Wasserstein gradient flow formulation of the time-fractional Fokker-Planck equation

Duong, Manh Hong; Jin, Bangti

DOI:
10.4310/CMS.2020.v18.n7.a6

Document Version
Peer reviewed version

Citation for published version (Harvard):
Duong, MH & Jin, B 2020, 'Wasserstein gradient flow formulation of the time-fractional Fokker-Planck equation', *Communications in Mathematical Sciences*, vol. 18, no. 7, pp. 1949–1975.
https://doi.org/10.4310/CMS.2020.v18.n7.a6

WASSERSTEIN GRADIENT FLOW FORMULATION OF THE TIME-FRACTIONAL FOKKER-PLANCK EQUATION*

MANH HONG DUONG† AND BANGTI JIN ‡

Abstract. In this work, we investigate a variational formulation for a time-fractional Fokker-Planck equation which arises in the study of complex physical systems involving anomalously slow diffusion. The model involves a fractional-order Caputo derivative in time, and thus inherently nonlocal. The study follows the Wasserstein gradient flow approach pioneered by [26]. We propose a JKO type scheme for discretizing the model, using the L1 scheme for the Caputo fractional derivative in time, and establish the convergence of the scheme as the time step size tends to zero. Illustrative numerical results in one- and two-dimensional problems are also presented to show the approach.

Keywords. Wasserstein gradient flow; time-fractional Fokker-Planck equation; convergence of time-discretization scheme.

AMS subject classifications. 35Q84; 65M12; 60G22

1. Introduction

In this work, we are interested in the following time-fractional Fokker-Planck equation (FPE):

$$\begin{cases} \partial_t^\alpha \rho = \operatorname{div}(\nabla \Psi \rho) + \Delta \rho, & \text{in } \mathbb{R}^d \\ \rho(0) = \rho_0, \end{cases} \tag{1.1}$$

where ρ_0 is the initial datum, and Ψ is the forcing term. Here, the notation $\partial_t^\alpha \varphi(t)$ denotes the Caputo fractional derivative of order $\alpha \in (0,1)$ in time, defined by [28, p. 91]

$$\partial_t^\alpha \varphi(t) = \frac{1}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1} \varphi'(s)\,\mathrm{d}s$$

where $\Gamma(z)$ is the Gamma function defined by $\Gamma(z) = \int_0^\infty s^{z-1} e^{-s}\,\mathrm{d}s$. The fractional derivative $\partial_t^\alpha \varphi(t)$ recovers the usual first-order derivative $\varphi'(t)$ as $\alpha \to 1^-$ for suitably smooth functions. Therefore, the model (1.1) can be regarded as a time-fractional analogue of the classical FPE.

The interest in the model (1.1) is motivated by an explosively growing list of practical applications involving anomalously slow diffusion processes (a.k.a. subdiffusion), which deviate from the classical diffusive behavior. The so-called subdiffusive process displays local motion occasionally interrupted by long sojourns and trapping effects, and it has been widely accepted to better describe transport phenomena in a number of practical applications in physics, biology and finance, e.g., the study of volatility of financial markets, bacterial motion and bird flight, etc. (see the review [37] for an extensive list with physical modelings). Model (1.1) can be viewed as the macroscopic limit of continuous time random walk with a heavy-tailed waiting time distribution (with a divergent mean) between consecutive jumps [7], in analogy with Brownian motion

*Received date, and accepted date (The correct dates will be entered by the editor).

†School of Mathematics, University of Birmingham, Birmingham B15 2TT, UK (h.duong@bham.ac.uk).

‡Department of Computer Science, University College London, Gower Street, London, WC1E 6BT, UK (b.jin@ucl.ac.uk,bangti.jin@gmail.com).

for normal diffusion. The evolution of the probability density function (PDF) associated with the subdiffusion process is governed by a time-fractional FPE, i.e., a FPE involving a fractional derivative in time, as given in (1.1). In the literature, there are also several works using fractional Laplacian to describe anomalously fast diffusion processes (i.e., superdiffusion), which leads to space fractional Fokker-Planck equations; see, e.g., [1,8,14] and references therein.

There have been several important studies on the model (1.1) from various different perspectives [5–7,10,20,27,32,35,38]. The physical modeling using time-fractional FPE has a long history; see [37] for in-depth detailed discussions. Barkai et al [7] derived the model (1.1) from the continuous time random walk model in order to describe anomalous diffusion in a time-independent external force field; see [5,20] for an extension to space- and time-dependent forcing. The well-posedness of the problem was discussed in [6], and the stochastic representation of the solutions was studied in [35,38]. Le et al [30] studied the numerical solution of the time-fractional FPE using the Galerkin finite element method. Camilli and De Maio [10] established the existence and uniqueness of a time-fractional mean field games system. Kemppainen and Zacher [27] investigated the long time behavior of a general class of nonlocal-in-time FPEs via an entropy argument, which is substantially different from that for the classical FPE. Li and Liu [32] described a discretization scheme for time-fractional gradient flow. However, none of these works has treated the gradient flow formulation for time-fractional FPE, which was recently pointed out by Kemppainen and Zacher [27] as "*an analogue of the celebrated theorem of Jordan, Kinderlehrer and Otto on the gradient flow structure of the classical FPE in the Wasserstein space* $\mathcal{P}_2(\mathbb{R}^d)$ *seems to be unknown for equation* (1.1) *and would be highly desirable.*"

The goal of this work is to discuss the time discretization of the model (1.1) via a JKO type scheme, thereby filling in an important missing piece on the time-fractional FPE pointed out by Kemppainen and Zacher [27]. This is carried out following the pioneering work of Jordan, Kinderlehrer and Otto [26] using the Wasserstein gradient flow for the classical FPE. Specifically, with a time step size τ, the scheme reads: given the initial datum ρ^0, find ρ^n, $n=1,2,\dots,N$ by minimizing

$$\frac{C_\alpha}{2\tau^\alpha} W_2^2(\rho,\overline{\rho}^{n-1}) + \mathcal{F}(\rho), \tag{1.2}$$

over the Wasserstein space $\mathcal{P}_2(\mathbb{R}^d)$, where $W_2(\cdot,\cdot)$ denotes the Wasserstein distance, $\overline{\rho}^{n-1}$ is a convex combination of of $\rho^0,\dots,\rho^{n-1}$ (with weights depending on the numerical approximation of the fractional derivative $\partial_t^\alpha \rho$), $C_\alpha = 1/\Gamma(2-\alpha)$ is a fixed constant and $\mathcal{F}(\rho)$ is the free energy; See Section 4 for details. The term $\overline{\rho}^{k-1}$ captures the nonlocal nature / memory effect of the mathematical model. The scheme recovers the classical JKO scheme [26] as $\alpha \to 1^-$, and thus it represents a fractional analogue of the latter. Numerically, it has comparable computational complexity as the classical JKO scheme, except the extra computation of the convex combination $\overline{\rho}^{n-1}$. The main result is given in Theorem 4.2, which shows that the piecewise constant interpolation converges weakly in $L^1((0,T)\times\mathbb{R}^d)$ to a weak solution of the model (1.1). Further, we numerically illustrate the performance of the approach, using recently developed powerful solvers for minimization problems involving Wassserstein distance based on entropy regularization [13,39,40].

The main technical challenge of the fractional extension (1.2) of the classical JKO scheme is to deal with the nonlocality of the fractional derivative $\partial_t^\alpha \rho$. Numerically, this is overcome by adopting one extremely popular fractional analogue of the backward Euler scheme (used in the JKO scheme) from the numerical analysis community, known

as the L1 scheme [34] for discretizing the Caputo derivative $\partial_t^\alpha \rho$, and its weights enter into the term $\overline{\rho}^{n-1}$. Naturally, the nonlocality of the term $\overline{\rho}^{n-1}$ also requires substantial adaptation of known techniques [26] for the convergence analysis. The gradient flow formulation and its convergence analysis represent the main contributions of this work.

The rest of the paper is organized as follows. In Section 2, we recall preliminaries on fractional calculus and describe the connection of the model (1.1) with stochastic process and related results on existence and uniqueness. Then in Section 3, we describe the L1 scheme, which is an extension of the backward Euler method to the fractional case, and derive relevant approximation properties, which are needed for constructing the scheme (1.2) and its convergence analysis. In Section 4, we describe the time-fractional JKO scheme, and state the main theorem, whose lengthy and technical proof is given in Section 5. Last, in Section 6, we present numerical results for one- and two-dimensional problems to illustrate features of the proposed JKO scheme. Below, C denotes a generic constant that depends on the parameters of the problem, on the initial datum ρ_0, and may change at each occurrence, but it is always independent of the time level n and of time step size τ.

2. Preliminaries

In this section we briefly recall preliminaries on fractional calculus, stochastic model for fractional FPEs and the concept of weak solution for problem (1.1).

2.1. Preliminaries on fractional calculus First, we recall basic concepts from fractional calculus [28]. Throughout, we always assume $\gamma \in [0,1)$, and $a<b$. Then for a function $f:(a,b)\to\mathbb{R}$, the left-sided and right-sided Riemann-Liouville fractional integrals of order γ, denoted by ${}_aI_t^\gamma f$ and ${}_tI_b^\gamma f$, are respectively defined by

$$ {}_aI_t^\gamma f(t) = \frac{1}{\Gamma(\gamma)}\int_a^t (t-s)^{\gamma-1} f(s)\,\mathrm{d}s \quad \text{and} \quad {}_tI_b^\gamma f(t) = \frac{1}{\Gamma(\gamma)}\int_t^b (s-t)^{\gamma-1} f(s)\,\mathrm{d}s. $$

These integral operators are well defined for $f\in L^1(a,b)$ and are bounded on $L^p(a,b)$ for any $p\geq 1$. The integral operators ${}_aI_t^\gamma$ and ${}_tI_b^\gamma$ are adjoint to each other with respect to $L^2(a,b)$:

$$ \int_a^b ({}_aI_t^\gamma f)(t) g(t)\,\mathrm{d}t = \int_a^b f(t)({}_tI_b^\gamma g)(t)\,\mathrm{d}t. \tag{2.1} $$

This relation can be verified directly by changing the order of integration.

The left-sided and right-sided Caputo derivative of order $\alpha\in(0,1)$ of a function $f:(a,b)\to\mathbb{R}$, denoted by ${}_aD_t^\alpha f$ and ${}_tD_b^\alpha f$, are respectively defined by

$$ {}_aD_t^\alpha f(t) = ({}_aI_t^{1-\alpha} f')(t) \quad \text{and} \quad {}_tD_b^\alpha f(t) = -({}_tI_b^{1-\alpha} f')(t). $$

Note that the definition of the Caputo derivative of order α requires the existence of a first-order derivative. Hence, the definition is more stringent. There have been several important efforts in relaxing the regularity requirement [18,31]. It can be verified that as $\alpha\to 1^-$, ${}_aD_t^\alpha f$ recovers the usual first-order derivative $\partial_t f$, when f is sufficiently smooth. Due to the nonlocality of the fractional derivatives, many useful rules in calculus are no longer available. The following integration by parts formula is useful.

LEMMA 2.1. *The following identity holds for $f,g\in C^1[a,b]$ with $g(b)=0$:*

$$ \int_a^b ({}_aD_t^\alpha f)(t) g(t)\,\mathrm{d}t = \int_a^b f(t)({}_tD_b^\alpha g)(t)\,\mathrm{d}t - \frac{f(a)}{\Gamma(1-\alpha)}\int_a^b (t-a)^{-\alpha} g(t)\,\mathrm{d}t. \tag{2.2} $$

Proof. Indeed, there holds

$$\int_a^b ({}_aD_t^\alpha f)(t)g(t)\,\mathrm{d}t = \int_a^b ({}_aI_t^{1-\alpha} f')(t)g(t)\,\mathrm{d}t = \int_a^b f'(t)({}_tI_b^{1-\alpha} g)(t)\,\mathrm{d}t$$
$$= -\int_a^b f(t)({}_tI_b^{1-\alpha} g)'(t)\,\mathrm{d}t + \big[f(t)({}_tI_b^{1-\alpha} g)(t)\big]_a^b.$$

where the first identity follows from the definition of the Caputo derivative ${}_aD_t^\alpha f$, the second identity follows from (2.1), and the third identity is obtained by integration by parts. Since $g(b)=0$, by the definition of the right-sided Caputo derivative, $-({}_tI_b^{1-\alpha} g)'(t) = {}_tD_b^\alpha g$ [28, (2.4.10), p. 91]. Then the desired assertion follows by

$$\big[f(t)({}_tI_b^{1-\alpha} g)(t)\big]_a^b = f(b)({}_tI_b^{1-\alpha} g)(b) - f(a)({}_tI_b^{1-\alpha} g)(a)$$
$$= -\frac{f(a)}{\Gamma(1-\alpha)}\int_a^b (s-a)^{-\alpha} g(s)\,\mathrm{d}s,$$

since $({}_tI_b^{1-\alpha} g)(b)=0$. This completes the proof of the lemma. □

Below we shall write $\partial_t^\alpha f$ and $D_t^\alpha f$ for ${}_0D_T^\alpha f$ and ${}_tD_T^\alpha f$, respectively, for notational simplicity.

2.2. From stochastic processes to time-fractional FPE It is well-known that the classical FPE

$$\partial_t f = \mathrm{div}(\nabla\Psi f) + \Delta f, \tag{2.3}$$

which corresponds to problem (1.1) with $\alpha=1$, is the Kolmogorov forward equation of the following stochastic differential equation (SDE):

$$\mathrm{d}X(t) = -\nabla\Psi(X(t))\,\mathrm{d}t + \sqrt{2}\mathrm{d}W(t), \quad \text{with } X(0)=X_0, \tag{2.4}$$

where $W(t)$ is a standard d-dimensional Wiener process and X_0 is a d-dimensional random vector distributed according to the density ρ_0. The SDE (2.4) describes the motion of a particle undergoing diffusion in an external field Ψ, where $X(t)$ is the position of the particle at time t, and the FPE (2.3) describes the time evolution of the PDF of the particle. Its solution $f(t,x)$ is the PDF of finding the particle at time t and at position x. The time-fractional FPE (1.1) can be viewed as the Kolmogorov forward equation of a stochastic process which is obtained from (2.4) under a time-changed process. Specifically, let $U_\alpha(t)$ be the α-stable subordinator with its Laplace transform given by $\mathbb{E}\big[e^{-kU_\alpha(\tau)}\big] = e^{-\tau k^\alpha}$, $0<\alpha<1$, and let $S_\alpha(t)$ be the inverse α-stable subordinator

$$S_\alpha(t) = \inf\{\tau>0\colon\ U_\alpha(\tau)>t\}.$$

Define the time-changed process

$$Y(t) = X(S_\alpha(t)).$$

Then the probability density function (PDF) $p(x,t)$ of $Y(t)$ satisfies the time-fractional FPE (1.1). In fact, the following theorem [19,35,36] describes a close connection between the solutions of (1.1) and (2.3).

