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

#### Abhishek Halder

Department of Applied Mathematics University of California, Santa Cruz Santa Cruz, CA 95064

#### Joint work with students and collaborators

















Aerospace Engineering Department Seminar, Iowa State University November 17, 2022



# Topic of this talk

Control theory and algorithms for measures/distributions and densities

### Intuition

measure a.k.a. distribution = mass

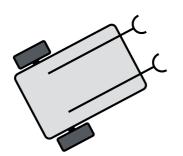
$$mass = density \times volume$$

conservation of mass:

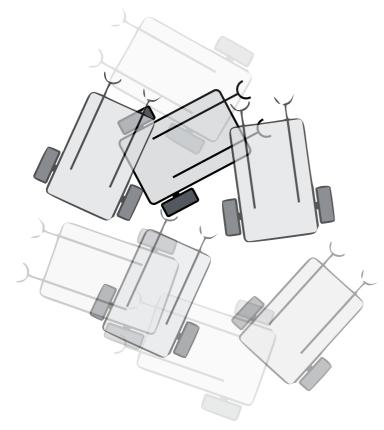
$$\int d(mass) = \int density \times d(volume)$$

= constant, say 1

# But what do we mean by density?



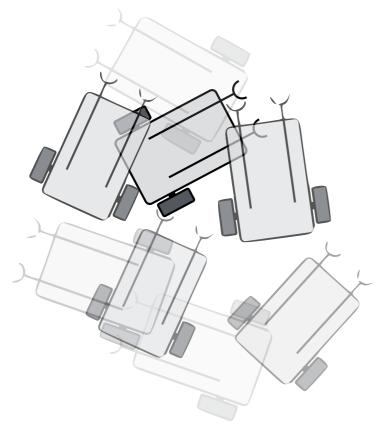
$$\mathbf{x}(t) = \begin{pmatrix} x \\ y \\ \theta \end{pmatrix} \in \mathcal{X} \equiv \mathbb{R}^2 \times \mathbb{S}^1$$



$$\mathbf{x}(t) = \begin{pmatrix} x \\ y \\ \theta \end{pmatrix} \in \mathcal{X} \equiv \mathbb{R}^2 \times \mathbb{S}^1$$

$$\rho\left(\mathbf{x},t\right):\mathcal{X}\times\left[0,\infty\right)\mapsto\mathbb{R}_{\geq0}$$

$$\int_{\mathcal{X}} d\mu = \int_{\mathcal{X}} \rho \, dx = 1 \quad \text{ for all } t \in [0, \infty)$$



$$\mathbf{x}(t) = \begin{pmatrix} x \\ y \\ \theta \end{pmatrix} \in \mathcal{X} \equiv \mathbb{R}^2 \times \mathbb{S}^1$$

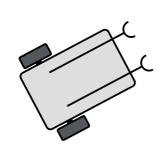
$$\rho(x,t): \mathcal{X} \times [0,\infty) \mapsto \mathbb{R}_{\geq 0}$$

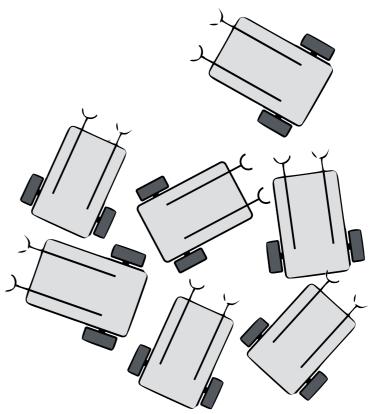
probability measure probability density function

$$\int_{\mathcal{X}} d\mu = \int_{\mathcal{X}} \rho \, dx = 1 \quad \text{for all } t \in [0, \infty)$$

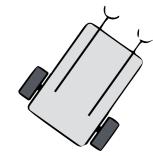
# Population Density Fn.





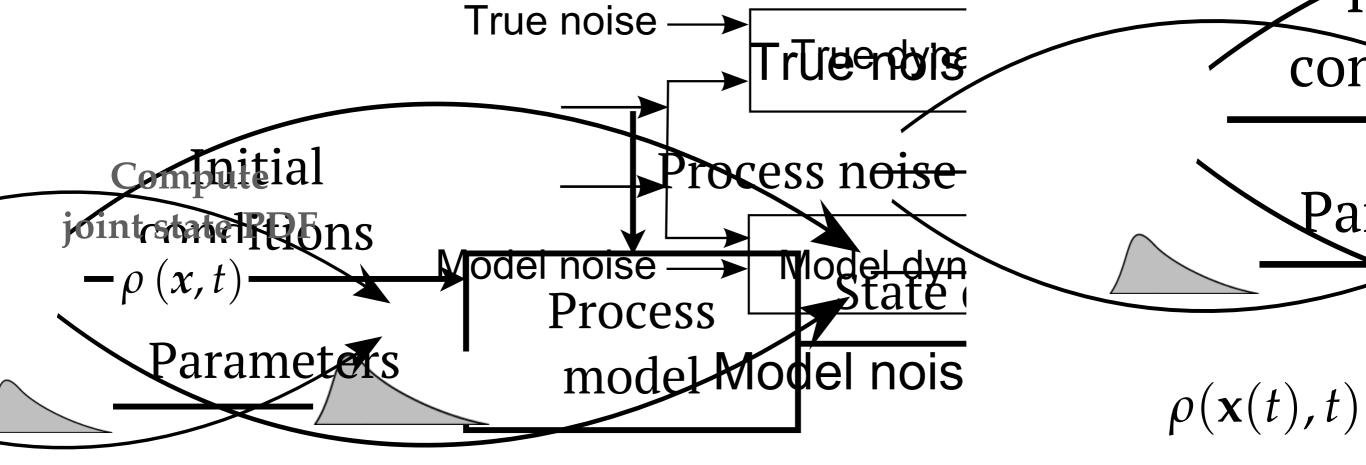


$$x(t) = \begin{pmatrix} x \\ y \\ \theta \end{pmatrix} \in \mathcal{X} \equiv \mathbb{R}^2 \times \mathbb{S}^1$$



