $S_{ij} = S_{\alpha,\beta}(\pi, \pi_j)$ and $\tilde{S}_{ij} = S_{\tilde{\alpha},\tilde{\beta}}(\pi, \pi_j)$ with $\alpha + \beta = \tilde{\alpha} + \tilde{\beta} = \hat{\alpha} + \hat{\beta}$. Then estimate* (4.5) *holds. In particular, we find the refined estimate*

$$\frac{1}{2\kappa^*}\|J^S U - J^{\tilde{S}}\tilde{U}\|^2_{L^2_{\hat{S}}(\mathcal{E})} = O\left(\pi_i - \pi_j\right)^6 \left(\left\|J^S U\right\|^2_{L^2_S(\mathcal{E})} + \left\|J^{\tilde{S}}\tilde{U}\right\|^2_{L^2_{\tilde{S}}(\mathcal{E})}\right).$$

In particular, the last result shows that convergence rates are similar up to order 3 for different α, β which satisfy $\alpha + \beta = \mathrm{const.}$

Corollary 4.4. *Let $\mathcal{T}$ be a mesh with right hand side $f \in L^2(\mathcal{P})$ and let u and $\tilde{u}$ be two solutions of the discrete FPE for different Stolarsky mean coefficients $S_{ij} = S_{\alpha,\beta}(\pi, \pi_j)$ and $\tilde{S}_{ij} = S_{\tilde{\alpha},\tilde{\beta}}(\pi, \pi_j)$ with $\alpha + \beta = \tilde{\alpha} + \tilde{\beta} = \hat{\alpha} + \hat{\beta}$. Then estimate* (4.5) *holds. For both S_{ij} and $\tilde{S}_{ij}$ let the quantities of Lemma 2.12 which depend on S be denoted by $\mathfrak{E}^S_{\mathcal{T},\mathrm{FPE},\kappa}(u;v)$ and $\mathfrak{E}^{\tilde{S}}_{\mathcal{T},\mathrm{FPE},\kappa}(u;v)$ as well as $\mathfrak{E}^S_{\mathcal{T},\kappa,\mathrm{conv}}(u;v)$ and $\mathfrak{E}^{\tilde{S}}_{\mathcal{T},\kappa,\mathrm{conv}}(u;v)$. If $\pi > c > 0$ is uniformly bounded from below then*

$$\left\|\mathfrak{E}^S_{\mathcal{T},\mathrm{FPE},\kappa}(u;v)\right\|^2_{H^*_{\mathcal{T},\omega}} \le 2\left\|\mathfrak{E}^{\tilde{S}}_{\mathcal{T},\mathrm{FPE},\kappa}(u;v)\right\|^2_{H^*_{\mathcal{T},\omega}} + O(h^6)\,.$$

Proof. We obtain from Lemma 2.12

$$\left\|\mathfrak{E}^S_{\mathcal{T},\mathrm{FPE},\kappa}(u;v)\right\|^2_{H^*_{\mathcal{T},\omega}} \le 2\sum_{i\sim j}\frac{h_{ij}}{m_{ij}}\omega_{ij}^{-1}\left(m_{ij}J^{\tilde{S}}_{ij}U - m_{ij}\overline{J}_{ij}U\right)^2 + 2\sum_{i\sim j}\frac{h_{ij}}{m_{ij}}\omega_{ij}^{-1}\left(m_{ij}J^{\tilde{S}}_{ij}U - m_{ij}J^S_{ij}U\right)^2$$

and from Corollary 4.3 we obtain the claim upon uniform boundedness of π. □

5 Convergence of the discrete FPE

In this section, we derive general estimates for the order of convergence of the Stolarsky FV operators. Throughout this section, we assume that the mesh satisfies the consistency property of Definition 2.14 with a suitable consistency function $\varphi : \mathbb{R}_{\ge 0} \to \mathbb{R}_{\ge 0}$ and discretization operator $\mathcal{R}_{\mathcal{T}} : H^1(\Omega) \supset B \to L^2(\mathcal{P})$. The parameters π_i and u_i below are then given in terms of

$$\pi_i = \left(\mathcal{R}_{\mathcal{T}}\pi\right)_i\,, \qquad u_i = \left(\mathcal{R}_{\mathcal{T}}u\right)_i\,, \qquad U_i = \left(\mathcal{R}_{\mathcal{T}}U\right)_i\,.$$

We derive consistency errors for U in Section (5.1) and consistency errors for u in Section (5.2). For both calculations we will need the following result.

Lemma 5.1. *Assume there exists a constant $C > 0$ such that for all cells Ω_i, Ω_j with $h_i = \mathrm{diam}\Omega_i$ it holds*

$$\forall f \in H^1(\Omega_i): \qquad \|f\|^2_{L^2(\sigma_{ij})} \le \frac{1}{h_i}C^2\|f\|^2_{H^1(\Omega_i)}\,, \tag{5.1}$$

$$\forall f \in H^1(\Omega_i)\cap C(\overline{\Omega}_i): \qquad \|f - f_i\|^2_{L^2(\sigma_{ij})} \le h_i C^2\|\nabla f\|^2_{L^2(\Omega_i)}\,. \tag{5.2}$$

Then for C^2-smooth Stolarsky means S_* and for every function $\varpi, U \in H^2(\Omega)$ with $\varpi_i := \varpi(x_i)$ and $S_{ij} := S_*(\varpi_i, \varpi_j)$ it holds

$$\left|\int_{\sigma_{ij}}\left(\varpi - S_{ij}\right)\kappa\nabla U\cdot\boldsymbol{\nu}_{ij}\right| \le C\begin{cases}\sum_{k=i,j}\|\nabla\varpi\|_{L^2(\Omega_k)}\|\kappa\nabla U\|_{H^1(\Omega_k)}\\ (m_{ij}h_i)^{\frac{1}{2}}\|\kappa\nabla U\|_{H^1(\Omega_i)}\\ \sum_{k=i,j}h_k^{\frac{1}{2}}\|\nabla\varpi\|_{L^2(\Omega_k)}\left(\int_{\sigma_{ij}}|\kappa\nabla U|^2\right)^{\frac{1}{2}}\end{cases} + O(h_{ij}^2)\,. \tag{5.3}$$

Remark. Note that (5.1)–(5.2) can be easily verified for convex sets with uniform bound on the relation $\frac{\mathrm{diam}_{\max}(\Omega_i)}{\mathrm{diam}_{\min}(\Omega_i)}$ between maximal and minimal diameter of a given cell. In particular, given $f \in H^1(\Omega_i)$ with $f_h(x) := f(\frac{x}{h})$ we find the scaled inequality

$$\frac{1}{h^{d-1}}\int_{h\partial\Omega_i}|f_h|^2 \le C\frac{1}{h^d}\int_{h\Omega_i}\left(|f_h|^2 + h^2|\nabla f_h|^2\right).$$

Furthermore, for $f \in H^1(\Omega_i)\cap C(\overline{\Omega}_i)$ one finds for a calculation similar to the Poincaré inequality for zero average functions (and for $x_i = 0$)

$$\int_{h\Omega_i}|f_h - f_i|^2 \le C\int_{h\Omega_i}h^2|\nabla f_h|^2 .$$

Proof. Observe that

$$\int_{\sigma_{ij}}|\varpi - S_{ij}|\,|\kappa\nabla U\cdot\boldsymbol{\nu}_{ij}| \le \left(\int_{\sigma_{ij}}|\varpi - S_{ij}|^2\right)^{\frac{1}{2}}\left(\int_{\sigma_{ij}}|\kappa\nabla U\cdot\boldsymbol{\nu}_{ij}|^2\right)^{\frac{1}{2}} \tag{5.4}$$

It remains to study $\frac{1}{m_{ij}}\int_{\sigma_{ij}}|\varpi - S_{ij}|^2$ in more detail. We have

$$S(\varpi_i,\varpi_j) - S\left(\frac{\varpi_i+\varpi_j}{2},\frac{\varpi_i+\varpi_j}{2}\right) = \frac{1}{2}(\varpi_i-\varpi_j)\nabla S\cdot(1,-1)^T + O(|\varpi_i-\varpi_j|)^2 = O(|\varpi_i-\varpi_j|)^2$$

and thus

$$\begin{aligned}\varpi - S_{ij} &= \frac{1}{2}(\varpi-\varpi_i) + \frac{1}{2}(\varpi-\varpi_j) + \left(\frac{\varpi_i+\varpi_j}{2} - S_{ij}\right)\\ &= \frac{1}{2}(\varpi-\varpi_i) + \frac{1}{2}(\varpi-\varpi_j) + O(|\varpi_i-\varpi_j|)^2 .\end{aligned}$$

The first term can be estimated by $|\varpi - \varpi_i| \le h_i\cdot\nabla\varpi + O(h_i^2)$ and a similar estimate holds for the second term. Using (5.1)–(5.2) we obtain in total

$$\int_{\sigma_{ij}}|\varpi - S_{ij}|\,|\kappa\nabla U\cdot\boldsymbol{\nu}_{ij}| \le C\sum_{k=i,j}h_k\,\|\nabla\varpi\|_{L^2(\Omega_k)}\left(\frac{1}{h_k}\,\|\kappa\nabla U\|^2_{H^1(\Omega_k)}\right)^{\frac{1}{2}} .$$

□

5.1 Error Analysis in U

In what follows, we assume that the discrete and the continuous solution satisfy homogeneous Dirichlet conditions. In view of the continuous and the discrete FPE given in the form (1.5) and (1.4) as well as formula (2.19) we observe that the natural variational consistency error for a given Stolarsky mean S equivalently takes the form

$$\begin{aligned}\mathfrak{E}_{\mathcal{T},\mathrm{FPE},\kappa}(u;v) &= \tilde{\mathfrak{E}}_{\mathcal{T},\mathrm{FPE},\kappa}(U;v)\\ &:= \sum_{i\sim j}(v_j - v_i)\left(\int_{\sigma_{ij}}\kappa\pi\nabla U\cdot\boldsymbol{\nu}_{ij} - \kappa_{ij}S_{ij}\frac{m_{ij}}{h_{ij}}\left((\mathcal{R}_{\mathcal{T}}U)_j - (\mathcal{R}_{\mathcal{T}}U)_i\right)\right).\end{aligned}$$

