Abstract—We develop a new method to solve the Fokker-Planck or Kolmogorov’s forward equation that governs the time evolution of the joint probability density function of a continuous-time stochastic nonlinear system. Numerical solution of this equation is fundamental for propagating the effect of initial condition, parametric and forcing uncertainties through a nonlinear dynamical system, and has applications encompassing but not limited to forecasting, risk assessment, nonlinear filtering and stochastic control. Our methodology breaks away from the traditional approach of spatial discretization for solving this second-order partial differential equation (PDE), which in general, suffers from the “curse-of-dimensionality”. Instead, we numerically solve an infinite dimensional proximal recursion in the space of probability density functions, which is theoretically equivalent to solving the Fokker-Planck-Kolmogorov PDE. We show that the dual formulation along with the introduction of an entropic regularization, leads to a smooth convex optimization problem that can be implemented via suitable block co-ordinate iteration and has fast convergence due to certain contraction property that we establish. This approach enables meshless implementation leading to remarkably fast computation.

I. INTRODUCTION

Given a deterministic or stochastic dynamical system in continuous time over some finite dimensional state space, say $\mathbb{R}^n$, we consider the problem of propagating the trajectory ensembles or densities subject to stochastic initial conditions—often referred to as the belief or uncertainty propagation problem. Mathematically, this amounts to solving an initial value problem associated with a partial differential equation (PDE) of the form

$$\frac{\partial \rho}{\partial t} = \mathcal{L}\rho, \quad \rho(x, t = 0) = \rho_0(x) \text{ given,}$$

(1)

describing the transport of the density function $\rho(x, t)$, which is a function of the state vector $x \in \mathbb{R}^n$, and time $t \geq 0$. Here, $\mathcal{L}$ is a spatial operator that guarantees $\rho \geq 0$, and $\int_{\mathbb{R}^n} \rho(x, t) dx = 1$ for all $t \geq 0$. Without loss of generality, one can interpret $\rho(x, t)$ as the joint probability density function (PDF) of the state vector $x$ at time $t$. We refer to (1) as transport PDE.

The structural form of $\mathcal{L}$ in (1) depends on the underlying trajectory level dynamics. For example, consider the case when the dynamics of $x(t) \in \mathbb{R}^n$ is governed by an Itô stochastic differential equation (SDE) $dx = f(x, t) dt + g(x, t) dw$, $x(t=0) = x_0 \sim \rho_0$ (given), the process noise $w(t) \in \mathbb{R}^m$ is Wiener and satisfy $\mathbb{E}[dw_i dw_j] = \delta_{ij} dt$ for all $i, j = 1, \ldots, n$, where $\delta_{ij} = 1$ for $i = j$, and zero otherwise. Then,

$$\mathcal{L}\rho \equiv -\nabla \cdot (\rho f) + \frac{1}{2} \sum_{i,j=1}^{n} \frac{\partial^2}{\partial x_i \partial x_j} (\rho g^T g)_{ij},$$

(2)

and the resulting transport PDE (2) is known as the Fokker-Planck or Kolmogorov’s forward equation. Henceforth, we will refer to it as the FPK PDE.

The problem of uncertainty propagation, that is, the problem of computing $\rho(x, t)$ that satisfies a PDE of the form (2), is ubiquitous across science and engineering. Representative applications include meteorological forecasting [1], dispersion analysis in spacecraft entry-descent-landing [2], orientation density evolution for liquid crystals in chemical physics [3]–[5], motion planning in robotics [6]–[8], computing the prior PDF in nonlinear filtering [9], [10], probabilistic model validation [11]–[13], and analyzing the statistical mechanics of macromolecules [14]. In all these applications, it is of importance to compute the joint PDF $\rho(x, t)$ in a scalable and unified manner, rather than employing specialized techniques in a case-by-case basis or developing discretization-based PDE solvers which suffer from the “curse of dimensionality” [15]. The Liouville PDE being first order, can be solved efficiently using the method-of-characteristics [2]. However, solving the second order FPK PDE in a manner that avoids both spatial discretization and function approximation, remains challenging to date.

