Generalized Gradient Flows for Stochastic Prediction, Estimation, Learning and Control

Abhishek Halder

Department of Aerospace Engineering Iowa State University Ames, IA 50011

Joint work with students and collaborators

















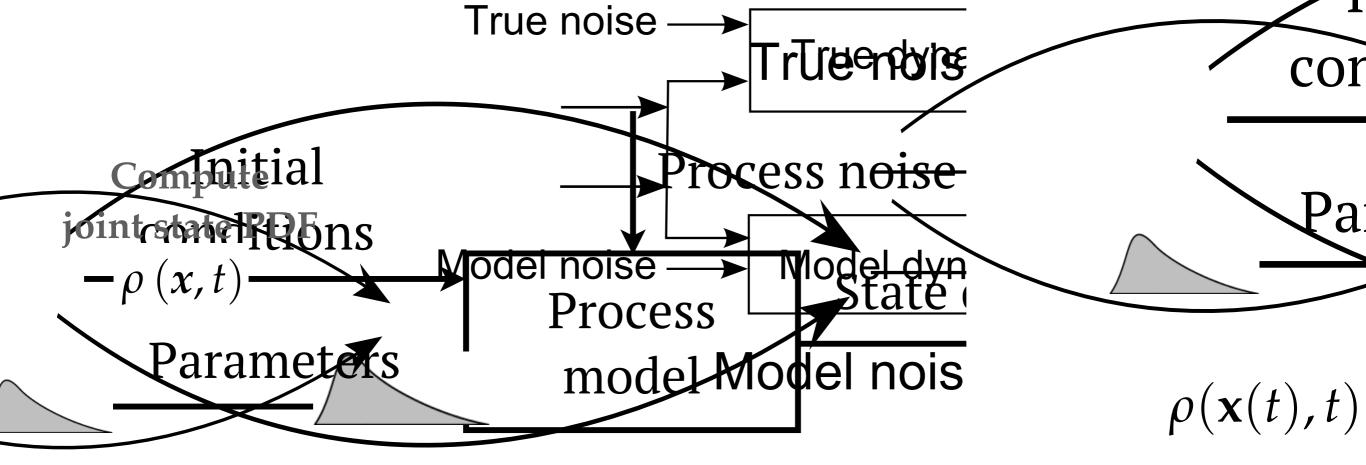
Translational AI Center Seminar, Iowa State University September 22, 2023



Topic of this talk

Control theory and algorithms for measures/distributions and densities

Why bother?

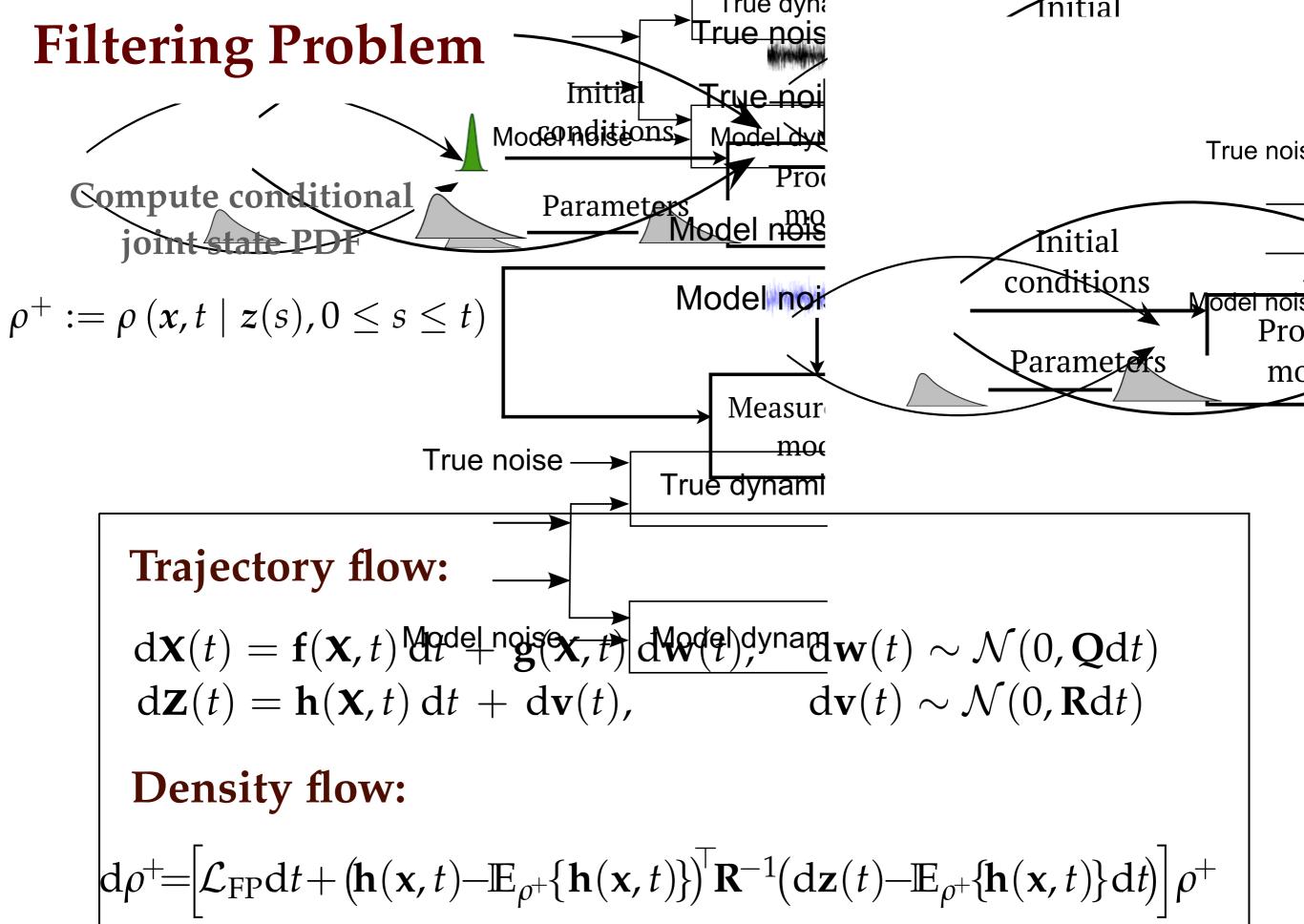


Trajectory flow:

$$d\mathbf{x}(t) = \mathbf{f}(\mathbf{x}, t) dt + \mathbf{g}(\mathbf{x}, t) d\mathbf{w}(t), \quad d\mathbf{w}(t) \sim \mathcal{N}(0, \mathbf{Q}dt)$$

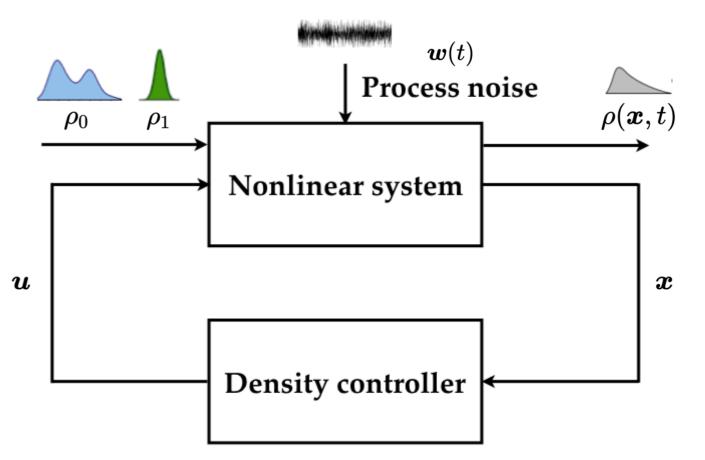
Density flow:

$$\frac{\partial \rho}{\partial t} = \mathcal{L}_{\text{FP}}(\rho) := -\nabla \cdot (\rho \mathbf{f}) + \frac{1}{2} \sum_{i,j=1}^{n} \frac{\partial^2}{\partial x_i \partial x_j} \left(\left(\mathbf{g} \mathbf{Q} \mathbf{g}^{\mathsf{T}} \right)_{ij} \rho \right)$$



Control Problem

Steer joint state PDF via feedback control over finite time horizon



minimize
$$\mathbb{E}\left[\int_0^1 \|u\|_2^2 dt\right]$$
 subject to $dx = f(x, u, t) dt + g(x, t) dw$, $x(t = 0) \sim \rho_0$, $x(t = 1) \sim \rho_1$

Mean Field Neural Network Learning Problem

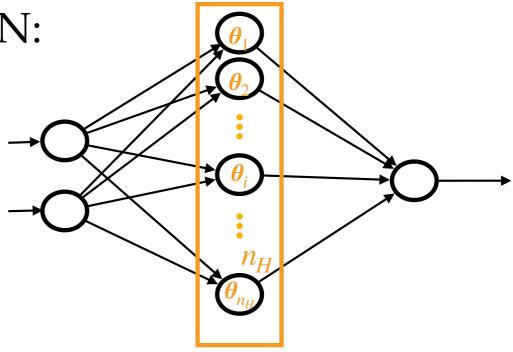
Infinite width limit of fully connected NN:

Mei, Montanari and Nguyen, *Proceedings of the National Academy of Sciences*, 2018

Chizat and Bach, NeurIPS, 2018

Rotskoff and Vanden-Eijnden, NeurIPS, 2018

Sirignano and Spiliopoulos, *Stochastic Processes* and their Applications, 2020



Mean field learning problem:

$$\inf_{
ho \in \mathcal{P}_2(\mathbb{R}^p)} R \left(\int \Phi(\boldsymbol{x}, \boldsymbol{ heta})
ho(\boldsymbol{ heta}) \mathrm{d} \boldsymbol{ heta} \right)$$
manifold of PDFs supported on \mathbb{R}^p with finite second moments

PDF dynamics:

$$\frac{\partial \rho}{\partial t} = -\nabla^W R \bigg(\int \Phi \rho \bigg) = \nabla \cdot \bigg(\rho \nabla \frac{\delta}{\delta \rho} R \bigg(\int \Phi \rho \bigg) \bigg)$$



. Mars Science Laborato

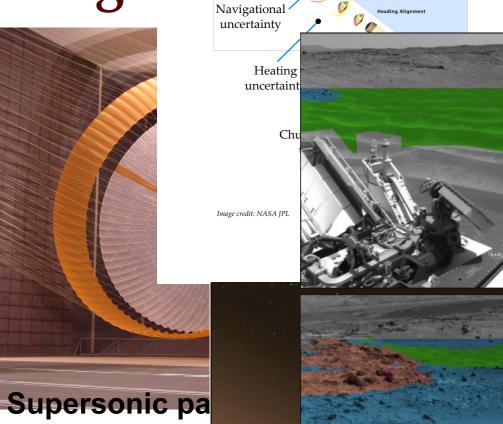
l-scale wind-tunnel testing

uch a large and capable r npelling site, which is rich

erve biomarkers, Presents

ever flown to Mars, fly th

generate a higher hyperso



GB parachute undersafely to a scientifically compelling site, which is rich in milieral street to a scientifically compelling site, which is rich in milieral street. and preserve biomarkers, presents a myriad of engineering 2 challenges. Not only is the payload mass significantly larger and preserve biomarkers, Presents a myriad of engineering than all previous Mars missions, the delivery accuracy and than all previous Mars missions, the delivery accuracy and Smooterral requirements are also nitrote stringent. In August of 1.0012, MSL will enter the Martian atmosphere will the largest aeroshell ever flown to Mars. fly the first guided lifting entry aeroshell ever flown to Mars, fly the first guided lifting entry ROCK Mars, generate a higher hypersonic lift-to-drag ratio than ROCK any previous Mars mission, and decelerate behind the largest supersonic parachute ever deployed at Mars. The MSL EDL System will also for the fNASA-TM-X-1451ftly 1

700 Pa [1]. However, Mach 2.1 i cessfully operating DBG parachute little flight test data above Mach 2 the amount of increased EDL sys the relevant flight tests and flight e the planned MSH payashue depice agree that higher Mach numbers r ity of failure, they have different o should be blaced. For example, Gil per bound of Mach 2 for parachute at Mars. However, Cruz [3] place range somewhere between two and Gale Crater This presents a challenge for ED Eberswalde Crater 23.86 S must then weigh the system perfor social devilarations de la la social de la la social de la social dela social de la social de la social de la social de la social dela social de la social de la social de la social de la social dela social de la social dela s altitudes and Mach numbers, again quantified, probability of parachu deploying a DGB at Mach 2.5 or 3 many sites were initially proposed

MAR

mgner Mach numbers result in mei of parachute structure, which can and (3) at Mach numbers above M

exhibit an instability, known as ar sult in multiple partial collapses

The chief concern with high Mach

parachute deployments in regions driving factor, is therefore, the in

The Viking parachute system was o Mach 1.4 and 2.1, and a dynamic

oscillations.

quirements Sare Tals 8-1150 fa SL will enter the Martian at

Parachut wire of the left are some of t iwidusa Maraning suswin ich redectured the glassification in Qui the trighte side the full classified panoramic mosaic is 1 to

Loose Sand

Bedrock

parachutatwas.

quantified, probability of parachlance and Small Recksntuitvely, we wish to avoid scale labisymeans btar Mach 2000 millet all bach do well the name of the post of proposing and selecting the part of the post of proposing and selecting the part of the post of the post of the part of t will often be identical in appearance to

tige whook read on exploit the place within inhall he had not been been a find the clear of the projective geometry

Learn surface feature from data

obshetainingarwe found that in practicus in practicus in practicus in the vergence. This was especially true given the small size this, we developed a pre-processing scheme for scale-nown

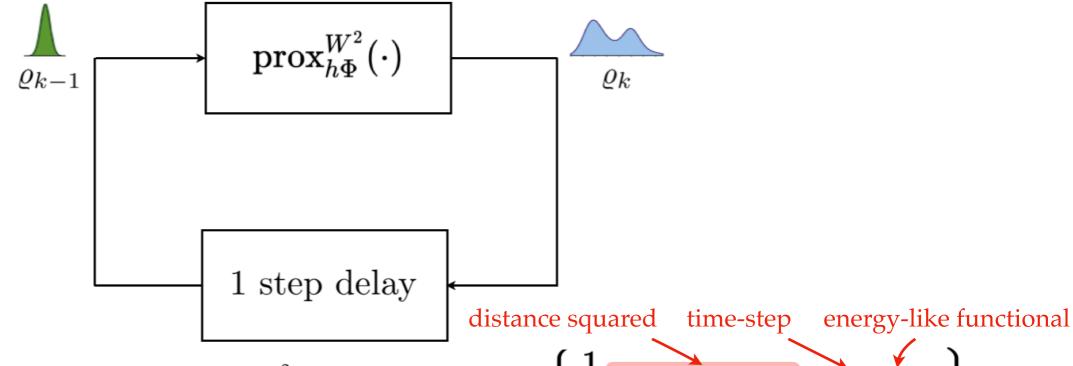
puse Marsing sion, and decelerate behind the largest is parachute ever deployed at Mars. The MSL EDL

Solving prediction problem as Wasserstein gradient flow

What's New?