population measure population density function

# Why care about densities?

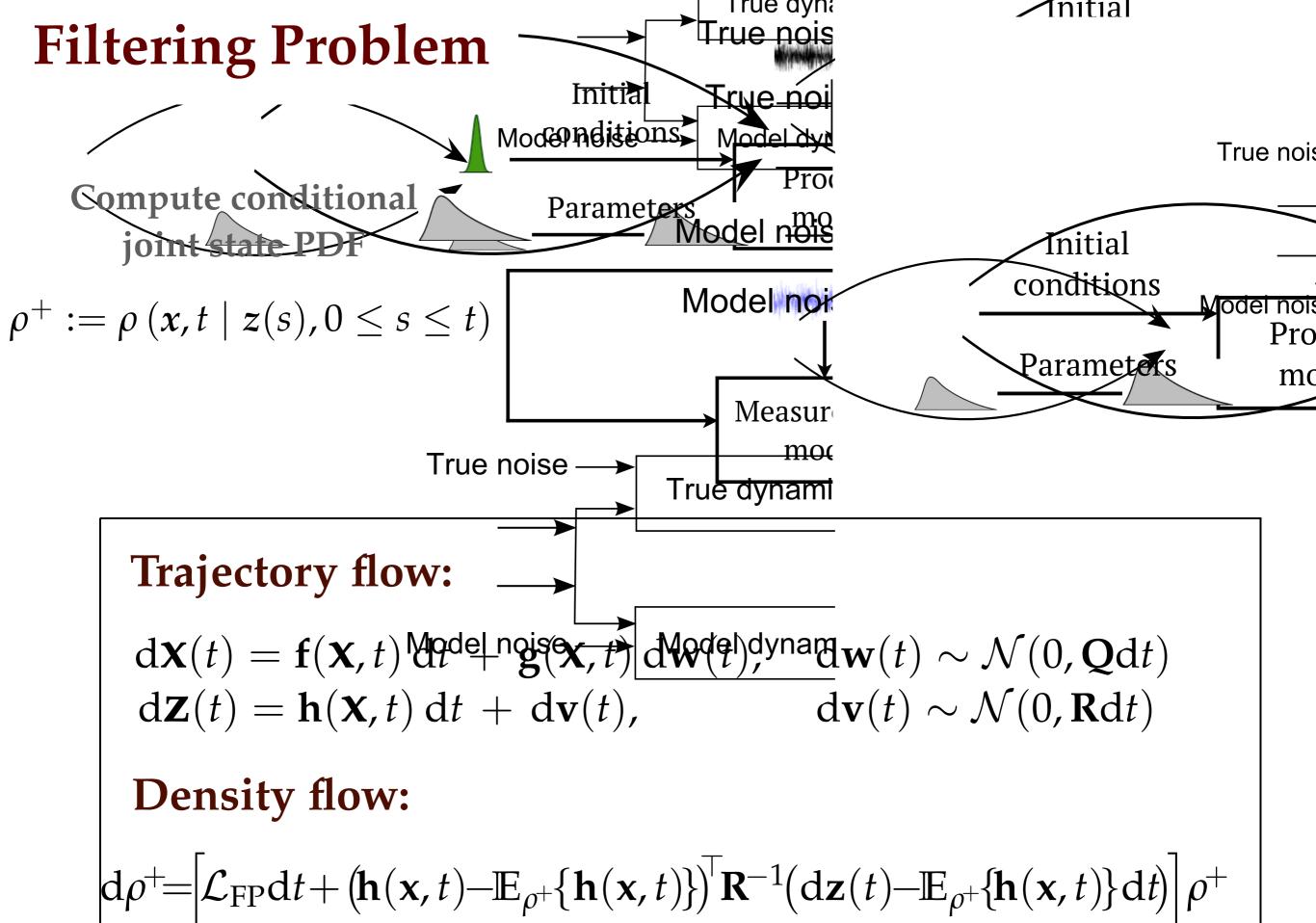


# **Trajectory flow:**

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

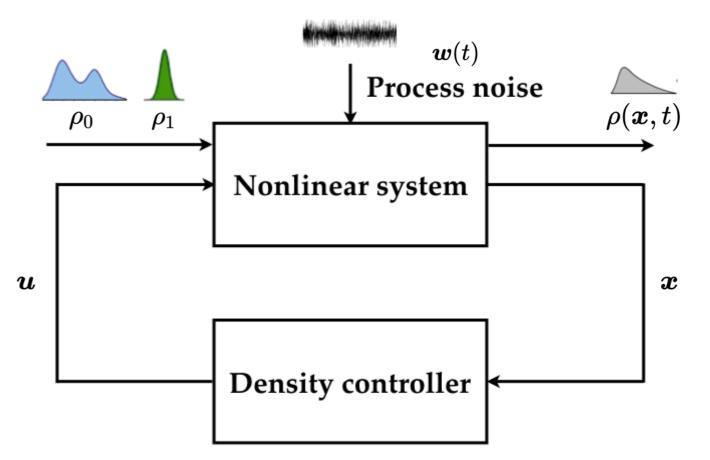
# **Density flow:**

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



#### **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

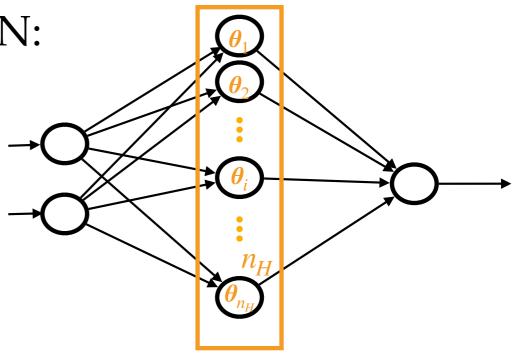
Infinite width limit of fully connected NN:

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:

$$\inf_{
ho \in \mathcal{P}_2(\mathbb{R}^p)} R \Biggl( \int \Phi(m{x}, m{ heta}) 
ho(m{ heta}) \mathrm{d}m{ heta} \Biggr)$$
 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

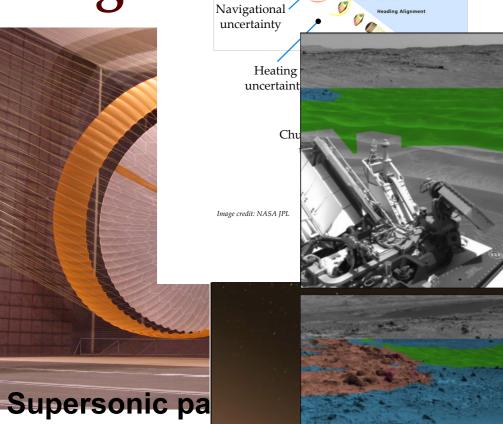
l-scale wind-tunnel testing

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 of the land of the part of the post o will often be identical in appearance to

tige whook read on exploit the place within inhall income the constant 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

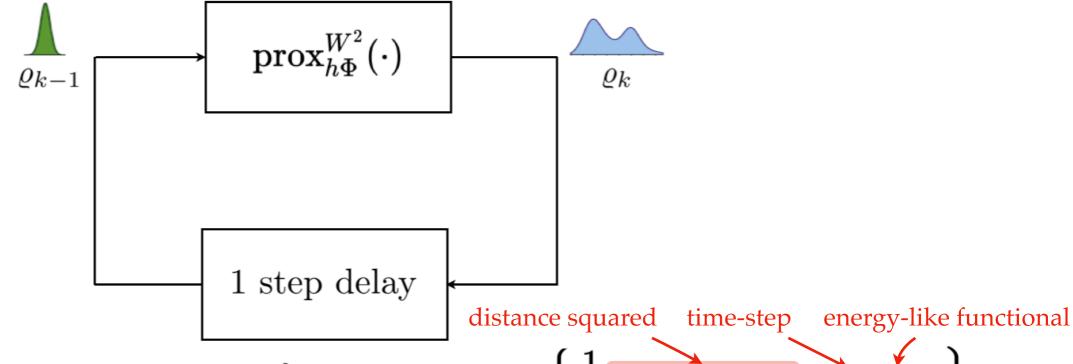
puse Marsing sion, and decelerate behind the largest is parachute ever deployed at Mars. The MSL EDL

# 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 $\mathcal{X}$

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

#### 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:**

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

#### **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:**

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

#### **Convergence:**

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

#### $\varphi$ as Lyapunov function:

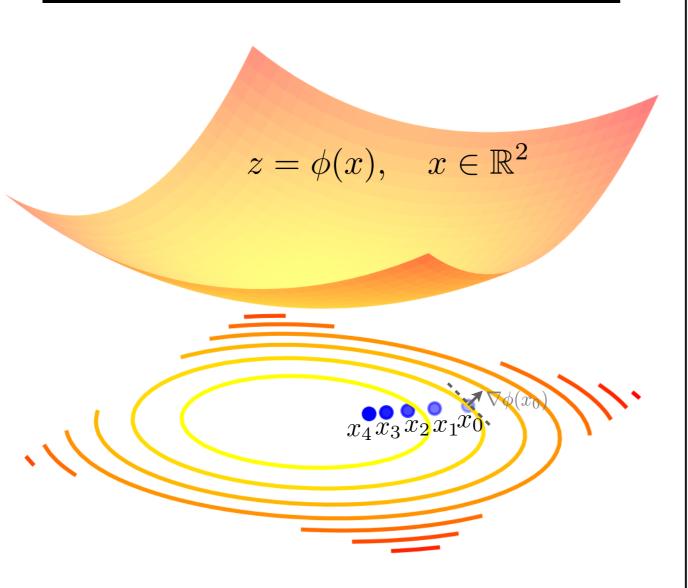
$$rac{\mathrm{d}}{\mathrm{d}t}arphi = -\parallel 
abla arphi \parallel_2^2 \ \le \ 0$$

#### $\Phi$ as Lyapunov functional:

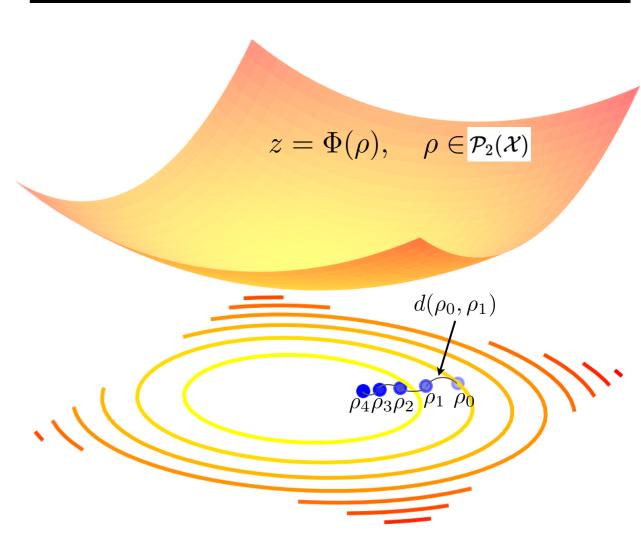
$$rac{\mathrm{d}}{\mathrm{d}t}\Phi = -\mathbb{E}_
hoigg[igg\|
ablarac{\delta\Phi}{\delta
ho}igg\|_2^2igg] \ \le \ 0$$