THEOREM 2.1. *Let $f(x,\tau)$ and $g(\tau,t)$ be respectively the PDFs of $X(\tau)$ and $S(t)$. The following assertions hold.*

(i) *The PDF* $p(x,t)$ *of* $Y(t)$ *is given by* $p(x,t)=\int_0^\infty f(x,\tau)g(\tau,t)\,\mathrm{d}\tau$.
(ii) *The Laplace transform of* p *and* f, *denoted by* $\hat{p}$ *and* $\hat{f}$, *respectively, satisfy* $\hat{p}(x,k)=k^{\alpha-1}\hat{f}(x,k^\alpha)$.
(iii) $p(x,t)$ *is a weak solution to the time-fractional FPE* (1.1) *in the sense of Definition* (2.1) *below.*

See the works [19, 35, 36] for further details on the stochastic representation of problem (1.1).

REMARK 2.1. *There are alternative equivalent reformulations of problem* (1.1). *One popular alternative reads*

$$\partial_t\rho={}^R\partial_t^{1-\alpha}(\nabla\cdot(\rho\nabla\Psi)+\Delta\rho), \tag{2.5}$$

where the ${}^R\partial_t^{1-\alpha}\varphi$ *denotes Riemann-Liouville fractional derivative of order* $1-\alpha$, *i.e.,* ${}^R\partial_t^{1-\alpha}\varphi(t)=\frac{\mathrm{d}}{\mathrm{d}t}({}_0I_t^\alpha\varphi)(t)$. *Formally, it can be obtained from* (1.1) *by applying* ${}^R\partial_t^{1-\alpha}$ *to both sides of* (1.1) *as*

$${}^R\partial_t^{1-\alpha}\partial_t^\alpha\varphi(t)=\frac{\mathrm{d}}{\mathrm{d}t}{}_0I_t^\alpha\,{}_0I_t^{1-\alpha}\varphi'(t)=\frac{\mathrm{d}}{\mathrm{d}t}{}_0I_t\varphi'(t)=\varphi'(t),$$

where the first identity is due to the definitions of the fractional derivatives and the second identity is due to the semigroup property of Riemann-Liouville fractional integral. Further, one may change the order the spatial and temporal derivative when the forcing Ψ *is time-independent. We refer to the work [20] for discussions on the proper formulation for a time-dependent forcing. In the present work, we focus on the formulation* (1.1), *and leave the study of other time-fractional FPE models to future works.*

2.3. Well-posedness Throughout, we only consider probability measures on $\mathbb{R}^d$ that are absolutely continuous with respect to Lebesgue measure, and often identify a probability measure with its density, as the classical setting [26]. We denote by $\mathcal{P}_2(\mathbb{R}^d)$ the set of all probability measures on $\mathbb{R}^d$ with a finite second moment, i.e.,

$$\mathcal{P}_2(\mathbb{R}^d):=\Big\{\rho:\mathbb{R}^d\to[0,\infty)\ \text{measurable},\ \int_{\mathbb{R}^d}\rho(x)\,\mathrm{d}x=1,\ M_2(\rho)<\infty\Big\},$$

where the second moment $M_2(\rho)$ is defined by

$$M_2(\rho)=\int_{\mathbb{R}^d}|x|^2\rho(x)\,\mathrm{d}x. \tag{2.6}$$

Now, we introduce a notion of weak solutions to problem (1.1). Similar to the classical setting, we multiply equation (1.1) by a smooth test function and using the integration by parts formula (2.2) in Lemma 2.1, which leads to the following notion of weak solution. Below we shall write a function $f(t,x)$ as a vector valued function $f(t)$.

DEFINITION 2.1. *A function* $\rho\in L^1(\mathbb{R}^+\times\mathbb{R}^d)$ *is called a weak solution of problem* (1.1) *with initial datum* $\rho_0\in\mathcal{P}_2(\mathbb{R}^d)$ *if it satisfies that for any* $\varphi\in C^\infty([0,T]\times\mathbb{R}^d)$ *with* $\varphi(T)=0$, *there holds*

$$\int_0^T\int_{\mathbb{R}^d}\Big({}_tD_T^\alpha\varphi(t)+\nabla\Psi\cdot\nabla\varphi(t)-\Delta\varphi(t)\Big)\rho(t)\,\mathrm{d}x\,\mathrm{d}t \tag{2.7}$$
$$=\frac{1}{\Gamma(1-\alpha)}\int_{\mathbb{R}^d}\int_0^T t^{-\alpha}\varphi(t)\,\mathrm{d}t\rho_0\,\mathrm{d}x.$$

Note that the formulation (2.7) of the weak solution involves a nonlocal term $\int_{\mathbb{R}^d} {}_t I_T^{1-\alpha}\varphi(0)\rho_0\,\mathrm{d}x$. This term appears due to the nonlocality of the Caputo derivative $\partial_t^\alpha \rho$, cf. Lemma 2.1. In the limit $\alpha \to 1^-$, it recovers the usual $\int_{\mathbb{R}^d}\varphi(0)\rho_0\,\mathrm{d}x$, in view of the identity $\lim_{\alpha\to 0^+} {}_0 I_t^{\alpha-1}\varphi(0) = \varphi(0)$, under suitable regularity assumptions. We are not are aware of any existing work directly investigating the existence and regularity of the solutions on problem (1.1). However, the existence and uniqueness of the weak solution of an equivalent formulation given in (2.5) of problem (1.1) were already proven in [10, Theorem 3.3]. See also [3, 42] for discussion on the well-posedness (existence and uniqueness) of abstract Volterra type evolution equations in a Hilbert space setting. It is also worth noting that the proper interpretation of the initial condition requires some care; see the works [18, 31] for in-depth discussions. We leave a detailed study on these important analytic issues (possibly in more general settings of metric spaces and spaces of probability measures) to future works.

3. Numerical approximation of Caputo derivative

Now we recall the numerical approximation of the Caputo derivative $\partial_t^\alpha \varphi(t)$. There are several different ways to construct a "fractional" analogue of the classical backward Euler method (see [23] for a concise overview), on which the classical JKO scheme [26] is based. We shall employ the so-called piecewise linear approximation, commonly known as the L1 approximation (due to Lin and Xu [34]) in the numerical analysis literature.

Consider a uniform partition of the time interval $[0,T]$, with a time step size $\tau = \frac{T}{N}$ and the grid $t_n = n\tau$, $n = 0,1,\ldots,N$. For any function $\varphi \in C[0,T]$, we use the shorthand notation $\varphi^n = \varphi(t_n)$. Further, we denote $C_\alpha = \Gamma(2-\alpha)^{-1}$. Then the L1 approximation [34] is constructed as follows. First we split the interval $[0,t_n]$ into n subintervals

$$\partial_t^\alpha \varphi^n = \frac{1}{\Gamma(1-\alpha)} \sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} (t_n - s)^{-\alpha} \varphi'(s)\,\mathrm{d}s,$$

and then by approximating φ by its linear interpolation over the subinterval $[t_{i-1}, t_i]$, i.e.,

$$\varphi(t) \approx \frac{t_i - t}{\tau}\varphi^{i-1} + \frac{t - t_{i-1}}{\tau}\varphi^i, \quad t \in [t_{i-1}, t_i], i = 1,\ldots,N,$$

or equivalently $\varphi'(t) \approx (\varphi^i - \varphi^{i-1})/\tau$ for $t \in [t_{i-1}, t_i]$, we obtain the following approximation to the Caputo derivative $\partial_t^\alpha \varphi$ at time $t = t_n$ by

$$\partial_t^\alpha \varphi^n = \frac{1}{\Gamma(1-\alpha)} \sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} (t_n - s)^{-\alpha} \frac{\varphi^i - \varphi^{i-1}}{\tau}\,\mathrm{d}s + r_\tau^n,$$

where r_τ^n is the local truncation error. It can be verified that r_τ^n takes the following form [34]

$$r_\tau^n \le c_\varphi \left[\frac{1}{\Gamma(1-\alpha)} \sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} \frac{t_i + t_{i-1} - 2s}{(t_n - s)^\alpha}\,\mathrm{d}s + O(\tau^2) \right],$$

with the constant c_φ depending only on $\|\varphi\|_{C^2[0,T]}$. Now the elementary integral

$$\int_{t_{i-1}}^{t_i} (t_n - s)^{-\alpha}\,\mathrm{d}s = (1-\alpha)^{-1}\tau^{1-\alpha}((n+1-i)^{1-\alpha} - (n-i)^{1-\alpha})$$

and simple algebraic manipulations (with $C_\alpha = 1/\Gamma(2-\alpha)$) lead to

$$\begin{aligned}\partial_t^\alpha \varphi^n &\approx C_\alpha \tau^{-\alpha} \sum_{i=1}^{n} (\varphi^i - \varphi^{i-1})((n+1-i)^{1-\alpha} - (n-i)^{1-\alpha}) \\ &= C_\alpha \tau^{-\alpha} \sum_{i=0}^{n} b_{n-i}^{(n)} \varphi^i := \bar\partial_\tau^\alpha \varphi^n, \end{aligned} \tag{3.1}$$

where the quadrature weights $b_i^{(n)}$ are given by

$$b_i^{(n)} = \begin{cases} 1, & i = 0, \\ (i+1)^{1-\alpha} + (i-1)^{1-\alpha} - 2i^{1-\alpha}, & i = 1, \ldots, n-1, \\ (n-1)^{1-\alpha} - n^{1-\alpha}, & i = n. \end{cases} \tag{3.2}$$

Note that the last weight $b_n^{(n)}$ depends on n differently than the preceding ones. In the special case $\alpha = 1$, the approximation reduces to the classical backward Euler method, since $b_0^{(n)} = 1$ and $b_1^{(n)} = -1$, and $b_i^{(n)} = 0$, for any $1 < i \le n$. In a similar manner, the L1 approximation $\overline{D}_\tau^\alpha \varphi^n$ to the right-sided Caputo fractional derivative ${}_tD_T^\alpha \varphi(t)$ at $t = t_n$ is given by

$$\overline{D}_\tau^\alpha \varphi^n = C_\alpha \tau^{-\alpha} \sum_{j=0}^{N-n} b_j^{(N-n)} \varphi^{n+j} = C_\alpha \tau^{-\alpha} \sum_{j=n}^{N} b_{j-n}^{(N-n)} \varphi^j. \tag{3.3}$$

This approximation can be obtained by a simple change of variables.

By construction, the L1 approximations $\bar\partial_\tau^\alpha \varphi^n$ and $\overline{D}_\tau^\alpha \varphi^n$ are essentially a weighted piecewise linear approximation, with respect to the weakly singular weight $t^{-\alpha}$. The discrete approximations are of convolution form, similar to the continuous fractional derivatives $\partial_t^\alpha \varphi$ and $D_t^\alpha \varphi$. The L1 approximation has been widely employed for solving time-fractional diffusion, due to its excellent empirical performance; see [22, 24, 25] for some relevant works on error analysis.

We will need the following auxiliary lemma.

LEMMA 3.1. *For $0 < \alpha < 1$ and a fixed $n \in \mathbb{N}$, for the weights $b_j^{(n)}$ given in (3.2), then there holds $b_i^{(n)} < 0$ for $i = 1, \ldots, n$ and $\sum_{i=0}^n b_i^{(n)} = 0$. Further,*

$$\sum_{n=1}^{k} (-b_n^{(n)}) = k^{1-\alpha} \quad \text{and} \quad \sum_{j=1}^{k-i} (-b_j^{(j+i)}) = 1 + (k-i)^{1-\alpha} - (k-i+1)^{1-\alpha}.$$

Proof. The first assertion is well known (see, e.g., [34, eq. (3.7)]), and we only give a proof for completeness. Consider the function $f(x) = x^{1-\alpha}$ for $x > 0$. Since $0 < \alpha < 1$, we have $f''(x) = -\alpha(1-\alpha)x^{-\alpha-1} < 0$, and hence f is strictly concave on $(0, \infty)$. By Jensen's inequality we have

$$\begin{aligned} i^{1-\alpha} = f(i) &= f\left(\tfrac{i+1+i-1}{2}\right) \\ &> \tfrac{1}{2} f(i+1) + \tfrac{1}{2} f(i-1) \\ &= \tfrac{1}{2}(i+1)^{1-\alpha} + \tfrac{1}{2}(i-1)^{1-\alpha}, \end{aligned}$$

which immediately implies that $b_i^{(n)} < 0$ for all $i = 1, \ldots, n$. Further, straight computations give

$$\sum_{i=0}^{n} b_i^{(n)} = 1 + \sum_{i=1}^{n-1} \left((i+1)^{1-\alpha} + (i-1)^{1-\alpha} - 2i^{1-\alpha}\right) + \left((n-1)^{1-\alpha} - n^{1-\alpha}\right) = 0.$$

This shows the second assertion. The rest follows from straightforward computation as:

$$\begin{aligned}\sum_{n=1}^{k}(-b_n^{(n)}) &= \sum_{n=1}^{k}(n^{1-\alpha}-(n-1)^{1-\alpha}) = k^{1-\alpha},\\ \sum_{j=1}^{k-i}(-b_j^{(j+i)}) &= -\sum_{j=1}^{k-i}\big((j+1)^{1-\alpha}+(j-1)^{1-\alpha}-2j^{1-\alpha}\big)\\ &= -\sum_{j=1}^{k-i}\big(((j+1)^{1-\alpha}-j^{1-\alpha})-(j^{1-\alpha}-(j-1)^{1-\alpha})\big)\\ &= 1+(k-i)^{1-\alpha}-(k-i+1)^{1-\alpha}.\end{aligned}$$

This completes the proof of the lemma. □

We will also need the following useful inequality of Gronwall type [33, Lemma 2.2].