In this paper, we pursue the solution of (1) through a variational viewpoint arising from the theory of optimal mass transport [16]. This viewpoint, first proposed in [17], interprets (1) as a gradient or steepest descent of certain functional $\Phi(\cdot)$ on the infinite dimensional manifold of PDFs with finite second (raw) moments, denoted as

$$\mathcal{D}_2 := \{ \rho : \mathbb{R}^n \mapsto \mathbb{R} \mid \rho \geq 0, \int_{\mathbb{R}^n} \rho = 1, \mathbb{E}_\rho[xx^T] < \infty \}. \quad \text{Specifically, let } k = 0, 1, 2, \ldots, \text{and for some fixed time-step } h > 0, \text{consider a variational recursion}
$$

$$\varrho_k(x) = \arg \inf_{\varrho \in \mathcal{D}_2} \frac{1}{2} d^2(\varrho, \varrho_{k-1}) + h \Phi(\varrho),$$

subject to the initial condition $\varrho_0(x) := \rho_0(x)$, i.e., the initial PDF of (1). Here, $d(\cdot, \cdot)$ is a distance metric on the manifold

$^1$We denote the expectation operator w.r.t. the measure $\rho(x) dx$ as $\mathbb{E}_\rho[\cdot]$.
Then, the idea is to design the metric \( d(\cdot,\cdot) \) and the functional \( \Phi(\cdot) \) in (3) such that \( \varrho_k(x) \to \rho(x,t=kh) \) as \( h \downarrow 0 \), i.e., in the small time-step limit, the solution of the variational recursion (3) converges (in strong \( L^1 \) sense) to that of (1). The main result in [17] was to show that for FPK operators of the form (2) with \( f \) being a gradient vector field and \( g \) being a scalar multiple of identity matrix, the distance \( d(\cdot,\cdot) \) can be taken as the Wasserstein-2 metric with \( \Phi(\cdot) \) as the free energy functional. We will make these ideas precise in Section II and III. The resulting variational recursion (3) has since been known as the Jordan-Kinderlehrer-Otto (JKO) scheme [18], and we will refer the FPK operator with such assumptions on \( f \) and \( g \) to be in “JKO canonical form”.

Similar gradient descent schemes have been derived for many other PDEs; see e.g., [19] for a recent survey.

To motivate gradient descent in infinite dimensional spaces, we appeal to a more familiar setting, i.e., gradient descent in \( \mathbb{R}^n \) associated with the flow

\[
\frac{dx}{dt} = -\nabla \varphi(x) \quad x(0) = x_0, \tag{4}
\]

where \( x, x_0 \in \mathbb{R}^n \) and \( \varphi : \mathbb{R}^n \to \mathbb{R}_+, \) and is continuously differentiable. The Euler discretization for (4) is given by

\[
x_k - x_{k-1} = -h \nabla \varphi(x_{k-1}), \tag{5}
\]

which can be rewritten as a variational recursion

\[
x_k = \arg \min_x \frac{1}{2} \left\| x - x_{k-1} \right\|^2 + h \varphi(x) + o(h). \tag{6}
\]

In the optimization literature, the mapping \( x_{k-1} \mapsto x_k \), given by

\[
\text{prox}_{h\varphi}(x_{k-1}) := \arg \min_x \frac{1}{2} \left\| x - x_{k-1} \right\|^2 + h \varphi(x), \tag{7}
\]

is called the “proximal operator” [20, p. 142]. The sequence \( \{x_k\} \) generated by the proximal recursion

\[
x_k = \text{prox}_{h\varphi}(x_{k-1}), \quad k = 0, 1, 2, \ldots \tag{8}
\]

converges to the flow of the ODE (4), i.e., the sequence satisfies \( x_k \to x(t=kh) \) as the step-size \( h \downarrow 0 \). Using the finite dimensional viewpoint (7), we define

\[
\text{prox}_{h\Phi}(g_{k-1}) := \arg \inf_{g \in \mathcal{G}_2} \frac{1}{2} d^2(g, g_{k-1}) + h \Phi(g), \tag{9}
\]

as an infinite dimensional proximal operator. As mentioned above, the sequence \( \{g_k\} \) generated by the proximal recursion (3) converges to the flow of the PDE (4), i.e., the sequence satisfies \( g_k(x) \to \rho(x,t=kh) \) as the step-size \( h \downarrow 0 \). We also note that in the finite dimensional case,