We recall that an estimate for $\tilde{\mathfrak{E}}_{\mathcal{T},\mathrm{FPE},\kappa}(U;\cdot)$ implies an order of convergence estimate by (2.15). Our main result of this section provides a connection between $\tilde{\mathfrak{E}}_{\mathcal{T},\mathrm{FPE},\kappa}(U;\cdot)$ and the variational consistency $\tilde{\mathfrak{E}}_{\mathcal{T},\kappa}(U;\cdot)$ (given by (2.21)) of the second order equation

$$-\nabla\cdot(\kappa\nabla U) = f$$

with the discretization scheme

$$\forall i: \qquad -\sum_{j:j\sim i}\kappa_{ij}\frac{m_{ij}}{h_{ij}}\left(U_j^{\mathcal{T}} - U_i^{\mathcal{T}}\right) = f_i\,.$$

Proposition 5.2. *Let $\mathcal{T} = (\mathcal{V},\mathcal{E},\mathcal{P})$ be a mesh. The variational consistency error $\mathfrak{E}_{\mathcal{T},\mathrm{FPE},\kappa}(U;\cdot)$ can be estimated by*

$$\left\|\tilde{\mathfrak{E}}_{\mathcal{T},\mathrm{FPE},\kappa}(U;\cdot)\right\|^2_{H^*_{\mathcal{T},\kappa S}} \le \|\pi\|_\infty\,|\mathfrak{E}|_{\mathcal{T}}(U;\cdot) + \sum_{i\sim j}\frac{h_{ij}}{m_{ij}}\kappa_{ij}^{-1}S_{ij}^{-1}\left(\int_{\sigma_{ij}}(\pi - S_{ij})\,\kappa\nabla U\cdot\boldsymbol{\nu}_{ij}\right)^2\,. \tag{5.5}$$

Proof. This is a direct consequence of Lemma 2.13. □

Using the above estimates, we can now show the main result of the section.

Theorem 5.3 (Localized order of convergence)**.** *Let $d \le 4$ and the mesh $\mathcal{T}$ be admissible in sense of Definition 2.1 and φ-consistent in sense of Definition 2.14. Let $u \in C_0^2(\Omega)$ be the solution to* (1.1)*. Let $f^{\mathcal{T}} := \mathcal{R}^*_{\mathcal{T}}f$ and let $u^{\mathcal{T}} \in \mathcal{S}^{\mathcal{T}}$ be the solution to* (2.4)*. Moreover, let $\kappa \le \kappa^*$, $b > 0$ and $S \in C^2(\mathbb{R}_{\ge 0}\times\mathbb{R}_{\ge 0})$. Then it holds*

$$\begin{aligned}\|\mathfrak{E}_{\mathcal{T},\mathrm{FPE},\kappa}(U;\cdot)\|^2_{H^*_{\mathcal{T},\kappa S}} &\le C(\kappa_*,\pi,d,\|U\|_{C^2})\times\left(\varphi(h)^2 + h^2\right).\\ \|u^{\mathcal{T}} - \mathcal{R}_{\mathcal{T}}u\|_{H_{\mathcal{T},\kappa S}} &\le C(\kappa_*,\pi,d,\|U\|_{C^2})\times\left(\varphi(h)^2 + h^2\right).\end{aligned}$$

Proof. Inserting estimate (5.3) into (5.5), we get

$$\begin{aligned}\|\mathfrak{E}_{\mathcal{T},\mathrm{FPE},\kappa}(U;\cdot)\|^2_{H^*_{\mathcal{T},\kappa S}} &\le \|\pi\|_\infty\,|\mathfrak{E}|_{\mathcal{T},\kappa S}(U;\cdot) + C\sum_{i\sim j}h_{ij}\kappa_{ij}^{-1}S_{ij}^{-1}h_i\|\kappa\nabla U\|^2_{H^1(\Omega_i)}\\ &\le \|\pi\|_\infty\,|\mathfrak{E}|_{\mathcal{T},\kappa S}(U;\cdot) + C(\kappa_*,\pi,d)\;h^2\sum_i\|\kappa\nabla U\|^2_{H^1(\Omega_i)}.\end{aligned}$$

Using (2.15) we obtain an estimate for the discretization error in the form

$$\|u^{\mathcal{T}} - \mathcal{R}_{\mathcal{T}}u\|^2_{H_{\mathcal{T},\kappa S}} \le \|\pi\|_\infty\,|\mathfrak{E}|_{\mathcal{T},\kappa S}(U;\cdot) + C(\kappa_*,\pi,d,\|U\|_{C^2})\,\mathrm{Size}(\mathcal{T})^2.$$

Using the consistency assumption on the discretization of the pure elliptic problem we obtain the desired estimate. □

5.2 Error Analysis in u

We will now derive an alternative estimate for the consistency error which accounts more for the convective aspect of the FPE. In Lemma 2.12 we have split the consistency error $\mathfrak{E}_{\mathcal{T},\mathrm{FPE},\kappa}(u;\cdot)$ into the two parts $\mathfrak{E}_{\mathcal{T},\kappa}(u;\cdot)$ and $\mathfrak{E}_{\mathcal{T},\mathrm{conv},\kappa}(u;\cdot)$. The error $\mathfrak{E}_{\mathcal{T},\kappa}(u;\cdot)$ relates to the elliptic part and is well understood in literature. Therefore, it remains to study the second part.

Proposition 5.4. *Using the notation of Lemma 2.12 it holds in $d \le 4$*

$$|\mathfrak{E}|_{\mathcal{T},\kappa,\mathrm{conv},\omega}(u) \le \sum_{i\sim j} \frac{h_{ij}}{m_{ij}} \omega_{ij}^{-1} \left(\int_{\sigma_{ij}} \kappa u \nabla V \cdot \boldsymbol{\nu}_{ij} - \frac{m_{ij}}{h_{ij}} \kappa_{ij} \frac{1}{2} \frac{S_{ij}}{\pi_i \pi_j} (\pi_i - \pi_j)(u_i + u_j) \right)^2 + C h^4 , \tag{5.6}$$

where C depends on $\|\pi\|_\infty$, $\|\nabla\pi\|_\infty$, $\|u\|_\infty$, $\|\nabla u\|_\infty$.

Proof. We use (2.18) to find

$$\frac{S_{ij} - \pi_j}{\pi_j} u_j - \frac{S_{ij} - \pi_i}{\pi_i} u_i = \frac{1}{2} \frac{S_{ij}}{\pi_i \pi_j} (\pi_i - \pi_j)(u_i + u_j) + \frac{1}{2} \frac{1}{\pi_i \pi_j} (S_{ij}\pi_i + S_{ij}\pi_j - 2\pi_i\pi_j)(u_i - u_j)$$

and in a next step we find on behalf of (3.2)

$$(S_{ij}\pi_i + S_{ij}\pi_j - 2\pi_i\pi_j) = \left(\frac{1}{2} + C_{\alpha,\beta} \left(\frac{\pi_j}{\pi_i} + \frac{\pi_i}{\pi_j} \right) \right) (\pi_i - \pi_j)^2 + O(\pi_i - \pi_j)^3$$

for $C_{\alpha,\beta} = \frac{1}{12}(\alpha + \beta - 3)$ and thus we conclude from

$$\begin{aligned} &\mathfrak{E}_{\mathcal{T},\mathrm{conv},\kappa}(u; v) = \\ &= \sum_{i\sim j} \left(\frac{m_{ij}}{h_{ij}} \kappa_{ij} \frac{1}{2} \frac{S_{ij}}{\pi_i\pi_j} (\pi_i - \pi_j)(u_i + u_j) - \int_{\sigma_{ij}} \kappa u \nabla V \cdot \boldsymbol{\nu}_{ij} \right) (v_j - v_i) \\ &+ \sum_{i\sim j} \left(\frac{m_{ij}}{h_{ij}} \kappa_{ij} \frac{1}{2} \frac{1}{\pi_i\pi_j} \left(\left(\frac{1}{2} + C_{\alpha,\beta} \left(\frac{\pi_j}{\pi_i} + \frac{\pi_i}{\pi_j} \right) \right) (\pi_i - \pi_j)^2 (u_i - u_j) + O(\pi_i - \pi_j)^3 \right) \right) (v_j - v_i) \end{aligned}$$

that (5.6)holds. □

Note that in general it holds

$$\frac{S_{ij}}{\pi_i\pi_j} (\pi_i - \pi_j) = \frac{1}{2} S_{ij} \left(\frac{1}{\pi_i} + \frac{1}{\pi_j} \right) (V_j - V_i) + O(h). \tag{5.7}$$

The Scharfetter Gummel scheme turns out to be special at this point.

Lemma 5.5 (SG is superior for large convection)**.** *In case of the Stolarsky mean $S_{0,-1}$ (the Scharfetter–Gummel case), it holds*

$$\frac{1}{2} \frac{S_{ij}}{\pi_i\pi_j} (\pi_i - \pi_j)(u_i + u_j) = (V_j - V_i) \frac{1}{2} (u_i + u_j) .$$

Proof. This follows immediately from $S_{0,-1}(x,y) = \frac{xy}{x-y} \log(x/y)$ and $\pi(x) = e^{-V(x)}$. □

The last observation plays an important role in the estimation of the right hand side of (5.6).