# Geometric Meaning of Gradient Flow

#### Gradient Flow in $\mathcal{X}$

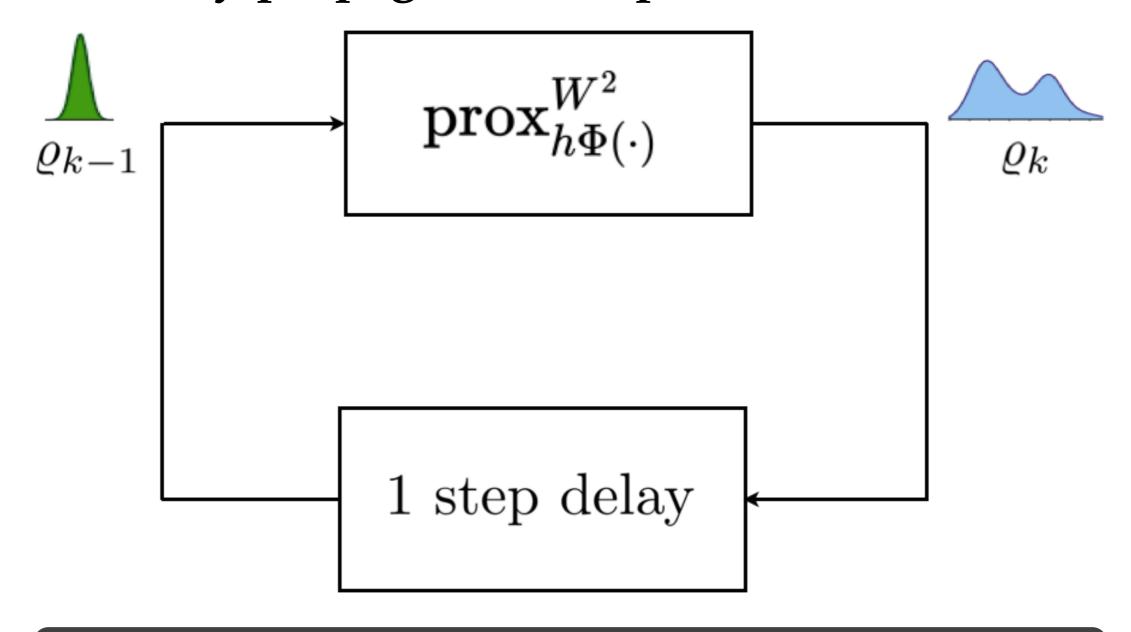


### Gradient Flow in $\mathcal{P}_2(\mathcal{X})$



# 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

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

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

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

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

$$\varrho_k = \underset{\rho}{\arg\min} \left\{ \underset{\boldsymbol{M} \in \Pi(\varrho_{k-1}, \varrho)}{\min} \frac{1}{2} \langle \boldsymbol{C}_k, \boldsymbol{M} \rangle + h \, \langle \psi_{k-1} + \beta^{-1} \rangle \right\}$$

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

#### **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  $\mathbb{R}^n_{\geq 0} \times \mathbb{R}^n_{\geq 0}$ 

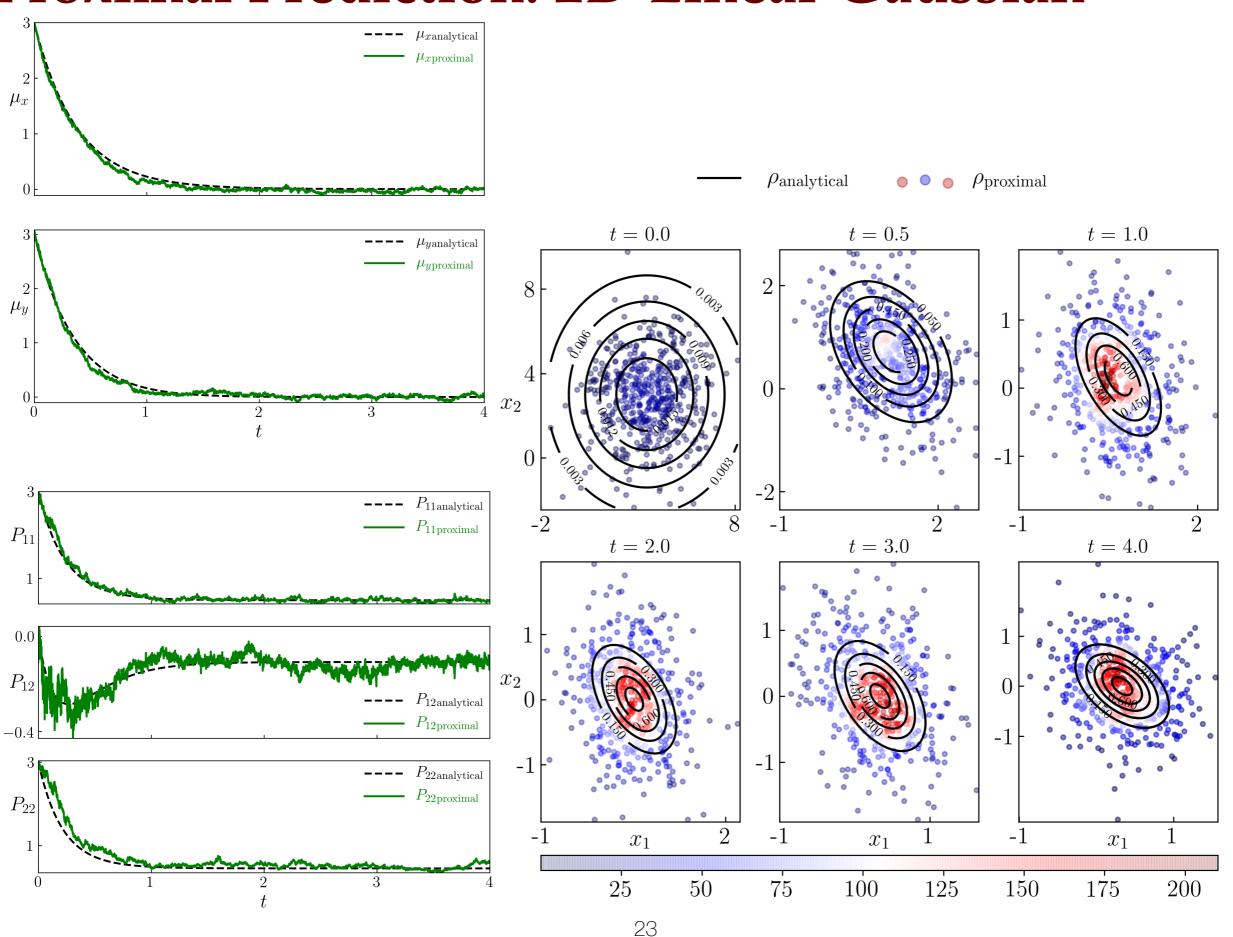
$$oldsymbol{y}\odot(oldsymbol{\Gamma}_koldsymbol{z})=oldsymbol{arrho}_{k-1},\quadoldsymbol{z}\odotoldsymbol{\left(\Gamma_k^{}}^{}oldsymbol{ au}oldsymbol{y}
ight)=oldsymbol{\xi}_{k-1}\odotoldsymbol{z}^{-rac{eta\epsilon}{h}},$$

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

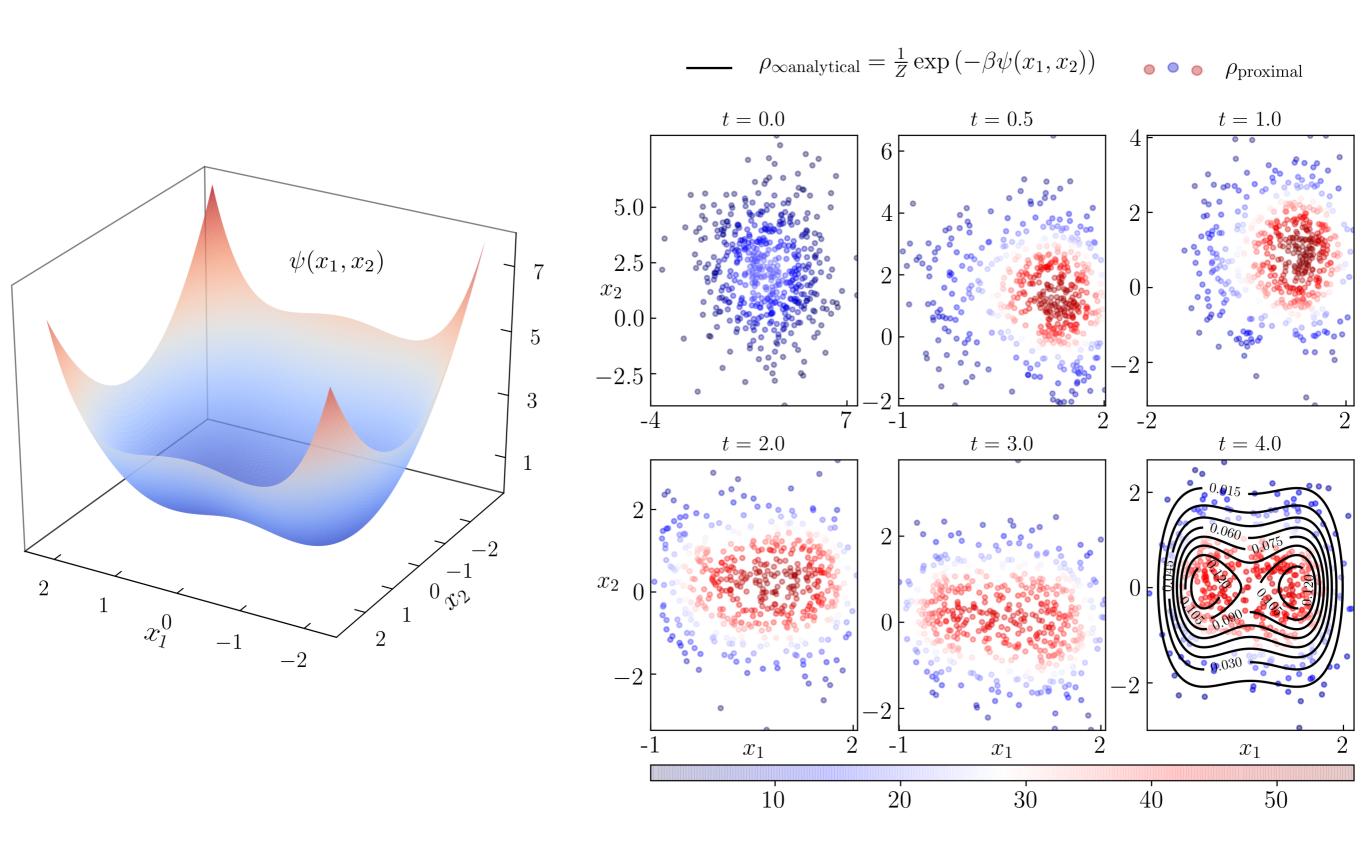
K.F. Caluya and A.H., Gradient flow algorithms for density propagation in stochastic systems, *IEEE TAC* 2019.