\[
\frac{d}{dt} \varphi = \left\langle \nabla \varphi, -\nabla \varphi \right\rangle = -\left\| \nabla \varphi \right\|^2 < 0 \tag{10}
\]

which implies \( \varphi \) decays along the flow of (4). As we will see next, the appeal of using (3) to solve the FPK PDE comes from the fact that the Euclidean gradient descent can be generalized to the manifold \( \mathcal{G}_2 \) by appropriately choosing the metric \( d(\cdot,\cdot) \) and the functional \( \Phi(\cdot) \) in (3), in parallel with the quantities \( \| \cdot \| \) and \( \varphi(\cdot) \) in (8), respectively.

\[\]
and its solution satisfies $\rho \geq 0$, $\int_{\mathbb{R}^n} \rho \, dx = 1$ for all $t \in [0, \infty)$. It is easy to verify that the unique stationary solution of (14) is the Gibbs PDF $\rho_\infty(x) = \kappa \exp \left( -\beta \psi(x) \right)$, where the normalizing constant $\kappa := \int_{\mathbb{R}^n} \exp(-\beta \psi(x)) \, dx$ is referred to as the partition function. A Lyapunov functional associated with the FPK PDE (14) is the free energy

$$ F(\rho) := \mathbb{E}_\rho \left[ \psi + \beta^{-1} \log \rho \right] = \beta^{-1} D_{KL} \left( \rho \| \exp(-\beta \psi(x)) \right) \geq 0, $$

(15)

that decays [17] along the solution trajectory of (14), i.e., $\frac{d}{dt} F < 0$. This follows from re-writing (14) as

$$ \frac{\partial \rho}{\partial t} = \nabla \cdot (\rho \nabla \zeta), \quad \text{where } \zeta := \beta^{-1} (1 + \log \rho) + \psi, \quad (17) $$

and consequently

$$ \frac{d}{dt} F = -\mathbb{E}_\rho \left[ \| \nabla \zeta \|^2 \right] < 0, \quad (18) $$

with equality achieved at the stationary solution $\rho_\infty = \kappa e^{-\beta \psi(x)}$. In our context, (18) serves as the infinite-dimensional analog of (10). The term free energy is motivated by noting that (15) can be seen as the sum of the potential energy $\int_{\mathbb{R}^n} \rho \psi(x) \, dx$ and the internal energy $\beta^{-1} \int_{\mathbb{R}^n} \rho \log \rho \, dx$. When $\psi = 0$, the PDE (14) reduces to the heat equation, which by (15), can then be interpreted as an entropy maximizing flow.

The seminal paper [17] establishes that the FPK PDE (14) can be seen as the gradient descent flow of the free energy functional $F(\cdot)$ w.r.t. the 2-Wasserstein Metric. Specifically, the solution of (14) can be recovered from the following proximal recursion of the form (3):

$$ \varrho_k = \text{prox}_{\beta^2\mathcal{F}(\cdot)}(\varrho_{k-1}) = \arg \inf_{\varrho \in \mathcal{Q}} \frac{1}{2} W_2^2(\varrho_{k-1}, \varrho) + h F(\varrho), \quad k = 1, 2, \ldots \quad (19a) $$

with $\varrho_0 \equiv \varrho_0(x)$ (from (14)) as $h \downarrow 0$. Next, we develop a framework to numerically solve (19).

IV. MAIN RESULTS

To solve (19), we discretize time as $t = 0, h, 2h, \ldots$, and develop an algorithm to solve (19) without making any spatial discretization. In other words, we would like to perform the recursion (19) on weighted scattered point cloud \( \{x_k, \varrho_k\}_{k=1}^N \) of cardinality $N$ at $t_k = kh$, $k \in \mathbb{N}$, where the location of the point $x_k \in \mathbb{R}^n$ denotes the state-space coordinate, and the corresponding weight $\varrho_k \in \mathbb{R}_{\geq 0}$ denotes the value of the joint PDF evaluated at that point at time $t_k$. Such weighted scattered point cloud representation of (19) results in the following problem:

$$ \varrho_k = \arg \min_{\varrho} \left\{ \min_{M \in \Pi(\varrho_{k-1}, \varrho)} \frac{1}{2} \langle C_k, M \rangle + h \langle \psi_{k-1}, \varrho \rangle + \beta^{-1} \log \varrho, \varrho \right\}, \quad (20) $$

to be solved for $k = 1, 2, \ldots$, where the drift potential vector $\psi_{k-1} \in \mathbb{R}^N$ is given by

$$ \psi_{k-1}(i) := \psi(x_{k-1}^i), \quad i = 1, 2, \ldots, N. $$

Similarly, the probability vectors $\varrho, \varrho_{k-1} \in \mathbb{R}^N$. Furthermore, for each $k = 1, 2, \ldots$, the matrix $C_k \in \mathbb{R}^{N \times N}$ is given by

$$ C_k(i, j) := \| x_k^i - x_{k-1}^j \|^2, \quad i, j = 1, 2, \ldots, N, $$

and $\Pi(\varrho_{k-1}, \varrho)$ stands for the set of all matrices $M \in \mathbb{R}^{N \times N}$ such that

$$ M \geq 0, \quad M1 = \varrho_{k-1}, \quad M^T 1 = \varrho. \quad (21) $$

Due to the nested minimization structure in (20), its numerical solution is far from obvious. Notice that the inner minimization in (21) is standard linear programming problem if it were to be solved for a given $\varrho$. However, the outer minimization in (20) precludes a direct numerical approach. To circumvent the aforesaid issues, following [22], we first regularize and then dualize (20). Specifically, adding an entropic regularization $H(M) := \langle M, \log M \rangle$ in (20) yields

$$ \varrho_k = \arg \min_{\varrho} \left\{ \min_{M \in \Pi(\varrho_{k-1}, \varrho)} \frac{1}{2} \langle C_k, M \rangle + \epsilon H(M) \right\}, \quad (22) $$

(23)

where $\epsilon > 0$ is a regularization parameter. The entropic regularization is standard in optimal mass transport literature [23], [24] and leads to efficient Sinkhorn iteration for the inner minimization. In our context, the entropic regularization “algebrizes” the inner minimization in the sense if $\lambda_0, \lambda_1$ are Lagrange multipliers associated with the equality constraints in (21), then the optimal coupling matrix $M^{opt} := [m^{opt}(i, j)]$ in (22) has the Sinkhorn form

$$ m^{opt}(i, j) = \exp(\lambda_0(i)/\epsilon) \exp(-C_k(i, j)/2\epsilon) \exp(\lambda_1(j)/\epsilon), \quad (23) $$

Since the objective in (22) is proper convex and lower semicontinuous in $\varrho$, the strong duality holds, and we consider the Lagrange dual of (22) given by:

$$ \lambda_0^{opt}, \lambda_1^{opt} = \arg \max_{\lambda_0, \lambda_1 \geq 0} \left\{ \langle \lambda_0, \varrho_{k-1} \rangle - F^*(-\lambda_1) \right\}, \quad (23) $$

(24)

where

$$ F^*(\varrho) := \sup_{x \in \mathbb{R}^n} \{ \langle \varrho, x \rangle - F(x) \} \quad (24) $$

is the Legendre-Fenchel transform of the free energy $F(\cdot)$ given by (15). Next, we derive the first order optimality conditions for (24), and then provide an algorithm to solve the same.
A. Conditions for Optimality

Given the vectors \( \varrho_{k-1}, \psi_{k-1} \), the matrix \( C_k \), and the positive scalars \( \beta, h, \epsilon \) in (24), let

\[
y := \exp(\lambda_0 h / \epsilon), \quad z := \exp(\lambda_1 h / \epsilon), \quad \Gamma_k := \exp(-C_k / 2\epsilon), \quad \xi_k := \exp(-\beta \psi_{k-1} - 1).
\]

(26) (27)

The following result provides a way of computing \( \lambda_0^{\text{opt}}, \lambda_1^{\text{opt}} \) in (24), and consequently \( \varrho_k \) in (22).

