ACADÉMIE
DES SCIENCES
INSTITUT DE FRANCE

Comptes Rendus
Mathématique

Angelo Iollo and Tommaso Taddei

Point-set registration in bounded domains via the Fokker–Planck equation

Volume 363 (2025), p. 809-824

Online since: 4 July 2025

https://doi.org/10.5802/crmath.753

The Comptes Rendus. Mathématique *are a member of the*
Mersenne Center for open scientific publishing
www.centre-mersenne.org — e-ISSN : 1778-3569

ACADÉMIE DES SCIENCES
INSTITUT DE FRANCE

Comptes Rendus. Mathématique
2025, Vol. 363, p. 809-824
https://doi.org/10.5802/crmath.753

Research article / *Article de recherche*
Numerical analysis / *Analyse numérique*

Point-set registration in bounded domains via the Fokker–Planck equation

Alignement de nuages de points dans des domaines bornés via l'équation de Fokker–Planck

Angelo Iollo [a,b] **and Tommaso Taddei** [a,b]

[a] IMB, UMR 5251, Univ. Bordeaux, 33400 Talence, France

[b] Centre Inria de l'Université de Bordeaux, Team MEMPHIS, 33400 Talence, France

E-mails: angelo.iollo@inria.fr, tommaso.taddei@inria.fr

Abstract. We present a point set registration method in bounded domains based on the solution to the Fokker–Planck equation. Our approach leverages (i) density estimation based on Gaussian mixture models; (ii) a stabilized finite element discretization of the Fokker–Planck equation; (iii) a specialized method for the integration of the particles. We review relevant properties of the Fokker–Planck equation that provide the foundations for the numerical method. We discuss two strategies for the integration of the particles and we propose a regularization technique to control the distance of the particles from the boundary of the domain. We perform extensive numerical experiments for two two-dimensional model problems to illustrate the many features of the method.

Résumé. Nous présentons une méthode d'alignement de nuages de points dans des domaines bornés basée sur la solution de l'équation de Fokker–Planck. Notre approche repose sur : (i) l'estimation de densité à l'aide de modèles de mélange gaussien ; (ii) une discrétisation par éléments finis stabilisée de l'équation de Fokker–Planck ; (iii) une méthode spécialisée pour l'intégration des particules. Nous examinons les propriétés pertinentes de l'équation de Fokker–Planck qui sous-tendent la méthode numérique. Nous discutons deux stratégies pour l'intégration des particules et proposons une technique de régularisation pour contrôler la distance des particules à la frontière du domaine. Nous réalisons de nombreuses expériences numériques pour deux problèmes modèles bidimensionnels afin d'illustrer l'efficacité de la méthode.

Keywords. Point set registration, Fokker–Planck, model order reduction.

Mots-clés. Alignement de nuages de points, équation de Fokker–Planck, réduction de modèle.

Manuscript received 12 December 2024, revised 13 March 2025 and 24 April 2025, accepted 21 April 2025.

1. Introduction

Point set registration (PSR) is the process of finding a spatial transformation that aligns two point clouds. Given the domain $\Omega \subset \mathbb{R}^n$ and the reference and the target point clouds $\{X_i^0\}_{i=1}^N, \{X_j^\infty\}_{j=1}^M \subset \Omega$, we seek a transformation $\Phi\colon \Omega \to \Omega$ such that the mapped point cloud $\{\Phi(X_i^0)\}_{i=1}^N$ is close — in the sense of Hausdorff — to the target point cloud. For several applications, it is also important to identify a (possibly non-autonomous) flow of diffeomorphisms $\Phi\colon \Omega \times \mathbb{R}_+ \to \Omega$ to coherently deform the reference point cloud $\{X_i^0\}_{i=1}^N$ into $\{X_j^\infty\}_{j=1}^M$. In this note,

we focus on the problem of PSR in moderate-dimensional ($n = 2$ or $n = 3$) bounded domains; we target applications in model order reduction (MOR) where coordinate transformations are employed to align coherent features of the solution field [5,10,11,26]; we also envision applications to data assimilation tasks that involve the interpolation between point clouds in complex geometries.

We propose a PSR method based on the solution to the Fokker–Planck (FP) equation. First, we model the reference and the target point clouds as independent identically distributed samples from two continuous distributions with densities ρ_0 and ρ_∞, respectively. Second, we estimate ρ_0 and ρ_∞ using Gaussian mixture models. Third, we solve a suitable FP equation to determine a flow of probability densities $t \mapsto \rho(\cdot, t)$ that is equal to ρ_0 at time $t = 0$, and converges to ρ_∞ as $t \to \infty$. Fourth, we exploit the properties of the FP equation to obtain the desired morphing based on the knowledge of $\rho(\cdot, t)$ at all times.

In this work, we aim to investigate the effectiveness of the FP equation for the transport of point clouds in bounded domains. The FP equation describes the evolution of a probability density ρ under diffusion and drift: it reads as a linear parabolic equation with Neumann boundary conditions on $\partial\Omega$, constant diffusion and time-independent advection. The FP equation has been extensively employed in Information Theory [8], stochastic processes in Physics and Chemistry [27], and more recently in machine learning for generative modeling [25]. The solution to the FP equation is tightly linked to the solution to optimal transportation (OT) problems [24]: however, while the FP equation is a linear parabolic equation with time-independent coefficients, the evolution equation associated to OT problems is highly nonlinear and much more difficult to handle, despite the many recent contributions to the field [18].

We remark that PSR is a well-established field in image processing [2,14,16]. Similarly to our approach, several PSR methods (e.g., [14,16]) interpret the point clouds as samples from unknown continuous probability distributions and then rely on (Gaussian) mixture models to estimate the densities. However, the vast majority of PSR techniques in computer vision including [14,16] are designed for registration in unbounded domains and are hence not directly applicable to the case of bounded domains. A notable exception is provided by the large deformation diffeomorphic metric mapping (LDDMM, [2]) method: similarly to our approach, LDDMM relies on a flow of diffeomorphisms to accomplish the registration task; however, while our method models the point clouds as densities and then resorts to the FP equation to determine the transformation, LDDMM resorts to the solution to a computationally-expensive optimization problem.

The outline of the note is as follows. In Section 2, we introduce the FP equation and we discuss well-known, elementary properties that are exploited in the remainder of the note. In Section 3, we present the numerical method: first, we introduce the discretization of the FP equation, which is based on the finite element method for space discretization and on the Crank–Nicolson method for time integration; second, we discuss two distinct methods for the transport of the point clouds; third, we present a regularization approach to control the distance of the particles from the boundary; fourth, we briefly review the problem of density estimation through Gaussian mixtures. In Section 4, we present several numerical results in two dimensions that illustrate the performance of the method. Section 5 concludes the paper.

2. Mathematical background

We review select properties of the Fokker–Planck (FP) equation that provide the mathematical foundation for the method. We emphasize that the FP equation has been the subject of extensive studies in Analysis and Calculus of Variations; in particular, the material of this section relies heavily on [8,12,17,24]. Below, Ω is a Lipschitz domain in $\mathbb{R}^n$, while $\mathscr{P}_2(\Omega)$ refers to the space of

probability density functions (pdfs) with bounded second-order moments; the acronym "a.s." (resp., "a.e.") stands for "almost surely" (resp., "almost every") with respect to the Lebesgue measure in Ω and the acronym "i.i.d." stands for "independent identically distributed".

2.1. *Kullback–Leibler divergence*

Given the densities $\rho, \rho_\infty \in \mathscr{P}_2(\Omega)$, we introduce the Kullback–Leibler (KL) divergence between ρ and ρ_∞ as [13]

$$D_{\mathrm{KL}}(\rho \parallel \rho_\infty) = \int_\Omega \rho(x) \log\left(\frac{\rho(x)}{\rho_\infty(x)}\right) \mathrm{d}x. \tag{1}$$

Note that $D_{\mathrm{KL}}(\rho \parallel \rho_\infty)$ is infinite if the support of ρ_∞ does not contain the support of ρ. Theorem 1 summarizes key properties of the KL divergence.