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  $\pi_1$ ,  $\pi_2$  because (4) and  $\pi_3$  because (4) and  $\pi_4$  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) are (4) and (4) and (4) are (4) and (4) and (4) are (4) are (4) and (4) are (4) are (4) and (4) are (4) are (4) are (4) and (4) are (4) are (4) and (4) are (4) are (4) and (4) are (4) are (4) are (4) are (4) and (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}) \\
(\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}) \\
(\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}) \\
(\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}) \\
(\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}) \\
(\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}) \\
(\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}) \\
(\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}) \\
(\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}) \\
(\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}) \\
(\varrho_{k}) d^{2} = \underset{\varrho \in \mathscr{D}_{2}}{\text{arg inf}} \frac{1}{2} d^{2}(\varrho, \varrho_{k}) \\
(\varrho_{k}) d^{2} = \underset{\varrho \in \mathscr{D}_{2}}{\text{arg inf}} \frac{1}{2} d^{2}(\varrho, \varrho_{k}) \\
(\varrho_{k}) d^{2} = \underset{\varrho \in \mathscr{D}_{2}}{\text{arg inf}} \frac{1}{2} d^{2}(\varrho, \varrho_{k}) \\
(\varrho_{k}) d^{2} = \underset{\varrho \in \mathscr{D}_{2}}{\text{arg inf}} \frac{1}{2} d^{2}(\varrho, \varrho_{k}) \\
(\varrho_{k}) d^{2} = \underset{\varrho \in \mathscr{D}_{2}}{\text{arg inf}} \frac{1}{2} d^{2}(\varrho, \varrho_{k}) \\
(\varrho_{k}) d^{2} = \underset{\varrho \in \mathscr{D}_{2}}{\text{arg inf}} \frac{1}{2} d^{2}(\varrho, \varrho_{k}) \\
(\varrho_{k}) d^{2} = \underset{\varrho \in \mathscr{D}_{2}}{\text{arg inf}} \frac{1}{2} d^{2}(\varrho, \varrho_{k}) \\
(\varrho_{k}) d^{2} = \underset{\varrho \in \mathscr{D}_{2}}{\text{arg inf}} \frac{1}{2} d^{2}(\varrho, \varrho_{k}) \\
(\varrho_{k}) d^{2} = \underset$  $W(\pi_1,\pi_2) :=$ 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 production of the paper opvious tip choosing us havitable personal the paper of the paper of the fact that the generalized production of the paper of th

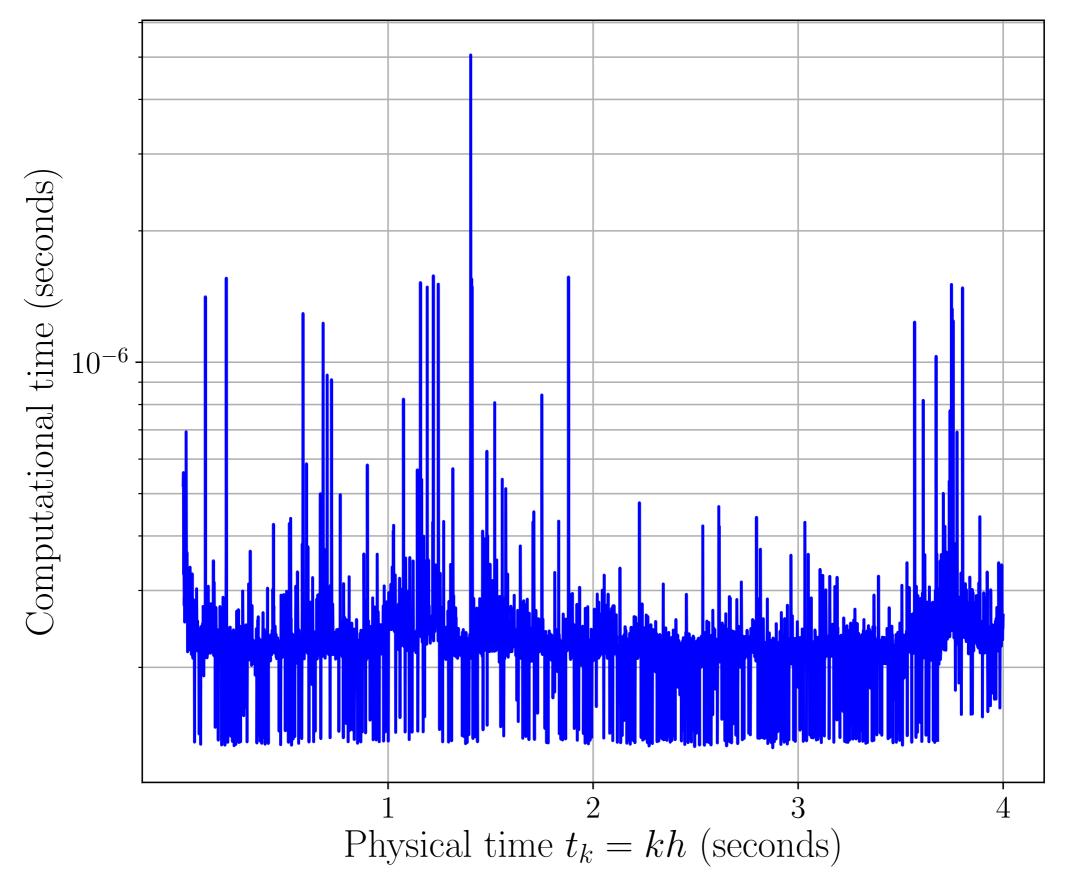
### Proximal Prediction: 2D Linear Gaussian



# Proximal Prediction: Nonlinear Non-Gaussian



# Computational Time: Nonlinear Non-Gaussian



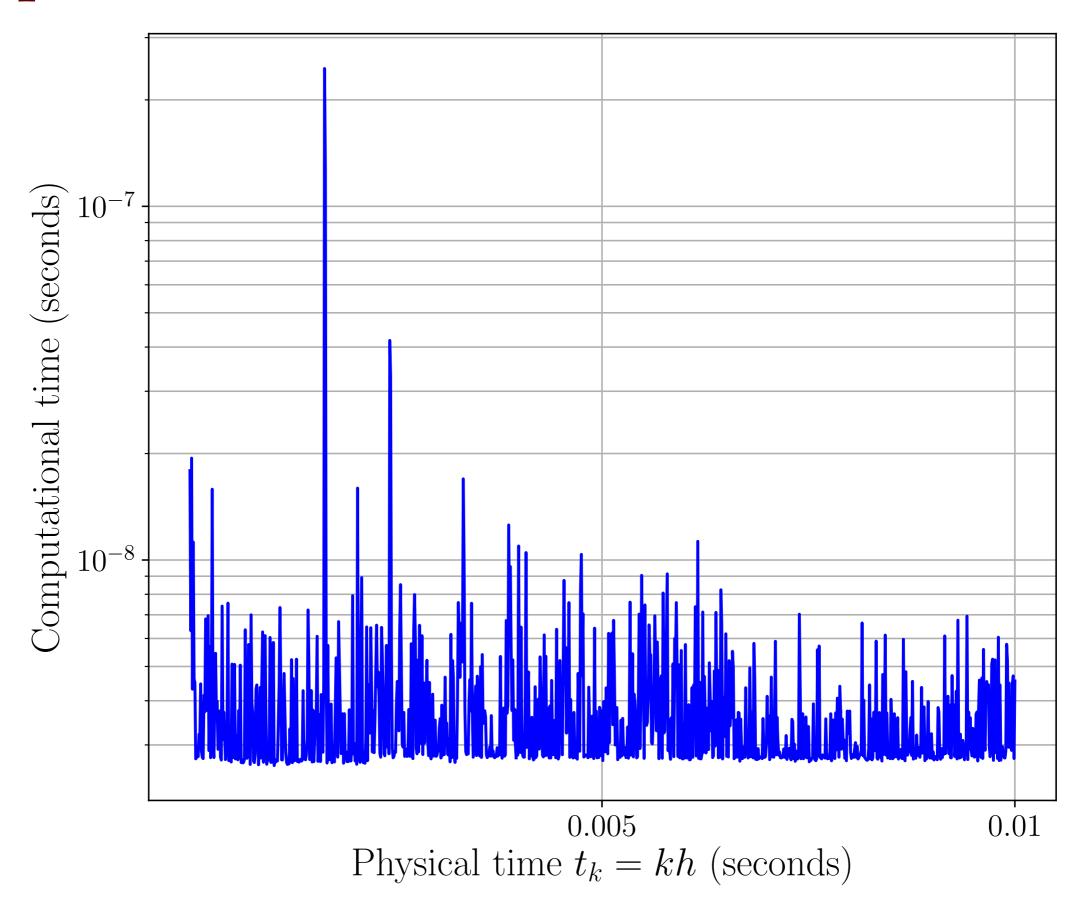
#### Proximal Prediction: Satellite in Geocentric Orbit

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

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

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

# Computational Time: Satellite in Geocentric Orbit



# **Extensions and Applications**

#### Networked nonlinear power system dynamics with O(100) states

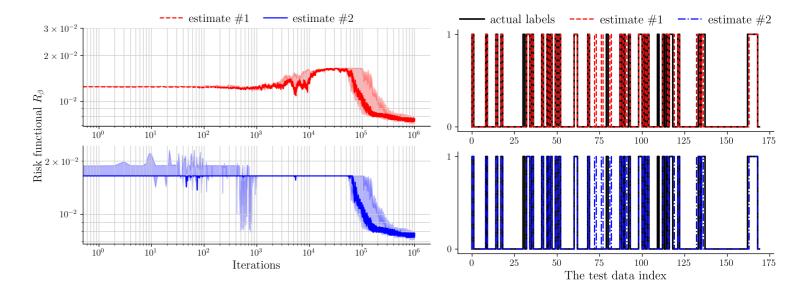
A.H., K.F. Caluya, P. Ojaghi, and X. Geng, Stochastic uncertainty propagation in power system dynamics with measure-valued proximal recursions, *IEEE Transactions on Power Systems*, 2022.