**Theorem 1:** (See [25, Section III.A]) The vectors \( \lambda_0^{\text{opt}}, \lambda_1^{\text{opt}} \) in (24) can be found by solving for \( y \) and \( z \) from the following system of equations:

\[
y \odot (\Gamma_k z) = \varrho_{k-1},
\]

\[
z \odot (\Gamma_k^T y) = \xi_k - z^{-\Delta},
\]

(28a) (28b)

and then inverting the maps (26). The vector \( \varrho_k \) in (22), i.e., the proximal update can then be obtained as

\[
\varrho_k = z^{\text{opt}} \odot (\Gamma_k^T y^{\text{opt}}),
\]

(29)

where \( (y^{\text{opt}}, z^{\text{opt}}) \) denotes the solution of (28).

B. Algorithm

1) Proximal recursion: We now propose a block coordinate iteration scheme to solve (28). Specifically, the proposed procedure, which we call PROXRECUR, and detail in Algorithm 1, takes \( \varrho_{k-1} \) as input and returns the proximal update \( \varrho_k \) as output for \( k = 1, 2, \ldots \). In addition to the data \( \varrho_{k-1}, \psi_{k-1}, C_k, \beta, h, \epsilon, N \), the Algorithm 1 requires two parameters as user input: numerical tolerance \( \delta \), and maximum number of iterations \( L \). The computation in Algorithm 1, as presented, involves making an initial guess for the vector \( z \) and then updating \( y \) and \( z \) until convergence.

**Algorithm 1**  Proximal recursion to compute \( \varrho_k \) from \( \varrho_{k-1} \)

1: procedure PROXRECUR(\( \varrho_{k-1}, \psi_{k-1}, C_k, \beta, h, \epsilon, N, \delta, L \))
2: \( \Gamma_k \leftarrow \exp(-C_k / 2\epsilon) \)
3: \( \xi \leftarrow \exp(-\beta \psi_{k-1} - 1) \)
4: \( \varrho_0 \leftarrow \text{rand}_{N \times 1} \)
5: \( \varrho \leftarrow \varrho_0, 0_{N \times (L-1)} \)
6: \( y \leftarrow \varrho_{k-1} \odot (\Gamma_k \varrho_0), 0_{N \times (L-1)} \)
7: \( \ell = 1 \)  \begin{aligned} &\text{iteration index} \end{aligned}
8: while \( \ell \leq L \) do
9: \( \text{end while} \)
10: \( \varrho^{(\ell)} := \varrho_{k-1} \odot (\Gamma_k^T y^{(\ell)}) \)
11: \( \text{if } \| y^{(\ell)} - y^{(\ell)} \| < \delta \) \begin{aligned} (z^{(\ell)}, \ell) &\| < \delta \) \begin{aligned} \text{error within tolerance} \end{aligned} \end{aligned}
12: \( \text{then} \)
13: \( \text{break} \)
14: \( \ell \leftarrow \ell + 1 \)
15: \( \text{end if} \)
17: \( \text{return } \varrho_k \leftarrow \varrho^{(\ell)} \odot (\Gamma_k^T y^{(\ell)}) \)
18: \( \text{end procedure} \)

Several questions arise: how can one ensure that such a procedure converges? Also, even if convergence can be guaranteed, is the rate fast in practice? The latter issue is important since the time-step \( h \) in the JKO scheme is small, and during the computation of Algorithm 1, the physical time is “frozen”. We will establish the convergence guaranteed by showing certain contractive properties of the recursion given in Algorithm 1. Before doing so, we now outline the overall algorithmic setup to implement the proximal recursion over probability weighted scattered point cloud data.

B. Algorithm

2) Overall scheme: Samples from the known initial joint PDF \( \rho_0 \) are generated as point cloud \( \{x_i, \phi_i \}_{i=1}^N \). Then for \( k = 1, 2, \ldots \), the state vectors \( \{x_i^k \}_{i=1}^N \) are updated via Euler-Maruyama scheme applied to the underlying SDE: the corresponding probability weights \( \{\phi_i^k \}_{i=1}^N \) are updated via Algorithm 1. Notice that computing \( C_k \) requires both \( \{x_i^k \}_{i=1}^N \) and \( \{\phi_i^k \}_{i=1}^N \), and that \( C_k \) needs to be passed as input to Algorithm 1. Thus, the execution of Euler-Maruyama scheme precedes that of Algorithm 1.