Theorem 1. *The following hold.*

(1) *Given the densities $\rho, \rho_\infty \in \mathscr{P}_2(\Omega)$, $D_{\mathrm{KL}}(\rho \parallel \rho_\infty) \geq 0$ and $D_{\mathrm{KL}}(\rho \parallel \rho_\infty) = 0$ if and only if $\rho = \rho_\infty$ a.s.*

(2) *The KL divergence* (1) *is asymmetric, that is in general $D_{\mathrm{KL}}(\rho \parallel \rho_\infty) \neq D_{\mathrm{KL}}(\rho_\infty \parallel \rho)$.*

(3) *The KL divergence* (1) *is strongly convex, that is, given $\rho_0, \rho_1, \rho_\infty \in \mathscr{P}_2(\Omega)$ and $\lambda \in (0,1)$, we have*

$$D_{\mathrm{KL}}\big(\lambda\rho_0 + (1-\lambda)\rho_1 \parallel \rho_\infty\big) \leq \lambda D_{\mathrm{KL}}(\rho_0 \parallel \rho_\infty) + (1-\lambda) D_{\mathrm{KL}}(\rho_1 \parallel \rho_\infty),$$

where the latter holds with equality if and only if $\rho_0 = \rho_1$ a.s.

(4) *(Pinsker's inequality) The KL divergence* (1) *satisfies the inequality:*

$$\|\rho - \rho_\infty\|_{L^1(\Omega)} \leq \min\big(\sqrt{2 D_{\mathrm{KL}}(\rho \parallel \rho_\infty)},\ \sqrt{2 D_{\mathrm{KL}}(\rho_\infty \parallel \rho)}\big).$$

We provide the proof of Theorem 1 in Appendix A. We notice that the KL divergence is minimized when $\rho = \rho_\infty$ (cf. property (1)); as discussed below, the strong convexity property and the Pinsker's inequality are the key to prove the existence of paths that link the initial and the target distributions (cf. Theorem 2). We further remark that the KL divergence is not a distance in $\mathscr{P}_2(\Omega)$ since it is not symmetric (cf. property (2)). If we denote by $V = -\log(\rho_\infty)$ the potential associated with the density ρ_∞, we can express (1) as:

$$D_{\mathrm{KL}}(\rho \parallel \rho_\infty) = \int_\Omega F[\rho]\,\mathrm{d}x = \underbrace{\int_\Omega \rho \log(\rho)\,\mathrm{d}x}_{=:\,(\mathrm{I})} + \underbrace{\int_\Omega \rho V\,\mathrm{d}x}_{=:\,(\mathrm{II})}, \quad \text{where } F[\rho] = \rho\log(\rho) + \rho V. \tag{2}$$

In Bayesian inference, the first term in (2) is interpreted as an entropy function, while the second term is interpreted as a potential energy, and the sum of the two terms is interpreted as a free-energy function (see, e.g., [6]).

2.2. *Fokker–Planck equation*

Let $\rho, \rho_\infty \in \mathscr{P}_2(\Omega)$ and let $V = -\log(\rho_\infty)$; then, we define the Fokker–Planck (FP) equation:

$$\begin{cases} \partial_t \rho + \nabla \cdot (-\rho \nabla V - \nabla \rho) = 0 & \text{in } \Omega \times \mathbb{R}_+, \\ \rho(\cdot, 0) = \rho_0 & \text{in } \Omega, \\ (\rho \nabla V + \nabla \rho) \cdot \mathbf{n} = 0 & \text{on } \partial\Omega \times \mathbb{R}_+, \end{cases} \tag{3}$$

where $\mathbf{n}$ denotes the outward normal to $\partial\Omega$. We assume that ρ_∞ is strictly positive in Ω and is sufficiently smooth to ensure that ∇V is Lipschitz continuous. We introduce the spaces

$\mathscr{H} = L^2(\Omega)$ and $\mathscr{V} = H^1(\Omega)$ and the weak formulation of the FP equation (3): find $\rho \in C(\mathbb{R}_+;\mathscr{H})$ such that $\rho(0) = \rho_0$ and

$$\int_\Omega \Big(\partial_t \rho(t)\cdot v + \big(\nabla\rho(t) + \rho(t)\nabla V\big)\cdot\nabla v\Big)\,\mathrm{d}x = 0 \quad \forall\, v\in\mathscr{V},\ \text{a.e. } t\in\mathbb{R}_+. \tag{4}$$

Theorem 2 reviews relevant properties of the FP equation; the proof is postponed to Appendix A.

Theorem 2. *Let Ω be a Lipschitz domain in $\mathbb{R}^n$; let $\rho_0, \rho_\infty \in \mathscr{P}_2(\Omega)$ satisfy:*

(i) $\inf_{x\in\Omega}\rho_\infty(x) > 0$;
(ii) ∇V *is Lipschitz-continuous in* Ω.

The following hold.

(1) *There exists a unique weak solution $\rho \in C(\mathbb{R}_+;\mathscr{H})$ to* (3).
(2) *The solution to* (3) *satisfies $\rho(\cdot,t)\in\mathscr{P}_2(\Omega)$ for all $t>0$.*
(3) *The solution to* (3) *converges to ρ_∞ for $t\to\infty$ in L^1, that is,* $\lim_{t\to\infty}\big\|\rho(\cdot,t)-\rho_\infty\big\|_{L^1(\Omega)} = 0$.
(4) *Let $u = -\nabla\big(\log(\rho)+V\big)$ and let $X_1^0,\dots,X_N^0 \overset{\text{iid}}{\sim} \rho_0$. Consider the trajectories*

$$\begin{cases} \dot X_i(t) = u\big(X_i(t),t\big) & t>0,\\ X_i(0) = X_i^0, \end{cases} \qquad i=1,\dots,N. \tag{5}$$

Then, for any $t>0$, we have $X_1(t),\dots,X_N(t) \overset{\text{iid}}{\sim} \rho(\cdot,t)$.

Theorem 2 provides the foundations for the use of the FP equation for point-set registration. The first statement ensures the existence and the uniqueness of the solution to (3); the second statement ensures that $\rho(\cdot,t)$ is a pdf for all $t>0$; the third statement shows that $\rho(\cdot,t)$ converges to the target distribution for $t\to\infty$; finally, the fourth statement provides a constructive way to sample from the target distribution and also to generate trajectories that link initial and target pdfs. Notice that the prescribed Neumann boundary conditions in (3) ensure that $u\cdot\mathbf{n}\big|_{\partial\Omega} = 0$; this condition ensures that the particles do not exit the domain.

2.3. *Fokker–Planck equation and gradient flows*

The authors of [12] showed that the FP equation can be interpreted as the gradient flow of the KL divergence with respect to the Wasserstein metric. This perspective forms the foundation of the "gradient flow" method proposed in Section 3 to determine the trajectories of particles. We can hence interpret the solution to the time-discrete FP equation as the solution to a sequence of (linearized) optimal transport problems. Below, we provide an informal proof of this statement for $\Omega = \mathbb{R}^n$; we refer to [12] for the rigorous derivation and for further details.

First, we recall the 2-Wasserstein metric between the densities ρ_0 and ρ_1:

$$W_2^2(\rho_0,\rho_1) := \inf_{X:\,\mathbb{R}^n\to\mathbb{R}^n}\int_{\mathbb{R}^n}\big|X(\xi)-\xi\big|^2\rho_0(\xi)\,\mathrm{d}\xi \quad \text{s.t. } \rho_1\big(X(\xi)\big)\det\nabla X(\xi) = \rho_0(\xi). \tag{6}$$

It is possible to show that the optimal transport map X is of the form $X = \mathtt{id}+\nabla\psi$ where $\mathtt{id}(\xi)=\xi$ is the identity map and $\psi:\mathbb{R}^n\to\mathbb{R}$ (cf. [3], see also [24, Chapter 1.3.1]). Therefore, we can restate (6) as

$$W_2^2(\rho_0,\rho_1) := \inf_{\psi:\,\mathbb{R}^n\to\mathbb{R}}\int_{\mathbb{R}^n}\big|\nabla\psi(\xi)\big|^2\rho_0(\xi)\,\mathrm{d}\xi \quad \text{s.t. } \rho_1\big(\xi+\nabla\psi(\xi)\big)\det\big(\mathbb{1}+H[\psi](\xi)\big) = \rho_0(\xi), \tag{7}$$

where $\mathbb{1}\in\mathbb{R}^{n\times n}$ is the identity matrix and $H[\psi]$ is the Hessian of ψ.