Theorem 5.6. *Let $d \le 4$ and $\mathcal{T}_h = (\mathcal{V}_h, \mathcal{E}_h, \mathcal{P}_h)$ be a family of meshes with $\mathrm{diam}\mathcal{T}_h \to 0$ as $h \to 0$ and let the assumptions of Lemma 5.1 hold. Using the notation of Lemma 2.12 let $u_{ij} := \frac{1}{2}(u_i + u_j)$. Then*

$$|\mathfrak{E}|_{\mathcal{T},\kappa,\mathrm{conv},\omega}(u) = 2\|u\|_\infty |\mathfrak{E}|_{\mathcal{T},\kappa,\omega}(V; \cdot) + 2 \sum_{i\sim j} \frac{h_{ij}}{m_{ij}} \omega_{ij}^{-1} \left(\int_{\sigma_{ij}} \kappa (u - u_{ij}) \nabla V \cdot \boldsymbol{\nu}_{ij} \right)^2 + O(h^2) .$$

In case $S_* = S_{0,-1}$ or $S_* = S_{\alpha,\beta}$ with $\alpha + \beta = -1$ the above can be improved to

$$|\mathfrak{E}|_{\mathcal{T},\kappa,\mathrm{conv},\omega}(u) = 2\,\|u\|_\infty\,|\mathfrak{E}|_{\mathcal{T},\kappa,\omega}(V;\cdot) + 2\sum_{i\sim j}\frac{h_{ij}}{m_{ij}}\omega_{ij}^{-1}\left(\int_{\sigma_{ij}}\kappa\,(u-u_{ij})\,\nabla V\cdot\boldsymbol{\nu}_{ij}\right)^2 + O(h^4)\,.$$

In all cases, $O(\,\cdot\,)$ depends on $\|\nabla V\|_\infty$.

Proof. We start from (5.6) applying (5.7). Defining $g := u$ and $g_{ij} := \frac{1}{4}S_{ij}\left(\frac{1}{\pi_i}+\frac{1}{\pi_j}\right)(u_i+u_j)$ applying Lemma 2.13 yields

$$|\mathfrak{E}|_{\mathcal{T},\kappa,\mathrm{conv},\omega}(u) \le 2\left(\sup_{i,j}|g_{ij}|\right)|\mathfrak{E}|_{\mathcal{T},\kappa,\omega}(V;\cdot)+2\sum_{i\sim j}\frac{h_{ij}}{m_{ij}}\omega_{ij}^{-1}\left(\int_{\sigma_{ij}}\kappa\,(u-g_{ij})\,\nabla V\cdot\boldsymbol{\nu}_{ij}\right)^2+O(h^2)\,.$$

We observe that $\frac{1}{2}S_{ij}\left(\frac{1}{\pi_i}+\frac{1}{\pi_j}\right) = 1 + O(h)$, where $O(h)$ depends on $\|\nabla V\|_\infty$ such that

$$\left|\int_{\sigma_{ij}}\kappa\,(u-g_{ij})\,\nabla V\cdot\boldsymbol{\nu}_{ij}\right| \le \left|\int_{\sigma_{ij}}\kappa\,(u-u_{ij})\,\nabla V\cdot\boldsymbol{\nu}_{ij}\right| + O(h)\,.$$

The claim now follows for general S_*. For $S_* = S_{0,-1}$ we apply Lemma 5.5 instead of (5.7). For general $S_* = S_{\alpha,\beta}$ with $\alpha+\beta=-1$ we apply Corollary 4.4. □

Corollary 5.7. *Under the assumptions of Theorem 5.6 it further holds*

$$\sum_{i\sim j}\frac{h_{ij}}{m_{ij}}\omega_{ij}^{-1}\left(\int_{\sigma_{ij}}\kappa\,(u-u_{ij})\,\nabla V\cdot\boldsymbol{\nu}_{ij}\right)^2 \le Ch^2\,\|V\|_{C^2}^2\,\|\nabla u\|_{L^2(\Omega)}^2\,.$$

In case $u\in C^1(\Omega)$ it even holds

$$\sum_{i\sim j}\frac{h_{ij}}{m_{ij}}\omega_{ij}^{-1}\left(\int_{\sigma_{ij}}\kappa\,(u-u_{ij})\,\nabla V\cdot\boldsymbol{\nu}_{ij}\right)^2 \le Ch^5\,\|V\|_{C^2}^2\,\|\nabla u\|_{\infty}^2\,.$$

Proof. In view of Lemma 5.1 we obtain in total

$$\left|\int_{\sigma_{ij}}\kappa\,(u-g_{ij})\,\nabla V\cdot\boldsymbol{\nu}_{ij}\right| \le C\,\|V\|_{C^2}\sum_{k=i,j}h^{\frac{3}{2}}\,\|\nabla u\|_{L^2(\Omega_k)} + O(h)$$

and $\|\nabla u\|_{L^2(\Omega_k)} \le h^{\frac{3}{2}}\,\|\nabla u\|_\infty$. □

6 Cubic Meshes

In view of Section 5 we consider the following specialization of Lemma 5.1 to cubic grids. Throughout this section we consider $d\le 3$ and a polygonal domain $\Omega\subset\mathbb{R}^d$ with a cubic mesh where $\Omega_i = x_i + [-h/2,h/2]^d$, $x_i\in h\mathbb{Z}\subset\Omega$.

Lemma 6.1. *Let $\Omega\subset\mathbb{R}^d$ be a polygonal domain with $d\le 4$ and a cubic mesh where $\Omega_i = x_i + [-h/2,h/2]^d$, $x_i\in h\mathbb{Z}\subset\Omega$. Then for every function $\varpi\in C^2$ with $\varpi_i := \varpi(x_i)$ and $S_{ij} := S_*(\varpi_i,\varpi_j)$ it holds*

$$\left|\int_{\sigma_{ij}}(\varpi - S_{ij})\,\kappa\nabla U\cdot\boldsymbol{\nu}_{ij}\right| = O(h^2)\,. \tag{6.1}$$

Proof. The following calculations are quite standard and, therefore, we shorten our considerations. We have for $x \in \sigma_{ij}$

$$\begin{aligned} S_{ij} - \varpi(x) &= S(\varpi_i, \varpi_j) - S(\varpi(x), \varpi(x)) = \\ &= \nabla S(x) \cdot \begin{pmatrix} \varpi_i - \varpi(x) \\ \varpi_j - \varpi(x) \end{pmatrix} + \begin{pmatrix} \varpi_i - \varpi(x) \\ \varpi_j - \varpi(x) \end{pmatrix} \cdot \nabla^2 S(x) \cdot \begin{pmatrix} \varpi_i - \varpi(x) \\ \varpi_j - \varpi(x) \end{pmatrix} + O(h^3). \end{aligned}$$

The gradient of S is given by $(1/2, 1/2)^T$ and hence, we $S_{ij} - \varpi(x) = \frac{\varpi_i + \varpi_j - 2\varpi(x)}{2} + O(h^2)$. We compute the first term in more detail. We have $\varpi_j - \varpi(x) = \nabla\varpi \cdot (x_j - x) + O(h^2)$ and correspondingly for $j \rightsquigarrow i$ and the sum yields

$$\varpi_i + \varpi_j - 2\varpi(x) = \nabla\varpi \cdot (x_i + x_j - 2x) + O(h^2) = \frac{1}{2}\nabla\varpi \cdot \tilde{x} + O(h^2)\,,$$

where $\tilde{x} = x - \frac{x_i + x_j}{2}$ the coordinate on the cell surface with respect to the middle point $\bar{x} = \frac{x_i + x_j}{2}$. Hence, we get

$$\int_{\sigma_{ij}} (\varpi - S_{ij}) \kappa \nabla U \cdot \nu_{ij} = \frac{1}{4} \int_{\sigma_{ij}} \nabla\varpi(x) \cdot \tilde{x} \kappa(x) \nabla U(x) \cdot \nu_{ij} \mathrm{d}\sigma(\tilde{x}) + O(h^2).$$

Now we can fix the function $s(x) = \kappa(x)\nabla U(x) \cdot \nu_{ij} \nabla\varpi(x)$ with respect to $\bar{x}$. We have $s(x) = s(\bar{x}) + (x - \bar{x})\nabla s(\bar{x}) + O(h^2)$, which implies (assuming that $U, \varpi \in \mathrm{C}^2$ and $\kappa \in \mathrm{C}^1$) that

$$\int_{\sigma_{ij}} (\varpi - S_{ij}) \kappa \nabla U \cdot \nu_{ij} = \frac{1}{4} \int_{\sigma_{ij}} (s(\bar{x}) + (x - \bar{x})\nabla s(\bar{x})) \cdot \tilde{x} \mathrm{d}\sigma(\tilde{x}) + O(h^2) = \frac{1}{4} \int_{\sigma_{ij}} s(\bar{x}) \cdot \tilde{x} \mathrm{d}\sigma(\tilde{x}) + O(h^2)\,.$$

But the first vanishes, since the interface σ_{ij} is symmetric w.r.t. the mid point $\bar{x}$ and we are integrating along $\tilde{x}$. Hence, we have (6.1). □

6.1 Consistency of purely elliptic operators on cubic meshes

Theorem 6.2 (Consistency on cubic meshes)**.** *Let* $\Omega \subset \mathbb{R}^d$ *with* $d \le 4$ *be a polygonal domain with a cubic mesh where* $\Omega_i = x_i + [-h/2, h/2]^d$, $x_i \in h\mathbb{Z} \subset \Omega$. *Then*

$$|\mathfrak{E}|_{\mathcal{T},\kappa,\omega}(u) \le Ch^4.$$

Proof. It holds

$$\left| m_{ij} \kappa_{ij} \frac{\hat{U}_j - \hat{U}_i}{h} - \int_{\sigma_{ij}} \kappa \nabla U \cdot \nu_{ij} \right| \le |\kappa_{ij}| \left| m_{ij} \frac{\hat{U}_j - \hat{U}_i}{h} - \int_{\sigma_{ij}} \nabla U \cdot \nu_{ij} \right| + \left| \int_{\sigma_{ij}} (\kappa_{ij} - \kappa) \nabla U \cdot \nu_{ij} \right|.$$

We have $\hat{U}_j = U(x) + \nabla U \cdot (x_j - x) + O(h^2)$ and $\hat{U}_i = U(x) + \nabla U \cdot (x_i - x) + O(h^2)$. Moreover, we can write $x_i - x = -\frac{h}{2}\nu_{ij} + \tilde{x}$ where $\tilde{x} \perp \nu_{ij}$ and $x_j - x = \frac{h}{2}\nu_{ij} + \tilde{x}$ (the normal ν_{ij} points outside or inside of Ω_i). Hence, we conclude

$$\begin{aligned} \hat{U}_j &= U(x) + \nabla U \cdot (\frac{h}{2}\nu_{ij} + \tilde{x}) + O(h^2) \\ \hat{U}_i &= U(x) + \nabla U \cdot (-\frac{h}{2}\nu_{ij} + \tilde{x}) + O(h^2). \end{aligned}$$

Subtracting both equations, we end up with $\frac{\hat{U}_j - \hat{U}_i}{h} = \nabla U \cdot \nu_{ij} + O(h^2)$, and hence,

$$\left| m_{ij} \frac{\hat{U}_j - \hat{U}_i}{h} - \int_{\sigma_{ij}} \nabla U \cdot \nu_{ij} \right| \le m_{ij} O(h^2).$$

The Theorem follows from Lemma 6.1, the definition of κ_{ij} and the cubic geometry. □

6.2 Quantitative estimate on cubic meshes in the diffusive representation

In view of Theorem 5.6 combined with Theorem 6.2 and Lemma 6.1 with $\frac{1}{2}(u_i + u_j) = S_{2,1}(u_i, u_j)$ we also obtain the following.