# 

#### 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%

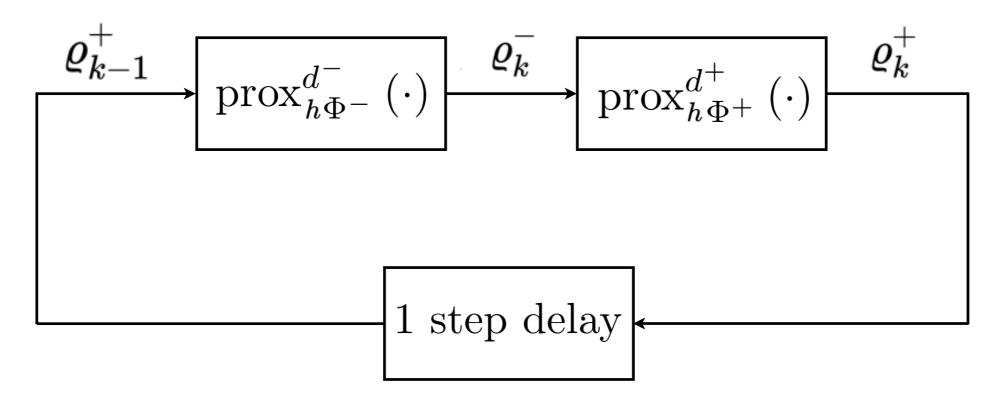
# Solving filtering as generalized gradient flow

#### 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:  $\varrho_k^+(h) o 
ho^+(x,t=kh)$  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:  $d^+ \equiv d_{ ext{FR}}^2 ext{ or } D_{ ext{KL}}, \quad \Phi^+ \underset{\scriptscriptstyle eta \cup}{\equiv} \; rac{1}{2} \mathbb{E}_{arrho^+} \Big[ (y_k - h(x))^ op R^{-1} (y_k - h(x)) \Big]$ 

# 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:

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

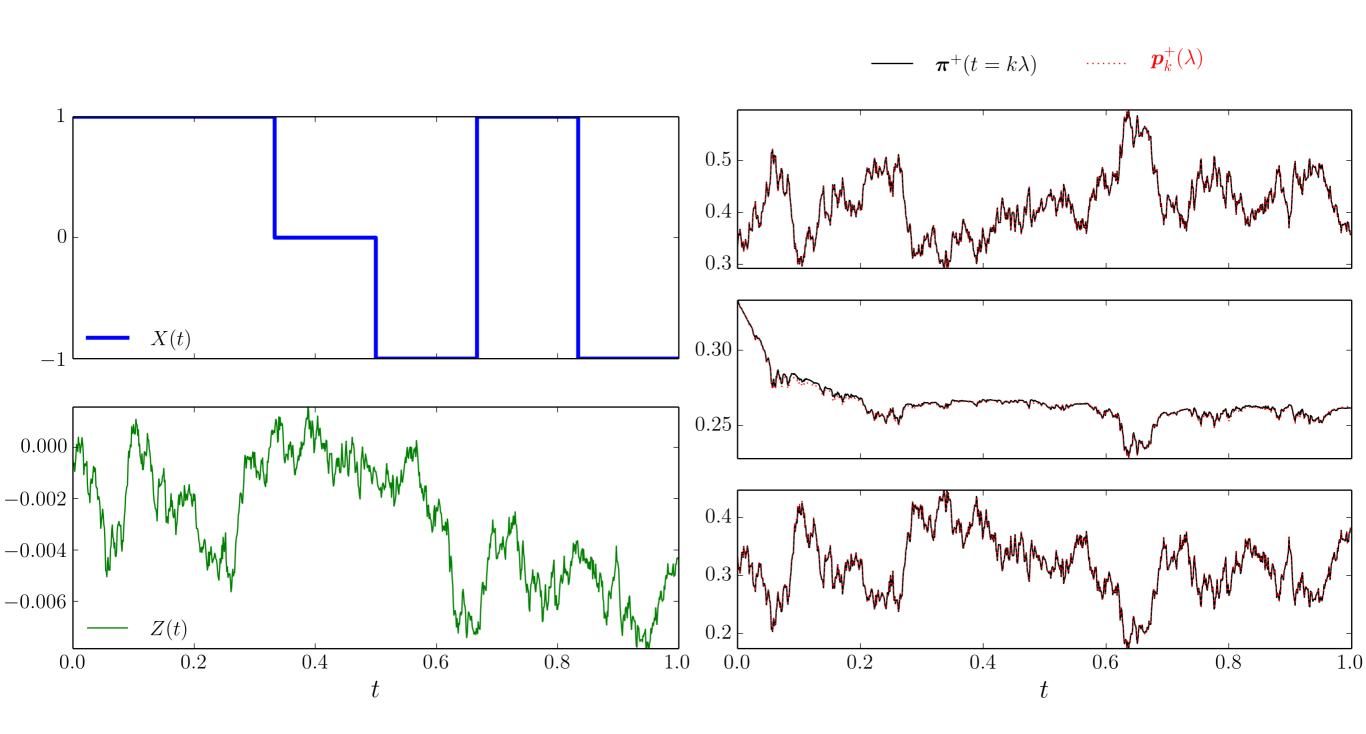
where 
$$H := \operatorname{diag}(h(a_1), \ldots, h(a_m)), \quad \widehat{h}(t) := \sum_{i=1}^m h(a_i) \pi_i^+(t),$$

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

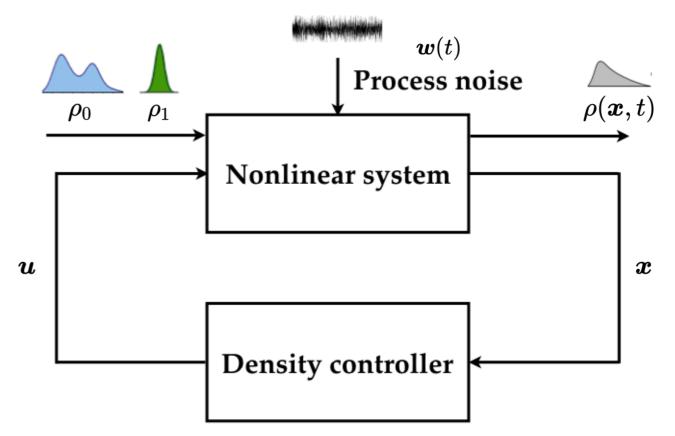


A.H. and T.T. Georgiou, Proximal Recursion for the Wonham Filter, CDC 2019.

# Solving density control as generalized gradient flow

# State Feedback Density Steering

Steer joint state PDF via feedback control over finite time horizon

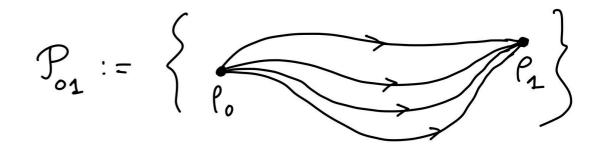


Common scenario:  $G \equiv B$ 

$$\begin{aligned} & \underset{u \in \mathcal{U}}{\text{minimize}} \quad \mathbb{E}\left[\int_0^1 \left(\frac{1}{2}\|u(t,x^u_t)\|_2^2 + q(t,x^u_t)\right) \, \mathrm{d}t\right] \\ & \text{subject to} \\ & \mathrm{d}x^u_t = \{f(t,x^u_t) + B(t,x^u_t)u\} \mathrm{d}t + \sqrt{2}G(t,x^u_t) \mathrm{d}w_t \\ & x^u_0 := x^u_t(t=0) \sim \rho_0, \quad x^u_1 := x^u_t(t=1) \sim \rho_1 \end{aligned}$$

# **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

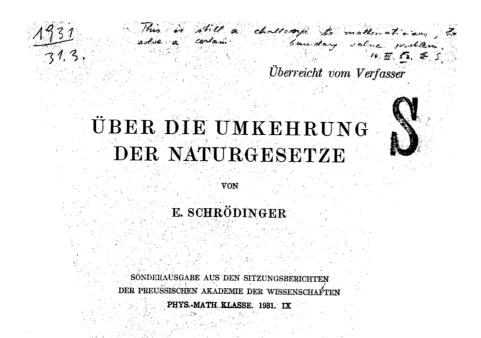
$$\frac{\partial \psi}{\partial t} + \langle \nabla \psi, f \rangle + \langle D, \text{Hess}(\psi) \rangle + \frac{1}{2} \langle \nabla \psi, D \nabla \psi \rangle = q$$