We denote by $\{t^{(k)} = k\Delta t\}_{k\in\mathbb{N}}$ the time grid, and we denote by $\rho^{(k)}$ the estimate of ρ at time $t^{(k)}$; we assume that $\rho^{(k+1)}$ results from the evaluation of $\rho^{(k)}$ subject to a potential velocity field $\varepsilon\nabla\psi$

with $\varepsilon \ll 1$. We express $\rho^{(k+1)} = \rho^{(k)} + \varepsilon\eta$. Since $\rho^{(k+1)}$ satisfies $\rho^{(k+1)}\big(X(\xi)\big)\det\nabla X(\xi) = \rho^{(k)}(\xi)$ with $X = \mathtt{id} + \varepsilon\nabla\psi$, exploiting the Jacobi formula for the derivative of the determinant, we find

$$\eta = -\nabla\cdot(\rho^{(k)}\nabla\psi) + \mathscr{O}(\varepsilon). \tag{8}$$

We seek ψ to maximize the first variation of the KL divergence for a fixed Wasserstein distance:

$$\psi^\star := \arg\inf_\psi \int_{\mathbb{R}^n} F'[\rho^{(k)}]\cdot\eta(\psi)\,\mathrm{d}x \quad \text{s.t.} \quad \begin{cases} \int_{\mathbb{R}^n} |\nabla\psi|^2\rho^{(k)}\,\mathrm{d}x = c, \\ \eta(\psi) = -\nabla\cdot(\rho^{(k)}\nabla\psi), \end{cases} \tag{9}$$

where c is a given constant and $F'[\rho] = \log(\rho) + 1 + V$. If we integrate by part the objective of (9), we find

$$\int_{\mathbb{R}^n} F'[\rho^{(k)}]\cdot\eta(\psi)\,\mathrm{d}x = -\int_{\mathbb{R}^n} \nabla\cdot\big(\rho^{(k)}F'[\rho^{(k)}]\big)\psi\,\mathrm{d}x$$

In conclusion, we can restate (9) as

$$\psi^\star := \arg\inf_\psi -\int_{\mathbb{R}^n} \nabla\cdot\big(\rho^{(k)}\nabla F'[\rho^{(k)}]\big)\psi\,\mathrm{d}\xi \quad \text{s.t.} \quad \int_{\mathbb{R}^n} |\nabla\psi|^2\rho^{(k)}\,\mathrm{d}x = c. \tag{10}$$

The optimality conditions for (10) are

$$\begin{cases} -\int_{\mathbb{R}^n} \nabla\cdot\big(\rho^{(k)}\nabla F'[\rho^{(k)}]\big)\cdot\delta\psi\,\mathrm{d}x + \lambda\int_{\mathbb{R}^n} \rho^{(k)}\nabla\psi\cdot\nabla\delta\psi\,\mathrm{d}x = 0 \quad \forall\,\delta\psi\in\mathcal{V}, \\ \int_{\mathbb{R}^n} |\nabla\psi|^2\rho^{(k)}\,\mathrm{d}x = c, \end{cases} \tag{11}$$

where $\mathcal{V} := H^1(\mathbb{R}^n)$. If we integrate by part the second term in the first equation of (11), we find[1]

$$\nabla\cdot(\rho^{(k)}\nabla\psi) = -\frac{1}{\lambda}\nabla\cdot\big(\rho^{(k)}F'[\rho^{(k)}]\big),$$

where the constant λ is a function of c that is determined by substituting the latter expression in (11). We conclude that for an appropriate choice of c in (9) we have

$$\rho^{(k+1)} = \rho^{(k)} - \Delta t\nabla\cdot\big(\rho^{(k)}\nabla V + \nabla\rho^{(k)}\big),$$

which is the explicit Euler discretization of the FP equation.

2.4. *Transport of Gaussian random variables*

We denote by $\mathcal{N}(\cdot,\mu,\sigma^2)$ the pdf associated with a univariate Gaussian random variable with mean μ and variance σ^2. We assume that $\rho_0 := \mathcal{N}(\cdot,\mu,\gamma^2)$ and $\rho_\infty = \mathcal{N}(\cdot,0,\sigma^2)$. It is possible to show that the solution to the FP equation in $\Omega = \mathbb{R}$ is given by (see, e.g., [27, Chapter VIII.4])

$$\rho^{\mathrm{fp}}(x,t) = \mathcal{N}\big(x,\mu(t),\Sigma(t)\big) \quad \text{where } \mu(t) = c(t)\,\mu,\ \Sigma(t) = c(t)\gamma^2 + \big(1-c(t)\big)\sigma^2, \tag{12}$$

with $c(t) := \exp\big(-\frac{2t}{\sigma^2}\big)$. For comparison, we recall that the McCann interpolation between the same two Gaussian random variables, which corresponds to the optimal transport (OT) geodesic (shortest path) in the Wasserstein metric (see, e.g., [18, Remark 2.31]), satisfies

$$\rho^{\mathrm{ot}}(x,t) = \mathcal{N}\big(x,\mu(t),\Sigma(t)\big), \quad \text{where } \mu(t) = (1-t)\mu,\ \Sigma(t) = \big((1-t)\gamma + t\sigma\big)^2,\ \ t\in[0,1]. \tag{13}$$

We observe that the solution to (3) connects ρ_0 to ρ_∞ in infinite time at non-constant speed, while the optimal transport map achieves the same goal in finite time and at constant speed. Note, nevertheless, that the convergence of ρ^{fp} to ρ_∞ is exponentially-fast: this observation motivates the use of adaptive time stepping strategies to integrate (4).

[1] We here exploit the fact that the space $\mathcal{V}$ is dense in L^2.

3. Methodology

3.1. *High-fidelity discretization*

We rely on the finite element (FE) method to discretize (4) in space, and on the Crank–Nicolson method for time integration. Given the time grid $\{t^{(k)} = k\Delta t\}_{k\in\mathbb{N}}$ and the FE space $\widehat{\mathcal{V}} \subset \mathcal{V}$, we define the sequence of approximate solutions $\{\widehat{\rho}^{(k)}\}_{k\in\mathbb{N}} \subset \widehat{\mathcal{V}}$ such that $\widehat{\rho}^{(0)} = \rho_0$ and

$$\int_\Omega \left(\frac{\widehat{\rho}^{(k+1)} - \widehat{\rho}^{(k)}}{\Delta t}\right)\cdot v + \frac{1}{2}\left(\nabla\widehat{\rho}^{(k+1)} + \widehat{\rho}^{(k+1)}\nabla V + \nabla\widehat{\rho}^{(k)} + \widehat{\rho}^{(k)}\nabla V\right)\cdot\nabla v\,\mathrm{d}x = 0 \quad \forall\, v\in\widehat{\mathcal{V}}, \tag{14}$$

for $k = 0, 1, \ldots$ We remark that problem (14) might feature strong advection: for this reason, in the numerical experiments, we resort to a SUPG stabilization of (14) to avoid spurious oscillations [4]; we refer to [19, Chapter 12.8.7] for the detailed presentation of the scheme and for the choice of the parameters. We further remark that our method is not guaranteed to preserve positivity; in the numerical experiments, we hence choose the time step and mesh size sufficiently small to ensure positivity through accuracy.

Given the particles $\{X_j^{(0)}\}_{j=1}^N$, we consider two distinct strategies to generate the approximate trajectories: the *"ODE" method* and the *"gradient flow (GF)" method*. The former is justified by property (4) in Theorem 2, while the latter is motivated by the discussion in Section 2.3.