C. Convergence

The following establish the convergence of Algorithm 1.

**Theorem 2:** (See [25, Section III.C]) Consider the notations in (26)-(27), and those in Algorithm 1. The iteration

\[
z^{(\ell, \ell+1)} := (\xi_k - \odot (\Gamma_k^T y^{(\ell)}))^{1+2h/\epsilon} \]

(30)

for \( \ell = 1, 2, \ldots \), is strictly contractive in the Thompson metric on \( \mathbb{R}_0^N \), and admits unique fixed point \( z^{\text{opt}} \in \mathbb{R}_0^N \).

**Corollary 3:** (See [25, Section III.C]) The Algorithm 1 converges to unique fixed point \( (y^{\text{opt}}, z^{\text{opt}}) \in \mathbb{R}_0^N \times \mathbb{R}_0^N \).

V. NUMERICAL SIMULATION

In this section, we apply the algorithmic setup proposed in Section IV.B to few examples illustrating the numerical approach. Our examples involve systems which are already in JKO canonical form (Section III), as well as those which can be transformed to such form by non-obvious change of coordinates.

A. Linear Gaussian System

For an Itô SDE of the form

\[
dx = Ax \, dt + B \, dw,
\]

it is well known that if \( x_0 := x(t = 0) \sim \mathcal{N}(\mu_0, \Sigma_0) \), then the transient joint PDFs \( \rho(x, t) = \mathcal{N}(\mu(t), \Sigma(t)) \) where the vector-matrix pair \( (\mu(t), \Sigma(t)) \) evolve according to the ODEs

\[
\dot{\mu}(t) = A \mu, \quad \mu(0) = \mu_0, \]

(32a)

\[
\dot{\Sigma}(t) = A\Sigma(t) + A\Sigma(t)^T + BB^T, \quad \Sigma(0) = \Sigma_0. \]

(32b)

We benchmark the numerical results produced by the proposed proximal algorithm vis-à-vis the above analytical solutions. We consider the following two sub-cases of (31).
1) Ornstein-Uhlenbeck Process: We consider the 1D system
\[ dx = -ax \, dt + \sqrt{2\beta} \, dw, \quad a, \beta > 0, \]  
(33)
which is in JKO canonical form with \( \psi(x) = \frac{1}{2}ax^2 \). We generate \( N = 400 \) samples from the initial PDF \( \rho_0 = \mathcal{N}(\mu_0, \sigma_0^2) \) with \( \mu_0 = 5 \) and \( \sigma_0^2 = 4 \times 10^{-2} \), and apply the proposed proximal recursion for (33) with time step \( h = 10^{-3} \), and with parameters \( a = 1, \beta = 1, \epsilon = 5 \times 10^{-2} \). For implementing Algorithm 1, we set tolerance \( \delta = 10^{-3} \) and maximum number of iterations \( L = 100 \). Fig. 1 shows that the PDF point clouds generated by the proximal recursion match with the analytical PDFs \( \mathcal{N}(\mu(t), \Sigma(t)) \) (left), \( \mathcal{N}(\mu(t) + \Sigma(t) \psi(x), \Sigma(t)) \) (middle), and \( \mathcal{N}(\mu(t), \Sigma_\infty(t)) \) (right). As mentioned in Section III, the stationary PDF is \( \rho_\infty(x) = \kappa \exp(-\beta \psi(x)) \), which for our choice of \( \psi \), is bimodal. The transient PDFs

2) Multivariate LTI: We next consider the multivariate case (31) where the pair \((A, B)\) is assumed to be controllable, and the matrix \( A \) is Hurwitz (not necessarily symmetric). Under these assumptions, the stationary PDF is \( \mathcal{N}(0, \Sigma_\infty) \) where \( \Sigma_\infty \) is the unique stationary solution of (32b) that is guaranteed to be symmetric positive definite. However, it is not apparent whether (31) can be expressed in the form (13), since for non-symmetric \( A \), there does not exist constant symmetric positive definite matrix \( \Psi \) such that \( A x = -\nabla x^T \Psi x \), i.e., the drift vector field does not admit a natural potential. Thus, implementing the JKO scheme for (31) is non-trivial in general.