### **Boundary conditions:**

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

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

### 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

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

$$\varphi(x,t) = \exp(\psi(x,t))$$
 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:**

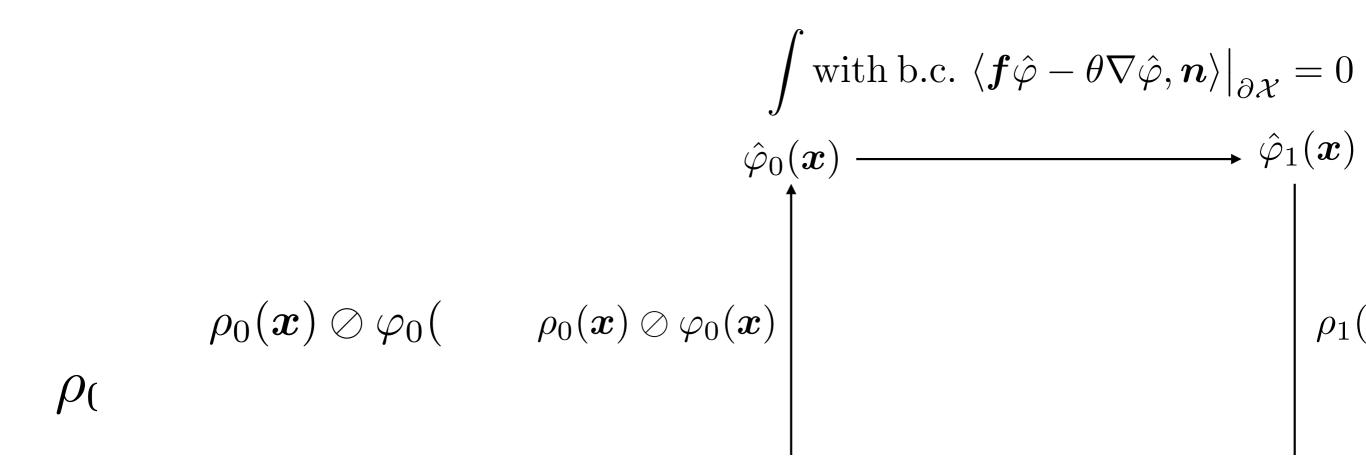
$$\frac{\partial \widehat{\varphi}}{\partial t} = -\nabla \cdot (\widehat{\varphi}f) + \langle \text{Hess}, \mathbf{D}\widehat{\varphi} \rangle - q\widehat{\varphi}, \qquad \widehat{\varphi}_0 \varphi_0 = \rho_0, 
\frac{\partial \varphi}{\partial t} = -\langle \nabla \varphi, f \rangle - \langle \text{Hess}(\varphi), \mathbf{D} \rangle + q\varphi, \qquad \widehat{\varphi}_1 \varphi_1 = \rho_1,$$

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

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

Fiz

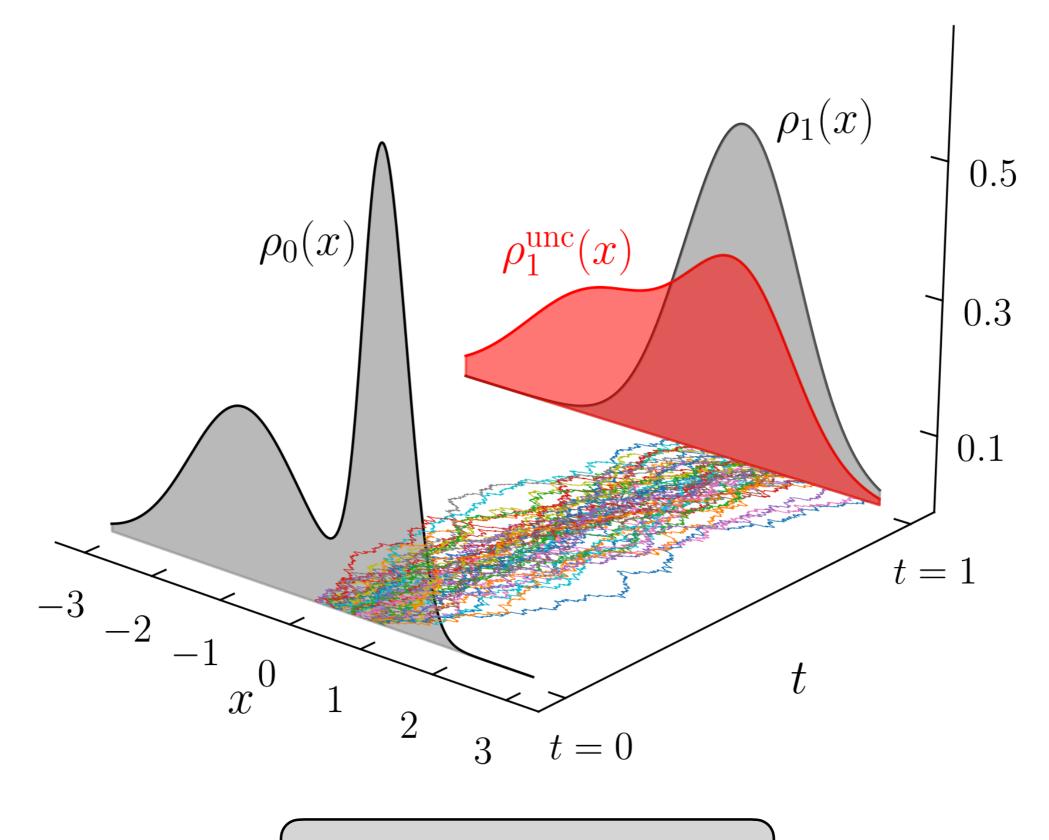
# Fixed Point Recursion for the pair



## Fixed Point Recursion for the pair

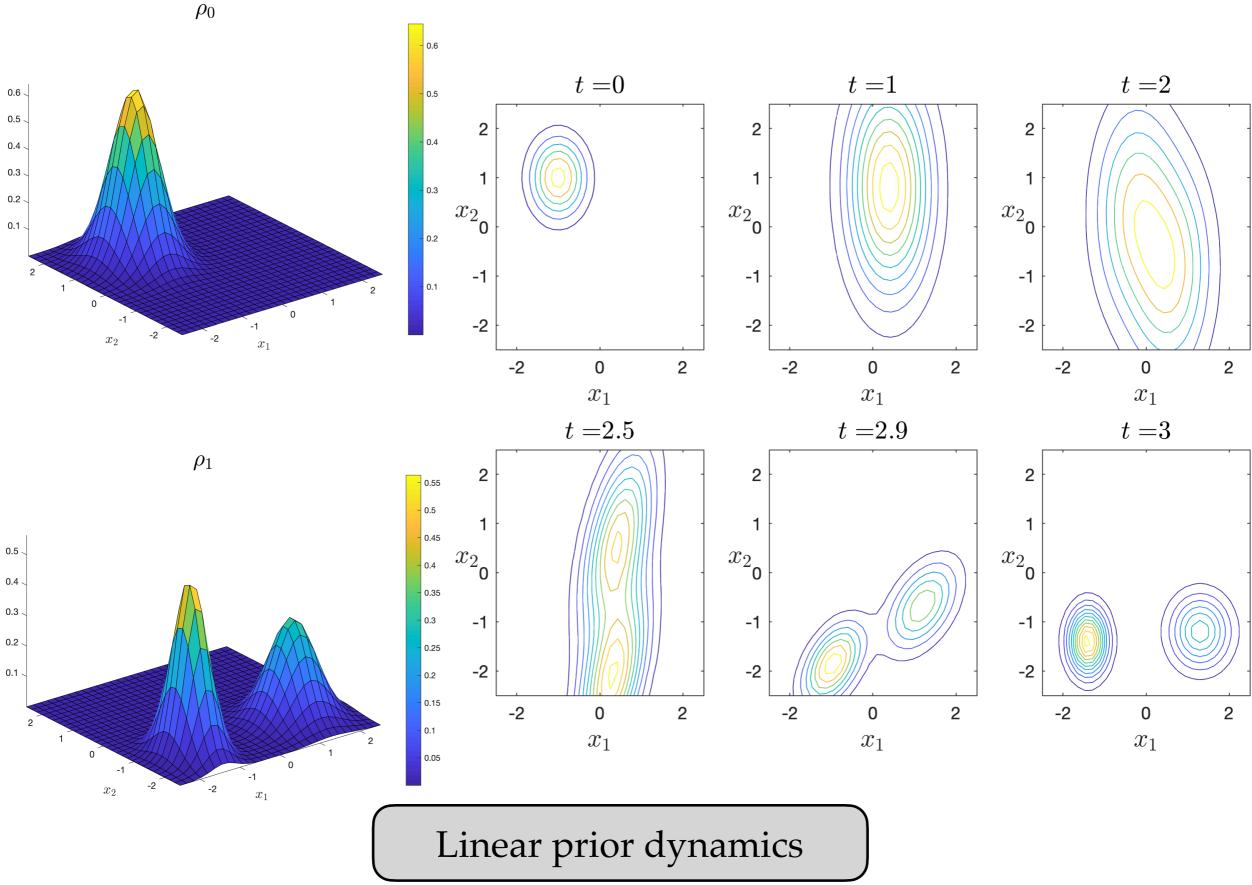
$$\int \text{with b.c. } \left\langle \boldsymbol{f} \hat{\varphi} - \theta \nabla \hat{\varphi}, \boldsymbol{n} \right\rangle \Big|_{\partial \mathcal{X}} = 0$$
 
$$\hat{\varphi}_0(\boldsymbol{x}) \xrightarrow{} \hat{\varphi}_1(\boldsymbol{x})$$
 This recursion is contractive in the Hilbert metric!!