ODE method. In this approach, we discretize the ODE system

$$\begin{cases} \dot{X}_i(\tau) = \widehat{\mathbf{u}}\big(X_i(\tau),\tau\big) & \tau\in(t^{(k)},t^{(k+1)}),\\ X_i(t^{(k)}) = X_i^{(k)}, \end{cases} \tag{15}$$

where $\widehat{\mathbf{u}}(\cdot,\tau) = -\nabla\big(V + \log\widehat{\rho}(\cdot,\tau)\big)$ and $\widehat{\rho}(\cdot,\tau)$ is a suitable approximation of the density in the interval $(t^{(k)}, t^{(k+1)})$. In the numerical experiments, we consider an explicit Euler discretization of (15) based on a single time step:

$$X_i^{(k+1)} = X_i^{(k)} + \Delta t\,\widehat{\mathbf{u}}^{(k)}(X_i^{(k)}), \quad \text{where } \widehat{\mathbf{u}}^{(k)} := -\nabla\big(V + \log(\widehat{\rho}^{(k)})\big),\ \ i = 1,\ldots,N,\ \ k = 0,1,\ldots \tag{16}$$

We also consider an explicit second-order Runge–Kutta (RK2) scheme for (15) based on the linear interpolation of the density field

$$\widehat{\rho}(\cdot,\tau) = \left(1 - \frac{\tau - t^{(k)}}{\Delta t}\right)\widehat{\rho}^{(k)} + \left(\frac{\tau - t^{(k)}}{\Delta t}\right)\widehat{\rho}^{(k+1)}, \quad \tau\in(t^{(k)},t^{(k+1)}).$$

To further increase accuracy, we also lower the time step used for (15).

GF method. Given the estimates $\widehat{\rho}^{(k)}, \widehat{\rho}^{(k+1)}$ obtained through (14), we compute the potential $\widehat{\psi}^{(k+1/2)} \in \widehat{\mathcal{V}}$ such that

$$\int_\Omega \left(\frac{\widehat{\rho}^{(k)} + \widehat{\rho}^{(k+1)}}{2}\right)\nabla\widehat{\psi}^{(k+1/2)}\cdot\nabla v\,\mathrm{d}x = \int_\Omega (\widehat{\rho}^{(k+1)} - \widehat{\rho}^{(k)})\,v\,\mathrm{d}x, \quad \forall\, v\in\widehat{\mathcal{V}}, \tag{17a}$$

and then we estimate the positions of the particles as

$$X_i^{(k+1)} = X_i^{(k)} + \widehat{\mathbf{u}}^{(k+1/2)}(X_i^{(k)}), \quad \text{where } \widehat{\mathbf{u}}^{(k+1/2)} := \nabla\psi^{(k+1/2)},\ \ i = 1,\ldots,N,\ \ k = 0,1,\ldots \tag{17b}$$

Note that (17a) corresponds to the linearization of the Monge–Ampère equation; in future works, we shall exploit this analogy to determine higher-order approximations of the trajectories of the particles. We observe that the linear system associated with (17a) is poorly conditioned when $\inf_{x\in\Omega}\widehat{\rho}^{(k)}$ is small. In the numerical examples, we consider the regularized problem:

$$\int_\Omega \left(\widehat{\rho}^{(k)}\nabla\widehat{\psi}^{(k)}\cdot\nabla v + \epsilon_{\mathrm{reg}}\,\widehat{\psi}^{(k)}\,v\right)\mathrm{d}x = \int_\Omega (\widehat{\rho}^{(k+1)} - \widehat{\rho}^{(k)})\,v\,\mathrm{d}x, \quad \forall\, v\in\widehat{\mathcal{V}}, \tag{18}$$

where $\epsilon_{\mathrm{reg}} > 0$ is a regularization parameter that is set equal to 10^{-10} in the numerical experiments.

3.2. *Treatment of boundaries*

For several applications, we envision that it is necessary to control the distance of the particles from the boundaries of the domain. In our setting, this can be achieved by introducing a regularization in the potential V at the price of introducing a controlled bias. In more detail, we consider the regularized potential

$$V_\epsilon(x) = -\log\big(\rho_\infty(x)\big) + \frac{\epsilon}{w_\delta(x)}, \tag{19}$$

where $w_\delta : \Omega \to \mathbb{R}_+$ is a (regularized) distance function from the boundary of Ω. The second term in (19) introduces a repulsive force that is active in the proximity of the boundary and prevents particles from approaching the boundary too closely. The constant $\epsilon > 0$ is a regularization parameter whose choice is investigated in the numerical examples, while the constant $\delta > 0$ is the smoothing parameter that is introduced below.

In order to define the function w_δ in (19), we first introduce the distance function $w(x) = \max\big(\mathrm{dist}(x, \partial\Omega), \mathtt{tol}\big)$, where the user-defined tolerance `tol` prevents division by zero in (19), for the nodes on the obstacle. Since the distance function w is in general only Lipschitz continuous, we apply a smoothing to ensure that ∇V is sufficiently smooth. We consider a non-conforming approximation of the biharmonic equation for smoothing [15]. We denote by $\mathtt{F}_j$ the j-th facet of the mesh, by $\mathbf{n}^+$ the positive normal to each facet, by $\mathtt{D}_k$ the k-th element of the mesh, by $\mathtt{I}_{\text{int}}$ the set of internal facets of the mesh. Given the field $w : \Omega \to \mathbb{R}$, we define the positive and the negative limits on the facets of the mesh

$$w^\pm(x) := \lim_{\epsilon \to 0^+} w(x \mp \epsilon \mathbf{n});$$

then, we introduce the normal jump of the gradient at each facet $[\![\nabla w]\!] = \mathbf{n}^+ \cdot (\nabla w^+ - \nabla w^-)$ and the facet average $\{w\} = \frac{1}{2}(w^+ + w^-)$. Exploiting the previous definitions, we define the smoothing problem: find $w_\delta \in \widehat{\mathcal{V}}$ such that $w_{\delta|_{\partial\Omega}} = \mathtt{tol}$ and

$$a(w_\delta, v) + \frac{1}{\delta}\int_\Omega (w_\delta - w)\cdot v \,\mathrm{d}x = 0 \quad \forall\, v \in \widehat{\mathcal{V}} : v\big|_{\partial\Omega} = 0, \tag{20a}$$

where

$$a(w, v) = \sum_{k=1}^{N_\mathrm{e}} \int_{\mathtt{D}_k} (\Delta w \cdot \Delta v)\,\mathrm{d}x + \sum_{j \in \mathtt{I}_{\text{int}}} \int_{\mathtt{F}_j} \Big(\beta_j [\![\nabla w]\!]\cdot[\![\nabla v]\!] + \tfrac{1}{\beta_j}\{\Delta w\}\cdot\{\Delta v\}\Big)\,\mathrm{d}s, \tag{20b}$$

and $\beta_j = \sigma_\beta \kappa^2 |\mathtt{F}_j|^{-1}$ with $\sigma_\beta = 10$, and κ equal to the polynomial degree of the FE basis. We refer to [15] for a detailed analysis of the formulation.

3.3. *Estimate of the target density*

The application of the FP equation to PSR requires the estimate of the densities ρ_0 and ρ_∞. Given the initial and the target point clouds $\{X_i^0\}_{i=1}^N$ and $\{X_j^\infty\}_{j=1}^M$, we fit two Gaussian mixture models (GMMs) to determine the estimates of ρ_0 and ρ_∞ that enter in (14) (see, e.g., [9, Chapter 8]). We resort to the Matlab function `fitgmdist` to determine the GMMs associated with the point clouds. We determine the number of mixture models using the Aikake information criterion (AIC, see [9, Chapter 7.5]); furthermore, we introduce a regularization of the covariance matrix to robustify the learning procedure[2].