Main idea: Solve
$$\frac{\partial \rho}{\partial t} = \mathcal{L}_{\mathrm{FP}} \rho, \; \rho(x,t=0) = \rho_0 \; \mathrm{as} \; \mathrm{gradient} \; \mathrm{flow} \; \mathrm{in} \; \mathcal{P}_2(\mathcal{X})$$

Infinite dimensional variational recursion:



$$\text{Proximal operator:} \ \ \varrho_k = \! \operatorname{prox}_{h\Phi}^{W^2}(\varrho_{k-1}) := \! \underset{\varrho \in \mathcal{P}_2(\mathcal{X})}{\operatorname{arg inf}} \bigg\{ \frac{1}{2} \underline{W^2(\varrho,\varrho_{k-1})} + \underline{h\Phi(\varrho)} \bigg\}$$

$$\textbf{Optimal transport cost:} \ W^2(\varrho,\varrho_{k-1}) := \inf_{\pi \in \Pi(\varrho,\varrho_{k-1})} \int_{\mathcal{X} \times \mathcal{X}} c(x,y) \ \mathrm{d}\pi(x,y)$$

Free energy functional:
$$\Phi(\varrho) := \int_{\mathcal{X}} \psi \varrho \, \mathrm{d}x + \beta^{-1} \int_{\mathcal{X}} \varrho \log \varrho \, \mathrm{d}x$$

Geometric Meaning of Gradient Flow

Gradient Flow in \mathcal{X}

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = -\nabla \varphi(\mathbf{x}), \quad \mathbf{x}(0) = \mathbf{x}_0$$

Recursion:

$$\begin{aligned} \mathbf{x}_{k} &= \mathbf{x}_{k-1} - h \nabla \varphi(\mathbf{x}_{k}) \\ &= \underset{\mathbf{x} \in \mathcal{X}}{\operatorname{arg min}} \left\{ \frac{1}{2} \|\mathbf{x} - \mathbf{x}_{k-1}\|_{2}^{2} + h \varphi(\mathbf{x}) \right\} \\ &= : \operatorname{prox}_{h\varphi}^{\|\cdot\|_{2}}(\mathbf{x}_{k-1}) \end{aligned}$$

Convergence:

$$\mathbf{x}_k \to \mathbf{x}(t = kh)$$
 as $h \downarrow 0$

φ as Lyapunov function:

$$rac{\mathrm{d}}{\mathrm{d}t}arphi = -\parallel
abla arphi \parallel_2^2 \ \le \ 0$$

Gradient Flow in $\mathcal{P}_2(\mathcal{X})$

$$\frac{\partial \rho}{\partial t} = -\nabla^W \Phi(\rho), \quad \rho(\mathbf{x}, 0) = \rho_0$$

Recursion:

$$\rho_{k} = \rho(\cdot, t = kh)$$

$$= \underset{\rho \in \mathcal{P}_{2}(\mathcal{X})}{\min} \left\{ \frac{1}{2} W^{2}(\rho, \rho_{k-1}) + h\Phi(\rho) \right\}$$

$$=: \underset{h\Phi}{\operatorname{prox}} \frac{W^{2}}{h^{\Phi}}(\rho_{k-1})$$

Convergence:

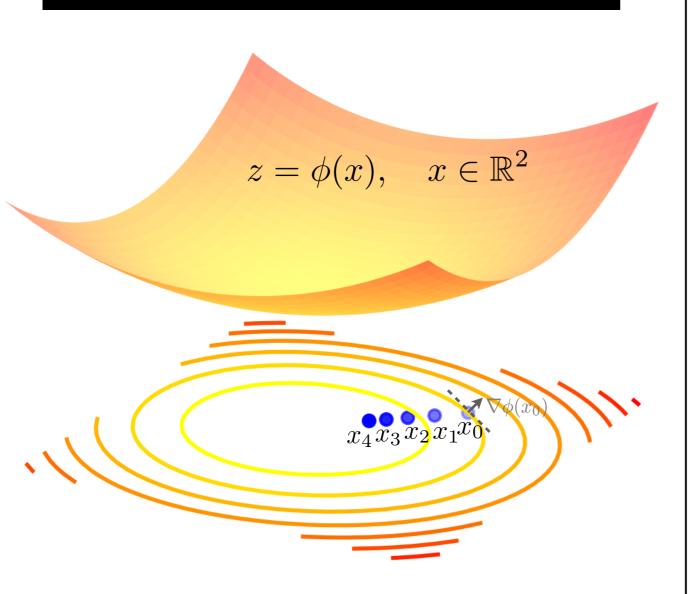
$$\rho_k \to \rho(\cdot, t = kh)$$
 as $h \downarrow 0$

Φ as Lyapunov functional:

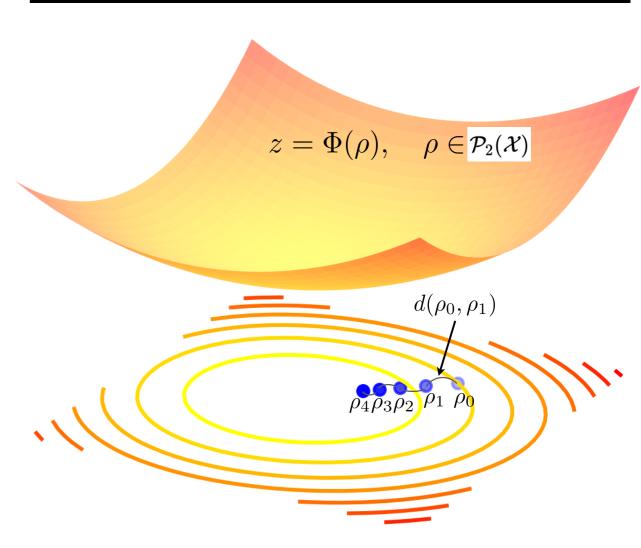
$$rac{\mathrm{d}}{\mathrm{d}t}\Phi = -\mathbb{E}_
hoigg[igg\|
ablarac{\delta\Phi}{\delta
ho}igg\|_2^2igg] \ \le \ 0$$

Geometric Meaning of Gradient Flow

Gradient Flow in ${\mathcal X}$

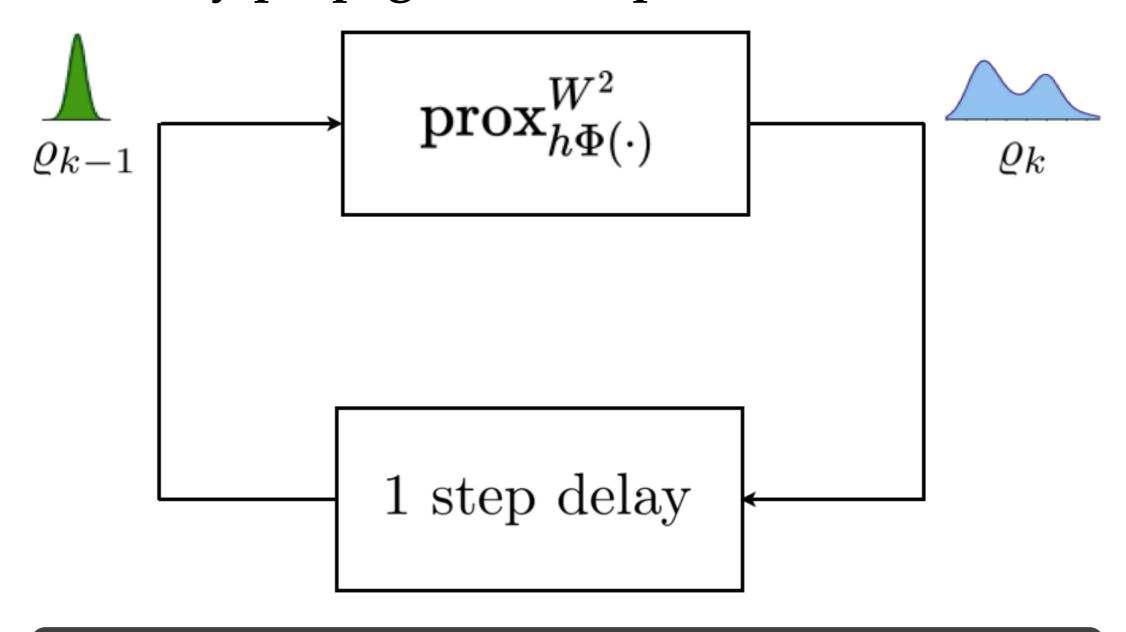


Gradient Flow in $\mathcal{P}_2(\mathcal{X})$



Algorithm: Gradient Ascent on the Dual Space

Uncertainty propagation via point clouds



No spatial discretization or function approximation

Algorithm: Gradient Ascent on the Dual Space

$$\frac{\partial \rho}{\partial t} = \nabla \cdot (\nabla \psi \rho) + \beta^{-1} \Delta \rho$$

$$\updownarrow \quad \text{Proximal Recursion}$$

$$\rho_k = \rho(\mathbf{x}, t = kh) = \underset{\rho \in \mathcal{P}_2(\mathbb{R}^n)}{\arg\inf} \left\{ \frac{1}{2} W^2(\rho, \rho_{k-1}) + h \Phi(\rho) \right\}$$

$$\Downarrow \quad \text{Discrete Primal Formulation}$$

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

Entropic Regularization

$$\varrho_{k} = \arg\min_{\varrho} \left\{ \min_{\mathbf{M} \in \Pi(\varrho_{k-1}, \varrho)} \frac{1}{2} \langle \mathbf{C}_{k}, \mathbf{M} \rangle + \epsilon H(\mathbf{M}) + h \langle \psi_{k-1} + \beta^{-1} \log \varrho, \varrho \rangle \right\} \quad \varrho \rangle$$

↑ Dualization

$$oldsymbol{\lambda}_0^{ ext{opt}}, oldsymbol{\lambda}_1^{ ext{opt}} = rg \max_{oldsymbol{\lambda}_0, oldsymbol{\lambda}_1 \geq 0} igg\{ \langle oldsymbol{\lambda}_0, oldsymbol{arrho}_{k-1}
angle - F^{\star}(-oldsymbol{\lambda}_1)$$

$$-\frac{\epsilon}{h} \left(\exp(\boldsymbol{\lambda}_0^\top h/\epsilon) \exp(-\boldsymbol{C}_k/2\epsilon) \exp(\boldsymbol{\lambda}_1 h/\epsilon) \right) \right\}$$

Recursion on the Cone

$$\mathbf{y} = e^{\frac{\lambda_0^*}{\epsilon}h} \qquad \qquad \mathbf{z} = e^{\frac{\lambda_1^*}{\epsilon}h}$$

Coupled Transcendental Equations in y and z

$$\Gamma_{k} = e^{\frac{-C_{k}}{2\epsilon}} \longrightarrow y \odot \Gamma_{k} z = \varrho_{k-1}$$

$$\varrho_{k-1} \longrightarrow \varrho_{k} = z \odot \Gamma_{k}^{\top} y$$

$$\xi_{k-1} = \frac{e^{-\beta \psi_{k-1}}}{e} \longrightarrow z \odot \Gamma_{k}^{\top} y = \xi_{k-1} \odot z^{-\beta \epsilon/2h}$$

Theorem: Consider the recursion on the cone $\mathbb{R}^n_{\geq 0} \times \mathbb{R}^n_{\geq 0}$

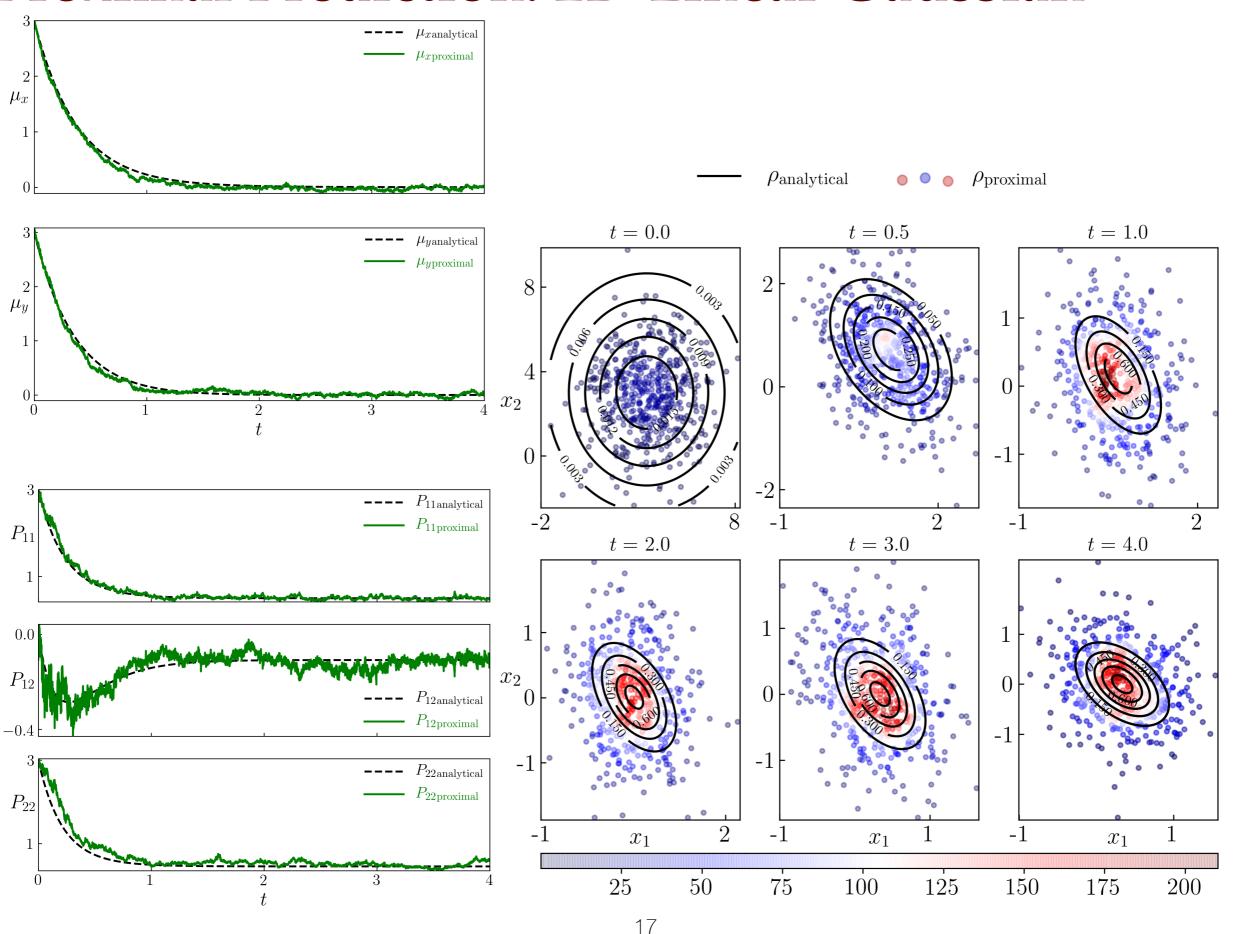
$$oldsymbol{y}\odot(oldsymbol{\Gamma}_koldsymbol{z})=oldsymbol{arrho}_{k-1},\quadoldsymbol{z}\odot\left(oldsymbol{\Gamma}_k^{\ op}oldsymbol{y}
ight)=oldsymbol{\xi}_{k-1}\odotoldsymbol{z}^{-rac{eta\epsilon}{h}},$$

Then the solution (\pmb{y}^*, \pmb{z}^*) gives the proximal update $\pmb{\varrho}_k = \pmb{z}^* \odot (\pmb{\Gamma}_k^{\ \ } \pmb{y}^*)$

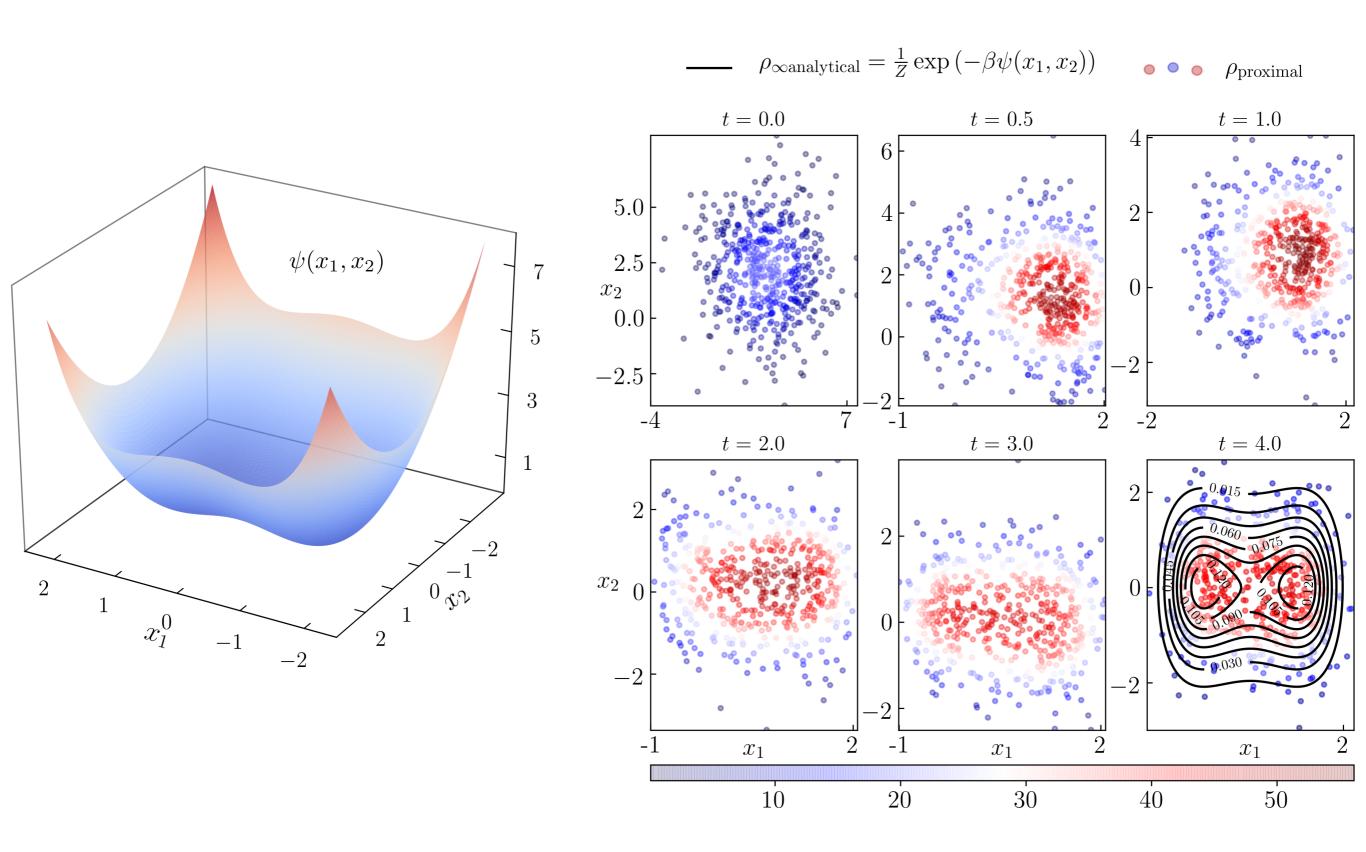
K.F. Caluya and A.H., Gradient flow algorithms for density propagation in stochastic systems, *IEEE TAC* 2019.