LEMMA 3.2. *Suppose* $\{\phi^n\}_{n=0}^N$ *are nonnegative, and satisfy the following inequality (with* $\tau=T/N$*)*

$$\bar\partial_\tau^\alpha \phi^n \le C_1 + C_2\phi^n,$$

where C_1, C_2 *are positive constants. Then there holds*

$$\phi^n \le 2E_\alpha(2C_2t_n^\alpha)\Big(\phi^0+\frac{C_1}{\Gamma(1+\alpha)}t_n^\alpha\Big),\quad \forall n=1,\dots,N$$

where E_α *denotes the Mittag-Leffler function* $E_\alpha(z)=\sum_{k=0}^\infty \frac{z^k}{\Gamma(k\alpha+1)}$.

The next result gives a "semi-discrete" version of the integration by parts formula in Lemma 2.1 for the L1 approximation $\bar\partial_\tau^\alpha\varphi^n$.

LEMMA 3.3. *Let* $\{\varphi^n\}_{n=0}^N$ *be a given sequence, and* $\phi(t)\in C^1[0,T]$ *with* $\phi(T)=0$. *The piecewise constant approximation* $\varphi_\tau(t)$ *is defined by* $\varphi_\tau(t)=\phi^n$ *for* $(n-1)\tau<t\le n\tau$, *with* $\phi_\tau(0)=\phi^0$. *Then the following identity holds*

$$\int_0^T(\bar\partial_\tau^\alpha\varphi^n)(t)\phi(t)\,\mathrm{d}t = \int_0^T\varphi_\tau(t)\overline{D}_\tau^\alpha\phi(t)\,\mathrm{d}t + C_\alpha\tau^{-\alpha}\varphi^0\sum_{n=1}^N b_n^{(n)}\int_{t_{n-1}}^{t_n}\phi(t)\,\mathrm{d}t,$$

where the function $\overline{D}_\tau^\alpha\phi(t)$ *is defined by (with zero extension on* ϕ*)*

$$\overline{D}_\tau^\alpha\phi(t)=\sum_{i=n}^N b_{i-n}^{(N-n)}\phi(t+(i-n)\tau),\quad \forall t\in(t_{n-1},t_n],\quad n=1,\dots,N.$$

Proof. By the definition of the L1 approximation in (3.1), we have

$$\begin{aligned}C_\alpha^{-1}\tau^\alpha\mathrm{LHS} &= \sum_{n=1}^N\int_{t_{n-1}}^{t_n}\Big[\varphi^n+\sum_{i=0}^{n-1}b_{n-i}^{(n)}\varphi^i\Big]\phi(t)\,\mathrm{d}t\\ &= \sum_{n=1}^N\int_{t_{n-1}}^{t_n}\varphi^n\phi(t)\,\mathrm{d}t+\sum_{n=1}^N\sum_{i=1}^{n-1}b_{n-i}^{(n)}\int_{t_{n-1}}^{t_n}\varphi^i\phi(t)\,\mathrm{d}t+\sum_{n=1}^N b_n^{(n)}\int_{t_{n-1}}^{t_n}\varphi^0\phi(t)\,\mathrm{d}t\\ &=: \mathrm{I}+\mathrm{II}+\mathrm{III}.\end{aligned}$$

By the definition of the interpolation $\varphi_\tau(t)$, the first term I can be rewritten as

$$\mathrm{I}=\sum_{n=1}^{N}\int_{t_{n-1}}^{t_n}\varphi_\tau(t)\phi(t)\,\mathrm{d}t.$$

Now we turn to the term II. Using the change of variables $t\mapsto t+(n-i)\tau$ and then applying the definition of the interpolation $\varphi_\tau(t)$, we deduce

$$\begin{aligned}\mathrm{II}&=\sum_{n=1}^{N}\sum_{i=1}^{n-1}b_{n-i}^{(n)}\int_{t_{i-1}}^{t_i}\varphi^i\phi(t+(n-i)\tau)\,\mathrm{d}t\\&=\sum_{n=1}^{N}\sum_{i=1}^{n-1}b_{n-i}^{(n)}\int_{t_{i-1}}^{t_i}\varphi_\tau(t)\phi(t+(n-i)\tau)\,\mathrm{d}t.\end{aligned}$$

Next we interchange the order of summation and relabel the indices (with the convention that the sum is zero when the lower index is greater than the upper index) to obtain

$$\begin{aligned}\mathrm{II}&=\sum_{i=1}^{N-1}\sum_{n=i+1}^{N}b_{n-i}^{(n)}\int_{t_{i-1}}^{t_i}\varphi_\tau(t)\phi(t+(n-i)\tau)\,\mathrm{d}t\\&=\sum_{n=1}^{N-1}\sum_{i=n+1}^{N}b_{i-n}^{(i)}\int_{t_{n-1}}^{t_n}\varphi_\tau(t)\phi(t+(i-n)\tau)\,\mathrm{d}t\\&=\sum_{n=1}^{N}\sum_{i=n+1}^{N}b_{i-n}^{(i)}\int_{t_{n-1}}^{t_n}\varphi_\tau(t)\phi(t+(i-n)\tau)\,\mathrm{d}t.\end{aligned}$$

Now recall the definition of the weights $b_{i-n}^{(i)}$ in (3.2), there holds

$$b_{i-n}^{(i)}=b_{i-n}^{(N-n)},\quad i=n+1,\ldots,N-1.$$

Further, since ϕ is supported on $(0,T)$, we may change $b_{N-n}^{(N)}$ to $b_{N-n}^{(N-n)}$, and thus obtain

$$\mathrm{II}=\sum_{n=1}^{N}\sum_{i=n+1}^{N}b_{i-n}^{(N-n)}\int_{t_{n-1}}^{t_n}\varphi_\tau(t)\phi(t+(i-n)\tau)\,\mathrm{d}t.$$

Consequently, since $b_0^{N-n}=1$ and using the definition of the notation $\overline{D}_\tau^\alpha\phi(t)$,

$$\begin{aligned}\mathrm{I}+\mathrm{II}&=\sum_{n=1}^{N}\int_{t_{n-1}}^{t_n}\varphi_\tau\Big(\phi(t)+\sum_{i=n+1}^{N}b_{i-n}^{(N-n)}\phi(t+(i-n)\tau)\Big)\,\mathrm{d}t\\&=\sum_{n=1}^{N}\int_{t_{n-1}}^{t_n}\varphi_\tau\overline{D}_\tau^\alpha\phi(t)\,\mathrm{d}t=\int_0^T\varphi_\tau\overline{D}_\tau^\alpha\phi(t)\,\mathrm{d}t.\end{aligned}$$

Then combining the preceding identities completes the proof of the lemma. □

The following result gives the error estimates of the L1 approximation for smooth functions.

THEOREM 3.1. *The following error estimates hold*

$$\bar\partial_\tau^\alpha\varphi^n=(\partial_t^\alpha\varphi)(t_n)+O(\tau^{2-\alpha}),\quad\forall\varphi\in C^2[0,T]$$

$$C_\alpha \tau^{-\alpha} \sum_{n=1}^{N} b_n^{(n)} \int_{t_{n-1}}^{t_n} \varphi(t)\,\mathrm{d}t = -({}_tI_T^\alpha \varphi)(0) + O(\tau), \quad \forall \varphi \in C^1[0,T].$$

Proof. The first estimate can be found at [34, equations (3.12) and (3.13)]. It suffices to show the second estimate. Using the expression of the weight $b_n^{(n)}$, we may rewrite the left hand side as (with $C'_\alpha = -C_\alpha(1-\alpha) = -\frac{1}{\Gamma(1-\alpha)}$

$$\begin{aligned}
\mathrm{LHS} &= C'_\alpha \sum_{n=1}^{N} \int_{t_{n-1}}^{t_n} t^{-\alpha}\,\mathrm{d}t \Big(\tau^{-1} \int_{t_{n-1}}^{t_n} \varphi(s)\,\mathrm{d}s\Big) \\
&= C'_\alpha \sum_{n=1}^{N} \int_{t_{n-1}}^{t_n} t^{-\alpha}\varphi(t)\,\mathrm{d}t - C'_\alpha \sum_{n=1}^{N} \int_{t_{n-1}}^{t_n} t^{-\alpha} \Big(\varphi(t) - \tau^{-1} \int_{t_{n-1}}^{t_n} \varphi(s)\,\mathrm{d}s\Big)\,\mathrm{d}t \\
&= C'_\alpha \int_0^T t^{-\alpha}\varphi(t)\,\mathrm{d}t - C'_\alpha \sum_{n=1}^{N} \int_{t_{n-1}}^{t_n} t^{-\alpha} \Big(\varphi(t) - \tau^{-1} \int_{t_{n-1}}^{t_n} \varphi(s)\,\mathrm{d}s\Big)\,\mathrm{d}t.
\end{aligned}$$

Next we bound the term in the bracket by

$$|\varphi(t) - \tau^{-1} \int_{t_{n-1}}^{t_n} \varphi(s)\,\mathrm{d}s| = |\tau^{-1} \int_{t_{n-1}}^{t_n} \varphi(t) - \varphi(s)\,\mathrm{d}s| \le \|\varphi\|_{C^1[0,T]}\tau.$$

Combining the last two estimates gives the desired estimate. ∎

4. Time-fractional JKO scheme

Now we construct a JKO type scheme for problem (1.1), and give the main result of the work.

4.1. Wasserstein distance The Wasserstein distance of order two, denoted by $W_2(\mu_1,\mu_2)$, between two (Borel) probability measures μ_1 and μ_2 on $\mathbb{R}^d$ is defined by

$$W_2(\mu_1,\mu_2)^2 = \inf_{p\in\mathcal{P}(\mu_1,\mu_2)} \int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^2 p(\mathrm{d}x\,\mathrm{d}y), \tag{4.1}$$

where $\mathcal{P}(\mu_1,\mu_2)$ is the set of all probability measures on $\mathbb{R}^d\times\mathbb{R}^d$ with the first marginal μ_1 and second marginal μ_2, and the symbol $|\cdot|$ denotes the usual Euclidean norm on $\mathbb{R}^d$. That is, a probability measure p is in $\mathcal{P}(\mu_1,\mu_2)$ if and only if for each Borel subset $A\subset\mathbb{R}^d$ there holds

$$p(A\times\mathbb{R}^d) = \mu_1(A) \quad \text{and} \quad p(\mathbb{R}^d\times A) = \mu_2(A).$$

It is well known that $W_2(\mu_1,\mu_2)$ defines a metric on the set of probability measure μ on $\mathbb{R}^d$ having finite second moments: $\int_{\mathbb{R}^d}|x|^2\mu(\mathrm{d}x) < \infty$ [17].

The variational problem (4.1) is an example of a Monge-Kantorovich mass transport problem with a cost function $c(x,y) = |x-y|^2$. In that context, an infimizer p^* is referred to as an optimal (transport) plan; see [17] for a probabilistic proof that the infimum in (4.1) exists when the measures μ_1 and μ_2 have finite second moments. Brenier [9] established the existence of a one-to-one optimal (transport) plan in the case that the measures μ_1 and μ_2 have bounded support and are absolutely continuous with respect to Lebesgue measure.

4.2. Time-fractional JKO scheme

Next we derive the fractional analogue of the JKO scheme for problem (1.1). The classical JKO scheme [26] for the FPE (2.3) is based on the backward Euler approximation of the first-order derivative $\partial_t \rho$ in time. Hence, naturally, the fractional analogue relies on a backward Euler type approximation to the Caputo derivative $\partial_t^\alpha \rho$. We shall employ the L1 approximation [34] described in Section 3. By combining the classical JKO scheme [26] and the L1 approximation of the fractional time derivative $\partial_t^\alpha \rho$, we obtain a JKO type scheme for problem (1.1) as follows.

SCHEME 4.1 (Discrete variational approximation scheme for the model (1.1)). *Let $\rho^0 := \rho_0$. Given ρ^0, find ρ^n, $n=1,2,\ldots,N$, as the unique minimizer of*

$$\frac{C_\alpha}{2\tau^\alpha} W_2^2(\rho, \overline{\rho}^{n-1}) + \mathcal{F}(\rho), \tag{4.2}$$

over $\rho \in \mathcal{P}_2(\mathbb{R}^d)$, where $\overline{\rho}^{n-1}$ and $\mathcal{F}(\rho)$ are defined respectively by

$$\overline{\rho}^{n-1} := \sum_{i=0}^{n-1} (-b_{n-i}^{(n)}) \rho^i \quad \text{and} \quad \mathcal{F}(\rho) := \mathcal{E}(\rho) + \mathcal{S}(\rho),$$

with

$$\mathcal{E}(\rho) = \int_{\mathbb{R}^d} \Psi \rho \,\mathrm{d}x \quad \text{and} \quad \mathcal{S}(\rho) = \int_{\mathbb{R}^d} \rho \log \rho \,\mathrm{d}x.$$

Given $\rho^0 \in \mathcal{P}_2(\mathbb{R}^d)$, the existence and uniqueness of a minimizer of Scheme 4.1 was proven in [26, Proposition 4.1]. In view of Lemma 3.1(i), $\sum_{i=0}^{n-1} (-b_{n-i}^{(n)}) = 1$, and thus $\overline{\rho}^{n-1}$ is a convex combination of all past approximations $\{\rho^i\}_{i=0}^{n-1}$. This property plays a crucial role in the convergence analysis. One distinct feature of the scheme is that instead of using only the immediate previous density ρ^{n-1} in (4.2) as in the classical JKO-scheme, it employs a *convex* combination $\overline{\rho}^{n-1}$ of all previous densities $\{\rho^i\}_{i=0}^{n-1}$. This is to capture the memory effect (and thus non-Markovian nature) of the continuous time-fractional FPE. In the limiting case $\alpha = 1$, it is identical with the classical JKO scheme (see the properties of the L1 approximation in Section 3).

Below we shall make one minor assumption on Ψ. Note that the assumption $\Psi(x) \geq 0$ can be relaxed to that Ψ is bounded from below.