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



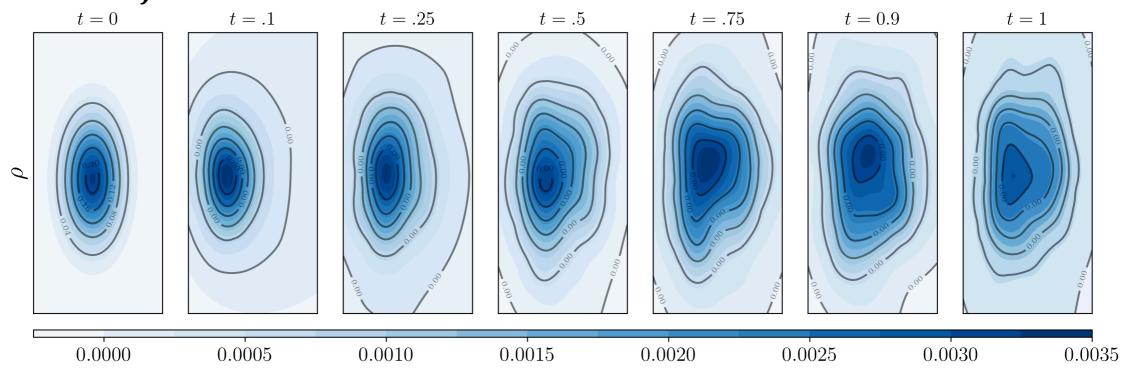
Zero prior dynamics

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

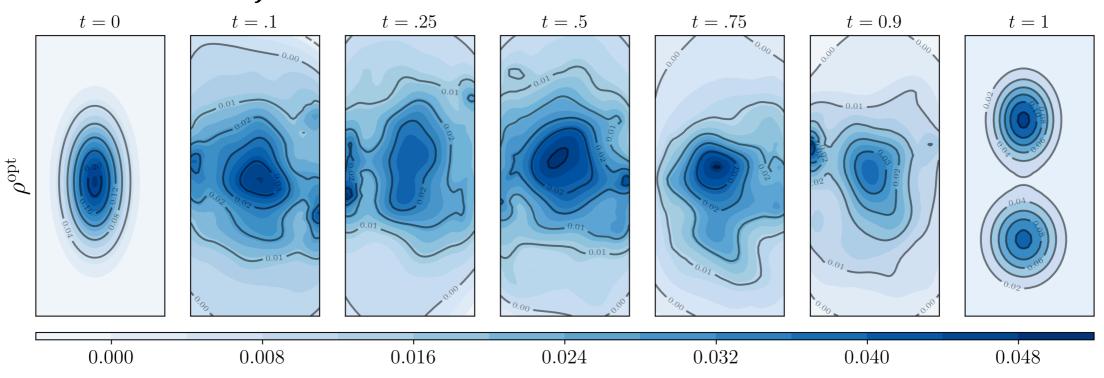


### Feedback Density Control: Nonlinear Grad. Drift

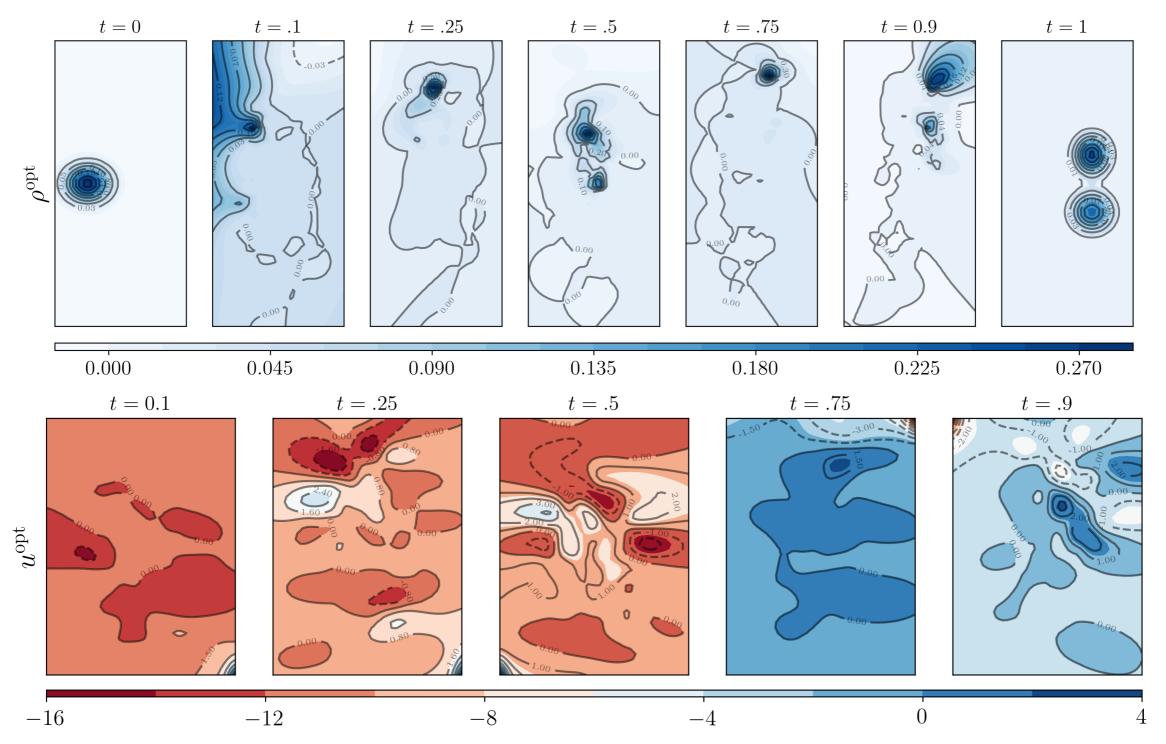
#### **Uncontrolled joint PDF evolution:**