ptimization literatures the imagine given by $\begin{array}{c|c} x & z \\ \hline D_{KI} & d\pi_1 & d\pi_2 \\ \hline \end{array} = \begin{array}{c|c} x & z \\ \hline D_{KI} & d\pi_1 & d\pi_2 \\ \hline \end{array} = \begin{array}{c|c} y & d\pi_2 & d\pi_2 \\ \hline \end{array} = \begin{array}{c|c} y & d\pi_2 & d\pi_2 \\ \hline \end{array} = \begin{array}{c|c} y & d\pi_2 & d\pi_2 \\ \hline \end{array} = \begin{array}{c|c} y & d\pi_2 & d\pi_2 \\ \hline \end{array} = \begin{array}{c|c} y & d\pi_2 & d\pi_2 \\ \hline \end{array} = \begin{array}{c|c} y & d\pi_2 & d\pi_2 \\ \hline \end{array} = \begin{array}{c|c} y & d\pi_2 & d\pi_2 \\ \hline \end{array} = \begin{array}{c|c} y & d\pi_2 & d\pi_2 \\ \hline \end{array} = \begin{array}{c|c} y & d\pi_2 & d\pi_2 \\ \hline \end{array} = \begin{array}{c|c} y & d\pi_2 & d\pi_2 \\ \hline \end{array} = \begin{array}{c|c} y & d\pi_2 & d\pi_2 \\ \hline \end{array} = \begin{array}{c|c} y & d\pi_2 & d\pi_2 \\ \hline \end{array} = \begin{array}{c|c} y & d\pi_2 & d\pi_2 \\ \hline \end{array} = \begin{array}{c|c} y & d\pi_2 & d\pi_2 \\ \hline \end{array} = \begin{array}{c|c} y & d\pi_2 & d\pi_2 \\ \hline \end{array} = \begin{array}{c|c} y & d\pi_2 & d\pi_2 \\ \hline \end{array} = \begin{array}{c|c} y & d\pi_2 & d\pi_2 \\ \hline \end{array} = \begin{array}{c|c} y & d\pi_2 & d\pi_2 \\ \hline \end{array} = \begin{array}{c|c} y & d\pi_2 & d\pi_2 \\ \hline \end{array} = \begin{array}{c|c} y & d\pi_2 & d\pi_2 \\ \hline \end{array} = \begin{array}{c|c} y & d\pi_2 & d\pi_2 \\ \hline \end{array} = \begin{array}{c|c} y & d\pi_2 & d\pi_2 \\ \hline \end{array} = \begin{array}{c|c} y & d\pi_2 & d\pi_2 \\ \hline \end{array} = \begin{array}{c|c} y & d\pi_2 & d\pi_2 \\ \hline \end{array} = \begin{array}{c|c} y & d\pi_2 & d\pi_2 \\ \hline \end{array} = \begin{array}{c|c} y & d\pi_2 & d\pi_2 \\ \hline \end{array} = \begin{array}{c|c} y & d\pi_2 & d\pi_2 \\ \hline \end{array} = \begin{array}{c|c} y & d\pi_2 & d\pi_2 \\ \hline \end{array} = \begin{array}{c|c} y & d\pi_2 & d\pi_2 \\ \hline \end{array} = \begin{array}{c|c} y & d\pi_2 & d\pi_2 \\ \hline \end{array} = \begin{array}{c|c} y & d\pi_2 & d\pi_2 \\ \hline \end{array} = \begin{array}{c|c} y & d\pi_2 & d\pi_2 \\ \hline \end{array} = \begin{array}{c|c} y & d\pi_2 & d\pi_2 \\ \hline \end{array} = \begin{array}{c|c} y & d\pi_2 & d\pi_2 \\ \hline \end{array} = \begin{array}{c|c} y & d\pi_2 & d\pi_2 \\ \hline \end{array} = \begin{array}{c|c} y & d\pi_2 & d\pi_2 \\ \hline \end{array} = \begin{array}{c|c} y & d\pi_2 & d\pi_2 \\ \hline \end{array} = \begin{array}{c|c} y & d\pi_2 & d\pi_2 \\ \hline = \begin{array}{c|c} y & d\pi_2 & d\pi_2 \\ \hline = \begin{array}{c|c} y & d\pi_2 & d\pi_2 \\ \hline = \begin{array}{c|c} y & d\pi_2 & d\pi_2 \\ \hline = \begin{array}{c|c} y & d\pi_2 & d\pi_2 \\ \hline = \begin{array}{c|c} y & d\pi_2 & d\pi_2 \\ \hline = \begin{array}{c|c} y & d\pi_2 & d\pi_2 \\ \hline = \begin{array}{c|c} y & d\pi_2 & d\pi_2 \\ \hline = \begin{array}{c|c} y & d\pi_2 & d\pi_2 \\ \hline = \begin{array}{c|c} y & d\pi_2 & d\pi_2 \\ \hline = \begin{array}{c|c} y & d\pi_2 & d\pi_2 \\ \hline = \begin{array}{c|c} y & d\pi_2 & d\pi_2 \\ \hline = \begin{array}{c|c} y & d\pi_2 & d\pi_2 \\ \hline = \begin{array}{c|c} y & d\pi_2 & d\pi_2 \\ \hline = \begin{array}{c|c} y & d\pi_2 & d\pi_2 \\ \hline = \begin{array}{c|c} y & d\pi_2 & d\pi_2 \\ \hline = \begin{array}{c|c} y & d\pi_2 & d\pi_2 \\ \hline = \begin{array}{c|c} y & d\pi_2 & d\pi_2 \\ \hline = \begin{array}{c|c} y & d\pi_2 & d\pi_2 \\ \hline = \begin{array}{c|c} y & d\pi_2 & d\pi_2 \\ \hline = \begin{array}{c|c} y & d\pi_2 & d\pi_2 \\ \hline = \begin{array}{c|c} y & d\pi_2 & d\pi_2 \\ \hline = \begin{array}{c|c} y & d\pi_2 & d\pi_2 \\ \hline = \begin{array}{c|c} y & d\pi_2 & d\pi_2 \\ \hline = \begin{array}{c|c} y & d\pi_2 & d\pi_2 \\ \hline = \begin{array}{c|c} y & d\pi_2 & d\pi_2 \\ \hline = \begin{array}{c|c$ given by given by $x_{k-1} := \arg\min_{\substack{x \in \mathcal{X}, x \in \mathcal{X}$ the "proximal obsessabled theo," proximal presence of and the second second the insetially the triangle in neisated by the proximat netates by the proximal recursion Definition 2: The 2-Wasteritiem in the between the $\{x_k\}$ generate $x_{k} = \operatorname{prox}_{h\varphi}^{\|\cdot\|}(\boldsymbol{x}_{k-1}), \quad x_{k} = 0 \operatorname{prox}_{h\varphi}^{\|\cdot\|}(\boldsymbol{x}_{k-1}), \quad x_{k} = 0 \operatorname{prox}_{h\varphi}^{\|\cdot\|}(\boldsymbol{x}_{k-1}), \quad x_{k-1} = 0 \operatorname{prox}_{h\varphi}^{\|\cdot$ $x_k = \frac{\text{supported respectively roll } \lambda^*, y^* = \mathbb{R}^2$, is denoted the flow of the objective of the flow of the sequence (4) with π_1 , π_2 because (4) and π_3 because (4) and π_4 because (4) and (4) are sequence (4) and (4) because (4) and (4) and (4) because (4) and (4) and (4) because (4) and (4) and (4) are (4) and (4) and (4) are (4) and (4) and (4) are (4) and (4) are (4) are (4) and (4) are (4) are (4) are (4) and (4) are (4) are (4) are (4) and (4) are (4)converges to satisfies $x_k \rightarrow x(t)$ to the step-size t_0 to the step state t_0 to the state t_0 to t $\begin{array}{c|c}
\text{Infite dimensit} \\
(\varrho_{k}) d^{2} = \underset{\varrho \in \mathscr{D}_{2}}{\text{arg inf}} \frac{1}{2} d^{2}(\varrho, \varrho_{k-1}) \\
(\varrho_{k}) d^{2} = \underset{\varrho \in \mathscr{D}_{2}}{\text{arg inf}} \frac{1}{2} d^{2}(\varrho, \varrho_{k-1}) \\
\vdots \\
+ \varrho \in \mathscr{D}_{2}(\varrho, \varrho_{k-1}) \\
+ \varrho \in \mathscr{D}_{2}(\varrho, \varrho_{k-1}) \\
\vdots \\
W(\pi_{1}, \pi_{2}) :=
\end{array}$ $W(\pi_1,\pi_2) :=$ as an infinite dimensional proximal operator. As mentioned inite adimensional proximal operator. As mentioned above, the sequence $\{\varrho_k\}$ generated by the proximal replacement $\{\varrho_k\}$ generated by the proximal replacement $\{\varrho_k\}$ generated by the proximal replacement $\{\varrho_k\}$ generated by the proximal step the $\{\varrho_k\}$ generated by the proximal step the $\{\varrho_k\}$ where $\{\varrho_k\}$ denotes the converges to the flow of the PDF (4) i.e., the $\{\varrho_k\}$ as the step-size. 3) curving energy sctosthe flow satisfies $PDE_{ck}(x)$ is p(x,t) = kh, as the step-size where $II(\pi_1,\pi_2)$ denotes the constitutions p(x) where p(x) denotes the satisfies p(x) denotes p(x) denotes the satisfies p(x) denotes p(xe lalsou **Theorem:** Block co-ordinate iteration of (y, z) recur- $\frac{\mathrm{d}}{\mathrm{d}t}\varphi =$ +20 tolk sion is contractive on $\mathbb{R}^n_{>0} \times \mathbb{R}^n_{>0}$. pWhich trom the fact that the **generalized that the least of the particles and the particles and the fact that the generalized that the least of the particles and the fact that the generalized to the generalized to the manifold metric of apprepriately undersing Particles and the generalized to the manifold metric of apprepriately undersing Particles and the generalized to the manifold metric of apprepriately undersing Particles and the generalized to the manifold metric of apprepriately undersing Particles and the general particles and the generalized to the manifold metric of apprepriately undersing Particles and the general particles and the general particles are the general particles are the general particles.**

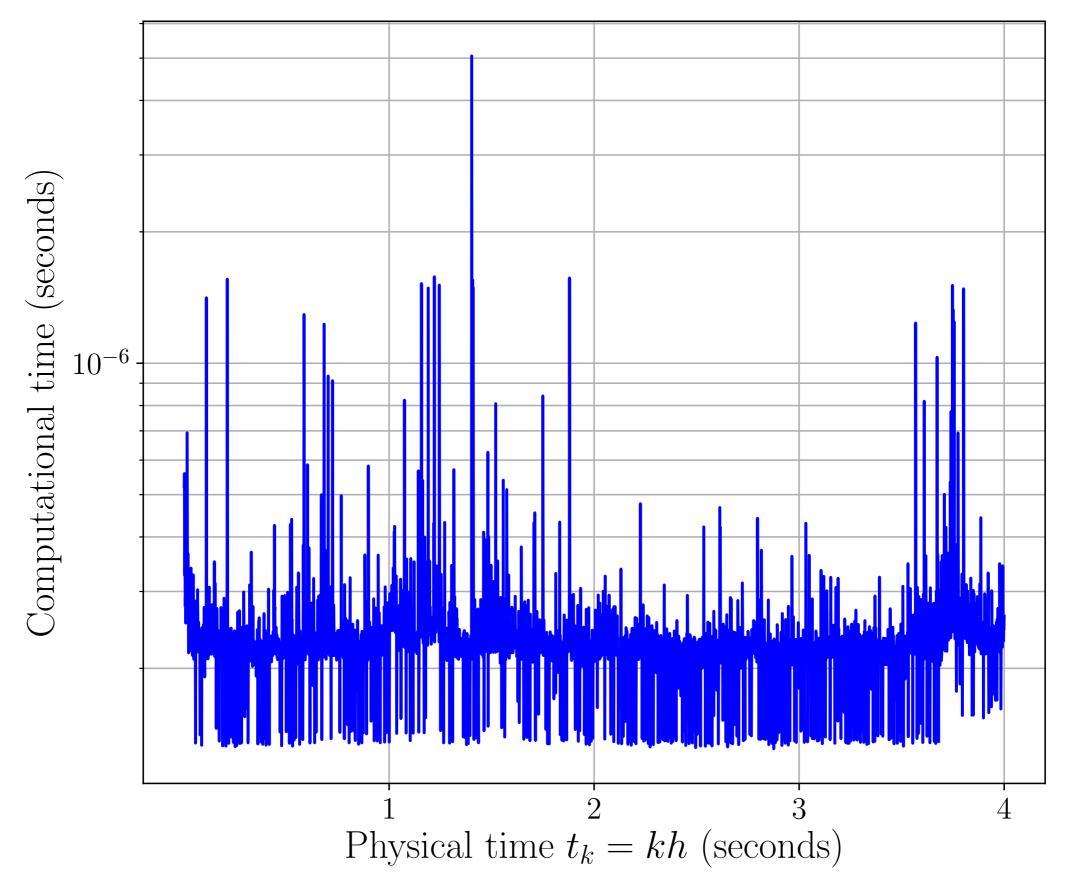
Proximal Prediction: 2D Linear Gaussian



Proximal Prediction: Nonlinear Non-Gaussian



Computational Time: Nonlinear Non-Gaussian



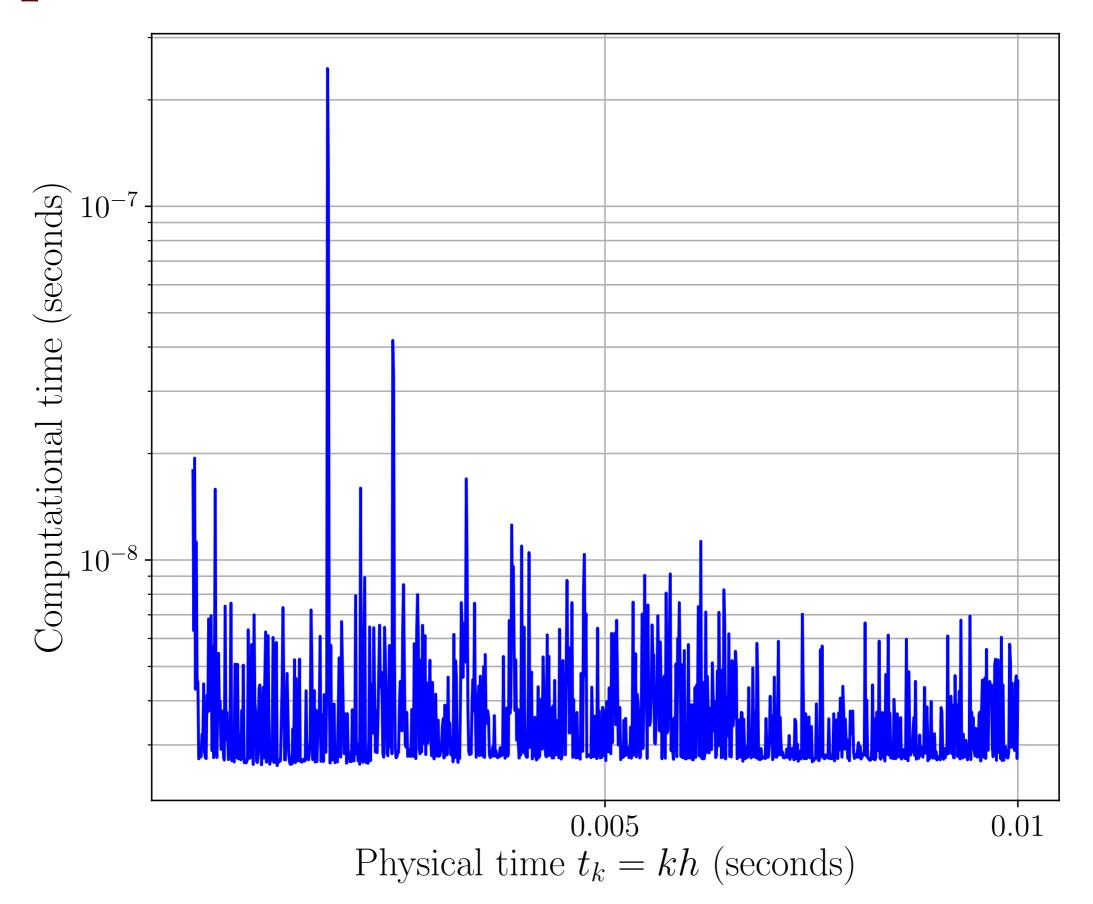
Proximal Prediction: Satellite in Geocentric Orbit