ASSUMPTION 4.1. *$\Psi(x) \in C^\infty(\mathbb{R}^d)$, $\Psi(x) \geq 0$ and $|\nabla \Psi(x)| \leq C(|x|+1)$ for all $x \in \mathbb{R}^d$.*

REMARK 4.1. *There are other possible formulations of JKO type schemes for the time-fractional FPE (1.1). For example, the following formulation seems also feasible. Given ρ^0, find ρ^n, $n=1,2,\ldots,N$, by minimizing over $\mathcal{P}_2(\mathbb{R}^d)$ the following functional*

$$\frac{C_\alpha}{2\tau^\alpha} \sum_{i=0}^{n-1} (-b_{n-i}^{(n)}) W_2^2(\rho, \rho^i) + \mathcal{F}(\rho).$$

By the convexity of the Wasserstein distance, this functional is an upper bound of the one in (4.2). However, it involves multiple Wasserstein distances and thus is computationally far less convenient. Thus it is not explored in this work.

The next result represents the main theoretical contribution of the paper, i.e., the convergence of the discrete approximations $\{\rho^n\}_{n=1}^N$. The proof of the theorem is lengthy and technical and will be given in Section 5.

THEOREM 4.2. *Let $\rho_0 \in \mathcal{P}_2(\mathbb{R}^d)$ satisfy $\mathcal{F}(\rho_0) < \infty$. For any fixed $\tau > 0$, let $\{\rho^n\}_{n=1}^N$ be the sequence of minimizers given by Scheme 4.1. For any $t \geq 0$, we define a picewise-constant time interpolation: with $\rho_\tau(0) = \rho^0$ and*

$$\rho_\tau(t,x) = \rho^n(x) \qquad \text{for} \quad (n-1)\tau < t \leq n\tau, \quad n = 1,\dots,N. \tag{4.3}$$

Then under Assumption 4.1, for any $T > 0$,

$$\rho_\tau \to \rho \quad \text{weakly in} \quad L^1\big((0,T)\times\mathbb{R}^d\big) \quad \text{as} \quad \tau \to 0, \tag{4.4}$$

where ρ is the unique weak solution to problem (1.1) in the sense of Definition 2.1.

5. Proof of Theorem 4.2

This section is devoted to the convergence analysis of Scheme 4.1, i.e., the proof of Theorem 4.2. First, we give the Euler-Lagrange equation for the sequence of minimizers.

LEMMA 5.1. *Let $\tau > 0$ and $\{\rho^n\}_{n=1}^N$ be the sequence of minimizer given by Scheme 4.1, and P^n the optimal plan for the Wasserstein distance $W_2(\rho^n, \overline{\rho}^{n-1})$ between $\overline{\rho}^{n-1}$ and ρ^n. Then for all $\varphi \in C_0^\infty(\mathbb{R}^d)$, there holds*

$$\frac{C_\alpha}{\tau^\alpha}\int_{\mathbb{R}^{2d}} (y-x)\cdot\nabla\varphi(y) P^n(\mathrm{d}x\,\mathrm{d}y) + \int_{\mathbb{R}^d}\Big(\nabla\Psi\cdot\nabla\varphi - \Delta\varphi\Big)\rho^n(x)\,\mathrm{d}x = 0. \tag{5.1}$$

Proof. The derivation of the Euler-Lagrange equation for the sequence $\{\rho^n\}_{n=1}^N$ of minimizers follows the now well-established procedure; see e.g., [15, 21, 26]. We only sketch the main steps below. Let $\xi \in C_c^\infty(\mathbb{R}^d, \mathbb{R}^d)$, and we define a flow $\Phi : [0,\infty)\times\mathbb{R}^d \to \mathbb{R}^d$ by

$$\frac{\partial \Phi_s}{\partial s} = \xi(\Phi_s), \quad \text{with } \Phi_0 = \mathrm{Id}.$$

For any $s \in \mathbb{R}$, let $\rho_s(y)\mathrm{d}y$ be the push-forward of the measure $\rho^n(y)\mathrm{d}y$ under Φ_s. That is, for any $\zeta \in C_0^\infty(\mathbb{R}^d)$, we have

$$\int_{\mathbb{R}^d} \rho_s(y)\zeta(y)\,\mathrm{d}y = \int_{\mathbb{R}^d} \rho^n(y)\zeta(\Phi_s(y))\,\mathrm{d}y.$$

Since $\Phi_0 = \mathrm{Id}$, it follows that $\rho_0(y) = \rho^n(y)$ and an explicit calculation yields

$$\partial_s \rho_s\big|_{s=0} = -\operatorname{div}(\rho^n \xi).$$

Following the computations in [26], we derive the following stationarity condition on ρ^n:

$$\frac{C_\alpha}{\tau^\alpha}\int_{\mathbb{R}^{2d}} (y-x)\cdot\xi(y) P^n(\mathrm{d}x\,\mathrm{d}y) + \int_{\mathbb{R}^d}\Big(\nabla\Psi\cdot\xi - \operatorname{div}\xi\Big)\rho^n(x)\,\mathrm{d}x = 0, \tag{5.2}$$

where P^n is the optimal plan in the definition of the Wasserstein distance $W_2(\rho^n, \overline{\rho}^{n-1})$ between $\bar{\rho}^{n-1}$ and ρ^n. For any $\varphi \in C_0^\infty(\mathbb{R}^d)$, by choosing $\xi = \nabla\varphi$ in (5.2), we get

$$\frac{C_\alpha}{\tau^\alpha}\int_{\mathbb{R}^{2d}} (y-x)\cdot\nabla\varphi(y) P^n(\mathrm{d}x\,\mathrm{d}y) + \int_{\mathbb{R}^d}\Big(\nabla\Psi\cdot\nabla\varphi - \Delta\varphi\Big)\rho^n(x)\,\mathrm{d}x = 0.$$

This completes the proof of the lemma. □

The next result is an immediate corollary of Lemma 5.1.

COROLLARY 5.1. *The solutions* $\{\rho^n\}_{n=1}^N$ *given by Scheme 4.1 satisfy for all* $\varphi\in C_c^\infty(\mathbb{R}^d)$ *and all* $n=1,\ldots,N$*:*

$$\left|\int_{\mathbb{R}^d}\frac{C_\alpha(\rho^n-\overline{\rho}^{n-1})}{\tau^\alpha}\varphi\,\mathrm{d}x+\int_{\mathbb{R}^d}(\nabla\Psi\cdot\nabla\varphi-\Delta\varphi)\rho^n\,\mathrm{d}x\right|\le\sup_x\frac{\|\nabla^2\varphi(x)\|}{2}\tau^{-\alpha}W_2^2(\rho^n,\overline{\rho}^{n-1}),$$

where $\nabla^2\varphi\in\mathbb{R}^{d\times d}$ *denotes the Hessian of* φ*, and* $\|\cdot\|$ *denotes the spectral norm of a matrix.*

Proof. The assertion follows identically with (5.15) and (5.16) below, and hence it is omitted. □

In the next few lemmas, we derive several important *a priori* estimates on the sequence $\{\rho^n\}_{n=1}^N$ of approximations. These estimates are analogous to (42)-(45) in [26]. However, due to the appearance of the convex combination density $\overline{\rho}^{n-1}$ instead of ρ^{n-1} in Scheme 4.1, the derivation of these estimates is more involved than that in [26].

First, we derive elementary inequalities for $\mathcal{F}(\overline{\rho}^n)$, using convexity of $\mathcal{F}(\rho)$.

LEMMA 5.2. *For any* n*, there holds*

$$\mathcal{F}(\overline{\rho}^{n-1})\le\sum_{i=0}^{n-1}(-b_{n-i}^{(n)})\mathcal{F}(\rho^i),\tag{5.3}$$

$$\sum_{i=1}^{n}\mathcal{F}(\overline{\rho}^{i-1})\le n^{1-\alpha}\mathcal{F}(\rho^0)+\sum_{i=1}^{n-1}\big(1+(n-i)^{1-\alpha}-(n-i+1)^{1-\alpha}\big)\mathcal{F}(\rho^i).\tag{5.4}$$

Proof. Since $\mathcal{F}(\overline{\rho}^{n-1})=\mathcal{E}(\overline{\rho}^{n-1})+\mathcal{S}(\overline{\rho}^{n-1})$, for the energy term $\mathcal{E}(\overline{\rho}^{n-1})$, we have

$$\begin{aligned}\mathcal{E}(\overline{\rho}^{n-1})&=\int_{\mathbb{R}^d}\Psi\overline{\rho}^{n-1}\,\mathrm{d}x=\int_{\mathbb{R}^d}\Psi\sum_{i=0}^{n-1}(-b_{n-i}^{(n)})\rho^i\,\mathrm{d}x\\&=\sum_{i=0}^{n-1}(-b_{n-i}^{(n)})\int_{\mathbb{R}^d}\Psi\rho^i\,\mathrm{d}x=\sum_{i=0}^{n-1}(-b_{n-i}^{(n)})\mathcal{E}(\rho^i).\end{aligned}$$

For the entropy term $\mathcal{S}(\overline{\rho}^{n-1})$: since the function $z\mapsto s(z)=z\log(z)$ is convex for $z\ge0$ and the identity $-\sum_{i=0}^{n-1}b_{n-i}^{(n)}=1$ (cf. Lemma 3.1(i)), Jensen's inequality implies

$$s(\overline{\rho}^{n-1})=s\Big(-\sum_{i=0}^{n-1}b_{n-i}^{(n)}\rho^i\Big)\le\sum_{i=0}^{n-1}(-b_{n-i}^{(n)})s(\rho^i),$$

which, upon integration, immediately implies

$$\mathcal{S}(\overline{\rho}^{n-1})=\int_{\mathbb{R}^d}s(\overline{\rho}^{n-1})\,\mathrm{d}x\le\sum_{i=0}^{n-1}(-b_{n-i}^{(n)})\int_{\mathbb{R}^d}s(\rho^i)\,\mathrm{d}x=\sum_{i=0}^{n-1}(-b_{n-i}^{(n)})\mathcal{S}(\rho^i).$$

Then the preceding two estimates imply

$$\begin{aligned}\mathcal{F}(\overline{\rho}^{n-1})&=\mathcal{E}(\overline{\rho}^{n-1})+\mathcal{S}(\overline{\rho}^{n-1})\\&\le\sum_{i=0}^{n-1}(-b_{n-i}^{(n)})(\mathcal{E}(\rho^i)+\mathcal{S}(\rho^i))=\sum_{i=0}^{n-1}(-b_{n-i}^{(n)})\mathcal{F}(\rho^i).\end{aligned}$$

This shows the first assertion. Next, summing the inequality over $i=1$ to $i=n\le N$, changing the order of summation and relabeling the indices yield

$$\sum_{i=1}^{n}\mathcal{F}(\bar{\rho}^{i-1})\le\sum_{i=1}^{n}\sum_{j=0}^{i-1}(-b_{j-i}^{(j)})\mathcal{F}(\rho^i)=\sum_{i=0}^{n-1}\sum_{j=i+1}^{n}(-b_{j-i}^{(j)})\mathcal{F}(\rho^i)$$
$$=\Big(\sum_{i=1}^{n}(-b_i^{(i)})\Big)\mathcal{F}(\rho^0)+\sum_{i=1}^{n-1}\Big(\sum_{j=1}^{n-i}(-b_j^{(j+i)})\Big)\mathcal{F}(\rho^i).$$

Upon noting the identities in Lemma 3.1, we obtain the second assertion. ∎

The next result gives useful bounds on the free energy $\mathcal{F}(\rho^n)$ and $\mathcal{F}(\bar{\rho}^{n-1})$.

LEMMA 5.3. *Suppose that $\mathcal{F}(\rho^0)$ is finite. Let $\{\rho^n\}_{n=1}^N$ be the sequence of minimizers given by Scheme 4.1. Then for any positive integer $1\le n\le N$,*

$$\mathcal{F}(\rho^n)\le\mathcal{F}(\rho^0)\quad\text{and}\quad\mathcal{F}(\bar{\rho}^{n-1})\le\mathcal{F}(\rho^0).\tag{5.5}$$

Proof. Since ρ^n is the minimizer of problem (4.2) and $\bar{\rho}^{n-1}$ is an admissible density, we have

$$\frac{C_\alpha}{2\tau^\alpha}W_2^2(\rho^n,\bar{\rho}^{n-1})+\mathcal{F}(\rho^n)\le\frac{C_\alpha}{2\tau^\alpha}W_2^2(\bar{\rho}^{n-1},\bar{\rho}^{n-1})+\mathcal{F}(\bar{\rho}^{n-1})=\mathcal{F}(\bar{\rho}^{n-1}),$$

which implies

$$W_2^2(\rho^n,\bar{\rho}^{n-1})\le 2C_\alpha^{-1}\tau^\alpha\big(\mathcal{F}(\bar{\rho}^{n-1})-\mathcal{F}(\rho^n)\big).\tag{5.6}$$

It follows from this inequality and Lemma 5.2 that

$$\mathcal{F}(\rho^n)\le\mathcal{F}(\bar{\rho}^{n-1})\le\sum_{i=0}^{n-1}(-b_{n-i}^{(n)})\mathcal{F}(\rho^i).$$

Then by mathematical induction, we claim $\mathcal{F}(\rho^n)\le\mathcal{F}(\rho^0)$. Indeed, the claim holds trivially for $n=0$. Now suppose it holds up to $n\le N-1$, then by the induction hypothesis and the facts that $b_{n+1-i}^{(n+1)}<0$ for $i=0,\dots,k$ and $\sum_{i=0}^{n}(-b_{n+1-i}^{(n+1)})=1$, cf. Lemma 3.1, it follows

$$\mathcal{F}(\rho^{n+1})\le\sum_{i=0}^{n}(-b_{n+1-i}^{(n+1)})\mathcal{F}(\rho^i)\le\sum_{i=0}^{n}(-b_{n+1-i}^{(n+1)})\mathcal{F}(\rho^0)=\mathcal{F}(\rho^0),$$

which shows directly the first assertion. Then the second assertion follows immediately as $\mathcal{F}(\bar{\rho}^{n-1})\le\sum_{i=0}^{n-1}(-b_{n-i}^{(n)})\mathcal{F}(\rho^i)\le\mathcal{F}(\rho^0)\sum_{i=0}^{n-1}(-b_{n-i}^{(n)})\le\mathcal{F}(\rho^0)$, cf. Lemma 5.2. ∎