In a recent work [26], two successive time-varying co-ordinate transformations were given which can bring (31) in the form (13), thus making it amenable to the JKO scheme. We apply these change-of-coordinates to (31) with
\[
A = \begin{pmatrix} -10 & 5 \\ -30 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 2 \\ 2.5 \end{pmatrix},
\]
which satisfy the stated assumptions on \((A, B)\), and implement the proposed proximal recursion on this transformed co-ordinates with \( N = 400 \) samples generated from the initial PDF \( \rho_0 = \mathcal{N}(\mu_0, \Sigma_0) \), where \( \mu_0 = (4, 4)^T \) and \( \Sigma_0 = 4I_2 \). As before, we set \( \delta = 10^{-3}, L = 100, h = 10^{-3}, \beta = 1, \epsilon = 5 \times 10^{-2} \). Once the proximal updates are done, we transform back the probability weighted scattered point cloud to the original state space co-ordinates via change-of-measure formula associated with the known co-ordinate transforms [26, Section III.B]. Fig. 2 shows the resulting point clouds superimposed with the contour plots for the analytical solutions \( \mathcal{N}(\mu(t), \Sigma(t)) \) given by (32).
Fig. 3: The proximal (weighted scattered point cloud) joint PDFs of the FPK PDE (14) with time step $h = 10^{-3}$, and with parameters $\beta = 1$, $\epsilon = 5 \times 10^{-2}$. Simulation details are given in Section V.B. The color (red = high, blue = low) denotes the joint PDF value obtained via proximal recursion at a point at that time (see colorbar).

have no known analytical solution but can be computed using the proposed proximal recursion. For doing so, we generate $N = 400$ samples from the initial PDF $\rho_0 = \mathcal{N}(\mu_0, \Sigma_0)$ with $\mu_0 = (2, 2)^T$ and $\Sigma_0 = 4I_2$, and set $\delta = 10^{-3}$, $L = 100$, $h = 10^{-3}$, $\beta = 1$, $\epsilon = 5 \times 10^{-2}$, as before. The resulting weighted point clouds are shown in Fig. 3; it can be seen that as time progresses, the joint PDFs computed via the proximal recursion, tend to the known stationary solution $\rho_\infty$ (contour plots in the right bottom sub-figure in Fig. 3).

VI. CONCLUSIONS

We proposed a variational recursion to numerically solve the transient Fokker-Planck or Kolmogorov’s forward equation by exploiting the underlying infinite-dimensional gradient flow structure in the manifold of PDFs. From a computational standpoint, this work develops a novel point cloud solver for performing the Otto calculus avoiding spatial discretization or function approximation. From systems-theoretic standpoint, this work contributes to an emerging research program [26], [27] in uncovering new geometric meanings of the equations of uncertainty propagation and filtering, and using the same to efficiently solve these equations via proximal algorithms [26].

REFERENCES

[1] M. Ehrendorfer, “The Liouville equation and its potential usefulness for the prediction of forecast skill. part I: Theory,” Monthly Weather Review, vol. 122, no. 4, pp. 703–713, 1994.

[2] A. Halder and R. Bhattacharya, “Dispersion analysis in hypersonic flight during planetary entry using stochastic Liouville equation,” Journal of Guidance, Control, and Dynamics, vol. 34, no. 2, pp. 459–474, 2011.

[3] S. Hess, “Fokker-Planck-equation approach to flow alignment in liquid crystals,” Zeitschrift für Naturforschung A, vol. 31, no. 9, pp. 1034–1037, 1976.

[4] W. Muschik and B. Su, “Mesoscopic interpretation of Fokker-Planck equation describing time behavior of liquid crystal orientation,” The Journal of Chemical Physics, vol. 107, no. 2, pp. 580–584, 1997.

[5] Y. P. Kalmykov and W. T. Coffey, “Analytical solutions for rotational diffusion in the mean field potential: application to the theory of dielectric relaxation in nematic liquid crystals,” Liquid crystals, vol. 25, no. 3, pp. 329–339, 1998.