Here, $\mathcal{X} \equiv \mathbb{R}^6$

$$\begin{pmatrix} \mathrm{d}x \\ \mathrm{d}y \\ \mathrm{d}z \\ \mathrm{d}v_x \\ \mathrm{d}v_y \\ \mathrm{d}v_z \end{pmatrix} = \begin{pmatrix} v_x \\ v_y \\ -\frac{\mu x}{r^3} + (f_x)_{\mathsf{pert}} - \gamma v_x \\ -\frac{\mu y}{r^3} + (f_y)_{\mathsf{pert}} - \gamma v_y \\ -\frac{\mu z}{r^3} + (f_z)_{\mathsf{pert}} - \gamma v_z \end{pmatrix} dt + \sqrt{2\beta^{-1}\gamma} \begin{pmatrix} 0 \\ 0 \\ 0 \\ \mathrm{d}w_1 \\ \mathrm{d}w_2 \\ \mathrm{d}w_3 \end{pmatrix},$$

$$\begin{pmatrix} f_x \\ f_y \\ f_z \end{pmatrix}_{\mathsf{pert}} = \begin{pmatrix} s\theta \ c\phi & c\theta \ c\phi & -s\phi \\ s\theta \ s\phi & c\theta \ s\phi & c\phi \\ c\theta & -s\theta & 0 \end{pmatrix} \begin{pmatrix} \frac{k}{2r^4} \left(3(s\theta)^2 - 1\right) \\ -\frac{k}{r^5}s\theta \ c\theta \\ 0 \end{pmatrix}, k := 3J_2R_{\mathrm{E}}^2, \mu = \mathsf{constant}$$

Computational Time: Satellite in Geocentric Orbit



Extensions and Applications

Networked nonlinear power system dynamics with O(100) states

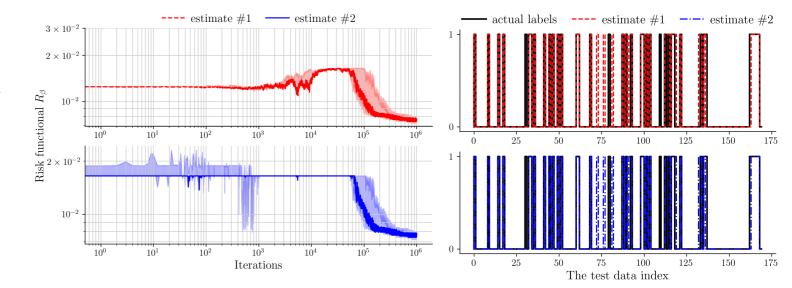
A.H., K.F. Caluya, P. Ojaghi, and X. Geng, Stochastic uncertainty propagation in power system dynamics with measure-valued proximal recursions, *IEEE Transactions on Power Systems*, 2022.

- Transmission Line #'s # - Bus #'s 19 10 11 11 11 11 12 13 14 15 16 17 4 18 19 10 10 11 10 10 11 10 10 11 10

Mean field learning in NN

A.M.H. Teter, I. Nodozi, and A.H., Proximal mean field learning in shallow neural networks, *arXiv*:2210:13879, 2022.

Case study: Wisconsin Breast Cancer Diagnostic (WBCD) Data Set



GPU: Jetson TX2 NVIDIA Pascal GPU 256 CUDA cores, 64 bit NVIDIA Denver + ARM Cortex A57 CPUs (≈ 2 hrs runtime)

Classification accuracy for the WBDC dataset		
β	Estimate #1	Estimate #2
0.03	91.17%	92.35%
0.05	92.94%	92.94%
0.07	78.23%	92.94%

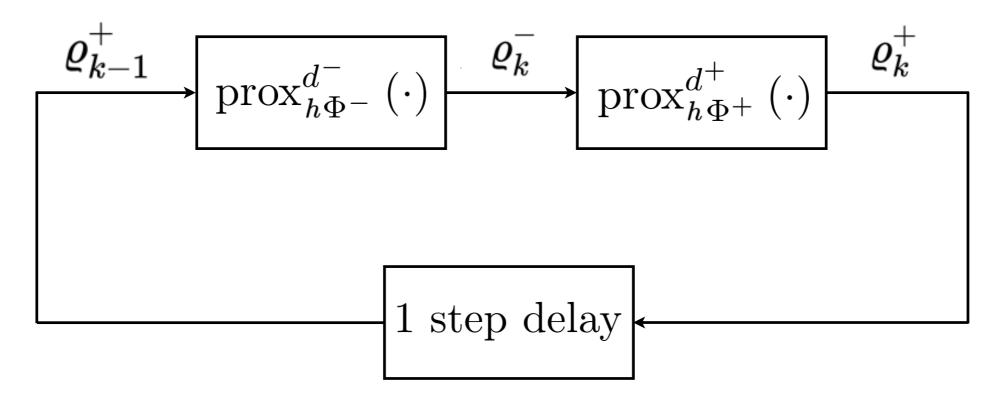
Solving filtering as generalized gradient flow

What's New?

Main idea: Solve the Kushner-Stratonovich SPDE

$$\mathrm{d}
ho^+ = ig[\mathcal{L}_{\mathrm{FP}}\mathrm{d}t + \mathcal{L}ig(\mathrm{d}z,\mathrm{d}t,
ho^+ig)ig]
ho^+, \;
ho(x,t=0) =
ho_0 ext{ as gradient flow in } \mathcal{P}_2(\mathcal{X})$$

Recursion of {deterministic o stochastic} proximal operators:



Convergence: $arrho_k^+(h) o
ho^+(x,t=kh)$ as $h\downarrow 0$

For prior, as before: $d^- \equiv W^2, \quad \Phi^- \equiv \ \mathbb{E}_{arrho} ig[\psi + eta^{-1} \log arrho ig]$

For posterior: $d^+ \equiv d_{ ext{FR}}^2 ext{ or } D_{ ext{KL}}, \quad \Phi^+ \underset{\scriptscriptstyle 24}{\equiv} \; rac{1}{2} \mathbb{E}_{arrho^+} \Big[(y_k - h(x))^ op R^{-1} (y_k - h(x)) \Big]$

Explicit Recovery of the Kalman-Bucy Filter

Model:

$$d\mathbf{x}(t) = \mathbf{A}\mathbf{x}(t)dt + \mathbf{B}d\mathbf{w}(t), \quad d\mathbf{w}(t) \sim \mathcal{N}(0, \mathbf{Q}dt)$$

$$d\mathbf{z}(t) = \mathbf{C}\mathbf{x}(t)dt + d\mathbf{v}(t), \qquad d\mathbf{v}(t) \sim \mathcal{N}(0, \mathbf{R}dt)$$

Given $\mathbf{x}(0) \sim \mathcal{N}(\mu_0, \mathbf{P}_0)$, want to recover:

$$\mathbf{P}^{+}\mathbf{C}\mathbf{R}^{-1}$$

$$\mathbf{d}\mu^{+}(t) = \mathbf{A}\mu^{+}(t)\mathbf{d}t + \mathbf{K}(t) \quad (\mathbf{d}\mathbf{z}(t) - \mathbf{C}\mu^{+}(t)\mathbf{d}t),$$

$$\dot{\mathbf{P}}^{+}(t) = \mathbf{A}\mathbf{P}^{+}(t) + \mathbf{P}^{+}(t)\mathbf{A}^{\top} + \mathbf{B}\mathbf{Q}\mathbf{B}^{\top} - \mathbf{K}(t)\mathbf{R}\mathbf{K}(t)^{\top}.$$

A.H. and T.T. Georgiou, Gradient Flows in Uncertainty Propagation and Filtering of Linear Gaussian Systems, CDC 2017.

A.H. and T.T. Georgiou, Gradient Flows in Filtering and Fisher-Rao Geometry, ACC 2018.

Explicit Recovery of the Wonham Filter

Model:

$$egin{aligned} x(t) \sim \operatorname{Markov}(Q), \ \operatorname{d}\!z(t) = h(x(t)) \operatorname{d}\!t \, + \, \sigma_v(t) \mathrm{d}v(t) \end{aligned}$$

State space: $\Omega := \{a_1, \ldots, a_m\}$

J.SIAM CONTROL Ser. A, Vol. 2, No. 3 Printed in U.S.A., 1965

SOME APPLICATIONS OF STOCHASTIC DIFFERENTIAL EQUATIONS TO OPTIMAL NONLINEAR FILTERING*

W. M. WONHAM†

Posterior $\pi^+(t) := \{\pi_1^+(t), \dots, \pi_m^+(t)\}$ solves the nonlinear SDE:

$$\mathrm{d}\pi^+(t) = \pi^+(t)Q\,\mathrm{d}t \;+\; rac{1}{\left(\sigma_v(t)
ight)^2}\pi^+(t)\Big(H-\widehat{h}(t)I\Big)\Big(\mathrm{d}z(t)-\widehat{h}(t)\mathrm{d}t\Big),$$

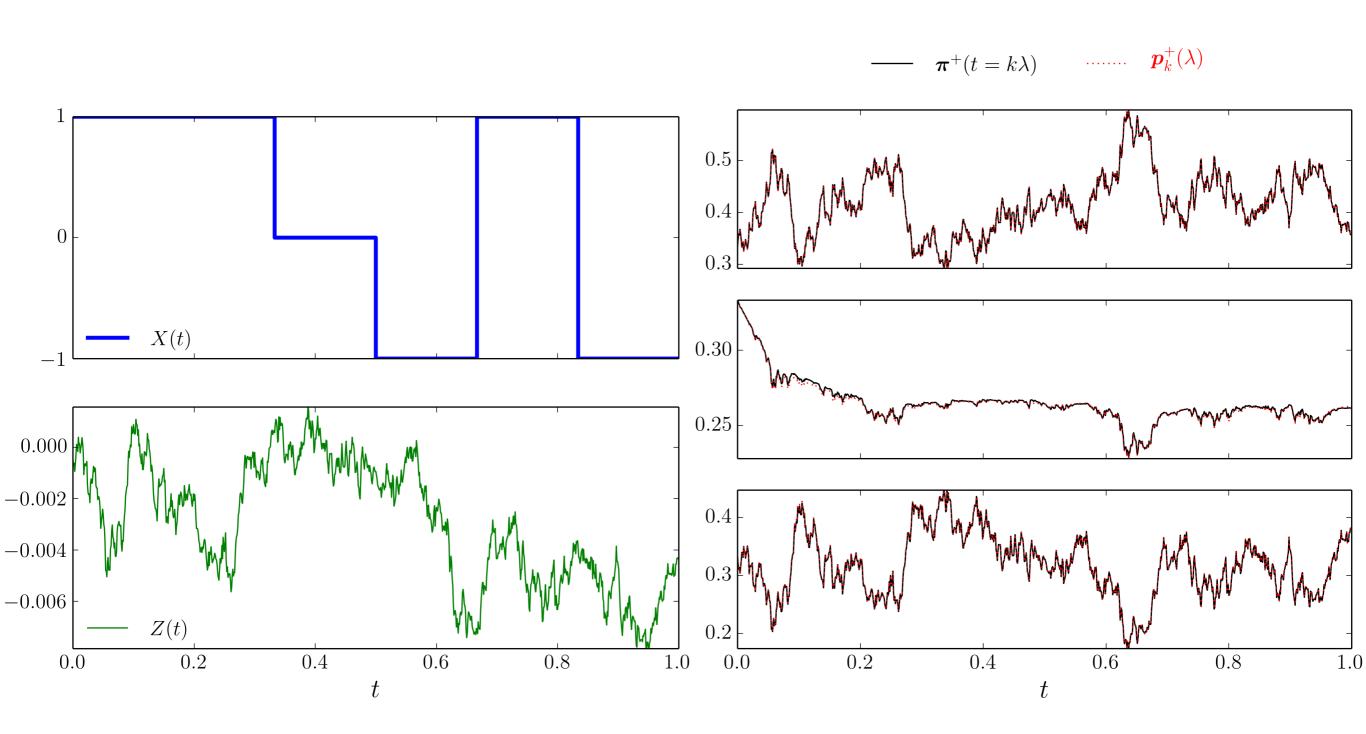
where
$$H := \operatorname{diag}(h(a_1), \ldots, h(a_m)), \quad \widehat{h}(t) := \sum_{i=1}^m h(a_i) \pi_i^+(t),$$

Initial condition: $\pi^+(t=0)=\pi_0,$

By defn.
$$\pi^+(t) = \mathbb{P}(x(t) = a_i \mid z(s), 0 \leq s \leq t)$$

A.H. and T.T. Georgiou, Proximal Recursion for the Wonham Filter, CDC 2019.

Numerical Results for the Wonham Filter

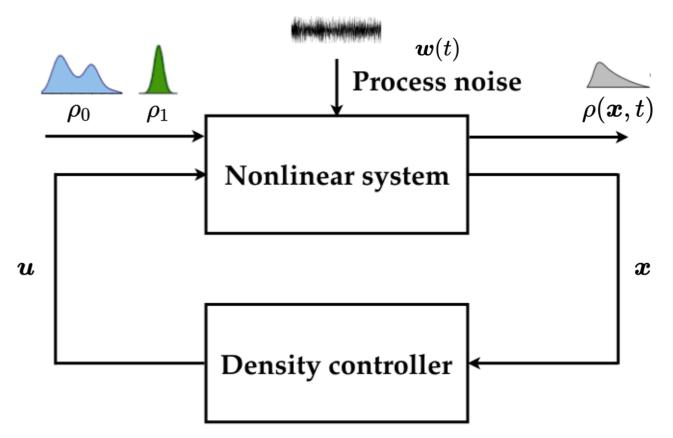


A.H. and T.T. Georgiou, Proximal Recursion for the Wonham Filter, CDC 2019.