The next result gives a uniform bound on the second moment $M_2(\rho^n)$ of the approximation ρ^n, which plays a crucial role in the convergence analysis. The proof crucially employs the property of the relative entropy. Recall that the relative entropy $\mathcal{H}(\mu,\nu)$ between two probability measures μ and ν is defined by

$$\mathcal{H}(\mu|\nu)=\begin{cases}\int\log\left(\frac{d\mu}{d\nu}\right)d\mu & \text{if } d\mu\ll d\nu\\ +\infty & \text{otherwises.}\end{cases}$$

By Jensen's inequality, $\mathcal{H}(\mu|\nu)\geq 0$ for all μ and ν. Taking $\mu\in\mathcal{P}_2(\mathbb{R}^d)$ and $\nu=Z^{-1}e^{-\frac{|x|^2}{2m}}$, where $Z=(2\pi m)^{-\frac{d}{2}}$ is the normalization constant ($m>0$ is to be chosen), gives

$$0\leq\mathcal{H}(\mu|Z^{-1}e^{-\frac{|x|^2}{2m}})=\int_{\mathbb{R}^d}\mu\log\mu\,\mathrm{d}x+\frac{1}{2m}\int_{\mathbb{R}^d}|x|^2\mu\,\mathrm{d}x+\log Z,$$

This implies the following useful inequality

$$-\int_{\mathbb{R}^d}\mu\log\mu\,\mathrm{d}x\leq\tfrac{1}{2m}M_2(\mu)-\tfrac{d}{2}\log(2\pi m). \tag{5.7}$$

Lemma 5.4. *Suppose that $\mathcal{F}(\rho^0)$ and $M_2(\rho^0)$ are finite. Let $\{\rho^n\}_{n=1}^N$ be the sequence of minimizers given by Scheme 4.1. Then for any positive integer $1\leq n\leq N$, there holds*

$$M_2(\rho^n)\leq C. \tag{5.8}$$

Proof. The proof of the lemma is inspired by [16, Lemma 3.7, (3.10)]. Let P^n be the optimal plan for the Wasserstein distance $W_2(\rho^n,\overline{\rho}^{n-1})$ between $\overline{\rho}^{n-1}$ and ρ^n. Then by the definition of the second moment $M_2(\rho^n)$, there holds

$$\begin{aligned}M_2(\rho^n)&=\int|y|^2\rho^n(\mathrm{d}y)=\int|y|^2P^n(\mathrm{d}x\,\mathrm{d}y)\\&\leq\int(2|y-x|^2+2x^2)P^n(\mathrm{d}x\,\mathrm{d}y)\\&=2W_2^2(\rho^n,\overline{\rho}^{n-1})+2M_2(\overline{\rho}^{n-1})\\&=2W_2^2(\rho^n,\overline{\rho}^{n-1})+2\sum_{i=0}^{n-1}(-b_i^{(n)})M_2(\rho^i).\end{aligned}$$

By means of mathematical induction, this estimate, the inequality (5.6) and the assumptions $\mathcal{F}(\rho^0)<\infty$ and $M_2(\rho^0)<0$ directly imply that the second moment $M_2(\rho^n)$ of each of the approximation ρ^n is indeed finite. To derive a uniform bound (with respect to n and τ), we estimate the "fractional" difference quotient of the second moment using Corollary 5.1 with $\varphi=|x|^2$. This choice is justified by the finiteness of the second moment of each of the ρ^n:

$$\begin{aligned}\frac{C_\alpha}{\tau^\alpha}(M_2(\rho^n)-M_2(\overline{\rho}^{n-1}))&=\frac{C_\alpha}{\tau^\alpha}\int_{\mathbb{R}^d}x^2(\rho^n-\overline{\rho}^{n-1})\,\mathrm{d}x\\&\leq\Big|\int_{\mathbb{R}^d}(2\nabla\Psi\cdot x-2)\rho^n(x)\,\mathrm{d}x\Big|+\tau^{-\alpha}W_2^2(\rho^n,\overline{\rho}^{n-1})\\&\leq 2\int_{\mathbb{R}^d}|\nabla\Psi||x|\rho^n\,\mathrm{d}x+2+\tau^{-\alpha}W_2^2(\rho^n,\overline{\rho}^{n-1}).\end{aligned}$$

Now by the growth condition (4.1) on Ψ, we have

$$\int_{\mathbb{R}^d}|\nabla\Psi||x|\rho^n\,\mathrm{d}x\leq C(1+M_2(\rho^n)).$$

It follows from these two estimates and (5.6) that

$$C_\alpha\tau^{-\alpha}(M_2(\rho^n)-M_2(\overline{\rho}^{n-1}))\leq C(1+M_2(\rho^n))+\tau^{-\alpha}W_2^2(\rho^n,\overline{\rho}^{n-1})$$

$$\leq C(1+M_2(\rho^n))+C_\alpha(\mathcal{F}(\overline{\rho}^{n-1})-\mathcal{F}(\rho^n)).$$

Next we bound the terms $\mathcal{F}(\overline{\rho}^{n-1})$ and $-F(\rho^n)$ on the right hand side. First, by Lemma 5.3, $\mathcal{F}(\overline{\rho}^{n-1})\leq\mathcal{F}(\rho^0)<\infty$. Meanwhile, since $\Psi\geq 0$ by Assumption 4.1, we obtain from (5.7) that

$$-\mathcal{F}(\mu)=-\int_{\mathbb{R}^d}\mu\log\mu\,\mathrm{d}x-\int_{\mathbb{R}^d}\Psi\mu\,\mathrm{d}x\leq-\int_{\mathbb{R}^d}\mu\log\mu\,\mathrm{d}x\leq\tfrac{1}{2m}M_2(\mu)-\tfrac{d}{2}\log(2\pi m).$$

Applying this inequality with $\mu=\rho^n$ and $m=1/2$ gives

$$-\mathcal{F}(\rho^n)\leq M_2(\rho^n)-\frac{d}{2}\log(\pi). \tag{5.9}$$

These estimates together imply

$$C_\alpha\tau^{-\alpha}(M_2(\rho^n)-M_2(\overline{\rho}^{n-1}))\leq C(1+M_2(\rho^n)).$$

Further, by the definition of $\overline{\rho}^{n-1}$,

$$M_2(\overline{\rho}^{n-1})=\int_{\mathbb{R}^d}|x|^2\overline{\rho}^{n-1}\,\mathrm{d}x=\int_{\mathbb{R}^d}|x|^2\sum_{i=0}^{n-1}(-b_{n-i}^{(n)})\rho^i\,\mathrm{d}x=\sum_{i=0}^{n-1}(-b_{n-i}^{(n)}M_2(\rho^i).$$

Together with the definition of the L1 scheme in (3.1), it implies

$$\bar{\partial}_\tau^\alpha M_2(\rho^n)\leq C(1+M_2(\rho^n)).$$

This and the discrete Gronwall's inequality from Lemma 3.2 immediately imply the desired assertion. □

The next result gives a uniform bound on the entropy and energy of the approximations $\{\rho^n\}_{n=1}^N$, which induces the necessary compactness needed in the proof of Theorem 4.2. The notation $[]_+$ denotes taking the positive part.

LEMMA 5.5. *Suppose that $\mathcal{F}(\rho^0)$ and $M_2(\rho^0)$ are finite. Let $\{\rho^n\}_{n=1}^N$ be the sequence of minimizers given by Scheme 4.1. Then for any positive integer $1\leq n\leq N$, there hold*

$$\int_{\mathbb{R}^d}[\rho^n\log\rho^n]_+\,\mathrm{d}x\leq C,\quad \mathcal{E}(\rho^n)\leq C,\quad \sum_{n=1}^k W_2^2(\rho^n,\overline{\rho}^{n-1})\leq C\tau^\alpha.$$

Proof. These estimates are analogous to (43), (44) and (45) in [26]. According to [26, Equations (14)-(15)], there exist $0<\gamma<1$ and $C<\infty$ such that for all $\rho\in\mathcal{P}_2(\mathbb{R}^d)$

$$S(\rho)\geq -C(M_2(\rho)+1)^\gamma\quad\text{and}\quad\int_{\mathbb{R}^d}|\min\{\rho\log\rho,0\}|\,dx\leq C(M_2(\rho)+1)^\gamma. \tag{5.10}$$

Now the first two estimates follow directly from (5.10) and (5.8), and Lemmas 5.3 and 5.4 as

$$\begin{aligned}\int\max\{\rho^n\log\rho^n,0\}\,dx&\leq S(\rho^n)+\int_{\mathbb{R}^d}|\min\{\rho^n\log\rho^N,0\}|\,dx\\&\leq S(\rho^n)+C(M_2(\rho^n)+1)^\gamma\\&\leq\mathcal{F}(\rho^n)+C(M_2(\rho^n)+1)^\gamma\leq C,\end{aligned}$$

$$\begin{aligned}\mathcal{E}(\rho^k)&=\mathcal{F}(\rho^n)-\mathcal{S}(\rho^n)\\&\le\mathcal{F}(\rho^n)+C(M_2(\rho^n)+1)^\gamma\le C.\end{aligned}$$

It remains to prove the last estimate. By summing (5.6) over n and using the inequality (5.4), we obtain

$$\begin{aligned}&\sum_{i=1}^n W_2^2(\rho^i,\bar\rho^{i-1})\le\frac{2\tau^\alpha}{C_\alpha}\sum_{i=1}^n\big(\mathcal{F}(\bar\rho^{i-1})-\mathcal{F}(\rho^i)\big)\\&\le\frac{2\tau^\alpha}{C_\alpha}\left[n^{1-\alpha}\mathcal{F}(\rho^0)+\sum_{i=1}^{n-1}\Big((n-i)^{1-\alpha}-(n-i+1)^{1-\alpha}\Big)\mathcal{F}(\rho^i)-\mathcal{F}(\rho^n)\right].\end{aligned}\tag{5.11}$$

Next we bound the summation in the square bracket. The inequality (5.7) (with m to be chosen below) implies

$$\begin{aligned}&\sum_{i=1}^{n-1}\Big((n-i)^{1-\alpha}-(n-i+1)^{1-\alpha}\Big)\mathcal{F}(\rho^i)\\&=\sum_{i=1}^{n-1}\Big((n-i+1)^{1-\alpha}-(n-i)^{1-\alpha}\Big)(-\mathcal{F}(\rho^i))\\&\le\sum_{i=1}^{n-1}\big((n-i+1)^{1-\alpha}-(n-i)^{1-\alpha}\big)\Big(\frac{1}{2m}M_2(\rho^i)-\frac{d}{2}\log(2\pi m)\Big)\\&\le\frac{M}{2m}\sum_{i=1}^{n-1}\big((n-i+1)^{1-\alpha}-(n-i)^{1-\alpha}\big)-\frac{d}{2}\log(2\pi m)(n^{1-\alpha}-1)\\&=n^{1-\alpha}\Big[\frac{M}{2m}-\frac{d}{2}\log(2\pi m)\Big]+\frac{d}{2}\log(2\pi m)-\frac{M}{2m},\end{aligned}$$

where $M>0$ is an upper bound of $M_2(\rho^i)$ for all $i=1,\dots,n-1$ derived in Lemma 5.4. Consequently,

$$\begin{aligned}&n^{1-\alpha}\mathcal{F}(\rho^0)+\sum_{i=1}^{n-1}\Big((n-i)^{1-\alpha}-(n-i+1)^{1-\alpha}\Big)\mathcal{F}(\rho^i)\\&\le n^{1-\alpha}\Big[\mathcal{F}(\rho^0)+\frac{M}{2m}-\frac{d}{2}\log(2\pi m)\Big]+\frac{d}{2}\log(2\pi m)-\frac{M}{2m}.\end{aligned}\tag{5.12}$$

It suffices to bound the right hand side uniformly with respect to n. To this end, let $g:(0,+\infty)\to\mathbb{R}$ be defined by $g(m):=\mathcal{F}(\rho^0)+\frac{M}{2m}-\frac{d}{2}\log(2\pi m)$. Then, simple computation shows $g'(m)=-\frac{M}{2m^2}-\frac{d}{2m}<0$, $\lim_{m\to0^+}g(m)=+\infty$ and $\lim_{m\to+\infty}g(m)=-\infty$. Thus, the equation $g(m)=0$ has a unique solution $m^*\in(0,+\infty)$ that depends only on $\mathcal{F}(\rho^0)$, d and M. Choosing $m=m^*$ in (5.12) gives

$$n^{1-\alpha}\mathcal{F}(\rho^0)+\sum_{i=1}^{n-1}\Big((n-i)^{1-\alpha}-(n-i+1)^{1-\alpha}\Big)\mathcal{F}(\rho^i)\le\frac{d}{2}\log(2\pi m^*)-\frac{M}{2m^*}=\mathcal{F}(\rho^0).$$

This estimate, (5.11) and (5.9) together imply

$$\sum_{i=1}^n W_2^2(\rho^i,\bar\rho^{i-1})\le C\tau^\alpha\left[n^{1-\alpha}\mathcal{F}(\rho^0)+\sum_{i=1}^{n-1}\Big((n-i)^{1-\alpha}-(n-i+1)^{1-\alpha}\Big)\mathcal{F}(\rho^i)-\mathcal{F}(\rho^n)\right]$$

$$\leq C\tau^{\alpha}\left[\mathcal{F}(\rho^0)+M(\rho^n)-\frac{d}{2}\log(\pi)\right]\leq C\tau^{\alpha},$$

where the last step follows from Lemma 5.4. This completes the proof of the lemma. □

Now we can state the proof of Theorem 4.2.

Proof. [Proof of Theorem 4.2] The proof follows the strategy in [15, 21, 26]. The key idea is to pass to the limit $\tau\to 0^+$ in the Euler-Lagrange equation for the sequence of minimizers (5.1) proved in Lemma 5.1. The *a priori* estimates in Lemmas 5.4 and 5.5 provide necessary compactness properties that allow us to extract a convergent subsequence.