[6] W. Park, J. S. Kim, Y. Zhou, N. J. Cowan, A. M. Okamura, and G. S. Chirikjian, “Diffusion-based motion planning for a nonholonomic flexible needle model,” in Robotics and Automation, 2005. ICRA 2005. Proceedings of the 2005 IEEE International Conference on. IEEE, 2005, pp. 4600–4605.

[7] W. Park, Y. Liu, Y. Zhou, M. Moses, and G. S. Chirikjian, “Kinematic state estimation and motion planning for stochastic nonholonomic systems using the exponential map,” Robotica, vol. 26, no. 4, pp. 419–434, 2008.

[8] H. Hamann and H. Wörn, “A framework of space–time continuous models for algorithm design in swarm robotics,” Swarm Intelligence, vol. 2, no. 2–4, pp. 299–239, 2008.

[9] S. Challa and Y. Bar-Shalom, “Nonlinear filter design using Fokker-Planck-Kolmogorov probability density evolutions,” IEEE Transactions on Aerospace and Electronic Systems, vol. 36, no. 1, pp. 309–315, 2000.

[10] F. Daum, “Nonlinear filters: beyond the Kalman filter,” IEEE Aerospace and Electronic Systems Magazine, vol. 20, no. 8, pp. 57–69, 2005.

[11] A. Halder and R. Bhattacharya, “Model validation: A probabilistic formulation,” in Decision and Control and European Control Conference (CDC-ECC), 2011 50th IEEE Conference on. IEEE, 2011, pp. 1692–1697.

[12] ——, “Further results on probabilistic model validation in Wasserstein metric,” in Decision and Control (CDC). IEEE, 2012 51st Annual Conference on. IEEE, 2012, pp. 5542–5547.

[13] ——, “Probabilistic model validation for uncertain nonlinear systems,” Automatica, vol. 50, no. 8, pp. 2038–2050, 2014.

[14] P. J. Flory and M. Volkenstein, Statistical mechanics of chain molecules. Wiley, 1969.

[15] R. E. Bellman, Dynamic Programming. Courier Dover Publications, 1957.

[16] C. Villani, Topics in optimal transportation. American Mathematical Soc., 2003, no. 58.

[17] R. Jordan, D. Kinderlehrer, and F. Otto, “The variational formulation of the Fokker–Planck equation,” SIAM Journal on Mathematical Analysis, vol. 29, no. 1, pp. 1–17, 1998.

[18] L. Ambrosio, N. Gigli, and G. Savaré, Gradient flows: in metric spaces and in the space of probability measures. Springer Science & Business Media, 2008.

[19] F. Santambrogio, “[Euclidean, metric, and Wasserstein] gradient flows: an overview,” Bulletin of Mathematical Sciences, vol. 7, no. 1, pp. 87–154, 2017.

[20] N. Parikh, S. Boyd et al., “Proximal algorithms,” Foundations and Trends® in Optimization, vol. 1, no. 3, pp. 127–239, 2014.

[21] J.-D. Benamou and Y. Brenier, “A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem,” Numerische Mathematik, vol. 84, no. 3, pp. 375–393, 2000.

[22] J. Karlsson and A. Ringh, “Generalized Sinkhorn iterations for regularizing inverse problems using optimal mass transport,” SIAM Journal on Imaging Sciences, vol. 10, no. 4, pp. 1935–1962, 2017.

[23] M. Cuturi, “Sinkhorn distances: Lightspeed computation of optimal transport,” in Advances in neural information processing systems, 2013, pp. 2292–2300.

[24] J.-D. Benamou, G. Carlier, M. Cuturi, L. Nenna, and G. Peyré, “Iterative Bregman projections for regularized transportation problems,” SIAM Journal on Scientific Computing, vol. 37, no. 2, A1111–A1138, 2015.

[25] K. F. Caluya and A. Halder, “Gradient flow algorithms for density propagation in stochastic systems,” preprint, 2018.

[26] A. Halder and T. T. Georgiou, “Gradient flows in uncertainty propagation and filtering of linear Gaussian systems,” in Decision and Control (CDC), 2017 IEEE 56th Annual Conference on. IEEE, 2017, pp. 3081–3088.

[27] ——, “Gradient flows in filtering and Fisher-Rao geometry,” in 2018 Annual American Control Conference (ACC). IEEE, 2018, pp. 4281–4286.