The use of GMMs for density estimation ensures that ρ_0 and ρ_∞ are strictly positive for all $x \in \Omega$, which is crucial to apply the results of Theorem 2. Furthermore, we recall that for Gaussian distributions $\rho = \mathcal{N}(\cdot; \mu, \Sigma)$ we have

$$\log\big(\mathcal{N}(x; \mu, \Sigma)\big) = -\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu). \tag{21}$$

[2] We consider the regularization value 10^{-2}; see the Matlab documentation for further details.

Since Σ is a $n \times n$ symmetric positive definite matrix, this implies that for GMMs the gradient of $\log(\rho)$ grows linearly with respect to the squared distance from the centers of the mixture models.

4. Numerical results

We present two numerical experiments that involve the transport of point clouds past a bluff body. The examples are designed to illustrate the ability of our method to deform point clouds past two-dimensional objects. As discussed in [11], this tool is central in the development of model reduction techniques based on coordinate transformations. In [11], the point clouds are associated to coherent structures of the solution field (e.g., shock waves) that are displaced due to parameter variations.

4.1. *Transport of Gaussian distributions across a cylinder*

4.1.1. *Problem setting*

We consider the problem of transporting the Gaussian density $\rho_0 = \mathcal{N}(\cdot;\mu_0,\Sigma_0)$ with $\mu_0 = [-2,0]$ and $\Sigma_0 = 0.2\mathbb{1}$ to $\rho_\infty = \mathcal{N}(\cdot;\mu_\infty,\Sigma_\infty)$ with $\mu_\infty = [2,0]$ and $\Sigma_\infty = \Sigma_0$ in the domain $\Omega = (-4,4)^2 \setminus \mathcal{B}_{r=0.5}(0)$. Figure 1 shows the initial and the target distributions.

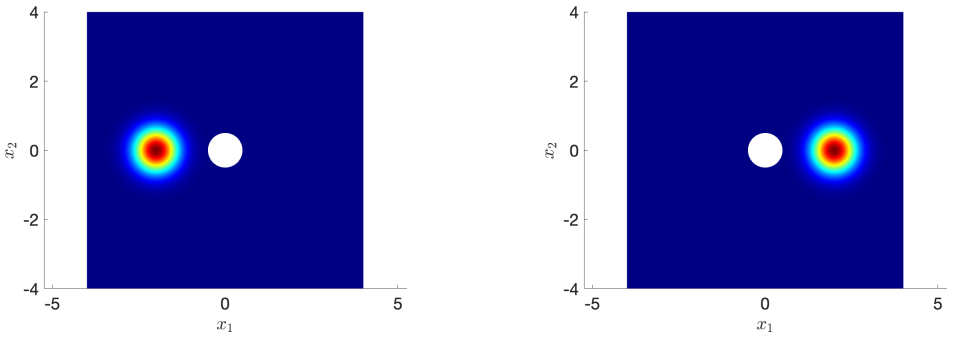


Figure 1. Transport of Gaussian distributions across a cylinder. (a) Initial distribution. (b) Target distribution.

We resort to a P2 FE discretization with $N_{\text{hf}} = 6469$ degrees of freedom; we integrate the equation in the time interval $(0, T = 5)$ and we consider the grid

$$\left\{ t^{(k)} = T\left(\frac{k}{K}\right)^{1.5} \right\}_{k=0}^{K}, \quad K = 3000. \tag{22}$$

The non-homogeneous time discretization is motivated by the observation that the dynamics is much faster at the onset of the simulation. The exponent 1.5 was determined empirically through systematic experimentation; in the future, we plan to implement more rigorous adaptive time stepping strategies. The regularized distance function in (19) is computed using $\delta = 10^{-2}$ and $\texttt{tol} = 10^{-2}$ and several values of ϵ.

We further remark that ∇V is computed by differentiating the FE approximation of $V = -\log(\rho_\infty)$: we notice that approximating the density ρ_∞ through its nodal representation ρ_∞^{fe} and then computing ∇V as $\nabla\rho_\infty^{\text{fe}}/\rho_\infty^{\text{fe}}$ leads to significantly less accurate results. Recalling (21), for Gaussian distributions, $V = -\log(\rho_\infty)$ is a second-order polynomial whose gradient can

be computed exactly. Furthermore, we empirically found that the computation of $\nabla\rho_\infty^{\text{fe}}/\rho_\infty^{\text{fe}}$ is numerically unstable when ρ_∞ is very small.

To validate the PSR procedure, we generate $X_1^0,\dots,X_N^0 \overset{\text{iid}}{\sim} \rho_0$ with $N = 100$ and we compare the trajectories obtained using our method with $M = 100$ i.i.d. samples from the target distribution. Time integration of the particles is performed using the ODE method based on the explicit Euler method (16) with the same time step that is used for the integration of (14). We also considered the GF method and the ODE method with RK2 time integration: since the results of the latter two methods are nearly equivalent to the ones obtained using (16), we do not report them here.

4.1.2. *Results*

Figure 2 shows the performance of our method. Figure 2a shows the behavior of the L^1 error $\|\rho(\cdot,t)-\rho_\infty\|_{L^1(\Omega)}$ for several values of ϵ: as we increase ϵ in (19), we introduce an increasing bias in the formulation that prevents convergence to the target density. Figure 2b compares the deformed point cloud at the final time $\{X_j^{(K)}\}_j$ for $\epsilon = 0$ with the "target" point cloud generated using ρ_∞: we observe that the results are extremely satisfactory for this model problem. Figure 2c shows the behavior of select trajectories, while Figure 2d shows the trajectory of one particle for several values of ϵ: we notice that, as we increase ϵ, the particle moves away from the cylinder due to the effect of the repulsive force.

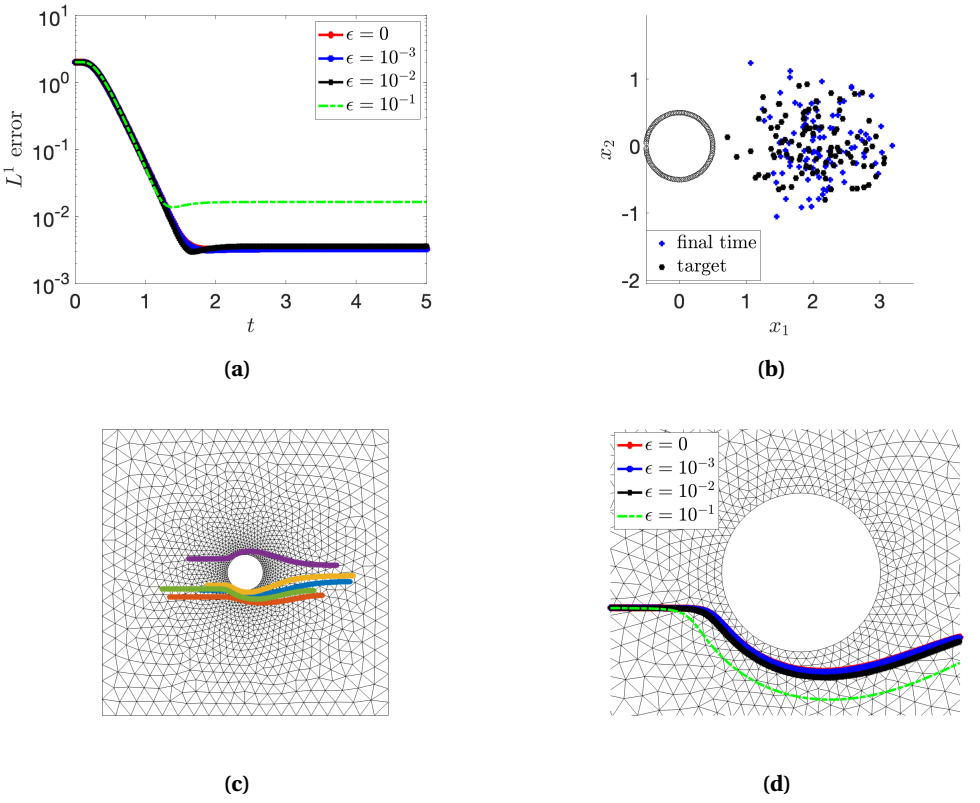


Figure 2. Transport of Gaussian distributions across a cylinder. (a) Behavior of the L^2 error $\|\rho(\cdot,t)-\rho_\infty\|_{L^1(\Omega)}$ for several choices of the regularization parameter ϵ. (b) Deformed point cloud at final time for $\epsilon = 0$ and target point cloud. (c) Behavior of several trajectories for $\epsilon = 0$. (d) Behavior of one trajectory for several values of $\epsilon = 0$.