Let $T>0$ be a given final time. For each fixed $\tau>0$, let $\{\rho^n\}_{n=1}^N$ be the sequence of minimizers given by Scheme 4.1 and let $t\mapsto\rho_\tau(t)$ be the approximation defined in (4.3). By Lemmas 5.4 and 5.5, we have

$$M_2(\rho_\tau(t))+\int_{\mathbb{R}^d}[\rho_\tau(t)\log\rho_\tau(t)]^+\,\mathrm{d}x\leq C,\qquad\text{for all}\quad 0\leq t\leq T. \tag{5.13}$$

Since the function $z\mapsto[z\log z]^+$ has super-linear growth, the bound (5.13) and Dunford-Pettis theorem [41] ensure that there exists a subsequence, denoted again by ρ_τ, and some $\rho\in L^1((0,T)\times\mathbb{R}^d)$ such that

$$\rho_\tau\to\rho\ \text{ weakly in }\ L^1((0,T)\times\mathbb{R}^d). \tag{5.14}$$

It remains to show that the limit ρ satisfies the weak formulation (2.7) of problem (1.1) in the sense of Definition 2.1. Fix any test function $\varphi\in C_c^\infty((-\infty,T)\times\mathbb{R}^d)$. Let $P^n\in\mathcal{P}(\overline{\rho}^{n-1},\rho^n)$ be the optimal plan for $W_2(\overline{\rho}^{n-1},\rho^n)$. For any $0<t<T$, we have

$$\begin{aligned}
&\int_{\mathbb{R}^d}\left[\rho^n(x)-\overline{\rho}^{n-1}(x)\right]\varphi(t,x)\,\mathrm{d}x\\
&=\int_{\mathbb{R}^d}\rho^n(y)\varphi(t,y)\,\mathrm{d}y-\int_{\mathbb{R}^d}\overline{\rho}^{n-1}(x)\varphi(t,x)\,\mathrm{d}x\\
&=\int_{\mathbb{R}^{2d}}\left[\varphi(t,y)-\varphi(t,x)\right]P^n(\mathrm{d}x\,\mathrm{d}y)\\
&=\int_{\mathbb{R}^{2d}}(y-x)\cdot\nabla\varphi(t,y)P^n(\mathrm{d}x\,\mathrm{d}y)+\varepsilon_n,
\end{aligned}\tag{5.15}$$

where in the last line, we have used Taylor expansion of φ. The error term ε_n depends on t through time-dependence of φ and can be bounded by

$$|\varepsilon_n(t)|\leq C\int_{\mathbb{R}^{2d}}|y-x|^2\,P^n(\mathrm{d}x\,\mathrm{d}y)\leq CW_2^2(\overline{\rho}^{n-1},\rho^n). \tag{5.16}$$

From Lemma 5.1 and the identity (5.15), we obtain

$$\frac{C_\alpha}{\tau^\alpha}\int_{\mathbb{R}^d}\left[\rho^n(x)-\overline{\rho}^{n-1}(x)\right]\varphi(t,x)\,\mathrm{d}x=\int_{\mathbb{R}^d}(-\nabla\Psi\cdot\nabla\varphi+\Delta\varphi)\rho^n(x)\,\mathrm{d}x-\frac{C_\alpha}{\tau^\alpha}\varepsilon_n, \tag{5.17}$$

which, upon integrating with respect to t from t_{n-1} to t_n, yields

$$\begin{aligned}
&\frac{C_\alpha}{\tau^\alpha}\int_{t_{n-1}}^{t_n}\int_{\mathbb{R}^d}[\rho^n(x)-\overline{\rho}^{n-1}(x)]\varphi(t,x)\,\mathrm{d}x\,\mathrm{d}t\\
&=\int_{t_{n-1}}^{t_n}\int_{\mathbb{R}^d}(-\nabla\Psi\cdot\nabla\varphi+\Delta\varphi)\,\rho^n(x)\,\mathrm{d}x\,\mathrm{d}t-\frac{C_\alpha}{\tau^\alpha}\int_{t_{n-1}}^{t_n}\varepsilon_n\,\mathrm{d}t
\end{aligned}$$

$$= \int_{t_{n-1}}^{t_n} \int_{\mathbb{R}^d} (-\nabla\Psi\cdot\nabla\varphi + \Delta\varphi)\,\rho_\tau(t,x)\,\mathrm{d}x\,\mathrm{d}t - \frac{C_\alpha}{\tau^\alpha}\int_{t_{n-1}}^{t_n} \varepsilon_n\,\mathrm{d}t,$$

where the last line follows from the definition of the piecewise constant interpolation $\rho_\tau(t,x)$. Summing the last identity from $n=1$ to N gives

$$\sum_{n=1}^{N} \frac{C_\alpha}{\tau^\alpha}\int_{t_{n-1}}^{t_n}\int_{\mathbb{R}^d} [\rho^n(x) - \overline{\rho}^{n-1}(x)]\varphi(t,x)\,\mathrm{d}x\,\mathrm{d}t$$
$$= \int_0^T \int_{\mathbb{R}^d} (-\nabla\Psi\cdot\nabla\varphi + \Delta\varphi)\,\rho_\tau(t,x)\,\mathrm{d}x\,\mathrm{d}t + e_\tau, \quad (5.18)$$

where the term e_τ is given by

$$e_\tau = -\frac{C_\alpha}{\tau^\alpha}\sum_{n=1}^{N}\int_{t_{n-1}}^{t_n} \varepsilon_n\,\mathrm{d}t. \quad (5.19)$$

Now recall that by the definition $\overline{\rho}^{n-1}$ of Scheme 4.1, $\overline{\rho}^{n-1} = \sum_{i=0}^{n-1}(-b_{n-i}^{(n)})\rho^i$. This and the definition of the L1 approximation in (3.1), we can rewrite the left hand side of the identity (5.18) as

$$\begin{aligned}
&\sum_{n=1}^{N} \frac{C_\alpha}{\tau^\alpha}\int_{t_{n-1}}^{t_n}\int_{\mathbb{R}^d} [\rho^n(x) - \overline{\rho}^{n-1}(x)]\varphi(t,x)\,\mathrm{d}x\,\mathrm{d}t\\
=&\sum_{n=1}^{N} \frac{C_\alpha}{\tau^\alpha}\int_{t_{n-1}}^{t_n}\int_{\mathbb{R}^d} \Big[\rho^n(x) + \sum_{i=0}^{n-1} b_{n-i}^{(n)}\rho^i(x)\Big]\varphi(t,x)\,\mathrm{d}x\,\mathrm{d}t\\
=&\sum_{n=1}^{N}\int_{t_{n-1}}^{t_n}\int_{\mathbb{R}^d} \bar{\partial}_\tau^\alpha \rho^n(x)\varphi(t,x)\,\mathrm{d}x\,\mathrm{d}t.
\end{aligned}$$

By Lemma 3.3, we obtain

$$\begin{aligned}
&\sum_{n=1}^{N}\int_{t_{n-1}}^{t_n}\int_{\mathbb{R}^d} (\bar{\partial}_\tau^\alpha\rho^n)(t,x)\varphi(t,x)\,\mathrm{d}t\,\mathrm{d}x\\
=&\int_{\mathbb{R}^d}\int_0^T \rho_\tau(t,x)\overline{D}_\tau^\alpha\varphi(t,x)\,\mathrm{d}t\,\mathrm{d}x + C_\alpha\tau^{-\alpha}\int_{\mathbb{R}^d}\rho^0(x)\Big(\sum_{n=1}^{N} b_n^{(n)}\int_{t_{n-1}}^{t_n}\varphi(t,x)\,\mathrm{d}t\Big)\,\mathrm{d}x.
\end{aligned}$$

Now by Theorem 3.1, the following two limits hold

$$\lim_{\tau\to 0^+} \overline{D}_\tau^\alpha\varphi(t) = {}_tD_T^\alpha\varphi(t),$$
$$\lim_{\tau\to 0^+} C_\alpha\tau^{-\alpha}\sum_{n=1}^{N} b_n^{(n)}\int_{t_{n-1}}^{t_n}\varphi(t)\,\mathrm{d}t = -({}_tI_T^\alpha\varphi)(0).$$

Thus, upon passing to limit $\tau\to 0^+$ and noting the weak convergence of the sequence ρ_τ to ρ in $L^1(\Omega)$, we deduce

$$\lim_{\tau\to 0^+}\sum_{n=1}^{N}\frac{C_\alpha}{\tau^\alpha}\int_{t_{n-1}}^{t_n}\int_{\mathbb{R}^d}[\rho^n(x) - \bar{\rho}^{n-1}(x)]\varphi(t,x)\,\mathrm{d}x\,\mathrm{d}t$$

$$= \lim_{\tau\to 0^+} \int_{\mathbb{R}^d}\int_0^T \rho_\tau(t,x)\overline{D}_\tau^\alpha\varphi(t,x)\,\mathrm{d}t\,\mathrm{d}x + C_\alpha\tau^{-\alpha}\int_{\mathbb{R}^d}\rho^0(x)\Big(\sum_{n=1}^N b_n^{(n)}\int_{t_{n-1}}^{t_n}\varphi(t,x)\,\mathrm{d}t\Big)\,\mathrm{d}x$$
$$= \int_{\mathbb{R}^d}\int_0^T \rho(t,x)\,{}_tD_T^\alpha\varphi(t,x)\,\mathrm{d}t\,\mathrm{d}x - \frac{1}{\Gamma(1-\alpha)}\int_{\mathbb{R}^d}\rho^0(x)\int_0^T t^{-\alpha}\varphi(t,x)\,\mathrm{d}t\,\mathrm{d}x.$$

Meanwhile, for the first term on the right-hand side of (5.18), using the weak convergence of ρ_τ to ρ in $L^1((0,T)\times\mathbb{R}^d)$, we obtain

$$\lim_{\tau\to 0^+}\int_0^T\int_{\mathbb{R}^d}(-\nabla\Psi\cdot\nabla\varphi+\Delta\varphi)\,\rho_\tau(t,x)\,\mathrm{d}x\,\mathrm{d}t = \int_0^T\int_{\mathbb{R}^d}(-\nabla\Psi\cdot\nabla\varphi+\Delta\varphi)\,\rho(t,x)\,\mathrm{d}x\,\mathrm{d}t.$$

It remains to consider the error term e_τ. Actually, by the estimates (5.19) and (5.16), we have

$$|e_\tau| \le \frac{C_\alpha}{\tau^\alpha}\sum_{n=1}^N\int_{t_{n-1}}^{t_n}|\varepsilon_n(t)|\,\mathrm{d}t \le C\tau^{-\alpha}\sum_{n=1}^N\int_{t_{n-1}}^{t_n} W_2^2(\bar\rho^{n-1},\rho^n)\,\mathrm{d}t$$
$$\le C\tau^{1-\alpha}\sum_{n=1}^N W_2^2(\bar\rho^{n-1},\rho^n) \le C\tau,$$

where the last step is due to the bound on $\sum_{n=1}^N W_2^2(\bar\rho^{n-1},\rho^n)$ from Lemma 5.5. This inequality implies that $e_\tau\to 0$ as $\tau\to 0^+$. Therefore, taking the limit $\tau\to 0^+$, we deduce that the limiting density $\rho(t,x)$ satisfies

$$\int_0^T\int_{\mathbb{R}^d}\rho(t,x)\,{}_tD_T^\alpha\varphi(t,x)\,\mathrm{d}t\,\mathrm{d}x - \frac{1}{\Gamma(1-\alpha)}\int_0^T t^{-\alpha}\int_{\mathbb{R}^d}\rho_0(x)\varphi(t,x)\,\mathrm{d}x\,\mathrm{d}t$$
$$= \int_0^T\int_{\mathbb{R}^d}(-\nabla\Psi\cdot\nabla\varphi+\Delta\varphi)\,\rho(t,x)\,\mathrm{d}x\,\mathrm{d}t$$

which is precisely the weak formulation (2.7) of (1.1) in Definition 2.1. This completes the proof of the theorem. □

6. Numerical experiments

The classical JKO scheme may be employed as a time-stepping scheme for solving FPEs [2,29], although not extensively studied due to relatively high computational cost associated with the Wasserstein distance. Following the setting in [29], we illustrate the fractional scheme 4.1 with the following time-fractional FPE:

$$\partial_t^\alpha u - \nabla\cdot(\nabla u+\nabla\Psi u) = 0 \quad \text{in} \quad \Omega, \tag{6.1}$$

subject to the following initial and boundary conditions

$$u\big|_{t=0} = u_0 \ge 0 \quad \text{and} \quad (\nabla u+\nabla\Psi u)\cdot\nu = 0 \quad \text{on} \quad \partial\Omega, \tag{6.2}$$

where ν is the unit outward normal direction, and u_0 is assumed to be a probability density in Ω, i.e., $\int_\Omega u_0(x)\,\mathrm{d}x = 1$.

6.1. Implementation details Based on Scheme 4.1, we employ the following time semi-discrete approximation: Let $u^0 := u_0$, and for $n\ge 1$, define u^n to be the unique minimizer over $\mathcal{A} := \{u:\Omega\to\infty : u\in L^1(\Omega) \text{ and } \int_\Omega u(x)\,\mathrm{d}x = \int_\Omega u^n(x)\,\mathrm{d}x\}$ of the following functional

$$J(u) := \frac{C_\alpha}{2\tau^\alpha}W_2^2(u,\overline{u}^{n-1}) + \int_\Omega(u\log u+\Psi u)\,\mathrm{d}x, \quad \text{with } \overline{u}^{n-1} = \sum_{i=0}^{n-1}(-b_{n-i}^{(n)})u^i. \tag{6.3}$$

Next we describe a spatial discretization of the function $J(u)$ for the one-dimensional case $\Omega=(0,1)$, and the discretization is similarly for the high-dimensional case, provided that one can have a regular decomposition (e.g., triangulation) of the domain Ω. The interval $\Omega=[0,1]$ is discretized into subintervals $[x_i,x_{i+1}]$, where $x_i=ih$ and $h=1/M$ denotes the mesh size and $M\in\mathbb{N}_+$. Similarly, the time interval $[0,T]$ is discretized as into $[t_n,t_{n+1}]$, where $t_n=n\tau$, $n=0,1,\ldots$, and $\tau=T/N$ denotes the time-step size and $M\in\mathbb{N}_+$. Following [2, 29], we approximate the solutions u^n with spatially piecewise constant functions. The initial data $u_0:\Omega\to\mathbb{R}$ and the forcing $\Psi:\Omega\to\mathbb{R}$ are taken to be piecewise constant functions whose values coincide with their function values at the midpoint $x_{i+\frac{1}{2}}=x_i+\frac{h}{2}$, i.e., u_0 by the sequence $\mathbf{u}^0=(u_{0i})_{i=0}^{M-1}\in\mathbb{R}^M$, with $u_{0i}=u_0(x_{i+\frac{1}{2}})$, and similarly for Ψ. Accordingly, the integral of a function $f:\Omega\to\mathbb{R}$ over the domain Ω is approximate by the following midpoint quadrature:

$$\int_\Omega f(x)\,\mathrm{d}x\simeq h\sum_{i=0}^{M-1} f_i.$$

By absorbing the mesh size h into the probability, i.e., $u_i=\int_{x_i}^{x_{i+1}}u\,\mathrm{d}x$, then $\mathbf{u}\in\mathbb{R}^M$ belongs to the discrete probability space. The discrete analogue $J_h(\mathbf{u})$ of the functional $J(u)$ is given by

$$J_h(\mathbf{u})=\frac{C_\alpha}{2\tau^\alpha}W_2^2(\mathbf{u},\overline{\mathbf{u}}^{n-1})+\langle\mathbf{u},\log\mathbf{u}+\boldsymbol{\psi}\rangle,\quad\text{with }\overline{\mathbf{u}}^{n-1}=\sum_{i=0}^{n-1}(-b_{n-i}^{(n)})\mathbf{u}^i,\tag{6.4}$$

where log (and exponential) of a vector is understood componentwise, $\langle\cdot,\cdot\rangle$ denotes the usual Euclidean inner product on $\mathbb{R}^M$ (or $\mathbb{R}^{M\times M}$). The minimization is over the probability simplex $\Sigma_M=\{\mathbf{u}\in\mathbb{R}^M:u_i\geq 0,\sum_{i=0}^{M-1}u_i=1\}$. In the functional, we have dropped the constant independent of $\log\mathbf{u}$, since it does not affect the minimization.