Solving density control as generalized gradient flow

State Feedback Density Steering

Steer joint state PDF via feedback control over finite time horizon



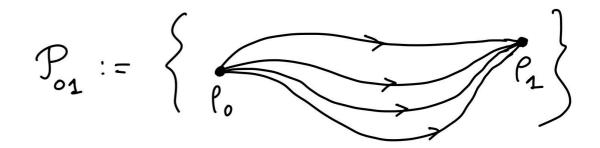
Common scenario: $G \equiv B$

minimize
$$\mathbb{E}\left[\int_0^1 \left(\frac{1}{2}\|\boldsymbol{u}(t,\boldsymbol{x}_t^{\boldsymbol{u}})\|_2^2 + q(t,\boldsymbol{x}_t^{\boldsymbol{u}})\right) \,\mathrm{d}t\right]$$
 subject to
$$\mathrm{d}\boldsymbol{x}_t^{\boldsymbol{u}} = \{f(t,\boldsymbol{x}_t^{\boldsymbol{u}}) + \boldsymbol{B}(t,\boldsymbol{x}_t^{\boldsymbol{u}})\boldsymbol{u}\}\mathrm{d}t + \sqrt{2}G(t,\boldsymbol{x}_t^{\boldsymbol{u}})\mathrm{d}\boldsymbol{w}_t$$

$$\boldsymbol{x}_0^{\boldsymbol{u}} := \boldsymbol{x}_t^{\boldsymbol{u}}(t=0) \sim \rho_0, \quad \boldsymbol{x}_1^{\boldsymbol{u}} := \boldsymbol{x}_t^{\boldsymbol{u}}(t=1) \sim \rho_1$$

Optimal Control Problem over PDFs

Diffusion tensor: $D := GG^{\top}$



Hessian operator w.r.t. state: Hess

$$\inf_{(\rho, \boldsymbol{u}) \in \mathcal{P}_{01} \times \mathcal{U}} \int_{\mathbb{R}^n} \int_0^1 \left(\frac{1}{2} \| \boldsymbol{u}(t, \boldsymbol{x}_t^{\boldsymbol{u}}) \|_2^2 + q(t, \boldsymbol{x}_t^{\boldsymbol{u}}) \right) \rho(t, \boldsymbol{x}_t^{\boldsymbol{u}}) \, \mathrm{d}t \, \mathrm{d}\boldsymbol{x}_t^{\boldsymbol{u}}$$
subject to
$$\frac{\partial \rho}{\partial t} + \nabla \cdot ((f + \boldsymbol{B}\boldsymbol{u}) \, \rho) = \langle \mathrm{Hess}, \boldsymbol{D} \rho \rangle$$

$$\rho(t = 0, \boldsymbol{x}_0^{\boldsymbol{u}}) = \rho_0, \quad \rho(t = 1, \boldsymbol{x}_1^{\boldsymbol{u}}) = \rho_1$$

Necessary Conditions of Optimality

(Assuming $G \equiv B$)

Coupled nonlinear PDEs + linear boundary conditions

Controlled Fokker-Planck or Kolmogorov's forward PDE

$$\frac{\partial \rho^{\text{opt}}}{\partial t} + \nabla \cdot ((f + D\nabla \psi) \rho^{\text{opt}}) = \langle \text{Hess}, D\rho \rangle$$

Hamilton-Jacobi-Bellman-like PDE

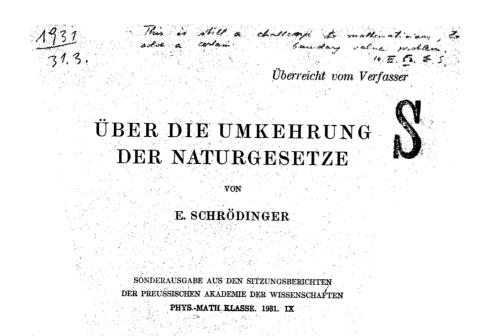
$$\frac{\partial \psi}{\partial t} + \langle \nabla \psi, f \rangle + \langle D, \text{Hess}(\psi) \rangle + \frac{1}{2} \langle \nabla \psi, D \nabla \psi \rangle = q$$

Boundary conditions:

$$\rho^{\text{opt}}(\cdot, t = 0) = \rho_0, \quad \rho^{\text{opt}}(\cdot, t = 1) = \rho_1$$

Optimal control:
$$u^{\mathrm{opt}} = B^{\top} \nabla \psi$$

Feedback Synthesis via the Schrödinger System



Sur la théorie relativiste de l'électron et l'interprétation de la mécanique quantique

E. SCHRÖDINGER

I — Introduction

J'ai l'intention d'exposer dans ces conférences diverses idées concernant la mécanique quantique et l'interprétation qu'on en donne généralement à l'heure actuelle; je parlerai principalement de la théorie quantique relativiste du mouvement de l'électron. Autant que nous pouvons nous en rendre compte aujourd'hui, il semble à peu près sûr que la mécanique quantique de l'électron, sous sa forme idéale, que nous ne possédons pas encore, doit former un jour la base de toute la physique. A cet intérêt tout à fait général, s'ajoute, ici à Paris, un intérêt particulier : vous savez tous que les bases de la théorie moderne de l'électron ont été posées à Paris par votre célèbre compatriote Louis de Broglie.



Hopf-Cole a.k.a. Fleming's logarithmic transform:

$$\left(
ho^{\mathrm{opt}},\psi
ight)\mapsto\left(\widehat{arphi},arphi
ight)$$
 — Schrödinger factors

$$\widehat{\varphi}(x,t) = \rho^{\mathrm{opt}}(x,t) \exp(-\psi(x,t))$$

$$\varphi(x,t) = \exp(\psi(x,t))$$
 for all $(x,t) \in \mathbb{R}^n \times [0,1]$

Feedback Synthesis via the Schrödinger System

2 coupled nonlinear PDEs → boundary-coupled linear PDEs!!

Uncontrolled forward-backward Kolmogorov PDEs:

$$\frac{\partial \widehat{\varphi}}{\partial t} = \boxed{-\nabla \cdot (\widehat{\varphi}f) + \langle \operatorname{Hess}, \mathbf{D}\widehat{\varphi} \rangle - q\widehat{\varphi}}, \quad \widehat{\varphi}_0 \varphi_0 = \rho_0,$$

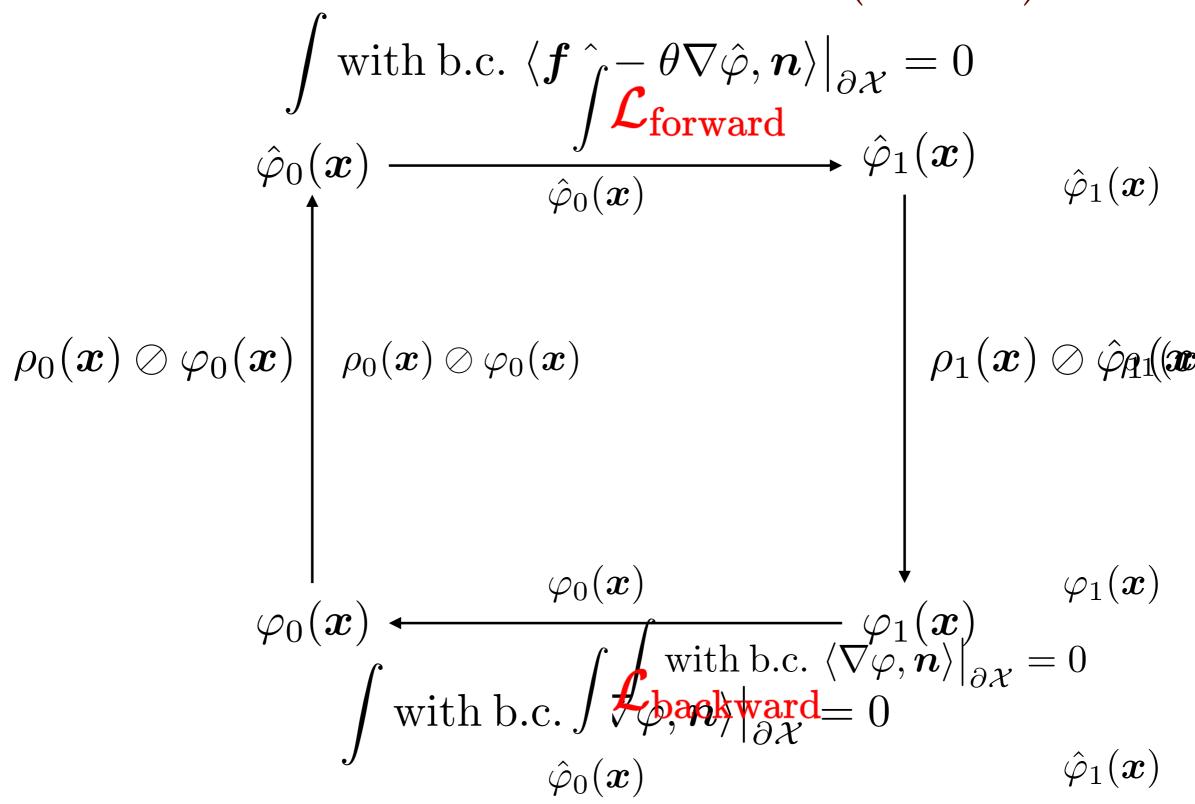
$$\frac{\partial \varphi}{\partial t} = \boxed{-\langle \nabla \varphi, f \rangle - \langle \operatorname{Hess}(\varphi), \mathbf{D} \rangle + q\varphi}, \quad \widehat{\varphi}_1 \varphi_1 = \rho_1,$$



Optimal controlled joint state PDF: $\rho^{\mathrm{opt}}(x,t) = \widehat{\varphi}(x,t) \varphi(x,t)$

Optimal control: $u^{\text{opt}}(x,t) = 2B^{\top} \nabla_x \log \varphi(x,t)$

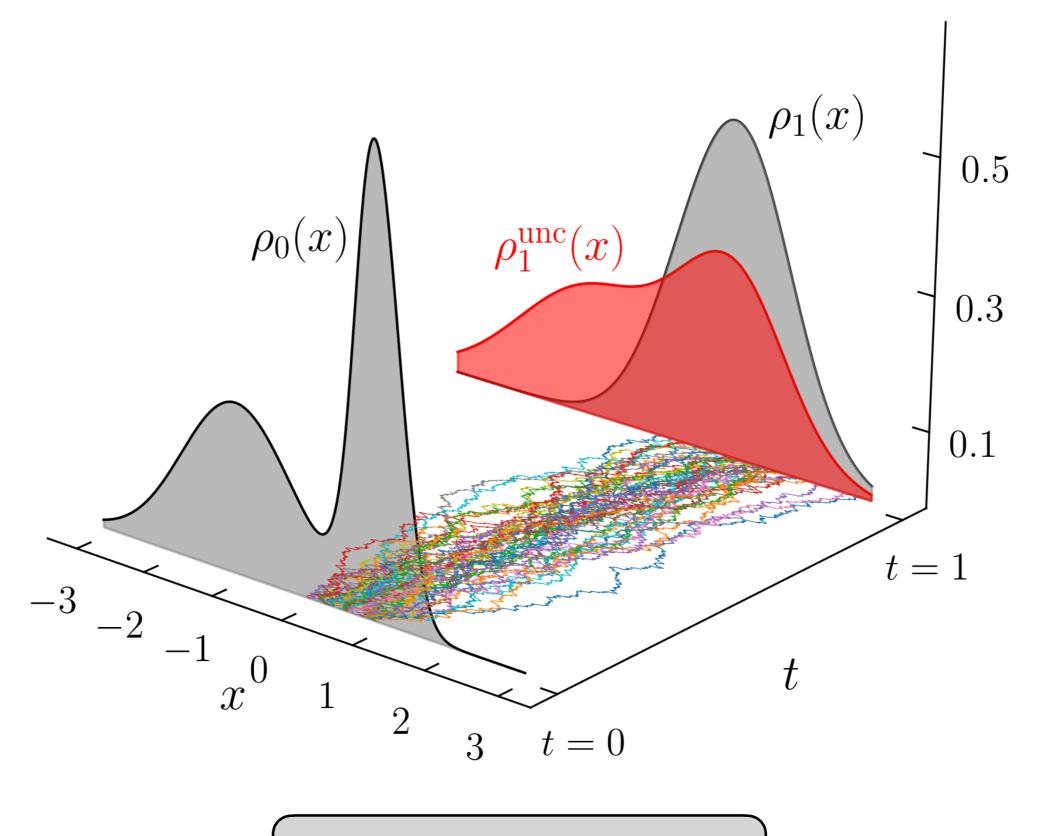
Fixed Point Recursion Over Pair $(\varphi_1, \hat{\varphi}_0)$



This recursion is contractive in the Hilbert metric!!

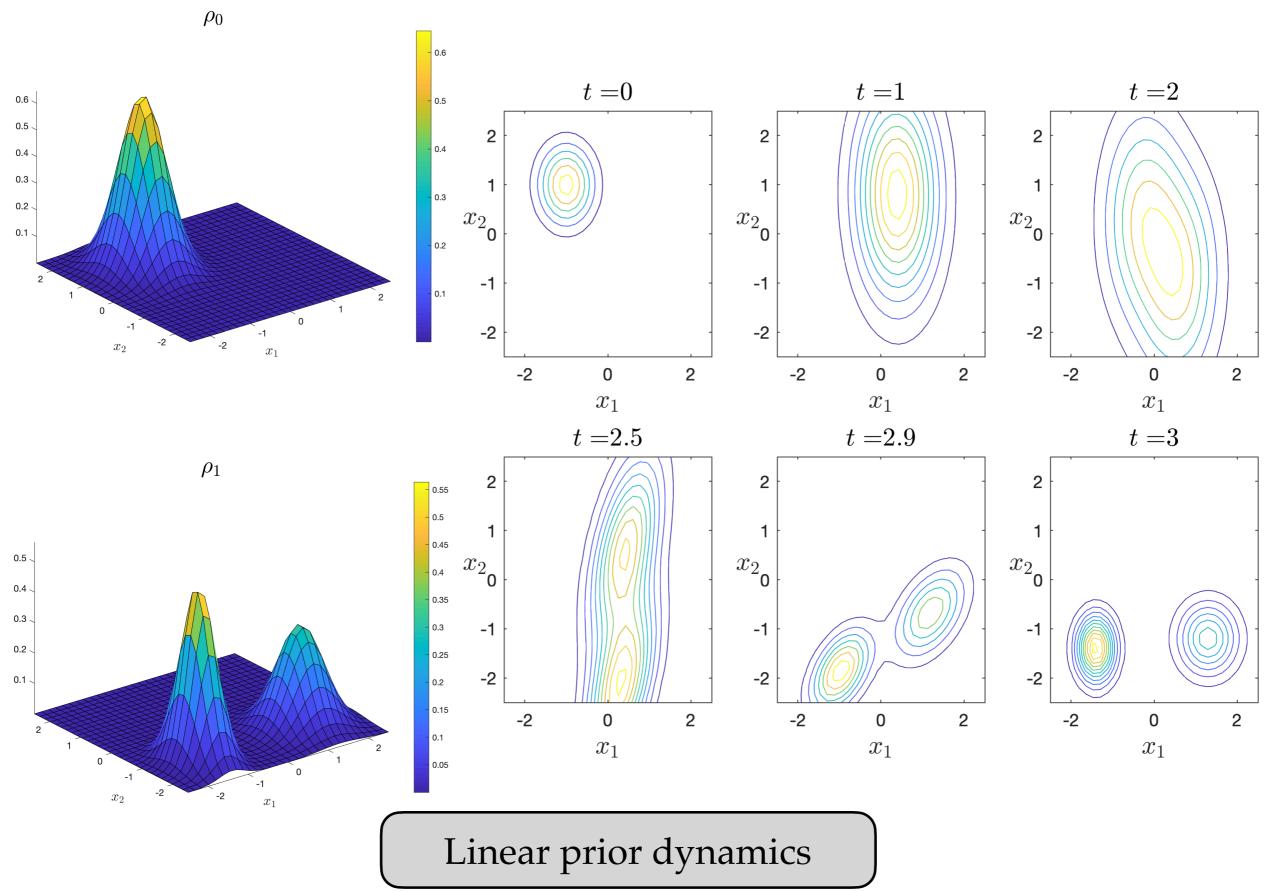
 $o_{1}(\mathbf{r}) \oslash o_{2}(\mathbf{r})$

Feedback Density Control: $f \equiv 0, B = G \equiv I, q \equiv 0$



Zero prior dynamics

Feedback Density Control: $f \equiv Ax, B = G, q \equiv 0$



In general ...

Need (uncontrolled) forward AND backward Kolmogorov solvers

Bad news: Need two different solvers

Good news: Sometimes one solver* suffices!!!

If not, use Feynman-Kac path integral for backward

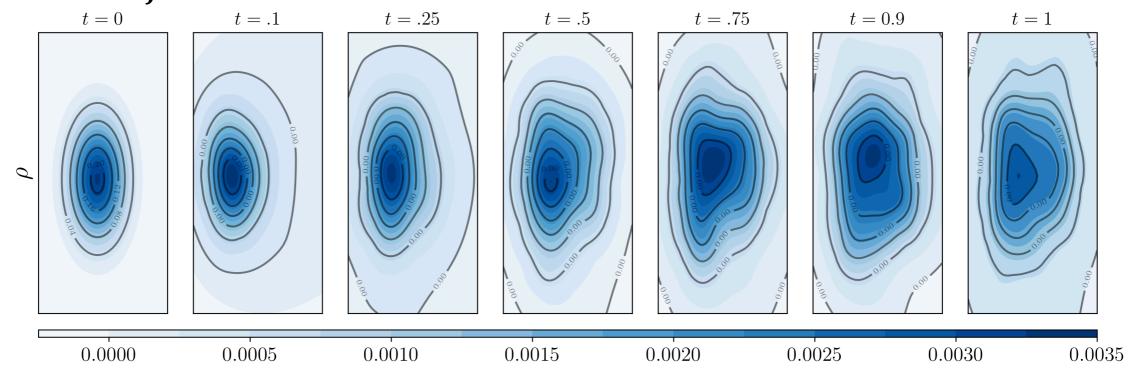
Even better: it is possible to design generalized gradient flow solvers based on point clouds!!