Theorem 6.3. *Let $d \leq 4$. On a polygonal domain $\Omega \subset \mathbb{R}^d$ with a cubic mesh where $\Omega_i = x_i + [-h/2, h/2]^d$, $x_i \in h\mathbb{Z} \subset \Omega$, it holds: Using the notation of Lemma 2.12 it holds*

$$|\mathfrak{E}|_{\mathcal{T},\omega,\mathrm{conv}}(u) = O(h^2)\,.$$

In case $S_* = S_{0,-1}$ or $S_* = S_{\alpha,\beta}$ with $\alpha + \beta = -1$ the above can be improved to

$$|\mathfrak{E}|_{\mathcal{T},\omega,\mathrm{conv}}(u) = O(h^4)\,.$$

7 Numerical simulation and convergence analysis

In this section, we provide a numerical convergence analysis of the flux discretization schemes based on Stolarsky means described in the previous sections. For the sake of simplicity, we restrict ourselves to one-dimensional examples with equidistant meshes, for which already non-trivial results can be observed.

Example 7.1. We consider the potential $V(x) = 2\sin(2\pi x)$, the right hand side $f(x) = x(1-x)$ on $x = (0,1)$ with $\kappa = 1$ and Dirichlet boundary conditions $u(0) = 0$ and $u(1) = 1$. The Stolarsky mean discretizations are compared point-wise with a numerically computed reference solution u_{ref} (and J_{ref}) that was obtained using a shooting method (involving a fourth order Runge–Kutta scheme) together with Brent's root finding algorithm [3] on a very fine grid with 136474 nodes.

The convergence results are summarized in Fig. 2. In Fig. 2 (a), the logarithmic error $\log_{10}(\|u - u_{\text{ref}}\|_{L_2})$ is shown in the (α, β)-plane of the Stolarsky mean parameters for an equidistant mesh with $2^{10} + 1 = 1025$ nodes. First, we note that the accuracy for a mean $S_{\alpha,\beta}$ is indeed practically invariant along $\alpha + \beta = $ const., which is consistent with our analytical result in Section 4. In this particular example, we observe optimal accuracy at about $\alpha + \beta \approx 4.2$. This coincides with the convergence results under mesh refinement shown in Fig. 2 (b), where the fastest convergence is obtained for the scheme involving the $S_{3.2,1}$-mean. The other considered schemes, however, show as well a quadratic convergence behavior with a slightly larger constant. Interestingly, for the same example, we find that the optimal mean for an accurate approximation of the flux J is on $\alpha + \beta = -3$, see Fig. 2 (c). This is further evidenced in Fig. 2 (d), where the harmonic mean $S_{-1,-2}$ converges significantly faster than the other schemes. Obviously, in the present example, the minimal attainable error for both u and J can not be achieved by the same discretization scheme.

Example 7.2. We consider the potential $V(x) = 5(x+1)x$ and keep the right hand side function, the diffusion constant and the boundary conditions as in the previous example. The problem has an exact solution involving the imaginary error function, that is related to the Dawson function, which has been obtained using Mathematica [50].

The numerical results are shown in Fig. 3. The discretization errors for both the density u and the flux J are depicted in Fig. 3 (a) and (c) show a sharp minimum for $\alpha + \beta = -1$. This involves the Scharfetter–Gummel mean $S_{0,-1}$, which converges fastest to the analytical solutions for u and J, as shown in 3 (b) and (d). The SQRA scheme, with geometric mean $S_{\alpha,-\alpha}$, is found to be second best in the present example.

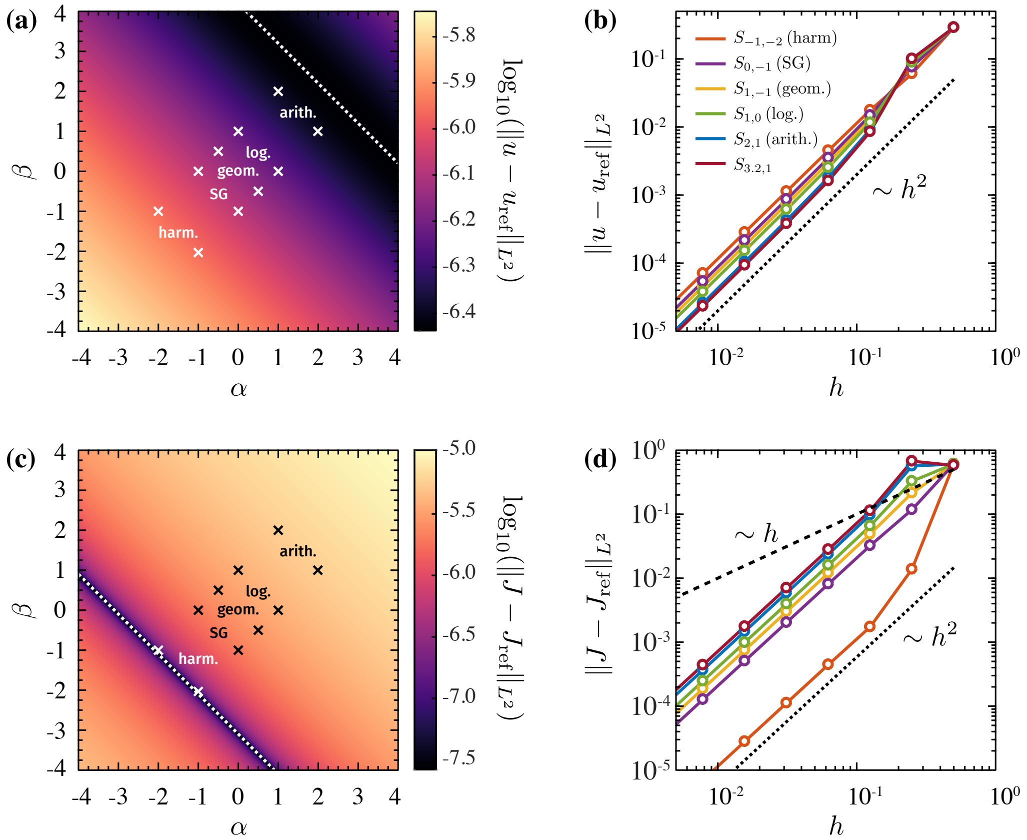

Fig. 2. Numerical results for example 7.1. **(a)** Discretization error $\log_{10}(\|u-u_{\text{ref}}\|_{L_2})$ in the (α,β)-plane on an equidistant mesh with $2^{10}+1$ nodes. The error is color-coded. Several special means an highlighted by crosses (notice the symmetry $S_{\alpha,\beta}(x,y)=S_{\beta,\alpha}(x,y)$). **(b)** Quadratic convergence of the discrete solution to the exact reference solution u_{exact} under mesh refinement in the L_2-norm. See the inset for a legend and color-coding of the considered means $S_{\alpha,\beta}$. In the present example, the best numerical result for u is achieved by $S_{3.2,1}$. **(c)** Logarithmic error of the numerically computed flux density $\log_{10}(\|j-j_{\text{ref}}\|_{L_2})$ in the (α,β)-plane on the same mesh as in (a). **(d)** Convergence of the numerically computed flux density to j_{exact}. In contrast to the convergence of u shown in (b), here the harmonic average $S_{-1,-2}$ performs best.

The numerical results are in line with Theorem 1.6: In the case of strong gradients ∇V, the Scharfetter–Gummel scheme provides the most accurate flux discretization, in particular, the SG mean $S_{0,-1}$ is the only Stolarsky mean that recovers the upwind scheme (1.9). Away from that drift-dominated regime, the situation is less clear and other averages $S_{\alpha,\beta}$ can be superior, see for instance Example 7.1.

A Appendix

A.1 A General Poincaré Inequality

We derive a general Poincaré inequality on meshes. The idea behind the proof seems to go back to Hummel [27] and has been adapted in a series of works e.g. [24, 25]. Let $e_0 = 0$ and $(e_i)_{i=1,\dots,n}$ be the canonical basis of $\mathbb{R}^n$. Define:

$$D^{d-1} := \{\nu \in \mathbb{S}^{d-1} \mid \exists m \in \{1,\cdots,d\} : \nu\cdot e_i = 0 \ \forall\ i \in \{0,1,\cdots,m-1\} \text{ and } \nu\cdot e_m > 0\}\,.$$

Every $\nu \in \mathbb{S}^{d-1}$ satisfies $\nu\cdot e_i \neq 0$ for at least one e_i. Thus, for every $\nu \in \mathbb{S}^{d-1}$ it holds $\nu \in D^{d-1}$ if and only if $-\nu \notin D^{d-1}$.