4.2. *Point-set registration across a cylinder*

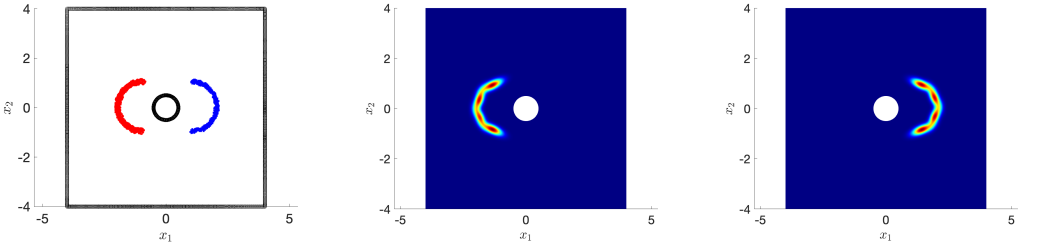
4.2.1. *Problem setting*

We consider the problem of continuously deforming the point cloud $\{X_i^0\}_{i=1}^N$ to match the target point cloud $\{X_j^\infty\}_{j=1}^M$. We consider the same domain Ω of the previous example and we define

$$\begin{cases} X_i^0 = \left[\cos\left(\theta_0 + \Delta\theta\,\frac{i-1}{N-1}\right) - 1, \sin\left(\theta_0 + \Delta\theta\,\frac{i-1}{N-1}\right)\right] + 0.1\,\widetilde{X}_i^0 & i = 1,\dots,N; \\ X_j^\infty = \left[\cos\left(\theta_\infty + \Delta\theta\,\frac{j-1}{M-1}\right) + 1, \sin\left(\theta_\infty + \Delta\theta\,\frac{j-1}{N-1}\right)\right] + 0.1\,\widetilde{X}_j^\infty & j = 1,\dots,M; \end{cases} \tag{23}$$

where $N = M = 141$, $\theta_0 = \frac{\pi}{2}$, $\theta_\infty = \frac{3\pi}{2}$, $\Delta\theta = \pi$ and $\widetilde{X}_i^0, \widetilde{X}_j^\infty \overset{\text{iid}}{\sim} \text{Uniform}\left((0,1)^2\right)$. Figure 3a shows the reference and the target point clouds. We consider two distinct discretizations of increasing size. The coarse discretization features a P2 FE mesh with $N_{\text{hf}} = 23\,409$ degrees of freedom and $K = 5 \cdot 10^3$ time steps; the fine discretization features a P2 FE mesh with $N_{\text{hf}} = 93\,134$ degrees of freedom and $K = 10^4$ time steps. In both cases, we integrate the system in the time interval $(0, T = 15)$ and we consider the non-uniform time grid (22).

We rely on the unregularized potential (i.e., $\epsilon = 0$ in (19)). We compare the ODE method based on forward Euler time integration with the GF method. We also run simulations for the ODE method with RK2 time integration with two time steps per time interval: since the results of the latter are comparable with the ones of the ODE method, they are not reported below. We model the point clouds (23) using two GMMs with four components — we recall that the number of components is selected by maximizing the AIC criterion. Figures 3b and 3c show the GMM densities that are obtained by the learning procedure.



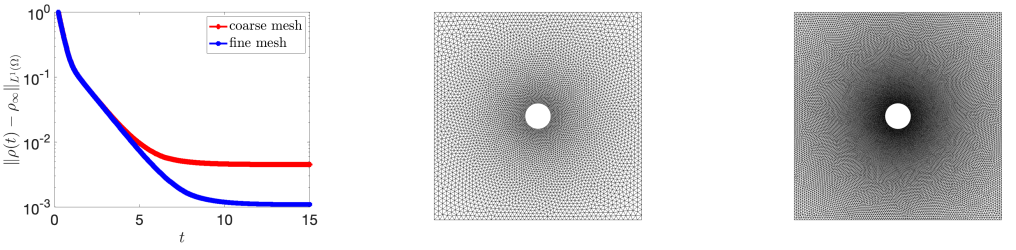
(a) **(b)** **(c)**

Figure 3. Point-set registration across a cylinder. (a) Reference (red) and target point (blue) clouds. (b) Initial condition ρ_0. (c) Target density ρ_∞.

4.2.2. *Results*

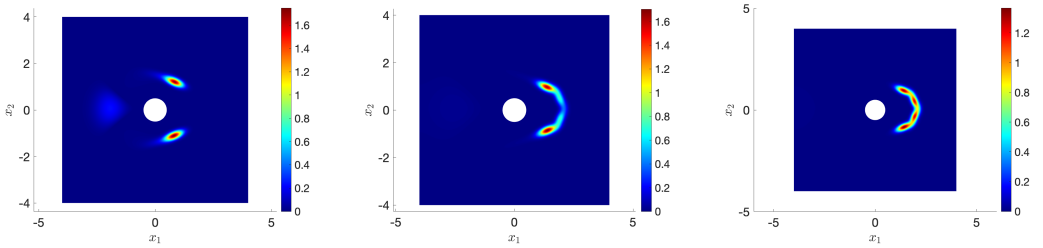
Figure 4a shows the behavior of the L^1 error $\left\|\rho(\cdot,t) - \rho_\infty\right\|_{L^1(\Omega)}$ over time for the two discretizations; Figures 4b and 4c show the two FE meshes. The error in the target density decreases by a factor 4: it is hence inversely proportional to the size of the mesh, as expected for second-order discretizations.

Figure 5 shows the behavior of the solution to (14) for three time instants. As expected, we notice that the solution is symmetric with respect to the cylinder. We further notice that for small time steps the solution has three peaks: the majority of the mass — corresponding to the top and the bottom peaks at $t = 0.1$ — rapidly goes through the cylinder, while a small portion of the mass — corresponding to the weaker peak on the left of the cylinder at $t = 0.1$ — reaches the target density later.



(a) **(b)** **(c)**

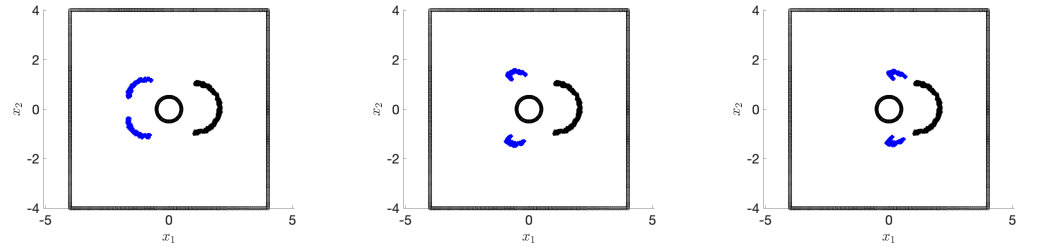
Figure 4. Point-set registration across a cylinder. (a) Behavior of the L^1 error $\|\rho(\cdot,t) - \rho_\infty\|_{L^1(\Omega)}$ for both coarse and fine meshes. (b) Coarse mesh ($N_{\text{hf}} = 23\,409$). (c) Fine mesh ($N_{\text{hf}} = 93\,134$).