*Details:

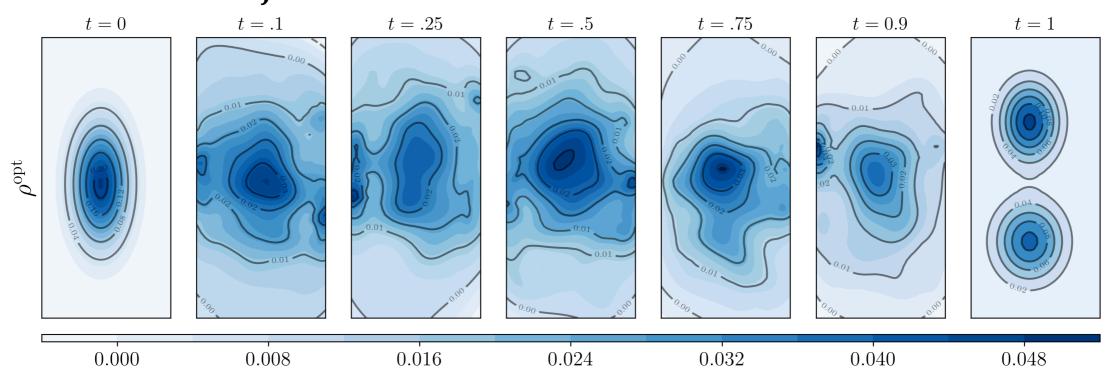
— K.F. Caluya, and A.H., Wasserstein Proximal Algorithms for the Schrödinger Bridge Problem: Density Control with Nonlinear Drift, *IEEE Trans. Automatic Control*, 2022.

Feedback Density Control: Nonlinear Grad. Drift

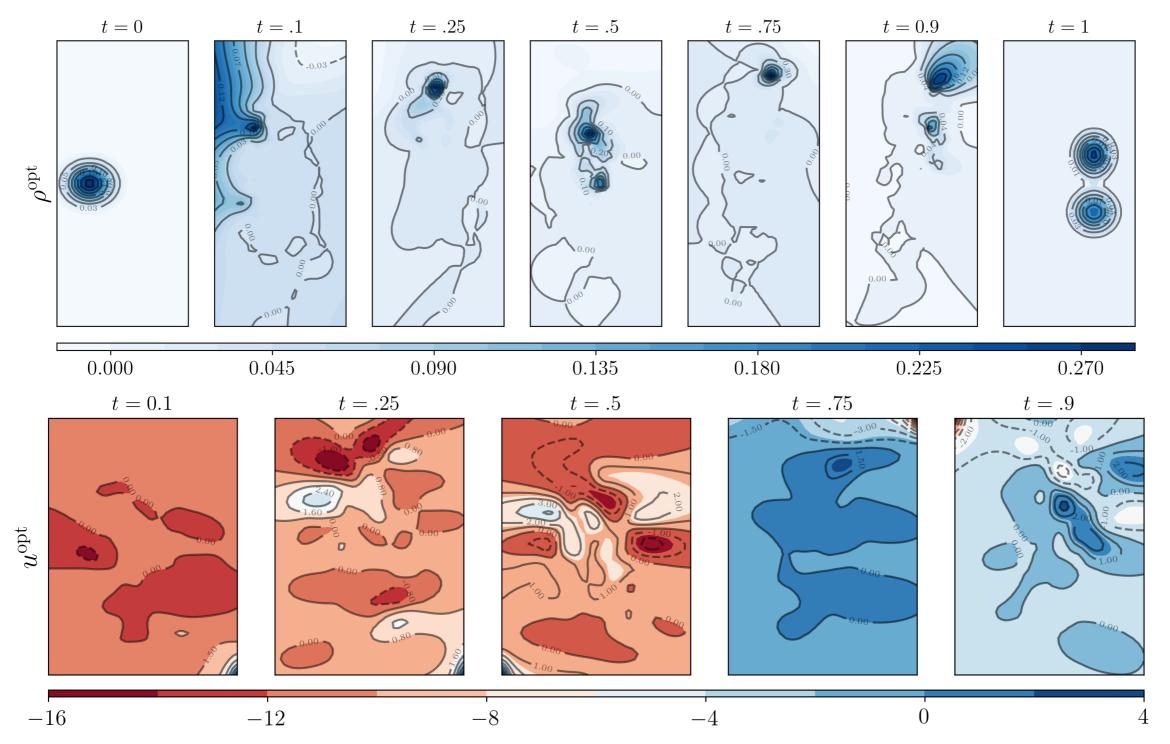
Uncontrolled joint PDF evolution:



Optimal controlled joint PDF evolution:

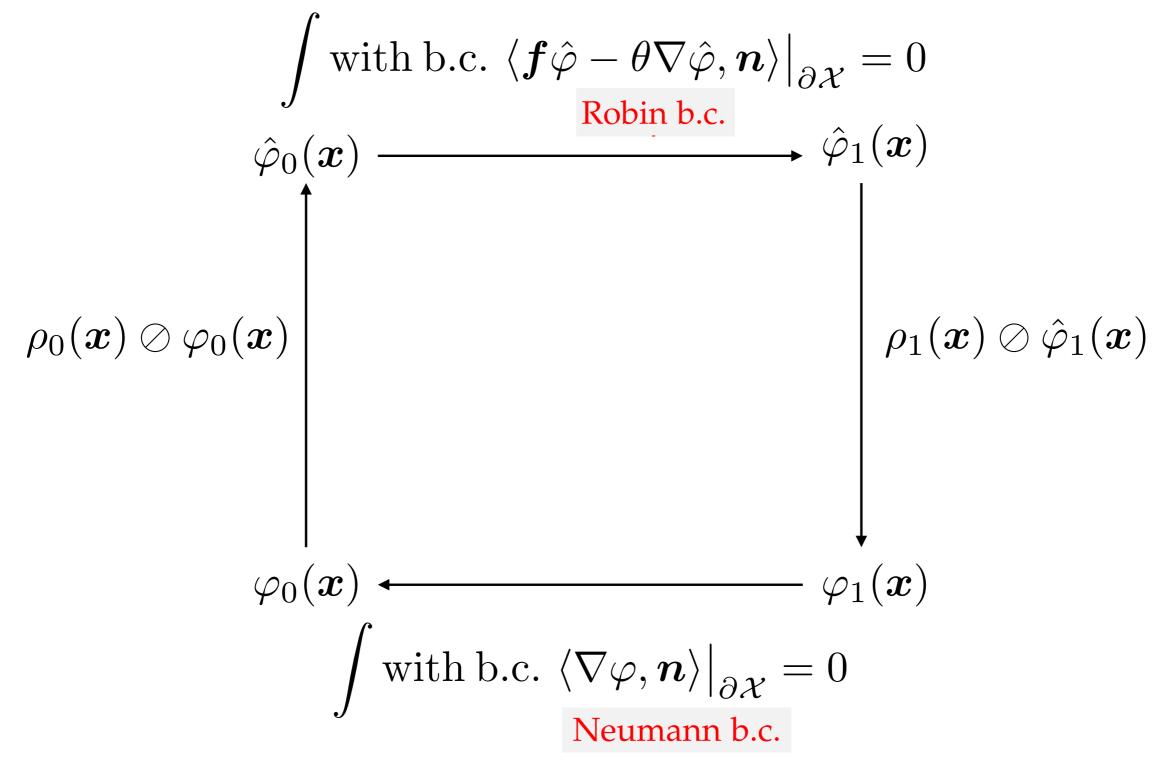


Feedback Density Control: Mixed Conservative-Dissipative Drift



K.F. Caluya and A.H., Wasserstein proximal algorithms for the Schrödinger bridge problem: density control with nonlinear drift, *IEEE TAC* 2021.

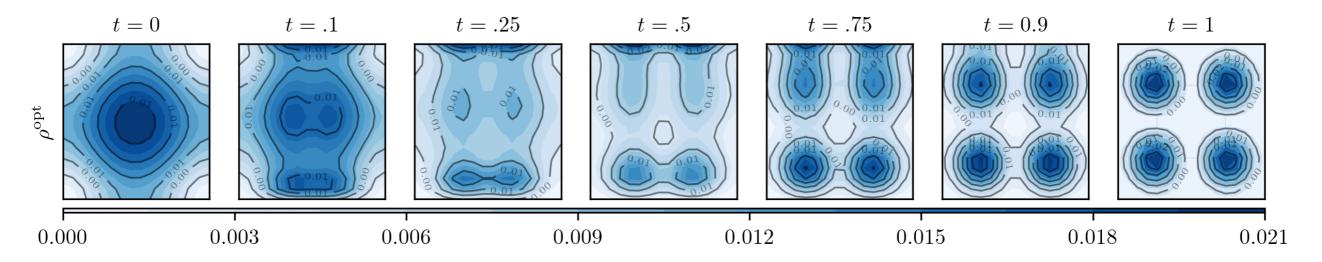
Nonlinear Density Steering with Deterministic Path Constraints



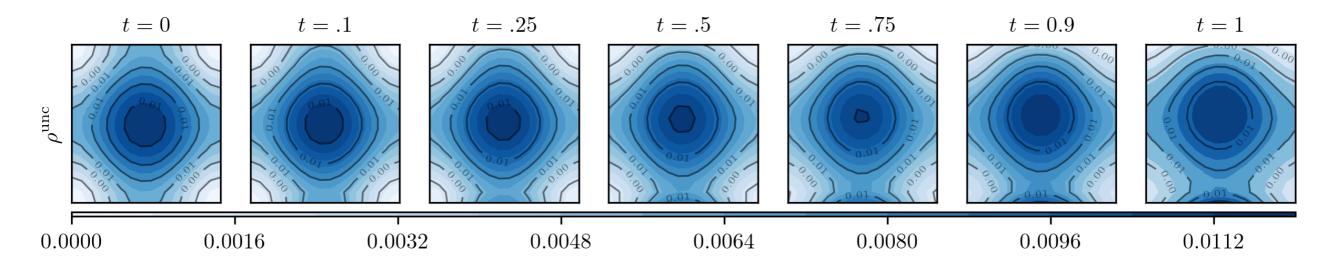
— K.F. Caluya, and A.H., Reflected Schrödinger Bridge: Density Control with Path Constraints, ACC 2021.

Nonlinear Density Steering with Deterministic Path Constraints

Optimal controlled state PDFs: $V(x_1, x_2) = (x_1^2 + x_2^3)/5$, $\overline{\mathcal{X}} = [-4, 4]^2$

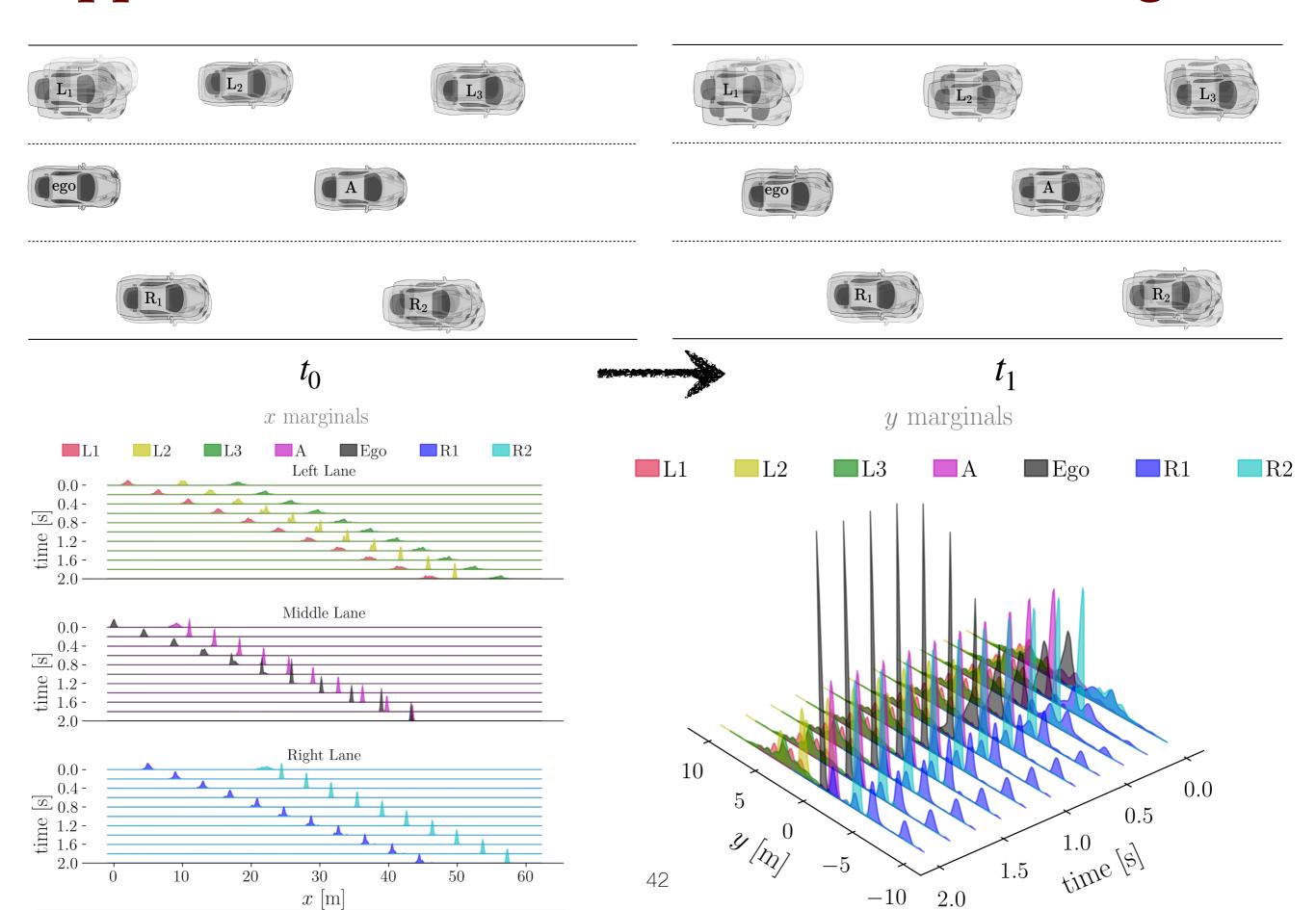


Uncontrolled state PDFs:

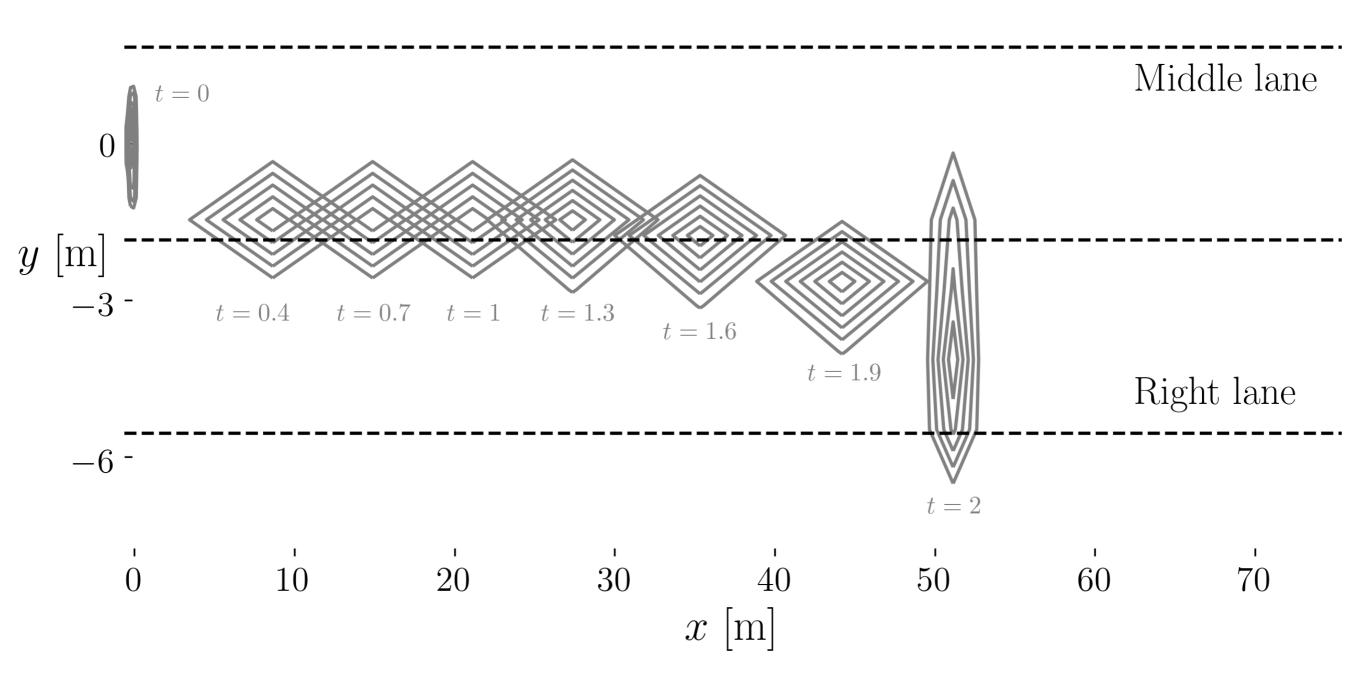


— K.F. Caluya, and A.H., Reflected Schrödinger Bridge: Density Control with Path Constraints, ACC 2021.