The discrete functional $J_h(\mathbf{u})$ involves the Wasserstein distance $W_2^2(\mathbf{u},\overline{\mathbf{u}}^{n-1})$, and thus its efficient minimization is nontrivial, which has restricted the computation of Wasserstein gradient flow traditionally to the one spatial dimensional case, for which the Wasserstein distance can be computed explicitly via an inverse cumulative distribution function [29]. Nonetheless, over the past few years, the computation of Wasserstein distance has witnessed significant progress, especially within the computer vision and machine learning communities; see the monograph [40] for an up-to-date account. In the numerical experiments below, we employ the Dykstra algorithm given in Peyre [39] for each JKO time-stepping. It is based on entropic approximation of Wasserstein distance [13], and easily extended to the multi-dimensional case, when compared with the relaxation algorithm and projected gradient descent employed in [2,29]. We describe the whole computational procedure for minimizing $J_h(\mathbf{u})$ in Appendix A for the convenience of readers. Note that the Wasserstein distance can also be approximated using the entropic regularization, leading to Sinkhorn algorithm [13]. This algorithm is employed below to compute the error in Wasserstein distance approximately. In the computation, the (crucial) relaxation parameter γ in the algorithms is fixed at $1/N$, where N is the number of time steps.

6.2. Numerical results and discussions

Now we present some numerical results. First we consider the one-dimensional case.

EXAMPLE 6.1. *The domain $\Omega=(0,1)$, the initial condition $u_0(x)=1$ in Ω, and the forcing Ψ is given by $\Psi(x)=x$ or $\Psi(x)=\frac{1}{2}x^2$.*

The numerical results are given in Tables 6.1 and 6.2, respectively, at the time $T=1$, for the forcing $\Psi(x)=x$ and $\Psi(x)=\frac{1}{2}x^2$, where the $L^1(\Omega)$ and $L^2(\Omega)$ error of the numerical solutions with respect to the reference solution, which is computed on a much finer temporal grid with a time step size $\tau=1/1280$. Note that the $L^1(\Omega)$ metric was employed in the prior studies [2, 29], whereas the $L^2(\Omega)$ metric is very common in numerical analysis [23]. In addition, we also present the error in the Wasserstein distance (indicated by W in the tables), computed using Sinkhorn algorithm [13]. The results show that the scheme is convergent in either norm, with the convergence rate in the L^1 norm slightly higher than that for the L^2 norm. The convergence rate is consistently observed to be sublinear for all fractional orders, and it is slower than the first-order convergence of the standard implementation of the L1 scheme (implemented with the Galerkin in space) [22]. Surprisingly, the convergence rate deteriorates as the fractional order α increases, however, the precise mechanism of the loss remains elusive. In sharp contrast, the convergence in Wasserstein distance is rather stable with respect to the fractional order α. The empirical rate is at 0.47, which is slower than the optimal first-order rate for the classical JKO scheme (under suitable conditions) [4, Theorem 4.0.4]; see also [11] and [12, Theorem 2.7] for a convergence rate $O(\tau^{\frac{1}{4}})$ and $O(\tau^{\frac{1}{2}})$, respectively. In view of these empirical observations, it is of enormous interest to rigorously derive sharp convergence rate in the fractional case. Note that for the two forcing terms, the convergence behavior of the scheme is very similar to each other; see Fig. 6.1 for the density function at $T=1$. Qualitatively, the plots also indicate that the convergence speed to the equilibrium differs significantly with the fractional order α, as recently established by Kemppainen and Zacher [27], i.e., the smaller is the fractional order α, the slower is the convergence to the equilibrium.

TABLE 6.1. *Numerical results for the forcing* $\Psi(x)=x$.

$\alpha\backslash N$		20	40	80	160	320	rate
0.6	L^1	2.22e-2	1.41e-2	8.85e-3	5.40e-3	2.98e-3	0.72
	L^2	3.22e-2	2.18e-2	1.41e-2	8.62e-3	4.71e-3	0.69
	W	1.44e-1	1.05e-1	7.57e-2	5.42e-2	3.86e-2	0.47
0.8	L^1	3.11e-2	2.18e-2	1.44e-2	9.42e-3	5.56e-3	0.62
	L^2	4.43e-2	3.32e-2	2.38e-2	1.59e-2	9.52e-3	0.55
	W	1.44e-1	1.05e-1	7.58e-2	5.42e-2	3.87e-2	0.47
1.0	L^1	3.37e-2	2.58e-2	1.83e-2	1.19e-2	6.81e-3	0.57
	L^2	4.45e-2	3.66e-2	2.86e-2	2.07e-2	1.32e-2	0.43
	W	1.44e-1	1.05e-1	7.58e-2	5.43e-2	3.87e-2	0.47

The next example is concerned with a two-dimensional problem.

EXAMPLE 6.2. *The domain* $\Omega=(0,1)^2$, *the initial data* $u_0(x)=1$ *in* Ω, *and the forcing* Ψ *is given by* $\Psi(x_1,x_2)=x_1+x_2$.

The numerical results for Example 6.2 are presented in Table 6.3 and Fig. 6.2. The empirical convergence rates are similar to the one-dimensional case, and the convergence is also very steady (but the computing time is much higher). The density profiles for the three fractional orders are largely comparable at $T=1$, with the main differences lie at the boundary, as observed in the one-dimensional case, cf. Fig. 6.1. This is possibly due to the difference in the long-time behavior for different fractional orders.

Appendix A. Dykstra algorithm.

TABLE 6.2. *Numerical results for the forcing* $\Psi(x)=x^2/2$.

$\alpha\backslash N$		20	40	80	160	320	rate
0.6	L^1	1.41e-2	8.74e-3	5.06e-3	2.98e-3	1.61e-3	0.78
	L^2	1.93e-2	1.30e-2	8.46e-3	5.14e-3	2.79e-3	0.69
	W	1.45e-1	1.05e-1	7.60e-2	5.44e-2	3.88e-2	0.47
0.8	L^1	2.18e-2	1.51e-2	9.81e-3	5.85e-3	3.28e-3	0.68
	L^2	2.88e-2	2.17e-2	1.56e-2	1.04e-2	6.22e-3	0.55
	W	1.45e-1	1.05e-1	7.60e-2	5.44e-2	3.87e-2	0.47
1.0	L^1	3.04e-2	2.34e-2	1.66e-2	1.07e-2	6.09e-3	0.58
	L^2	3.81e-2	3.15e-2	2.46e-2	1.78e-2	1.14e-2	0.43
	W	1.45e-1	1.05e-1	7.60e-2	5.44e-2	3.88e-2	0.47

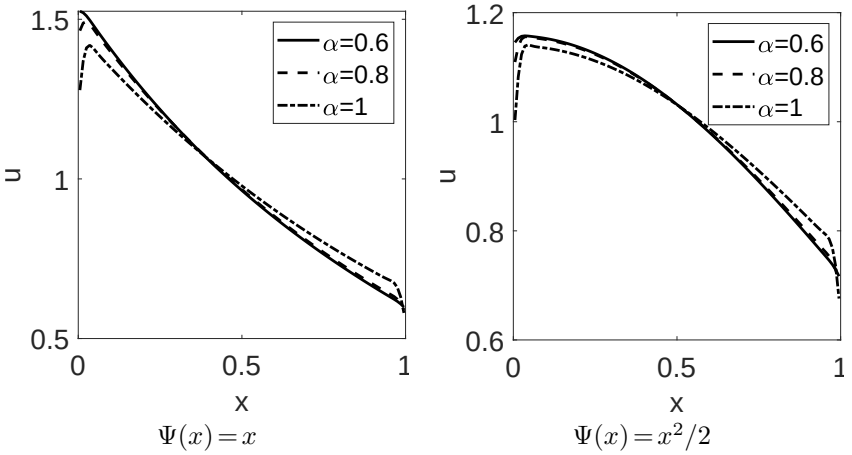

FIG. 6.1. *The probability density function at* $t=1$.

TABLE 6.3. *Numerical results for the forcing* $\Psi(x)=x_1+x_2$.

$\alpha\backslash N$		20	40	80	160	320	rate
0.6	L^1	3.28e-2	2.15e-2	1.35e-2	8.14e-3	4.43e-3	0.72
	L^2	4.81e-2	3.27e-2	2.13e-2	1.31e-2	7.24e-3	0.68
	W	2.04e-1	1.48e-1	1.07e-1	7.66e-2	5.47e-2	0.47
0.8	L^1	4.57e-2	3.31e-2	2.27e-2	1.46e-2	8.58e-3	0.60
	L^2	6.63e-2	5.01e-2	3.62e-2	2.46e-2	1.49e-2	0.53
	W	2.04e-1	1.48e-1	1.07e-1	7.67e-2	5.47e-2	0.47
1.0	L^1	4.93e-2	3.96e-2	2.92e-2	1.94e-2	1.09e-2	0.54
	L^2	6.57e-2	5.43e-2	4.27e-2	3.13e-2	2.05e-2	0.42
	W	2.04e-1	1.49e-1	1.07e-1	7.68e-2	5.48e-2	0.47

In this appendix, we describe the Dykstra algorithm for JKO stepping, originally developed in [39]. Let $\mathbf{C}=[c_{ij}]\in\mathbb{R}^{M\times M}$ be the cost, with $c_{ij}=|x_i-x_j|^2$. The entropic regularization of Wasserstein distance between two discrete probability measures $\mathbf{p},\mathbf{q}\in$

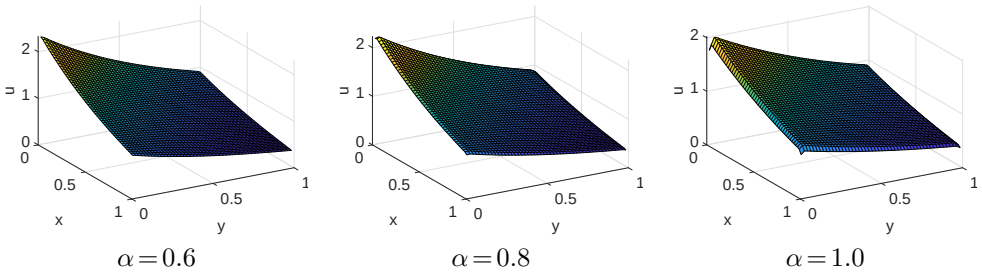

FIG. 6.2. *The probability density function for Example 6.2 at* $t=1$.

Σ_M for a cost $\mathbf{C}\in\mathbb{R}^{M\times M}$ is given by

$$W_{2,\gamma}(\mathbf{p},\mathbf{q})^2=\min_{\boldsymbol{\pi}\in\mathcal{P}(\mathbf{p},\mathbf{q})}\langle\mathbf{C},\boldsymbol{\pi}\rangle+\gamma\langle\boldsymbol{\pi},\log\boldsymbol{\pi}-\mathbf{1}\rangle+\langle\mathbf{1},\ell_{\mathbb{R}_+^{M\times M}}(\boldsymbol{\pi})\rangle,$$

where $\gamma>0$ is a small number, controlling the tradeoff between accuracy and computational efficiency, and $\mathcal{P}(\mathbf{p},\mathbf{q})$ is the set of couplings between $\mathbf{p}$ and $\mathbf{q}$, i.e., $\mathcal{P}(\mathbf{p},\mathbf{q})=\{\boldsymbol{\pi}\in\mathbb{R}_+^{M\times M}:\boldsymbol{\pi}\mathbf{1}=\mathbf{p},\boldsymbol{\pi}^T\mathbf{1}=\mathbf{q}\}$, with $\mathbf{1}$ being a vector or a matrix with all entries equal to unit. Accordingly, the entropic regularization of the fractional JKO functional is given by

$$\langle\mathbf{C},\boldsymbol{\pi}\rangle+\gamma\langle\boldsymbol{\pi},\log\boldsymbol{\pi}-\mathbf{1}\rangle+\langle\mathbf{1},\ell_{\mathbb{R}_+^{M\times M}}(\boldsymbol{\pi})\rangle+\tau' f(\boldsymbol{\pi}\mathbf{1})+\ell_{\mathcal{C}_q}(\boldsymbol{\pi}),\tag{A.1}$$

where ℓ_C is an indicator function, $\tau'=\frac{2\tau^\alpha}{C_\alpha}$, $\mathcal{C}_q=\{\boldsymbol{\pi}\in\mathbb{R}^{M\times M}:\boldsymbol{\pi}^T\mathbf{1}=q\}$ and $f(\mathbf{q})=\langle\mathbf{q},\log\mathbf{q}-\mathbf{1}+\boldsymbol{\psi}\rangle$. This functional can be recast into

$$\min_{\boldsymbol{\pi}}\mathrm{KL}(\boldsymbol{\pi}|\boldsymbol{\xi})+\varphi_1(\boldsymbol{\pi})+\varphi_2(\boldsymbol{\pi}),$$

with KL being the classical KL divergence, and

$$\varphi_1(\boldsymbol{\pi})=\ell_{\mathcal{C}_q}(\boldsymbol{\pi})\quad\text{and}\quad\varphi_2(\boldsymbol{\pi})=\tfrac{\tau'}{\gamma}f(\boldsymbol{\pi}\mathbf{1}),$$

and the Gibbs kernel $\boldsymbol{\xi}$ is given by

$$\boldsymbol{\xi}=e^{-\mathbf{C}/\gamma}\in\mathbb{R}_{+,*}^{M\times M}.$$

The update is obtained using $\mathbf{p}=\boldsymbol{\pi}\mathbf{1}$. It remains to minimize (A.1) with respect to the coupling $\boldsymbol{\pi}\in\mathcal{P}(\mathbf{p},\mathbf{q})$. This can be carried out using the Dykstra algorithm developed in [39]; see Algorithm 1 for the complete procedure, where the notation $\circ$ denotes componentwise product between two vectors. It is noteworthy that the algorithm operates only on vectors $\mathbf{a},\mathbf{b},\mathbf{u},\mathbf{v}$ instead of the coupling $\boldsymbol{\xi}$ directly, due to the fact that the optimal coupling satisfies $\boldsymbol{\pi}=\mathrm{diag}(\mathbf{a})\boldsymbol{\xi}\mathrm{diag}(\mathbf{b})$, for some $\mathbf{a},\mathbf{b}\in\mathbb{R}_+^M$, like the classical entropic regularization of optimal transport [13].