(a) $t = 0.1$ **(b)** $t = 0.5$ **(c)** $t = 1.5$

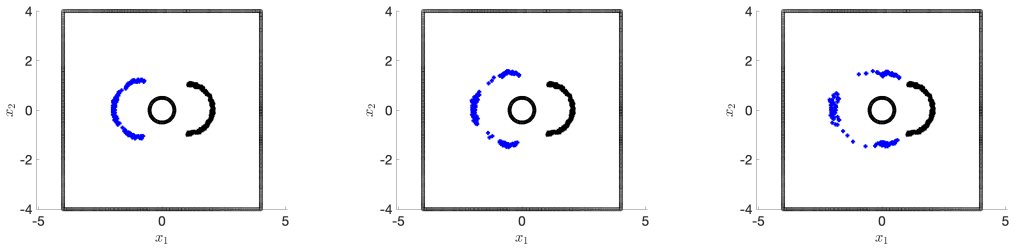
Figure 5. Point-set registration across a cylinder; density profiles for three time instants.

Figures 6 and 7 show the behaviors of the point clouds for the two particle integration methods discussed in Section 3: we emphasize that both approaches consider the same particles and the same estimates of the initial and target distributions. The results correspond to the fine discretization. Interestingly, the qualitative behavior of the two methods is significantly different, particularly for small values of t: in the ODE method the point cloud is nearly split in two parts at $t = 0.1$ and all the points proceed towards the target cloud with similar speeds; instead, in the GF method, we can clearly identify three clusters whose elements move towards the target cloud with different speeds. The latter behavior appears to be more consistent with the results for the density in Figure 5.



(a) $t = 0.1$ **(b)** $t = 0.5$ **(c)** $t = 1.5$

Figure 6. Point-set registration across a cylinder; deformed point clouds $\{X_i(t)\}_{i=1}^N$ for three time instants. ODE method.



(a) $t = 0.1$ **(b)** $t = 0.5$ **(c)** $t = 1.5$

Figure 7. Point-set registration across a cylinder; deformed point clouds $\{X_i(t)\}_{i=1}^N$ for three time instants. GF method.

Figure 8 compares the point clouds obtained with the two methods at the final time. We notice that the ODE method does not provide satisfactory results for this test case, while the GF method is significantly more accurate.

(a) ODE **(b)** GF

Figure 8. Point-set registration across a cylinder; deformed point clouds $\{X_i(t)\}_{i=1}^N$ for $t = T_{\max}$ for the two time integration methods.

5. Conclusions and perspectives

We developed a PSR method for the interpolation of point clouds in bounded domains based on the Fokker–Planck equation. Our method relies on GMMs for density estimation and on a standard (Eulerian) discretization of the FP equation for the evaluation of the density. We also proposed two distinct strategies for the estimation of the trajectories of the particles, which are both rigorously justified by the analysis of Section 2.

The numerical experiments of Section 4 showed the potential of the method for PSR. The regularized potential (19) provides a simple yet effective way to control the distance of the particles from the boundaries at all times; furthermore, in both numerical experiments, the convergence to the target distribution is rapid and monotonic. As regards the motion of particles, the results of Section 4.2 clearly showed the superiority of the GF approach based on the solution to a linearized OT problem over the ODE approach based on the explicit integration of the characteristics.

Despite the linearity of the equation, the solution to (4) using the FE method on a fixed grid is expensive and demands further investigations. We shall devise a specialized solver for the FP equation with emphasis on the accurate approximation of the particles' dynamics. Specifically,

we plan to consider (i) adaptive time stepping schemes and (ii) adaptive spatial discretizations to reduce the overall costs: we expect to borrow ideas from deterministic particle methods for the FP equation [20–22]. In the future, we also plan to investigate variants of (19) to effectively control the distance of the particles from the boundaries at a reduced impact on the limit distribution. Finally, we plan to assess the robustness of our approach with respect to the choice of the densities ρ_0, ρ_∞.

Appendix A. Proofs

A.1. *Proof of Theorem 1*

Proof of Theorem 1(1). We observe that

$$g(t) := t - 1 - \log(t) \geq 0 \quad \forall\, t \in \mathbb{R}_+. \tag{24}$$

To prove (24), we notice that $t = 1$ is a global minimum of g in $\mathbb{R}_+$ and $g(1) = 0$. If we define the set $A := \{x \in \Omega : \rho(x) > 0\} \subset \Omega$, we find

$$-D_{\mathrm{KL}}(\rho \parallel \rho_\infty) = \int_\Omega \rho \log\left(\frac{\rho_\infty}{\rho}\right) \mathrm{d}x = \int_A \rho \log\left(\frac{\rho_\infty}{\rho}\right) \mathrm{d}x;$$

then, exploiting (24) with $t = \frac{\rho_\infty}{\rho}$, we find

$$-D_{\mathrm{KL}}(\rho \parallel \rho_\infty) = \int_A \rho \log\left(\frac{\rho_\infty}{\rho}\right) \leq \int_A \rho \left(\frac{\rho_\infty}{\rho} - 1\right) = \int_A \rho_\infty - 1 \leq \int_\Omega \rho_\infty - 1 = 0,$$

which proves $D_{\mathrm{KL}}(\rho \parallel \rho_\infty) \geq 0$.

Let $\rho \in \mathscr{P}_2(\Omega)$ satisfy $D_{\mathrm{KL}}(\rho \parallel \rho_\infty) = 0$. If we denote by $\phi\colon t \mapsto -\log(t)$ and $f\colon x \mapsto \frac{\rho(x)}{\rho_\infty(x)}$, we find

$$\phi\left(\int_\Omega f(x)\rho(x)\,\mathrm{d}x\right) = \phi(1) = 0, \qquad \int_\Omega \phi \circ f(x)\rho(x)\,\mathrm{d}x = -D_{\mathrm{KL}}(\rho \parallel \rho_\infty) = 0.$$

Notice that $\phi\left(\int_\Omega f(x)\rho(x)\,\mathrm{d}x\right) = \int_\Omega \phi \circ f(x)\rho(x)\,\mathrm{d}x$: since f is convex, recalling Jensen's inequality, we must have that f is constant, which implies that $\rho = \rho_\infty$ a.s. □

Proof of Theorem 1(2). It suffices to show an example for which $D_{\mathrm{KL}}(\rho \parallel \rho_\infty) \neq D_{\mathrm{KL}}(\rho_\infty \parallel \rho)$. Towards this end, consider $\Omega = (-1, 1)$ and

$$\rho(x) \equiv \frac{1}{2}, \quad \rho_\infty(x) = \begin{cases} 0.1 & x \in (-1, 0), \\ 0.9 & x \in (0, 1). \end{cases}$$

By straightforward calculations, we find $D_{\mathrm{KL}}(\rho \parallel \rho_\infty) = 0.5108$ and $D_{\mathrm{KL}}(\rho_\infty \parallel \rho) = 0.3681$. □

Proof of Theorem 1(3). We define $\rho_\lambda := \lambda\rho_0 + (1-\lambda)\rho_1$. Since the function $t \mapsto t\log(t/c)$ is strictly convex for any $c > 0$, we have

$$\rho_\lambda(x)\log\left(\frac{\rho_\lambda(x)}{\rho_\infty(x)}\right) \leq \lambda\rho_0(x)\log\left(\frac{\rho_0(x)}{\rho_\infty(x)}\right) + (1-\lambda)\rho_1(x)\log\left(\frac{\rho_1(x)}{\rho_\infty(x)}\right), \quad \forall\, x \in \Omega,\ \ \lambda \in (0,1);$$

furthermore, the latter holds with equality if and only if $\rho_0(x) = \rho_1(x)$. Recalling the expression of the KL divergence, we obtain the desired result. □

Proof of Theorem 1(4). We can exploit [7, Theorem 3] to show that $\|\rho - \rho_\infty\|_{L^1(\Omega)} \leq \sqrt{2D_{\mathrm{KL}}(\rho \parallel \rho_\infty)}$; then, the desired result follows from the symmetry of the norm. □

A.2. *Proof of Theorem 2*

(1) The existence and uniqueness of the solution to (4) can be studied using the theory of abstract parabolic problems. We refer to [23, Chapter 10, Theorem 10.11] for a detailed proof.