Application: Multi-lane Automated Driving

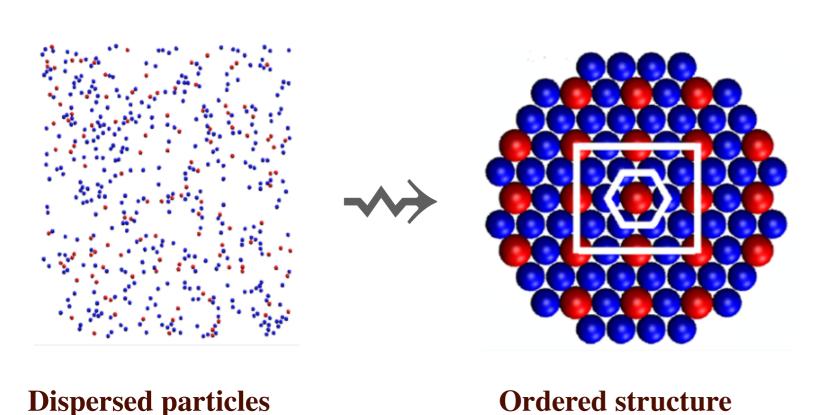






- S. Haddad, K.F. Caluya, A.H., and B. Singh, Prediction and Optimal Feedback Steering of Probability Density Functions for Safe Automated Driving, *IEEE Control Systems Letters*, 2020.
- S. Haddad, A.H., and B. Singh, Prediction and Optimal Feedback Steering of Probability Density Functions for Safe Automated Driving, *IEEE Trans. Control Systems Technology*, 2022.

Application: Controlled SA



Applications:

Precision (e.g., sub nm scale) manufacturing of materials with advanced electrical, magnetic or optical properties

Typical state variable: $\langle C_6 \rangle \in (0,6)$

Average number of hexagonally close packed neighboring particles in 2D assembly --> measure of crystallinity order

Typical control variable: U

Electric field voltage

Technical challenge:

Nonlinear + noisy molecular dynamics



 $\langle C_6 \rangle$ is a controlled stochastic process

Controlled SA as PDF Steering

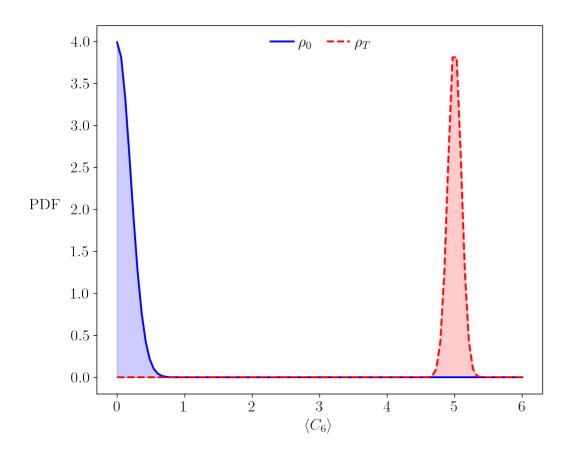
Intuition:

$$\langle C_6 \rangle \approx 0 \Leftrightarrow \text{Crystalline disorder}$$



 $\langle C_6 \rangle \approx 5 \Leftrightarrow \text{Crystalline order}$

Steer the PDF of the stochastic state $\langle C_6 \rangle$ from disordered at $t = t_0 \equiv 0$ to ordered at $t = T \equiv 200 \, \text{s}$



Typical prescribed finite horizon for controlled self-assembly

Endpoint PDF constraints: $\langle C_6 \rangle (t = t_0) \sim \rho_0$ (given)

$$\langle C_6 \rangle (t = T) \sim \rho_T \text{ (given)}$$

Control policy to accomplish the PDF steering:

$$u = \pi\left(\langle C_6 \rangle, t\right)$$

Underdetermined

Minimum Effort SA

$$\inf_{u \in \mathcal{U}} \quad \mathbb{E}_{\mu^u} \left[\int_0^T \frac{1}{2} u^2 \, \mathrm{d}t \right], \quad \mu^u \ll \mathrm{d}x^u$$

drift diffusion free energy inf $\lim_{u \in \mathcal{U}} \mathbb{E}_{\mu^u} \left[\int_0^T \frac{1}{2} u^2 \, \mathrm{d}t \right], \quad \mu^u \ll \mathrm{d}x^u$ $\lim_{u \in \mathcal{U}} \mathbb{E}_{\mu^u} \left[\int_0^T \frac{1}{2} u^2 \, \mathrm{d}t \right], \quad \mu^u \ll \mathrm{d}x^u$ $\lim_{u \in \mathcal{U}} \mathbb{E}_{\mu^u} \left[\int_0^T \frac{1}{2} u^2 \, \mathrm{d}t \right], \quad \mu^u \ll \mathrm{d}x^u$

either from model or learnt from MD simulation data

subject to
$$dx^u = D_1(x^u, u) dt + \sqrt{2D_2(x^u, u)} dw$$
,
$$\langle C_6 \rangle$$
 standard Wiener process
$$x^u(t=0) \sim d\mu_0 = \rho_0 dx^u, \quad x^u(t=T) \sim d\mu_T = \rho_T dx^u$$

Nonlinear in state, non-affine in control

Conditions for Optimality

$$\frac{\partial \psi}{\partial t} = \frac{1}{2} \left(\pi^{\text{opt}} \right)^2 - \frac{\partial \psi}{\partial x} D_1 - \frac{\partial^2 \psi}{\partial x^{u^2}} D_2$$

$$\frac{\partial \rho^{u}}{\partial t} = -\frac{\partial}{\partial x^{u}} \left(D_{1} \rho^{u} \right) + \frac{\partial^{2}}{\partial x^{u2}} \left(D_{2} \rho^{u} \right)$$

$$\pi^{\text{opt}}(x^u, t) = \frac{\partial \psi}{\partial x^u} \frac{\partial D_1}{\partial u} + \frac{\partial^2 \psi}{\partial x^{u2}} \frac{\partial D_2}{\partial u}$$

$$\rho^{u}(x^{u}, t = 0) = \rho_{0}, \quad \rho^{u}(x^{u}, t = T) = \rho_{T}$$

HJB PDE

Controlled FPK PDE

Optimal policy

Boundary conditions

value optimally optimal function controlled PDF policy

to be solved for the triple:
$$\psi(x^u, t)$$
, $\rho^u(x^u, t)$, $\pi^{\text{opt}}(x^u, t)$

Solve via PINN: Losses for Training

Loss term for HJB PDE

$$\mathcal{L}_{\psi} = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{\partial \psi}{\partial t} \bigg|_{x_i} - \frac{1}{2} (\pi^{\text{opt}})^2 \bigg|_{x_i^u} - + \frac{\partial \psi}{\partial x^u} D_1 \bigg|_{x_i^u} - + \frac{\partial^2 \psi}{\partial x^{u2}} D_2 \bigg|_{x_i^u} \right)^2$$

Loss term for FPK PDE

$$\mathcal{L}_{\rho^{u}} = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{\partial \rho^{u}}{\partial t} \bigg|_{\mathbf{x}_{i}^{u}} + \frac{\partial}{\partial x^{u}} \left(D_{1} \rho^{u} \right) \bigg|_{\mathbf{x}_{i}^{u}} - \frac{\partial^{2}}{\partial x^{u2}} \left(D_{2} \rho^{u} \right) \bigg|_{\mathbf{x}_{i}^{u}} \right)^{2}$$

Loss term for policy equation

$$\mathcal{L}_{\pi^{\text{opt}}} = \frac{1}{n} \sum_{i=1}^{n} \left(\pi^{\text{opt}} \Big|_{x_i^u} - \frac{\partial \psi}{\partial x^u} \frac{\partial D_1}{\partial u} \Big|_{x_i^u} - \frac{\partial^2 \psi}{\partial x^{u2}} \frac{\partial D_2}{\partial u} \Big|_{x_i^u} \right)^2$$

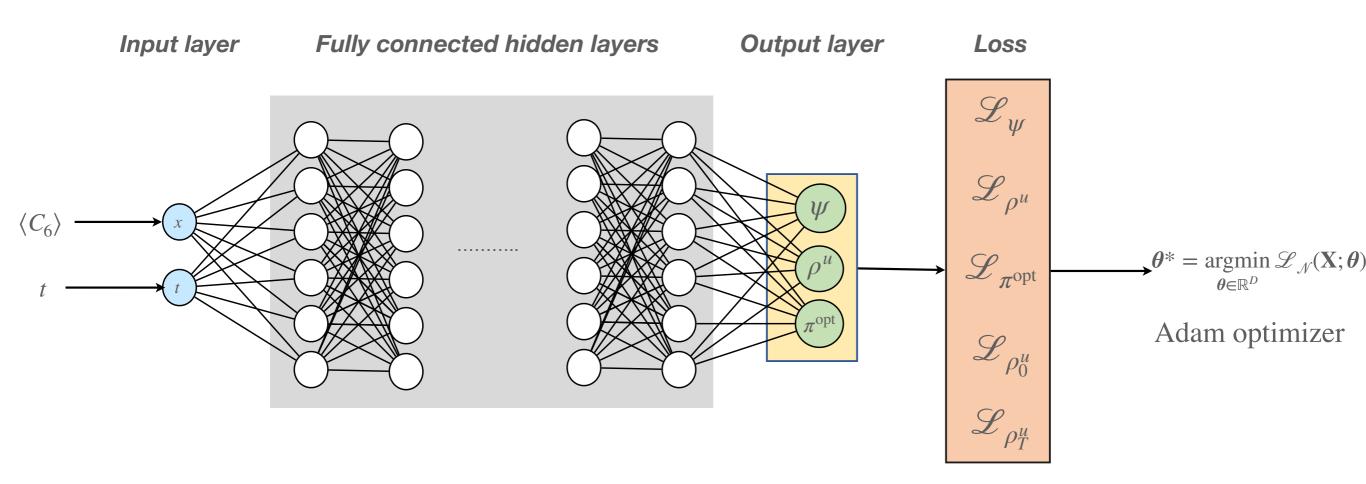
Loss term for initial condition

$$\mathcal{L}_{\rho_0^u} = \frac{1}{n} \sum_{i=1}^n \left(\rho^u \Big|_{t=0} - \rho_0^u(x) \right)^2$$

Loss term for terminal condition

$$\mathcal{L}_{\rho_T^u} = \frac{1}{n} \sum_{i=1}^n \left(\rho^u \Big|_{t=T} - \rho_T^u(x) \right)^2$$

PINN Architecture

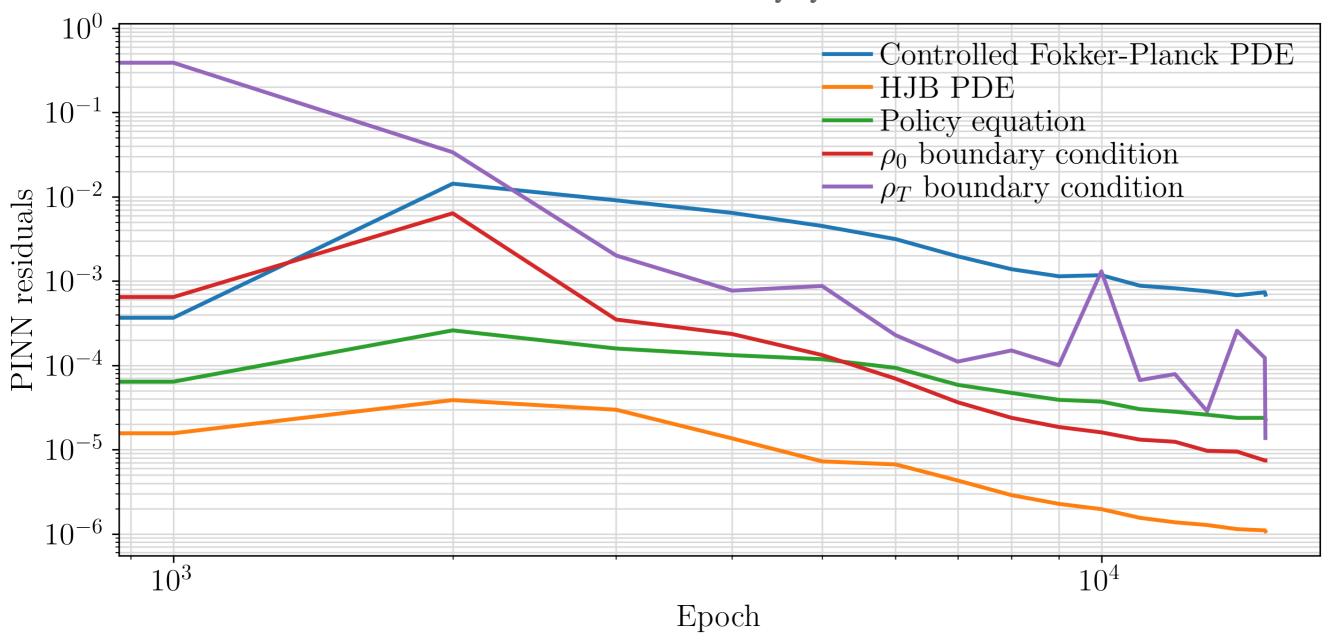


$$\mathcal{L}_{\mathcal{N}} = \mathcal{L}_{\psi} + \mathcal{L}_{\rho^{u}} + \mathcal{L}_{\pi^{\mathrm{opt}}} + \mathcal{L}_{\rho^{u}_{0}} + \mathcal{L}_{\rho^{u}_{T}}$$