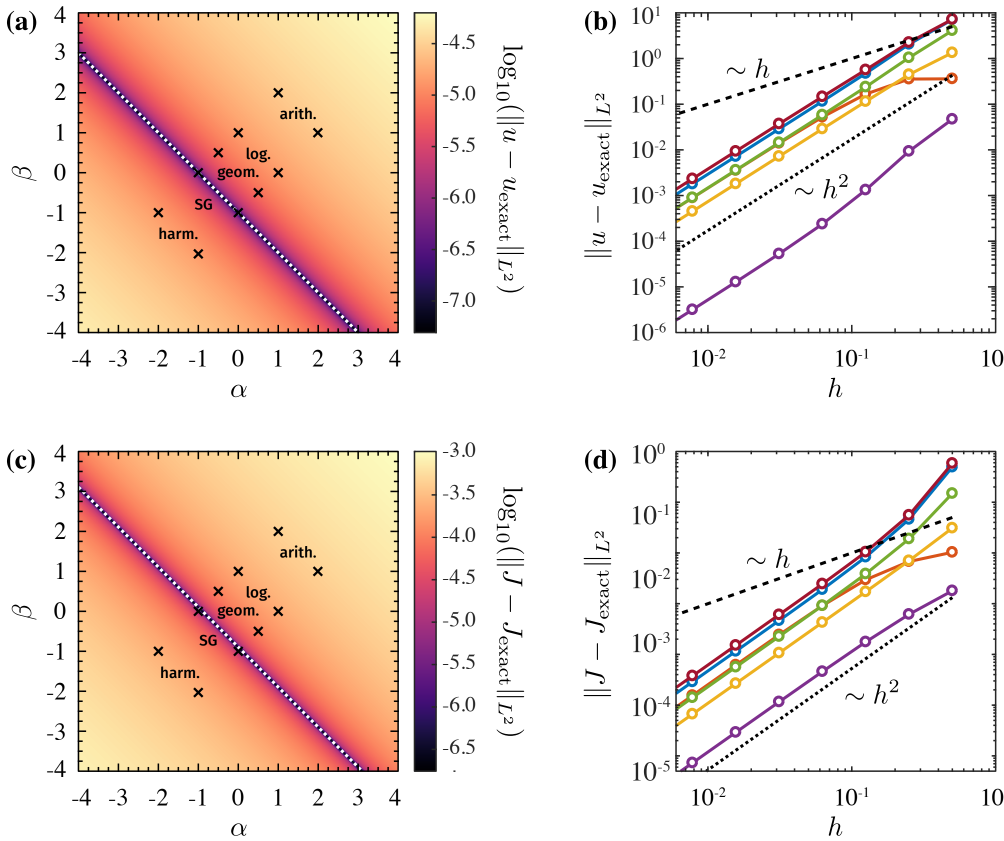

Fig. 3. Discretization errors and convergence behavior of the numerically computed u and J in example 7.2 using the Stolarsky mean schemes. The errors in **(a)** and **(c)** are color-coded. The coloring of the means in **(b)** and **(d)** is the same as in Fig. 2 (b). The plots clearly show a superior performance of the Scharfetter–Gummel scheme (i.e., the Stolarsky mean $S_{0,-1}$) for the approximation of both the density u and the flux J.

We denote $\Gamma = \bigcup_{\sigma\in\mathcal{E}_\Omega} \sigma$ and say that $x \in \Gamma$ is a Lipschitz point if Γ is a Lipschitz graph in a neighborhood of x. The set of Lipschitz-Points is called $\Gamma_L \subset \Gamma$ and we note that for the $(d-1)$-dimensional Hausdorff-measure of $\Gamma\backslash\Gamma_L$ it holds $\mathcal{H}^{d-1}\left(\Gamma\backslash\Gamma_L\right) = 0$.

For $x \in \Gamma_L$, we denote $\nu_x \in D^{d-1}$ the normal vector to Γ in x.. Let

$$\mathcal{C}^1_0(\Omega;\Gamma) := \left\{u \in C(\Omega\backslash\Gamma) \;:\; u|_{\partial\Omega} \equiv 0\,,\; \forall i\, \exists v_i \in C^1\left(\overline{\Omega_i}\right) : u|_{\Omega_i} = v_i\right\}$$

and for $u \in \mathcal{C}^1_{K,0}(\Omega)$ define in Lipschitz points $x \in \Gamma_L$

$$u_\pm(x) := \lim_{h\to 0}\left(u\left(x \pm h\nu_x\right)\right)\,, \qquad [\![u]\!](x) := u_+(x) - u_-(x)\,.$$

For two points $x, y \in \mathbb{R}^n$ denote (x, y) the closed straight line segment connecting x and y and for $\xi \in (x, y) \cap \Gamma_L$ denote

$$[\![u]\!]_{x,y}(\xi) := \lim_{h\to 0}\left(u\left(\xi + h(y-x)\right) - u\left(\xi - h(y-x)\right)\right)$$

the jump of the function u at ξ in direction $(y - x)$, i.e. $[\![u]\!]_{x,y}(\xi) \in \pm[\![u]\!]\,(\xi)$. We can extend $[\![u]\!]$ to Γ by $[\![u]\!]\,(x) = 0$ for $x \in \Gamma\backslash\Gamma_L$ and define

$$\|u\|_{H^1(\Omega;\Gamma)} := \left(\int_{\Omega\backslash\Gamma} |\nabla u|^2 + \int_\Gamma [\![u]\!]^2\right)^{\frac{1}{2}}\,,$$

$$H^1_0\left(\Omega;\Gamma\right) := \overline{\mathcal{C}^1_0\left(\Omega;\Gamma\right)}^{\|\cdot\|_{H^1(\Omega;\Gamma)}}\,.$$

Then we find the following result:

Lemma A.1 (Semi-discrete Poincaré inequality). *Let $\Omega \subset \mathbb{R}^d$ be a bounded domain. The space $H^1_0(\Omega;\Gamma)$ is linear and closed for every $s \in [0,\frac{1}{2})$ and there exists a positive constant $C_s > 0$ such that the following holds: Suppose there exists a constant $C_\# > 0$ such that for almost all $(x,y) \in \Omega^2$ it holds $\#((x,y)\cap\Gamma) \le C_\#$.. Then for every $u \in H^1_0(\Omega;\Gamma)$ it holds*

$$\|u\|^2_{H^s(\Omega)} \le C_s\left(C_\# \int_\Gamma [\![u]\!]^2 + \|\nabla u\|^2_{L^2(\Omega\backslash\Gamma)}\right). \tag{A.1}$$

Furthermore, for every $u \in H^1(\Omega;\Gamma)$ and every $\boldsymbol{\eta} \in \mathbb{R}^d$ it holds

$$\int_\Omega |u(x) - u(x+\boldsymbol{\eta})|^2\,dx \le |\boldsymbol{\eta}|\left(C_\# \int_\Gamma [\![u]\!]^2 + \|\nabla u\|^2_{L^2(\Omega\backslash\Gamma)}\right). \tag{A.2}$$

Proof. In what follows, given $u \in \mathcal{C}^1_0(\Omega;\Gamma)$, we write $\widehat{\nabla u}(x) := \nabla u(x)$ if $x \in \Omega\backslash\Gamma$ and $\widehat{\nabla u}(x) = 0$ else. For $y \in \mathbb{R}^d$ we denote $(x,y) = \{x + s(y-x) \,:\, s \in [0,1]\}$. Using $2ab < a^2 + b^2$, we infer for $u \in \mathcal{C}^1_0(\Omega;\Gamma)$ and $x, y \in \overline{\Omega}\backslash\Gamma$ such that $(x,y)\cap\Gamma$ is finite the inequality

$$\begin{aligned}|u(x)-u(y)|^2 &\le \left(\sum_{\xi\in(x,y)\cap\Gamma} [\![u]\!]_{x,y}(\xi) + \int_0^1 \widehat{\nabla u}\,(x+s(y-x))\cdot(x-y)\,ds\right)^2 \\ &< |x-y|^2 \int_0^1 \left|\widehat{\nabla u}\,(x+s(y-x))\right|^2 ds + \left(\sum_{\xi\in(x,y)\cap\Gamma} [\![u]\!]_{x,y}(\xi)\right)^2\end{aligned}$$

Since $[\![u]\!]_{x,y} = [\![u]\!]$ we compute

$$\left(\sum_{\xi\in(x,y)\cap\Gamma} [\![u]\!]_{x,y}(\xi)\right)^2 \le \#((x,y)\cap\Gamma) \sum_{\xi\in(x,y)\cap\Gamma} [\![u]\!]^2(\xi)$$

and obtain

$$\begin{aligned}|u(x)-u(y)|^2 < |x-y|^2 &\int_0^1 \left|\widehat{\nabla u}\,(x+s(y-x))\right|^2 ds \\ &+ \#((x,y)\cap\Gamma) \sum_{\xi\in(x,y)\cap\Gamma} [\![u]\!]^2(\xi).\end{aligned} \tag{A.3}$$

We fix $\eta > 0$ and consider the orthonormal basis $(e_i)_{i=1,\dots,d}$ of $\mathbb{R}^d$. The determinant of the first fundamental form of Γ is bigger than 1 almost everywhere. Hence we can observe that

$$\begin{aligned}\int_\Omega \sum_{\xi\in(x,x+\eta e_1)\cap\Gamma} [\![u]\!]^2(\xi)\,\mathrm{d}x &= \int_\mathbb{R}\left(\int_{\mathbb{R}^{d-1}} \sum_{\xi\in(x,x+\eta e_1)\cap\Gamma} [\![u]\!]^2(\xi)\,\mathrm{d}x_2\dots\mathrm{d}x_d\right)\mathrm{d}x_1 \\ &\le \int_\mathbb{R}\int_{\Gamma\cap((x_1,x_1+\eta)\times\mathbb{R}^{d-1})} [\![u]\!]^2(x)\,\mathrm{d}\sigma\,\mathrm{d}x_1 \\ &\le \eta\int_\Gamma [\![u]\!]^2(x)\,\mathrm{d}x\,,\end{aligned}$$

where we used that the surface elements are bigger than 1.. Furthermore, we have

$$\eta^2 \int_0^1 \left|\widehat{\nabla u}\,(x+s\eta e_1)\right|^2 ds = \eta \int_0^\eta \left|\widehat{\nabla u}\,(x+se_1)\right|^2 ds\,.$$