(2) Exploiting [23, Theorem 10.11], we find that $\rho(\cdot,t) \in L^2(\Omega)$ for all $t>0$; it hence suffices to show that $\int_\Omega \rho(x,t)\,\mathrm{d}x = 1$ for all $t>0$ and that $\rho(x,t)\geq 0$ for a.e. $x\in\Omega$ and $t>0$. If we consider $v=1$ in (4), we find

$$0 = \int_\Omega \partial_t \rho(x,t)\,\mathrm{d}x = \frac{\mathrm{d}}{\mathrm{d}t}\int_\Omega \rho(x,t)\,\mathrm{d}x \quad \Longrightarrow \quad \int_\Omega \rho(x,t)\,\mathrm{d}x = \int_\Omega \rho_0(x)\,\mathrm{d}x = 1,$$

which proves conservation of mass. The proof of positivity is significantly more technical and is provided in [17, Corollary 4.3].

(3) Recalling notation introduced in (2), since $\rho\geq 0$ a.e., we find that

$$\frac{\mathrm{d}}{\mathrm{d}t} D_{\mathrm{KL}}\big(\rho(\cdot,t)\,\|\,\rho_\infty\big) = -\int_\Omega \rho(x,t)\big|\nabla F'\big[\rho(x,t)\big]\big|^2\,\mathrm{d}x.$$

Provided that $c := \inf_{x\in\Omega}\rho_\infty(x)$ is strictly positive and $\rho_0\in L^2(\Omega)$, we have that $D_{\mathrm{KL}}(\rho_0\,\|\,\rho_\infty)<\infty$: we have indeed

$$\begin{aligned} D_{\mathrm{KL}}(\rho_0\,\|\,\rho_\infty) &\leq \int_\Omega \rho_0 \log\Big(\frac{\rho_0}{c}\Big) \\ &= \int_\Omega \rho_0 \log\rho_0\,\mathrm{d}x - \log(c) \\ &\leq \int_\Omega (\rho_0^2+\rho_0)\,\mathrm{d}x - \log(c) \\ &= \|\rho_0\|^2_{L^2(\Omega)} + 1 - \log(c), \end{aligned}$$

where in the second inequality we used the identity $t\log(t)\leq t^2+t$ that is valid for all $t\in\mathbb{R}_+$. Therefore, since $\rho(\cdot,t)\in\mathscr{P}_2(\Omega)$ for all $t>0$, we find that the KL divergence is monotonically-decreasing and uniformly bounded. Exploiting the Pinsker's inequality, we also find that $t\mapsto \big\|\rho(\cdot,t)\big\|_{L^1(\Omega)}$ is uniformly bounded. Then, we can apply [1, Theorem 2.3] to conclude that $\rho(\cdot,t)\to\rho_\infty$ in L^1 for $t\to\infty$.

(4) The result exploits the Lagrangian interpretation of the FP equation, which is extensively discussed in [24, Chapter 4] — the precise result is provided in [24, Theorem 4.4].

Declaration of interests

The authors do not work for, advise, own shares in, or receive funds from any organization that could benefit from this article, and have declared no affiliations other than their research organizations.

References

[1] N. J. Alves, J. Skrzeczkowski and A. E. Tzavaras, "Strong convergence of sequences with vanishing relative entropy", 2024. Online at https://arxiv.org/abs/2409.16892.

[2] M. F. Beg, M. I. Miller, A. Trouvé and L. Younes, "Computing large deformation metric mappings via geodesic flows of diffeomorphisms", *Int. J. Comput. Vision* **61** (2005), pp. 139–157.

[3] Y. Brenier, "Polar factorization and monotone rearrangement of vector-valued functions", *Commun. Pure Appl. Math.* **44** (1991), no. 4, pp. 375–417.

[4] A. N. Brooks and T. Hughes Jr, "Streamline upwind/Petrov–Galerkin formulations for convection dominated flows with particular emphasis on the incompressible Navier–Stokes equations", *Comput. Methods Appl. Mech. Eng.* **32** (1982), no. 1-3, pp. 199–259.

[5] S. Cucchiara, A. Iollo, T. Taddei and H. Telib, "Model order reduction by convex displacement interpolation", *J. Comput. Phys.* **514** (2024), article no. 113230 (22 pages).

[6] C. W. Fox and S. J. Roberts, "A tutorial on variational Bayesian inference", *Artif. Intell. Rev.* **38** (2012), pp. 85–95.

[7] G. L. Gilardoni, "On Pinsker's and Vajda's type inequalities for Csiszár's f-divergences", *IEEE Trans. Inf. Theory* **56** (2010), no. 11, pp. 5377–5386.

[8] R. M. Gray, *Entropy and information theory*, Second edition, Springer, 2011, pp. xxviii+409.

[9] T. Hastie, R. Tibshirani and J. Friedman, *The elements of statistical learning*, Second edition, Springer Series in Statistics, Springer, 2009, pp. xxii+745.

[10] A. Iollo and D. Lombardi, "Advection modes by optimal mass transfer", *Phys. Rev. E* **89** (2014), no. 2, article no. 022923 (33 pages).

[11] A. Iollo and T. Taddei, "Mapping of coherent structures in parameterized flows by learning optimal transportation with Gaussian models", *J. Comput. Phys.* **471** (2022), article no. 111671 (24 pages).

[12] R. Jordan, D. Kinderlehrer and F. Otto, "The variational formulation of the Fokker–Planck equation", *SIAM J. Math. Anal.* **29** (1998), no. 1, pp. 1–17.

[13] S. Kullback, *Information theory and statistics*, Dover Publications, 1997, pp. xvi+399.

[14] J. Ma, J. Wu, J. Zhao, J. Jiang, H. Zhou and Q. Z. Sheng, "Nonrigid point set registration with robust transformation learning under manifold regularization", *IEEE Trans. Neural Netw. Learn. Syst.* **30** (2018), no. 12, pp. 3584–3597.

[15] I. Mozolevski, E. Süli and P. R. Bösing, "hp-version a priori error analysis of interior penalty discontinuous Galerkin finite element approximations to the biharmonic equation", *J. Sci. Comput.* **30** (2007), no. 3, pp. 465–491.

[16] A. Myronenko and X. Song, "Point set registration: coherent point drift", *IEEE Trans. Pattern Anal. Mach. Intell.* **32** (2010), no. 12, pp. 2262–2275.

[17] E.-M. Ouhabaz, *Analysis of heat equations on domains*, London Mathematical Society Monographs, Princeton University Press, 2009, pp. xi+284.

[18] G. Peyré and M. Cuturi, *Computational optimal transport: with applications to data science*, now publishers, 2019, 272 pages.

[19] A. Quarteroni, *Numerical models for differential problems*, MS&A. Modeling, Simulation and Applications, Springer, 2009, pp. xvi+601.

[20] G. Russo, "A particle method for collisional kinetic equations. I. Basic theory and one-dimensional results", *J. Comput. Phys.* **87** (1990), no. 2, pp. 270–300.

[21] G. Russo, "Deterministic diffusion of particles", *Commun. Pure Appl. Math.* **43** (1990), no. 6, pp. 697–733.

[22] G. Russo, "A deterministic vortex method for the Navier–Stokes equations", *J. Comput. Phys.* **108** (1993), no. 1, pp. 84–94.

[23] S. Salsa, *Partial differential equations in action*, Third edition, Unitext, Springer, 2016, pp. xviii+686.

[24] F. Santambrogio, *Optimal transport for applied mathematicians*, Progress in Nonlinear Differential Equations and their Applications, Birkhäuser/Springer, 2015, pp. xxvii+353.

[25] Y. Song and S. Ermon, "Generative modeling by estimating gradients of the data distribution", in *Advances in Neural Information Processing Systems 32 (NeurIPS 2019)* (H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox and R. Garnett, eds.), Advances in Neural Information Processing Systems, vol. 32, NeurIPS, 2019, 13 pages.

[26] T. Taddei, "A registration method for model order reduction: data compression and geometry reduction", *SIAM J. Sci. Comput.* **42** (2020), no. 2, pp. A997–A1027.
[27] N. G. Van Kampen, *Stochastic processes in physics and chemistry*, Elsevier, 1992, 480 pages.