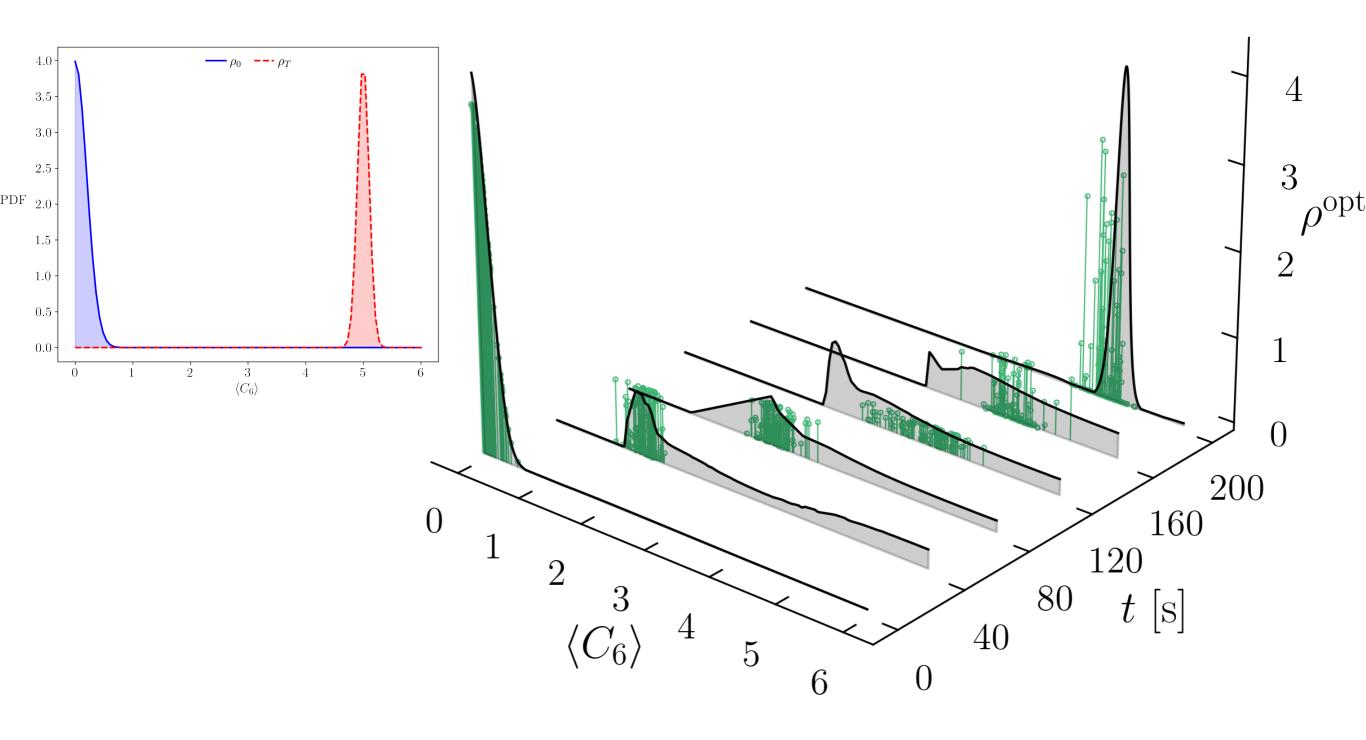
[Lu Lu, et al, 2021] [Niaki, et al, 2021]

Training of the PINN

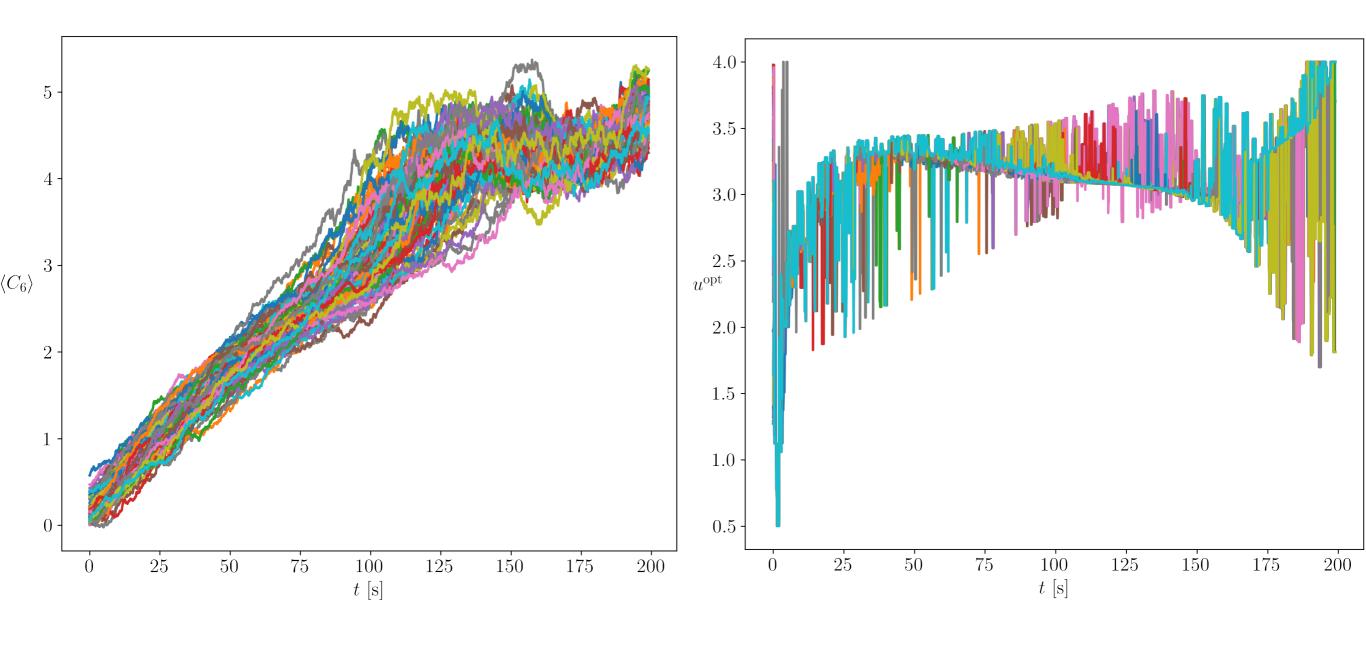
Benchmark controlled self-assembly system: [Y Xue, et al, IEEE Trans. Control Sys. Technology, 2014]



Optimally Controlled State PDFs



Optimal State and Optimal Control Sample Paths



GSBP Conditions for Optimality with m Inputs

m + 2 coupled PDEs with endpoint boundary conditions:

$$\frac{\partial \psi}{\partial t} = \frac{1}{2} \|\boldsymbol{u}_{\mathrm{opt}}\|_{2}^{2} - \langle \nabla_{\boldsymbol{x}} \psi, \boldsymbol{f} \rangle - \langle \boldsymbol{G}, \mathbf{Hess}(\psi) \rangle,$$

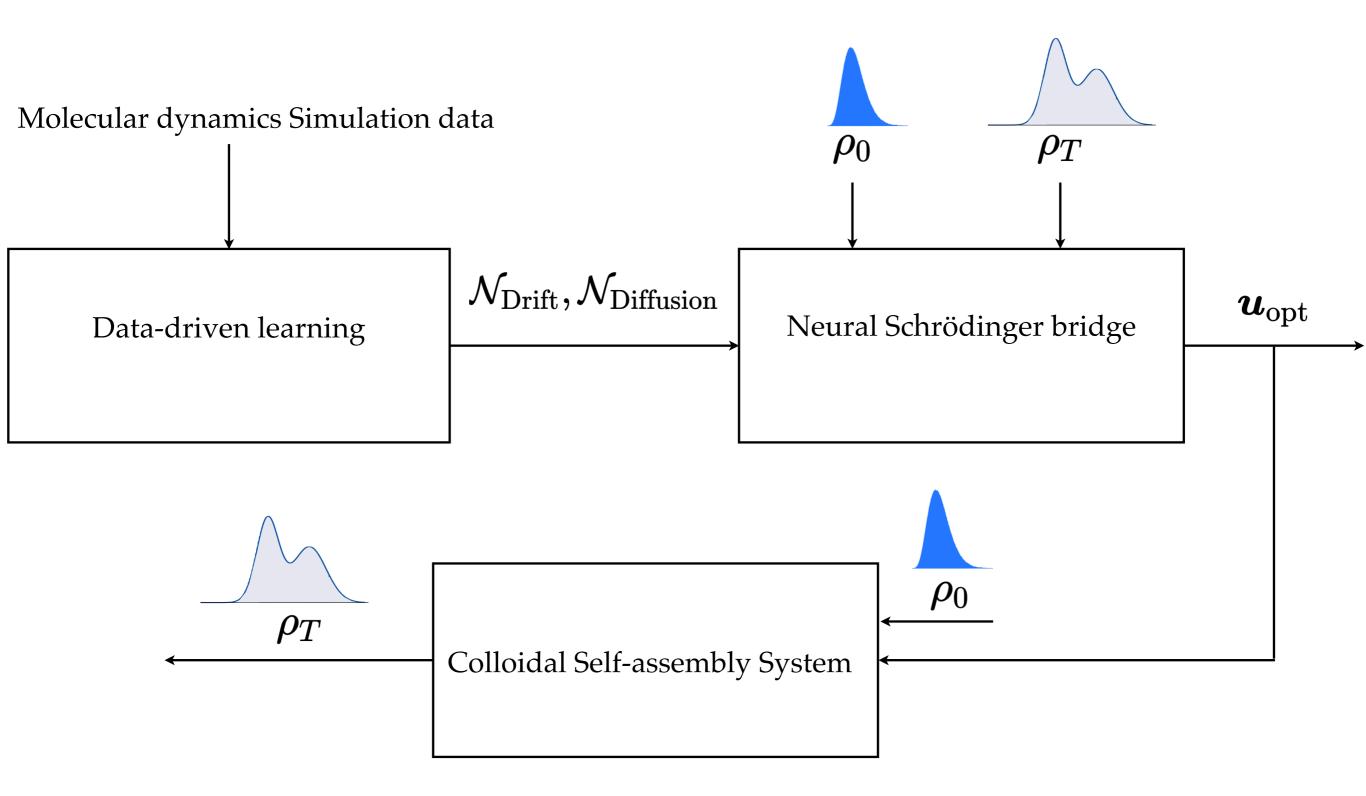
$$\frac{\partial \rho_{\mathrm{opt}}^{\boldsymbol{u}}}{\partial t} = -\nabla \cdot (\rho_{\mathrm{opt}}^{\boldsymbol{u}} \boldsymbol{f}) + \langle \boldsymbol{G}, \mathbf{Hess}(\rho_{\mathrm{opt}}^{\boldsymbol{u}}) \rangle,$$

$$\boldsymbol{u}_{\mathrm{opt}} = \nabla_{\boldsymbol{u}_{\mathrm{opt}}} \left(\langle \nabla_{\boldsymbol{x}} \psi, \boldsymbol{f} \rangle + \langle \boldsymbol{G}, \mathbf{Hess}(\phi) \rangle \right),$$

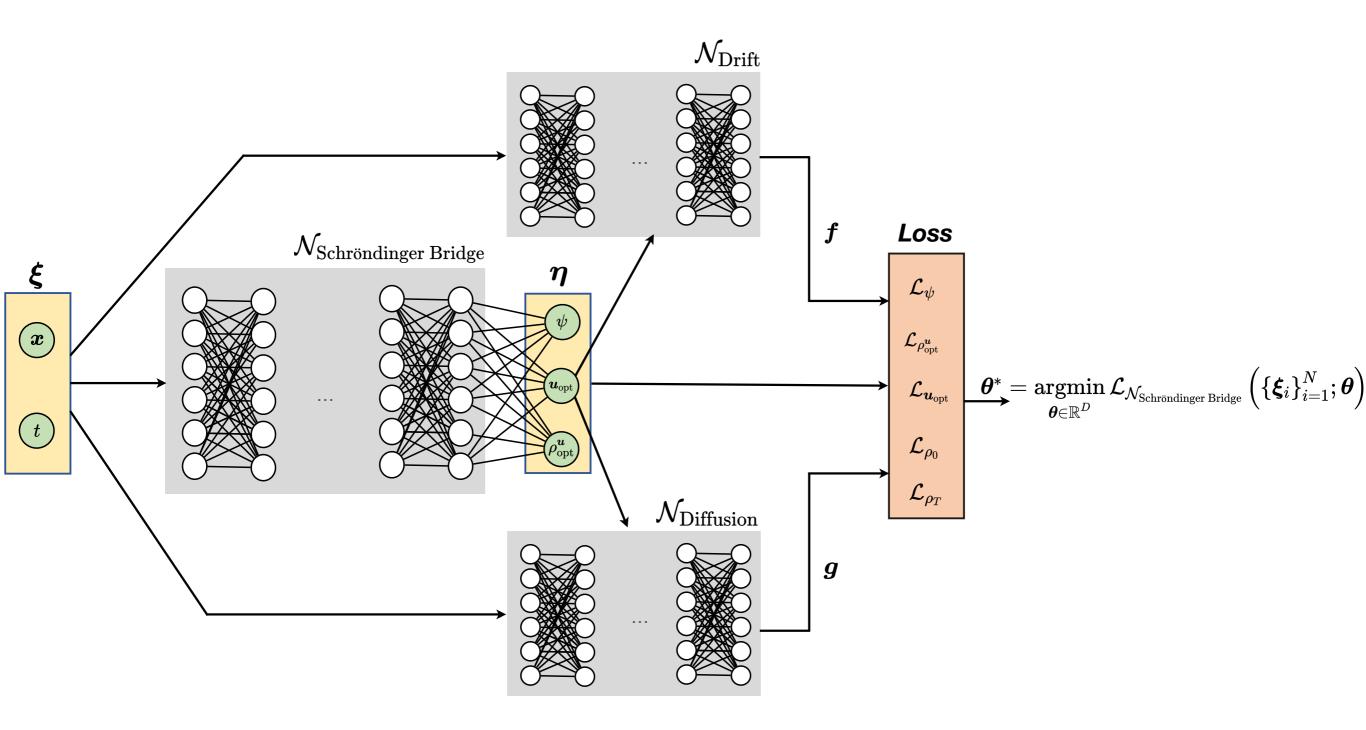
$$\rho_{\mathrm{opt}}^{\boldsymbol{u}}(0, \boldsymbol{x}) = \rho_{0}, \quad \rho_{\mathrm{opt}}^{\boldsymbol{u}}(T, \boldsymbol{x}) = \rho_{T},$$

Cf. classical SBP: two coupled PDEs + optimal policy explicit in value fn ψ

Data-driven GSBP for Colloidal SA



Architecture for Data-driven GSBP



Sinkhorn Losses for Boundary Conditions

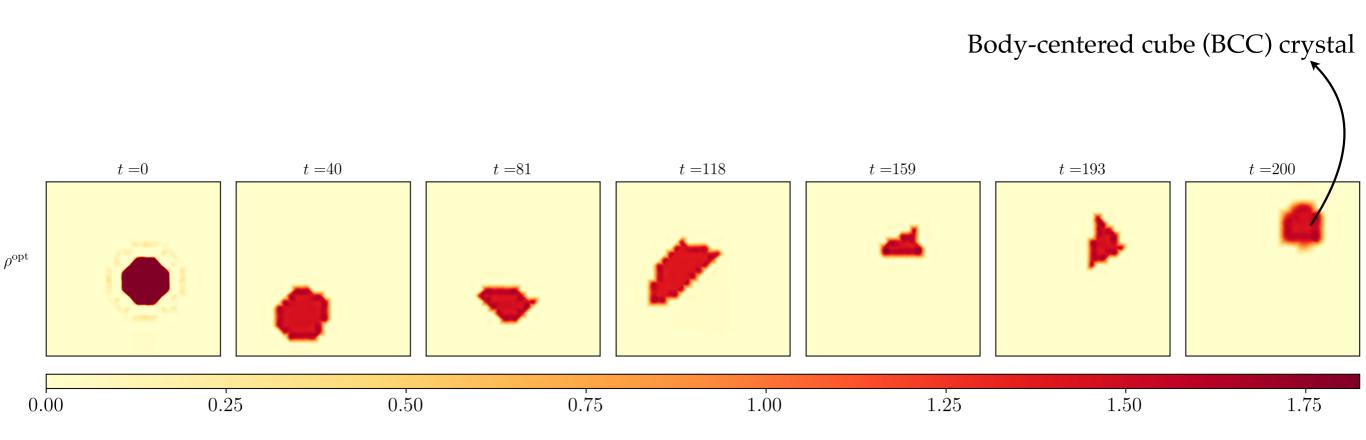
$$W^2_arepsilon(\mu_0,\mu_1) := \inf_{\pi \in \Pi_2(\mu_0,\mu_T)} \int_{\mathbb{R}^n imes \mathbb{R}^n} ig\{ \|oldsymbol{x} - oldsymbol{y}\|_2^2 + arepsilon \log \pi(oldsymbol{x},oldsymbol{y}) ig\} \mathrm{d}\pi(oldsymbol{x},oldsymbol{y})$$

For boundary conditions, use Sinkhorn losses:
$$\mathcal{L}_{
ho_i} := W_arepsilon^2 \Big(
ho_i,
ho_i^{ ext{epoch index}}(oldsymbol{ heta}) \Big)$$

Implementation friendly for PINN training:

$$\mathtt{Autodiff}_{m{ heta}}W^2_arepsilon\Big(
ho_i,
ho_i^{\mathrm{epoch\;index}}(m{ heta})\Big) \quad orall i \in \{0,T\}$$

Case Study: Synthesize BCC Crystalline Structure by PDF Steering in $(\langle C_{10} \rangle, \langle C_{12} \rangle)$ Space





Uses PINN with Sinkhorn losses + the drift-diffusion are themselves NNs

Take Home Message

 Lots of interesting theory, algorithms and applications to be done

- Excellent intersections with related communities: ML, robotics, systems biology, smart manufacturing

Hiring: students and postdocs to work in Schrödinger bridge, stochastic control, stochastic ML

Thank You