Replacing e_1 in the above calculations with any unit vector e, we obtain from integration of (A.3) with $y = x + \boldsymbol{\eta}$, $\boldsymbol{\eta} = \eta e$, over Ω that

$$\int_\Omega |u(x) - u(x+\boldsymbol{\eta})|^2\,dx \le |\boldsymbol{\eta}|\left(C_\# \int_\Gamma [\![u]\!]^2 + \|\nabla u\|^2_{L^2(\Omega\backslash\Gamma)}\right).$$

Dividing by $|\boldsymbol{\eta}|$ and integrating over $\boldsymbol{\eta} \in \mathbb{R}^d$, we obtain that for every $s \in [0, \frac{1}{2})$ there exists a positive constant $C_s > 0$ independent from u and K such that

$$\|u\|^2_{H^s(\Omega)} \leq C_s \left(C_\# \int_\Gamma [\![u]\!]^2 + \|\nabla u\|^2_{L^2(\Omega\setminus\Gamma)} \right) . \tag{A.4}$$

Hence, by approximation, the last two estimates hold for all $u \in H^1_0(\Omega;\Gamma)$.. □

A.2 Physical relevance of the geometric mean

Theorem A.2. *Let $S_{ij} = S_*(\pi_i, \pi_j)$ be a Stolarsky mean and let ψ^* be a symmetric strictly convex function with $\psi^*(0) = 0$. If $\partial_\pi(S_{ij}a_{ij}) = 0$ then $S_{ij} = \sqrt{\pi_i\pi_j}$ and ψ^* is proportional to C^*.*

Proof of Theorem A.2. The case $S_{ij} = \sqrt{\pi_i\pi_j}$ and $\psi^*(\xi) = \cosh\xi - 1$ was explained in detail in [24].

In the general case, symmetry of ψ^* in $\xi_i - \xi_j$ implies $\psi^*(\xi_i - \xi_j) = \psi^*(|\xi_i - \xi_j|)$. We make use of the fact that the original $\mathsf{C}^*(\xi) = \cosh\xi - 1$ is a bijection on $[0,\infty)$ and suppose that hence $\psi^*(\xi_i - \xi_j) = \theta(\mathsf{C}^*(\xi_i - \xi_j))$. This implies particularly that

$$0 \leq x\,\partial_x(\theta(\mathsf{C}^*(x))) = x\,\partial_\xi\theta(\mathsf{C}^*(x))\,\partial_x\mathsf{C}^*(x) .$$

Furhtermore, the symmetry of ψ^* implies by the last inequality that $\partial_\xi\theta(\mathsf{C}^*(x)) > 0$. Inserting this information in (3.7) and (3.8) we observe that

$$S_{ij}\left(\frac{u_i}{\pi_i} - \frac{u_j}{\pi_j}\right)\partial_\xi\theta\left(\mathsf{C}^*\left(\ln\left(\frac{u_i}{\pi_i}\right) - \ln\left(\frac{u_j}{\pi_j}\right)\right)\right)^{-1}\sinh\left(\ln\left(\frac{u_i}{\pi_i}\right) - \ln\left(\frac{u_j}{\pi_j}\right)\right)^{-1}$$

has to be independent from π_i and π_j. From the above case $S_{ij} = \sqrt{\pi_i\pi_j}$, we know that

$$\sqrt{\pi_i\pi_j}\left(\frac{u_i}{\pi_i} - \frac{u_j}{\pi_j}\right)\sinh\left(\ln\left(\frac{u_i}{\pi_i}\right) - \ln\left(\frac{u_j}{\pi_j}\right)\right)^{-1}$$

is constant in π_i and π_j. Hence it remains to show that

$$f(\pi_i, \pi_j) := S_{ij}\sqrt{\pi_i\pi_j}^{-1}\partial_\xi\psi\left(\frac{u_i}{u_j}\frac{\pi_j}{\pi_i} + \frac{u_j}{u_i}\frac{\pi_i}{\pi_j}\right)^{-1}$$

is independent from π_i and π_j if and only if $\partial_\xi\psi = const$ and $S_{ij} = \sqrt{\pi_i\pi_j}$.

Assume first that $S_{ij}\sqrt{\pi_i\pi_j}^{-1} = const$. Then for $p = \frac{\pi_i}{\pi_j}$ we obtain that

$$\partial_p\left(\partial_\xi\theta\left(\frac{u_i}{u_j}p^{-1} + \frac{u_j}{u_i}p\right)^{-1}\right) = 0$$

has to hold. This implies that $\partial_\xi\psi = const$.

If $S_{ij}\sqrt{\pi_i\pi_j}^{-1} \neq const$, we use the definition of the weighted Stolarsky means given in (1.3) and note that

$$S_{ij} := S(\pi_i, \pi_j) = \left(\frac{\beta(\pi_i^\alpha - \pi_j^\alpha)}{\alpha(\pi_i^\beta - \pi_j^\beta)}\right)^{\frac{1}{\alpha-\beta}} = \pi_j\left(\frac{\beta(p^\alpha - 1)}{\alpha(p^\beta - 1)}\right)^{\frac{1}{\alpha-\beta}} ,$$

where again $p = \frac{\pi_i}{\pi_j}$. Hence we obtain that

$$f(\pi_i,\pi_j) = \tilde{f}(p) := \sqrt{\frac{1}{p}}\left(\frac{\beta(p^\alpha - 1)}{\alpha(p^\beta - 1)}\right)^{\frac{1}{\alpha-\beta}} \partial_\xi\theta\left(\frac{u_i}{u_j}p^{-1} + \frac{u_j}{u_i}p\right)^{-1}$$
$$= \left(\frac{\beta\left(p^{\frac{\alpha}{2}} - p^{-\frac{\alpha}{2}}\right)}{\alpha\left(p^{\frac{\beta}{2}} - p^{-\frac{\beta}{2}}\right)}\right)^{\frac{1}{\alpha-\beta}} \partial_\xi\theta\left(\frac{u_i}{u_j}p^{-1} + \frac{u_j}{u_i}p\right)^{-1}$$

has to be independent of π_i and π_j. But then, $\tilde{f}$ is independent of p. Now, we define $a = \frac{u_j}{u_i}$ and observe that

$$\tilde{f}\left(\frac{1}{a^2p}\right) = \left(\frac{\beta\left((a^2p)^{-\frac{\alpha}{2}} - (a^2p)^{\frac{\alpha}{2}}\right)}{\alpha\left((a^2p)^{-\frac{\beta}{2}} - (a^2p)^{\frac{\beta}{2}}\right)}\right)^{\frac{1}{\alpha-\beta}} \partial_\xi\theta\left(\frac{u_i}{u_j}p^{-1} + \frac{u_j}{u_i}p\right)^{-1}.$$

We assume for $\alpha \neq \beta$. The case $\alpha = \beta$ can follows by continuity. For any p it should holds $\tilde{f}\left(\frac{1}{a^2p}\right) = \tilde{f}(p)$, which implies

$$\left(\frac{\beta\left(p^{\frac{\alpha}{2}} - p^{-\frac{\alpha}{2}}\right)}{\alpha\left(p^{\frac{\beta}{2}} - p^{-\frac{\beta}{2}}\right)}\right)^{\frac{1}{\alpha-\beta}} = \left(\frac{\beta\left((a^2p)^{-\frac{\alpha}{2}} - (a^2p)^{\frac{\alpha}{2}}\right)}{\alpha\left((a^2p)^{-\frac{\beta}{2}} - (a^2p)^{\frac{\beta}{2}}\right)}\right)^{\frac{1}{\alpha-\beta}},$$

or equivalently, after introducing $q^2 = p$,

$$\left(a^\alpha - a^\beta\right)q^{\alpha+\beta} + \left(a^\beta - a^{-\alpha}\right)q^{\beta-\alpha} + \left(a^{-\beta} - a^\alpha\right)q^{\alpha-\beta} + \left(a^{-\alpha} - a^{-\beta}\right)q^{-\beta-\alpha} = 0.$$

Since $\alpha \neq \beta$, one of the terms $q^{\pm\alpha\pm\beta}$ grows faster than the other. Hence we conclude that $a^\alpha = a^{\pm\beta}$ which means, $a = 1$, a contradiction. □

A.3 Properties of the Stolarsky mean

Lemma A.3. *For every of the above Stolarsky means $S_*(x,y)$ it holds*

$$\partial_x S_*(x,x) = \partial_y S_*(x,x) = \frac{1}{2} \quad \textit{and} \quad \partial_x^2 S_*(x,x) = \partial_y^2 S_*(x,x) = -\partial_{xy}^2 S_*(x,x) = -\partial_{yx}^2 S_*(x,x).$$

Proof. Since $S_*(x,x) = x$ and S_* is symmetric in x and y, we find from differentiating $\partial_x S_* = \partial_y S_* = \frac{1}{2}$. From the last equality, we find $\partial_x S_*(x,x) - \partial_y S_*(x,x) = 0$ as well as $\partial_x S_*(x,x) + \partial_y S_*(x,x) = 1$ and differentiation yields

$$\partial_x^2 S_*(x,x) - \partial_y^2 S_*(x,x) - \partial_{xy}^2 S_*(x,x) + \partial_{yx}^2 S_*(x,x) = 0, \tag{A.5}$$
$$\partial_x^2 S_*(x,x) + \partial_y^2 S_*(x,x) + \partial_{xy}^2 S_*(x,x) + \partial_{yx}^2 S_*(x,x) = 0. \tag{A.6}$$

Since $-\partial_{xy}^2 S_*(x,x) + \partial_{yx}^2 S_*(x,x) = 0$, equation (A.5) yields $\partial_x^2 S_*(x,x) = \partial_y^2 S_*(x,x)$. Inserting the last two relations into (A.6) yields $\partial_{xy}^2 S_*(x,x) = \partial_{yx}^2 S_*(x,x) = -\partial_x^2 S_*(x,x)$. □

Lemma A.4. *It holds* (3.2)$\partial_x^2 S_{\alpha,\beta}(\pi,\pi) = \frac{1}{12\pi}(\alpha+\beta-3)$.