The (Kullback-Leibler) KL proximal operator $\mathrm{Prox}_{\sigma f}^{\mathrm{KL}}(\mathbf{q})$ in (A.4) for any $\mathbf{q}\in\mathbb{R}_+^M$ is defined by

$$\mathrm{Prox}_{\sigma f}^{\mathrm{KL}}(\mathbf{q})=\arg\min_{\mathbf{p}\in\mathbb{R}_+^M}\langle\mathbf{p},\log\tfrac{\mathbf{p}}{\mathbf{q}}-\mathbf{1}\rangle+\sigma\langle\mathbf{p},\log\mathbf{p}-\mathbf{1}+\boldsymbol{\psi}\rangle.$$

Algorithm 1 Dykstra algorithm for JKO stepping.

1: Set $\mathbf{a}^0=\mathbf{b}^0=\mathbf{u}^0=\mathbf{v}^0=\mathbf{1}$, and specify the tolerance ϵ.
2: **for** $\ell=1,\dots,L$ **do**
3: **if** ℓ is odd **then**
4: update $\mathbf{a}^\ell$ and $\mathbf{b}^\ell$ by

$$\mathbf{a}^\ell=\mathbf{a}^{\ell-1}\circ\mathbf{u}^{\ell-2}\quad\text{and}\quad\mathbf{b}^\ell=\frac{\mathbf{q}}{\boldsymbol{\xi}^T(\mathbf{a}^\ell)};\tag{A.2}$$

5: **else**
6: update $\mathbf{a}^\ell$ and $\mathbf{b}^\ell$ by

$$\mathbf{b}^\ell=\mathbf{b}^{\ell-1}\circ\mathbf{v}^{\ell-2}\quad\text{and}\quad\mathbf{a}^\ell=\frac{\mathbf{p}^\ell}{\boldsymbol{\xi}(\mathbf{b}^\ell)},\tag{A.3}$$

where $\mathbf{p}^\ell$ is given by

$$\mathbf{p}^\ell=\mathrm{Prox}^{\mathrm{KL}}_{\frac{\tau}{\gamma}f}(\mathbf{a}^{\ell-1}\circ\mathbf{u}^{\ell-2}\circ\boldsymbol{\xi}(\mathbf{b}^\ell));\tag{A.4}$$

7: **end if**
8: update $\mathbf{u}^\ell$ and $\mathbf{v}^\ell$ by

$$\mathbf{u}^\ell=\mathbf{u}^{\ell-2}\circ\frac{\mathbf{a}^{\ell-1}}{\mathbf{a}^\ell}\quad\text{and}\quad\mathbf{v}^\ell=\mathbf{v}^{\ell-2}\circ\frac{\mathbf{b}^{\ell-1}}{\mathbf{b}^\ell};$$

9: if $\|\mathbf{b}^\ell\circ\boldsymbol{\xi}^T(\mathbf{a}^\ell)-\mathbf{q}\|<\epsilon$ and ℓ is even, terminate the iteration;
10: **end for**
11: Output $\mathbf{p}^\ell$ defined in (A.4).

Due to the separability of the optimization problem, it suffices to minimize the one-dimensional function $g(s)=s\log\frac{s}{t}-s+t+\sigma(s\log s-s+s\psi)$. Differentiating with respect to s and setting it to zero gives $\log s-\log t+\sigma(\log s+\psi)=0$, i.e., $\log s=\frac{1}{1+\sigma}\log t-\frac{\sigma\psi}{1+\sigma}$, and $s^*=t^{\frac{1}{1+\sigma}}e^{-\frac{\sigma}{1+\sigma}\psi}$. Thus the proximal operator is given by

$$\mathrm{Prox}^{\mathrm{KL}}_{\sigma f}(\mathbf{q})=\mathbf{q}^{\frac{1}{1+\sigma}}\circ e^{-\frac{\sigma}{1+\sigma}\boldsymbol{\psi}}.$$

The stopping criterion at line 9 employs the violation of the constraint $\mathcal{C}_q$.

REFERENCES

[1] P. Aceves-Sanchez and L. Cesbron. Fractional diffusion limit for a fractional Vlasov-Fokker-Planck equation. *SIAM J. Math. Anal.*, 51(1):469–488, 2019. 1

[2] M. Agueh and M. Bowles. One-dimensional numerical algorithms for gradient flows in the p-Wasserstein spaces. *Acta Appl. Math.*, 125:121–134, 2013. 6, 6.1, 6.1, 6.2

[3] G. Akagi. Fractional flows driven by subdifferentials in Hilbert spaces. *Israel J. Math.*, 234(2):809–862, 2019. 2.3

[4] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient Flows in Metric Spaces and in the Space of Probability Measures*. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, second edition, 2008. 6.2

[5] C. N. Angstmann, I. C. Donnelly, B. I. Henry, T. A. M. Langlands, and P. Straka. Generalized continuous time random walks, master equations, and fractional Fokker-Planck equations. *SIAM J. Appl. Math.*, 75(4):1445–1468, 2015. 1

[6] B. Baeumer and P. Straka. Fokker-Planck and Kolmogorov backward equations for continuous time random walk scaling limits. *Proc. Amer. Math. Soc.*, 145(1):399–412, 2017. 1

[7] E. Barkai, R. Metzler, and J. Klafter. From continuous time random walks to the fractional Fokker-Planck equation. *Phys. Rev. E (3)*, 61(1):132–138, 2000. 1

[8] M. Bowles and M. Agueh. Weak solutions to a fractional Fokker-Planck equation via splitting and Wasserstein gradient flow. *Appl. Math. Lett.*, 42:30–35, 2015. 1

[9] Y. Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Comm. Pure Appl. Math.*, 44(4):375–417, 1991. 4.1

[10] F. Camilli and R. De Maio. A time-fractional mean field game. *Adv. Differential Equations*, 24(9/10):531–554, 2019. 1, 2.3

[11] P. Clément and W. Desch. A Crandall-Liggett approach to gradient flows in metric spaces. *J. Abstr. Differ. Equ. Appl.*, 1(1):46–60, 2010. 6.2

[12] K. Craig. The exponential formula for the Wasserstein metric. *ESAIM Control Optim. Calc. Var.*, 22(1):169–187, 2016. 6.2

[13] M. Cuturi. Sinkhorn distances: lightspeed computation of optimal transport. In *Advances in Neural Information Processing Systems*, pages 2292–2300, 2013. 1, 6.1, 6.2, A

[14] M. H. Duong and Y. Lu. An operator splitting scheme for the fractional kinetic Fokker-Planck equation. *Discrete Contin. Dyn. Syst.*, 39(10):5707–5727, 2019. 1

[15] M. H. Duong, M. A. Peletier, and J. Zimmer. Conservative-dissipative approximation schemes for a generalized Kramers equation. *Math. Methods Appl. Sci.*, 37(16):2517–2540, 2014. 5, 5

[16] S. Eberle, B. Niethammer, and A. Schlichting. Gradient flow formulation and longtime behaviour of a constrained Fokker-Planck equation. *Nonlinear Anal.*, 158:142–167, 2017. 5

[17] C. R. Givens and R. M. Shortt. A class of Wasserstein metrics for probability distributions. *Michigan Math. J.*, 31(2):231–240, 1984. 4.1

[18] R. Gorenflo, Y. Luchko, and M. Yamamoto. Time-fractional diffusion equation in the fractional Sobolev spaces. *Fract. Calc. Appl. Anal.*, 18(3):799–820, 2015. 2.1, 2.3

[19] M. G. Hahn, K. Kobayashi, and S. Umarov. Fokker-Planck-Kolmogorov equations associated with time-changed fractional Brownian motion. *Proc. Amer. Math. Soc.*, 139(2):691–705, 2011. 2.2, 2.2

[20] B. I. Henry, T. A. M. Langlands, and P. Straka. Fractional Fokker-Planck equations for subdiffusion with space- and time-dependent forces. *Phys. Rev. Lett.*, 105:170602, 2010. 1, 2.1

[21] C. Huang. A variational principle for the Kramers equation with unbounded external forces. *J. Math. Anal. Appl.*, 250(1):333–367, 2000. 5, 5

[22] B. Jin, R. Lazarov, and Z. Zhou. An analysis of the L1 scheme for the subdiffusion equation with nonsmooth data. *IMA J. Numer. Anal.*, 36(1):197–221, 2016. 3, 6.2

[23] B. Jin, R. Lazarov, and Z. Zhou. Numerical methods for time-fractional evolution equations with nonsmooth data: a concise overview. *Comput. Methods Appl. Mech. Engrg.*, 346:332–358, 2019. 3, 6.2

[24] B. Jin, B. Li, and Z. Zhou. Numerical analysis of nonlinear subdiffusion equations. *SIAM J. Numer. Anal.*, 56(1):1–23, 2018. 3

[25] B. Jin and Z. Zhou. An analysis of Galerkin proper orthogonal decomposition for subdiffusion. *ESAIM Math. Model. Numer. Anal.*, 51(1):89–113, 2017. 3

[26] R. Jordan, D. Kinderlehrer, and F. Otto. The variational formulation of the Fokker-Planck equation. *SIAM J. Math. Anal.*, 29(1):1–17, 1998. (document), 1, 1, 2.3, 3, 4.2, 4.2, 5, 5, 5, 5

[27] J. Kemppainen and R. Zacher. Long-time behavior of non-local in time Fokker-Planck equations via the entropy method. *Math. Models Methods Appl. Sci.*, 29(2):209–235, 2019. 1, 6.2

[28] A. A. Kilbas, H. M. Srivastava, and J. J. Trujillo. *Theory and Applications of Fractional Differential Equations*. Elsevier Science B.V., Amsterdam, 2006. 1, 2.1, 2.1

[29] D. Kinderlehrer and N. J. Walkington. Approximation of parabolic equations using the Wasserstein metric. *M2AN Math. Model. Numer. Anal.*, 33(4):837–852, 1999. 6, 6.1, 6.1, 6.2

[30] K. N. Le, W. McLean, and K. Mustapha. Numerical solution of the time-fractional Fokker-Planck equation with general forcing. *SIAM J. Numer. Anal.*, 54(3):1763–1784, 2016. 1

[31] L. Li and J.-G. Liu. A generalized definition of Caputo derivatives and its application to fractional ODEs. *SIAM J. Math. Anal.*, 50(3):2867–2900, 2018. 2.1, 2.3

[32] L. Li and J.-G. Liu. A discretization of Caputo derivatives with application to time fractional SDEs and gradient flows. *SIAM J. Numer. Anal.*, 57(5):2095–2120, 2019. 1

[33] H.-l. Liao, D. Li, and J. Zhang. Sharp error estimate of the nonuniform L1 formula for linear reaction-subdiffusion equations. *SIAM J. Numer. Anal.*, 56(2):1112–1133, 2018. 3

[34] Y. Lin and C. Xu. Finite difference/spectral approximations for the time-fractional diffusion equation. *J. Comput. Phys.*, 225(2):1533–1552, 2007. 1, 3, 3, 3, 4.2

[35] M. Magdziarz, J. Gajda, and T. Zorawik. Comment on fractional Fokker-Planck equation with

space and time dependent drift and diffusion. *J. Stat. Phys.*, 154(5):1241–1250, 2014. 1, 2.2, 2.2

[36] M. Magdziarz, A. Weron, and K. Weron. Fractional fokker-planck dynamics: Stochastic representation and computer simulation. *Phys. Rev. E*, 75:016708, Jan 2007. 2.2, 2.2

[37] R. Metzler and J. Klafter. The random walk's guide to anomalous diffusion: a fractional dynamics approach. *Phys. Rep.*, 339(1):1–77, 2000. 1

[38] E. Nane and Y. Ni. Stochastic solution of fractional Fokker-Planck equations with space-time-dependent coefficients. *J. Math. Anal. Appl.*, 442(1):103–116, 2016. 1

[39] G. Peyré. Entropic approximation of Wasserstein gradient flows. *SIAM J. Imaging Sci.*, 8(4):2323–2351, 2015. 1, 6.1, A, A

[40] G. Peyré and M. Cuturi. Computational optimal transport. *Foundations and Trends in Machine Learning*, 11(5–6):355–607, 2019. 1, 6.1

[41] K. Yosida. *Functional Analysis*. Springer-Verlag, Berlin-New York, sixth edition, 1980. 5

[42] R. Zacher. Weak solutions of abstract evolutionary integro-differential equations in Hilbert spaces. *Funkcial. Ekvac.*, 52(1):1–18, 2009. 2.3