Proof. We know from Lemma A.3 that $\partial_x S_{\alpha,\beta}(x,x) = \frac{1}{2}$ and $\partial_x^2 S_{\alpha,\beta}(x,x) = -\partial_y\partial_x S_{\alpha,\beta}(x,x)$. Hence we find

$$\partial_x S_{\alpha,\beta}(x+h,x-h) - \frac{1}{2} = \begin{pmatrix} h \\ -h \end{pmatrix}\begin{pmatrix} \partial_x^2 S_{\alpha,\beta}(x,x) \\ \partial_y\partial_x S_{\alpha,\beta}(x,x) \end{pmatrix} = 2h\partial_x^2 S_{\alpha,\beta}(x,x)\ .$$

We make use of the explicit form

$$\partial_x S_{\alpha,\beta}(x,y) = \left(\frac{\beta}{\alpha}\right)^{\frac{1}{\alpha-\beta}} \frac{(x^\alpha - y^\alpha)^{\frac{1}{\alpha-\beta}-1}}{(x^\beta - y^\beta)^{\frac{1}{\alpha-\beta}-1}} \frac{\alpha\left(x^\beta - y^\beta\right)x^\alpha - \beta\left(x^\alpha - y^\alpha\right)x^\beta}{(\alpha-\beta)\ x\ \left(x^\beta - y^\beta\right)^2}$$

for $x \neq y$. We insert $x = x+h$ and $y = x-h$ and make use of the following expansions

$$\begin{aligned}
\left((x+h)^\alpha - (x-h)^\alpha\right)^c &= \left(\alpha h x^{\alpha-1}\right)^c\left(2^c + O\left(h^2\right)\right) \\
\beta\left((x+h)^\alpha - (x-h)^\alpha\right)(x+h)^\beta &= 2\alpha\beta h x^{\alpha+\beta-1} + 2\alpha\beta^2 h^2 x^{\alpha+\beta-2} \\
&\quad + \frac{1}{3}\alpha\beta h^3\left(\alpha^2 - 3\alpha + 3\beta^2 - 3\beta + 2\right) + O\left(h^4\right) \\
\alpha\left((x+h)^\beta - (x-h)^\beta\right)(x+h)^\alpha &= 2\alpha\beta h x^{\alpha+\beta-1} + 2\alpha^2\beta h^2 x^{\alpha+\beta-2} \\
&\quad + \frac{1}{3}\alpha\beta h^3\left(\beta^2 - 3\beta + 3\alpha^2 - 3\alpha + 2\right) + O\left(h^4\right) \\
(x+h)\ \left((x+h)^\beta - (x-h)^\beta\right)^2 &= 4\beta^2 h^2 x^{2\beta-1} + 4\beta^2 h^3 x^{2\beta-2} + O\left(h^4\right)
\end{aligned}$$

$$\begin{aligned}
&\alpha\left((x+h)^\beta - (x-h)^\beta\right)(x+h)^\alpha - \beta\left((x+h)^\alpha - (x-h)^\alpha\right)(x+h)^\beta \\
&\qquad = 2\alpha\beta(\alpha-\beta)h^2 x^{\alpha+\beta-2} + \frac{\alpha\beta}{3}h^3 x^{\alpha+\beta-3}\left(2\alpha^2 - 2\beta^2\right) + O\left(h^4\right)
\end{aligned}$$

to obtain

$$\frac{\beta\left(x^\alpha - y^\alpha\right)x^\beta - \alpha\left(x^\beta - y^\beta\right)x^\alpha}{(\alpha-\beta)\ x\ \left(x^\beta - y^\beta\right)^2} = \frac{\alpha\left(x^{\alpha+\beta-2} + h\frac{1}{3}x^{\alpha+\beta-3}(\alpha+\beta) + O\left(h^2\right)\right)}{2\beta\left(x^{2\beta-1} + hx^{2\beta-2} + O\left(h^2\right)\right)}$$

and

$$\frac{(x^\alpha - y^\alpha)^{\frac{1}{\alpha-\beta}-1}}{(x^\beta - y^\beta)^{\frac{1}{\alpha-\beta}-1}} \approx \left(\frac{\alpha}{\beta}\right)^{\frac{1}{\alpha-\beta}-1}\left(\frac{x^{\alpha-1}\left(1+O\left(h^2\right)\right)}{x^{\beta-1}\left(1+O\left(h^2\right)\right)}\right)^{\frac{1}{\alpha-\beta}-1}.$$

Together with

$$\begin{aligned}
\frac{a+bh}{c+dh} &= \frac{a}{c} + \frac{bc-ad}{c^2}h + O\left(h^2\right) \\
\left(\frac{1+ah^2}{1+bh^2}\right)^c &= 1 + ch^2(a-b) + O\left(h^4\right)
\end{aligned}$$

we find

$$\begin{aligned}
\partial_x S_{\alpha,\beta}(x+h,x-h) &= \left(\frac{\left(1+O\left(h^2\right)\right)}{\left(1+O\left(h^2\right)\right)}\right)^{\frac{1}{\alpha-\beta}-1}\left(\frac{\left(1+h\frac{1}{3}x^{-1}(\alpha+\beta)+O\left(h^2\right)\right)}{2\left(1+hx^{-1}+O\left(h^2\right)\right)}\right) \\
&= \left(\frac{1}{2} + \frac{\frac{2}{3}(\alpha+\beta)-2}{4\,x}h\right) + O\left(h^2\right)
\end{aligned}$$

and hence (3.2). □

A.4 Approximation of potential to get the SQRA mean

The aim of this section is to provide a class of potentials which are easy to handle and which generate the SQRA-mean $S_{-1,1}(\pi_0, \pi_h)$ by $\pi_{\text{mean}} = \left(\frac{1}{h}\int_0^h \pi^{-1}\right)^{-1}$. Clearly, choosing the constant potential $V(x) := V_c := -\log S_{-1,1}(\pi_0, \pi_h)$ we obtain right mean. Although this works for any means, this has two drawbacks

1 The potential jumps and hence the gradient is somewhere infinite, which means that at these points the force on the particles is infinitely high which is not physical.

2 Approximating a general function by piecewise constants, on each interval the accuracy is only of order h. However, approximating a function by affine interpolation the accuracy is of order h^2 on each interval (see below for the calculation).

So we want to get a potential which may be used as a good approximation (i.e. approximating of order h^2), is physical (i.e. continuous) and generates the SQRA-mean. Note, that most considerations below also work for other Stolarsky means. For simplicity we focus on the SQRA mean $S_{-1,1}$.

A.4.1 Approximation order for linear approximation

Let us first realize that a linear interpolation provides an approximation of order h^2. Let $V : [0, h] \to \mathbb{R}$ be a general C^2-potential. We define with $V(0) = V_0$ and $V(h) = V_h$

$$\tilde{V}(x) = V_0 + \frac{V_h - V_0}{h} x.$$

Then one easily checks that

$$V(x) = V_0 + \partial_x V(0)x + \frac{1}{2}\partial_x^2 V(0)x^2 + O(h^3)$$

and hence,

$$V(x) - \tilde{V}(x) = \left(\partial_x V(0) - \frac{V_h - V_0}{h}\right)x + \frac{1}{2}\partial_x^2 V(0)x^2 + O(h^3).$$

Clearly, we also have

$$V_h = V_0 + \partial_x V(0)h + \frac{1}{2}\partial_x^2 V(0)h^2 + O(h^3)$$

which yields

$$V(x) - \tilde{V}(x) = -\frac{1}{2}\partial_x^2 V(0)hx + \frac{1}{2}\partial_x^2 V(0)x^2 + O(h^3) = \frac{1}{2}\partial_x^2 V(0)(x-h)x + O(h^3) = O(h^2).$$

A.4.2 Definition of potentials $\hat{V}$ which generate the SQRA mean

We consider a piecewise linear potential of the form

$$\hat{V}(x) = \begin{cases} \frac{V_c - V_0}{x_1}x + V_0 & , x \in [0, x_1] \\ V_c & , x \in [x_1, x_2] \\ \frac{V_h - V_c}{h - x_2}(x - x_2) + V_c & , x \in [x_2, h] \end{cases}.$$

where $x_1, x_2 \in [0,h]$ are firstly arbitrary and $V_c = -\log S_{-1,1}(\pi_0, \pi_h) = \frac{1}{2}(V_h + V_0)$. The potential is clearly continuous. Then

$$\frac{1}{h}\int_0^h \mathrm{e}^{\hat{V}(x)}\mathrm{d}x = \frac{x_1}{h}\frac{\mathrm{e}^{V_c} - \mathrm{e}^{V_0}}{V_c - V_0} + \frac{x_2 - x_1}{h}\mathrm{e}^{V_c} + \frac{h - x_2}{h}\frac{\mathrm{e}^{V_h} - \mathrm{e}^{V_c}}{V_h - V_c}.$$

Introducing the ratios $\alpha = \frac{x_1}{h}$ and $\beta = \frac{h-x_2}{h}$ (which are in $[0, 1/2]$) , we want to solve $\frac{1}{h}\int_0^h \mathrm{e}^{\hat{V}(x)}\mathrm{d}x = \mathrm{e}^{\frac{1}{2}(V_h+V_0)}$. Indeed, introducing the difference of the difference of the potentials $\bar{V} = V_h - V_0$, we obtain

$$\lambda = \frac{\alpha}{\beta} = \frac{\mathrm{e}^{\bar{V}/2} - \bar{V}/2 - 1}{\mathrm{e}^{-\bar{V}/2} + \bar{V}/2 - 1} \approx 1 + \frac{1}{3}\bar{V} + \frac{1}{18}\bar{V}^2.$$

Hence, any value α, β satisfying this ratio generates a potential with the SQRA-mean.

A.4.3 Proof that the potential approximates an arbitrary potential of order h^2

Since the linear potentials approximates a general potential of order h^2 it suffices to approximate the linear potential $\tilde{V}$ by $\hat{V}$. We show that there are α, β satisfying $\frac{\alpha}{\beta} = \lambda$, such that $\|\hat{V} - \tilde{V}\|_{C([x_i, x_{i+1}])} = O(h^2)$. The difference of $\hat{V}$ and $\tilde{V}$ is the largest at $x = x_1$ or $x = x_2$. We estimate both differences. We have

$$\tilde{V}(x_1) = V_0 + \frac{V_h - V_0}{h}x_1 = V_0 + \alpha\bar{V},\ \ \tilde{V}(x_2) = V_0 + \frac{V_h - V_0}{h}x_2 = V_0 + (1-\beta)\bar{V}.$$

Hence we have to estimate

$$\Delta_1 := |V_0 - V_c + \alpha\bar{V}|,\ \ \Delta_2 := |V_0 - V_c + (1-\beta)\bar{V}|.$$

In the case of SQRA, one possible choice for α, β is $\alpha + \beta = 1$. Then $\Delta_1 = \Delta_2 = |V_0 - V_c + \alpha\bar{V}| = |V_0 - V_c + \frac{\lambda}{1+\lambda}\bar{V}| = \frac{1}{1+\lambda}|(1+\lambda)(V_0 - V_c) + \lambda\bar{V}|$. We have $V_0 - V_c = -\bar{V}/2$, and hence

$$\Delta_1 = \Delta_2 = \frac{1}{1+\lambda}\frac{\bar{V}}{2}|\lambda - 1|.$$

One can check that $\lambda \approx 1 + \bar{V}/3$ and hence, $\Delta_1 + \Delta_2 \approx \frac{\bar{V}^2}{6} \approx O(h^2)$.

References

[1] D. N. d. G. Allan and R. V. Southwell. Relaxation methods applied to determine the motion in two dimensions of a viscous fluid past a fixed cylinder. *Q. J. Mech. Appl. Math.*, 8(2):129–145, 1955.

[2] R. Bank, W. Coughran, and L. C. Cowsar. The finite volume scharfetter-gummel method for steady convection diffusion equations. *Computing and Visulaization in Science*, 1(123-136), 1998.

[3] R. P. Brent. An algorithm with guaranteed convergence for finding a zero of a function. *Comput. J.*, 14(4):422–425, 1971.

[4] F. Brezzi, L. D. Marini, and P. Pietra. Numerical simulation of semiconductor devices. *Comput. Methods Appl. Mech. Eng.*, 75(1-3):493–514, 1989.

[5] C. Chainais-Hillairet and J. Droniou. Finite-volume schemes for noncoercive elliptic problems with neumann boundary conditions. *IMA Journal of Numerical Analysis*, 31(1):61–85, 2011.

[6] J. Chang and C. G. A practical difference scheme for fokker-planck equations. *Journal of Computational Physics*, 6:1–16, 1970.

[7] S.-N. Chow, W. Huang, Y. Li, and H. Zhou. Fokker-Planck equations for a free energy functional or Markov process on a graph. 203(3):969–1008, 2012.

[8] D. A. Di Pietro and J. Droniou. A third strang lemma and an aubin–nitsche trick for schemes in fully discrete formulation. *Calcolo*, 55(3):40, 2018.

[9] 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.

[10] P. D. Dixit, A. Jain, G. Stock, and K. A. Dill. Inferring transition rates of networks from populations in continuous-time markov processes. *Journal of chemical theory and computation*, 11(11):5464–5472, 2015.

[11] L. Donati, M. Heida, M. Weber, and B. Keller. Estimation of the initesimal generator by square-root approximation. *In preparation*.

[12] P. Dondl, T. Frenzel, and A. Mielke. A gradient system with a wiggly energy and relaxed EDP-convergence. *ESAIM Control Optim. Calc. Var.*, 2018. To appear. WIAS preprint 2459.

[13] M. Erbar and J. Maas. Ricci curvature of finite Markov chains via convexity of the entropy. 206(3):997–1038, 2012.

[14] L. Evans. *Partial Differential Equations*. AMS, 1998.

[15] R. Eymard, J. Fuhrmann, and K. Gärtner. A finite volume scheme for nonlinear parabolic equations derived from one-dimensional local dirichlet problems. *Numer. Math.*, 102(3):463–495, 2006.

[16] R. Eymard, T. Gallouët, and R. Herbin. Finite volume methods. *Handbook of numerical analysis*, 7:713–1018, 2000.

[17] K. Fackeldey, P. Koltai, P. Névir, H. Rust, A. Schild, and M. Weber. From metastable to coherent sets in time-discretization schemes. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 29(1):012101, 2019.

[18] P. Farrell, T. Koprucki, and J. Fuhrmann. Computational and analytical comparison of flux discretizations for the semiconductor device equations beyond Boltzmann statistics. *Journal of Computational Physics*, 346:497–513, 2017.

[19] P. Farrell, N. Rotundo, D. H. Doan, M. Kantner, J. Fuhrmann, and T. Koprucki. Drift-Diffusion Models. In J. Piprek, editor, *Handbook of Optoelectronic Device Modeling and Simulation: Lasers, Modulators, Photodetectors, Solar Cells, and Numerical Methods*, volume 2, chapter 50, pages 731–771. CRC Press, Taylor & Francis Group, Boca Raton, 2017.

[20] D. Forkert, J. Maas, and L. Portinale. Evolutionary γ-convergence of entropic gradient flow structures for fokker-planck equations in multiple dimensions. *arXiv:2008.10962*, 2020.

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

[22] T. Gallouët, R. Herbin, and M. H. Vignal. Error estimates on the approximate finite volume solution of convection diffusion equations with general boundary conditions. *SIAM Journal on Numerical Analysis*, 37(6):1935–1972, 2000.

[23] P. Gladbach, E. Kopfer, J. Maas, and L. Portinale. Homogenisation of one-dimensional discrete optimal transport. *arXiv:1905.05757*, 2019.

[24] M. Heida. Convergences of the squareroot approximation scheme to the Fokker–Planck operator. *Mathematical Models and Methods in Applied Sciences*, 28(13):2599–2635, 2018.

[25] M. Heida, R. Kornhuber, and J. Podlesny. Fractal homogenization of multiscale interface problems. *arXiv preprint arXiv:1712.01172*, 2017.

[26] M. Heida, J. Màlek, and K. Rajagopal. On the development and generalizations of Allen-Cahn and Stefan equations within a thermodynmic framework. *to be submitted to Zeitschrift für Angewandte Mathematik und Physik (ZAMP)*, 2011.

[27] H. Hummel. *Homogenization of Periodic and Random Multidimensional Microstructures*. PhD thesis, Technische Universität Bergakademie Freiberg, 1999.

[28] A. M. Il'in. Differencing scheme for a differential equation with a small parameter affecting the highest derivative. *Mathematical notes of the Academy of Sciences of the USSR*, 6(2):237–248, 1969. Translated from Mat. Zametki, Vol. 6, No. 2, pp. 237–248 (1969).

[29] R. Jordan, D. Kinderlehrer, and F. Otto. The variational formulation of the fokker–planck equation. *SIAM journal on mathematical analysis*, 29(1):1–17, 1998.

[30] M. Kantner. Generalized Scharfetter–Gummel schemes for electro-thermal transport in degenerate semiconductors using the Kelvin formula for the Seebeck coefficient. *Journal of Computational Physics*, 402:109091, 2020.

[31] R. Lazarov, I. D. Mishev, and P. S. Vassilevski. Finite volume methods for convection-diffusion problems. *SIAM Journal on Numerical Analysis*, 33(1):31–55, 1996.

[32] H. C. Lie, K. Fackeldey, and M. Weber. A square root approximation of transition rates for a markov state model. *SIAM Journal on Matrix Analysis and Applications*, 34:738–756, 2013.

[33] 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.

[34] L. Lu and J.-G. Liu. Large time behaviors of upwind schemes and b-schemes for fokker-planck equations on $\mathbb{R}$ by jump processes. *Mathematics of Computation*, 89(325):2283–2320, 2020.

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

[36] R. Marcelin. Contribution a l'étude de la cinétique physico-chimique. *Annales de Physique*, III:120–231, 1915.

[37] P. A. Markowich. *The stationary Semiconductor device equations*. Springer, Vienna, 1986.

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

[39] A. Mielke. Geodesic convexity of the relative entropy in reversible markov chains. *Calculus of Variations and Partial Differential Equations*, 48(1):1–31, 2013.

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

[41] 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.

[42] 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.

[43] 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.

[44] A. Mielke and A. Stephan. Coarse-graining via EDP-convergence for linear fast-slow reaction systems. *WIAS preprint 2643*, 2019.

[45] J. J. H. Miller and S. Wang. An analysis of the Scharfetter–Gummel box method for the stationary semiconductor device equations. *ESAIM: Mathematical Modelling and Numerical Analysis*, 28(2):123–140, 1994.

[46] D. Scharfetter and H. Gummel. Large-signal analysis of a silicon read diode oscillator. *IEEE Trans. Electron Devices*, 16(1):64–77, 1969.

[47] K. B. Stolarsky. Generalizations of the logarithmic mean. *Mathematics Magazine*, 48(2):87–92, 1975.

[48] W. W. van Roosbroeck. Theory of the flow of electrons and holes in germanium and other semiconductors. *Bell Syst. Tech. J.*, 29(4):560–607, Oct 1950.

[49] M. Weber and N. Ernst. A fuzzy-set theoretical framework for computing exit rates of rare events in potential-driven diffusion processes. *arXiv preprint arXiv:1708.00679*, 2017.

[50] I. Wolfram Research. Mathematica, 2017.

[51] J. Xu and L. Zikatanov. A monotone finite element scheme for converction-diffusion equations. *Math. Comp.*, 68(228):1429–1446, 1999.