#### **Optimal controlled joint PDF evolution:**

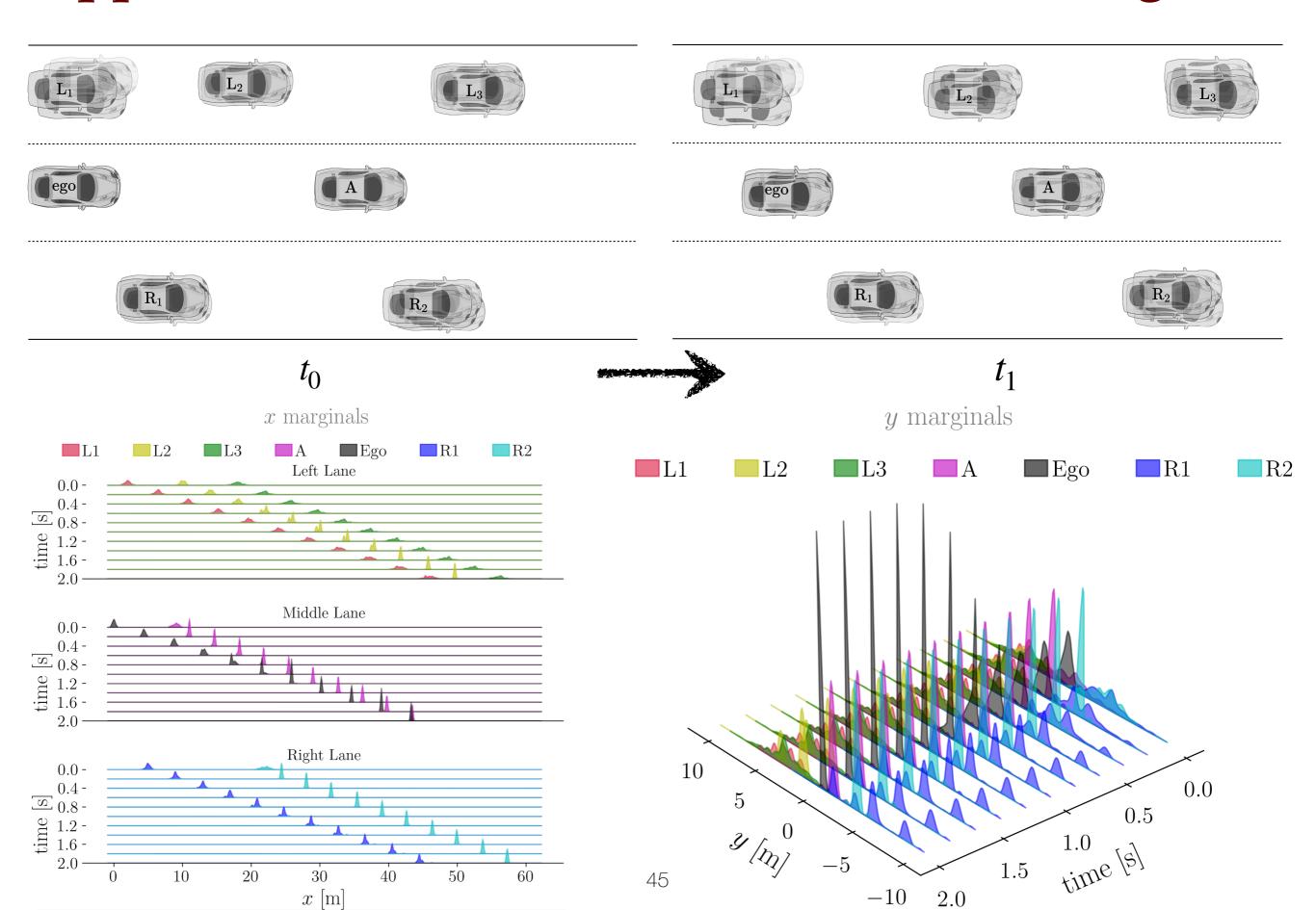


# Feedback Density Control: Mixed Conservative-Dissipative Drift

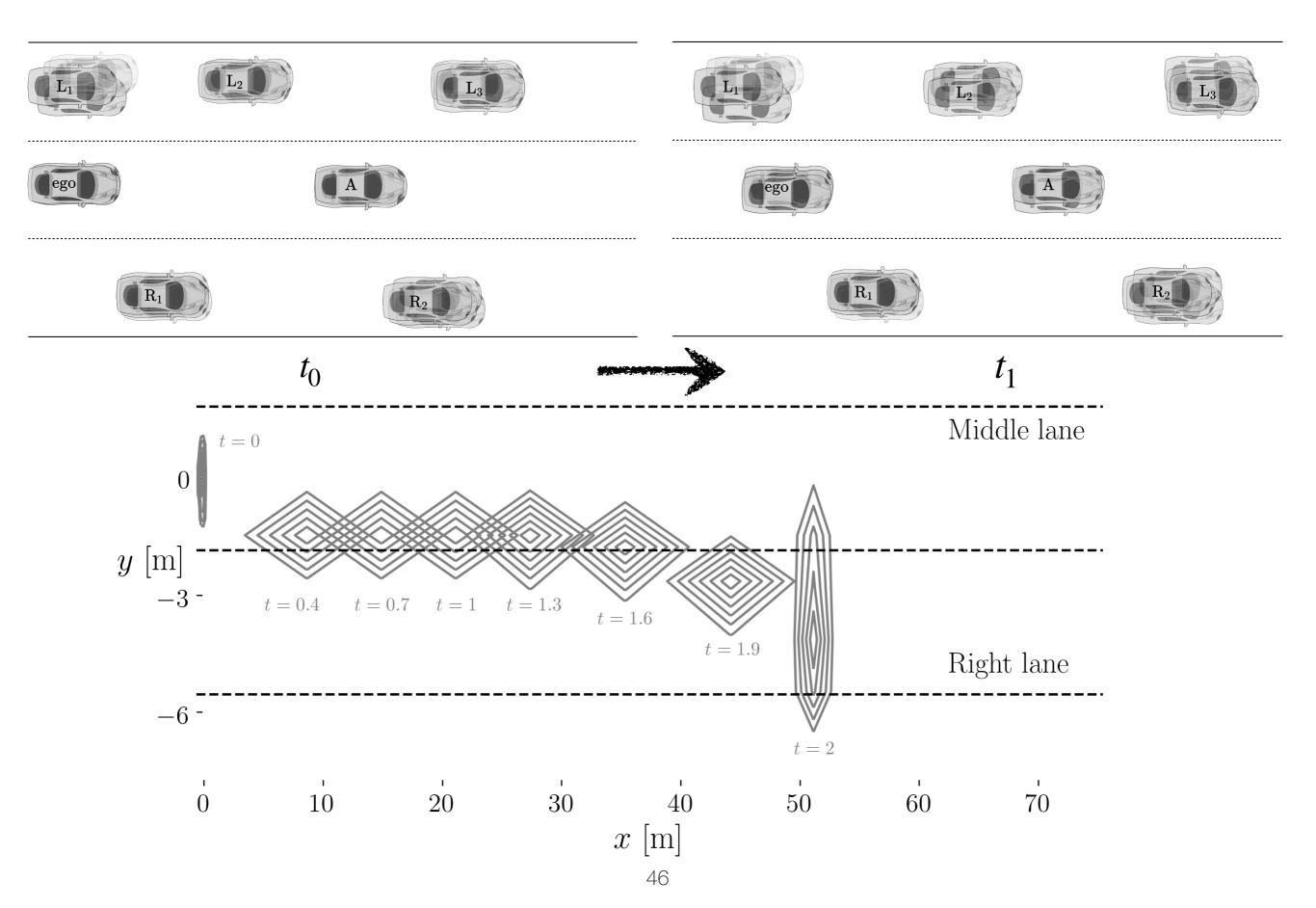


K.F. Caluya and A.H., Wasserstein proximal algorithms for the Schrödinger bridge problem: density control with nonlinear drift, *IEEE TAC* 2021.

### Application: Multi-lane Automated Driving



### Application: Multi-lane Automated Driving



### **Extensions and Applications**

#### Multi-lane automated driving

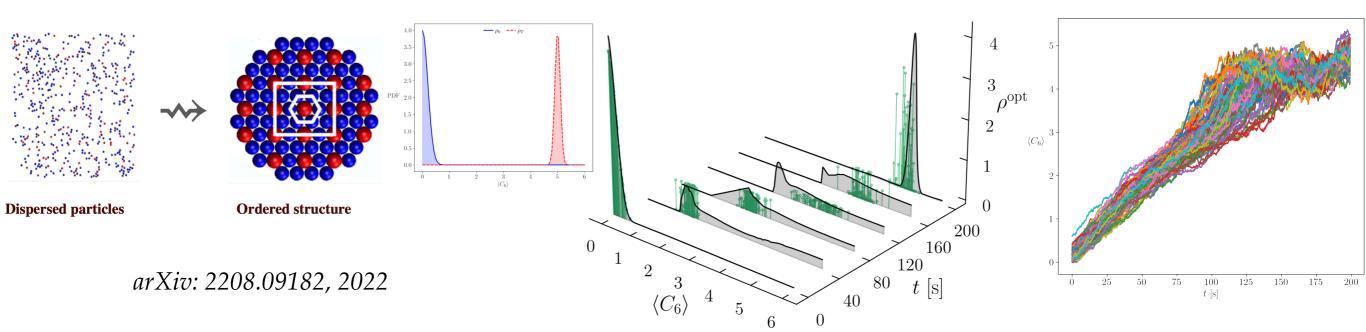
- S. Haddad, A.H., and B. Singh, Density-based stochastic reachability computation for occupancy prediction in automated driving, *IEEE Transactions on Control Systems Technology*, 2022.
- S. Haddad, K.F. Caluya, A.H., and B. Singh, Prediction and optimal feedback steering of probability density functions for safe automated driving, *IEEE Control Systems Letters*, 2021.

#### Hard path constraints, non-standard nonlinear drift and diffusion

K.F. Caluya, and A.H., Reflected Schrödinger bridge: density control with path constraints, *ACC* 2021.

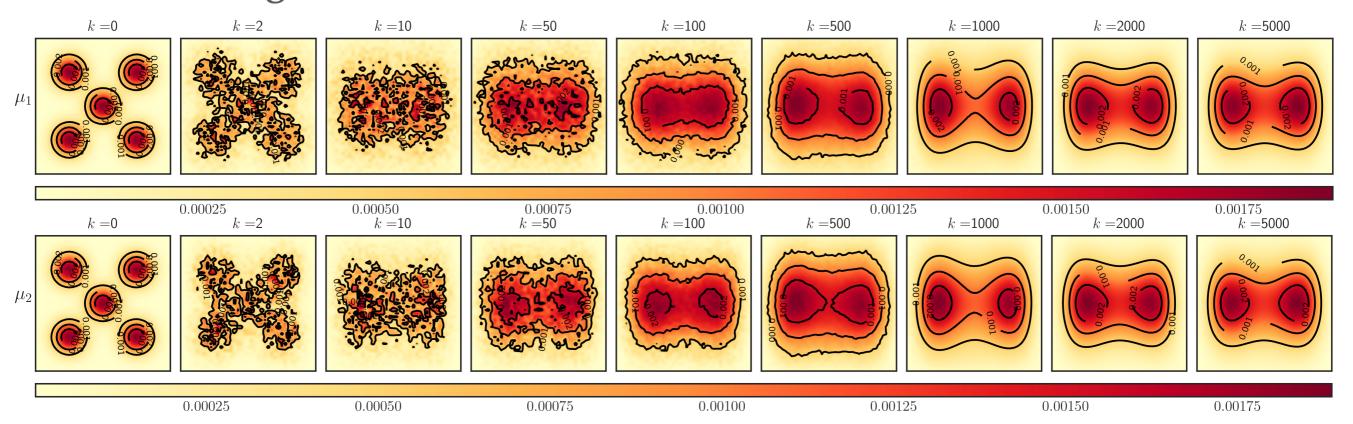
I. Nodozi, and A.H., Schrödinger meets Kuramato via Feynman-Kac: minimum effort distribution steering for noisy nonuniform Kuramoto oscillators, *CDC* 2022.

#### Controlled colloidal self-assembly



## **Ongoing Efforts**

#### Distributed algorithms

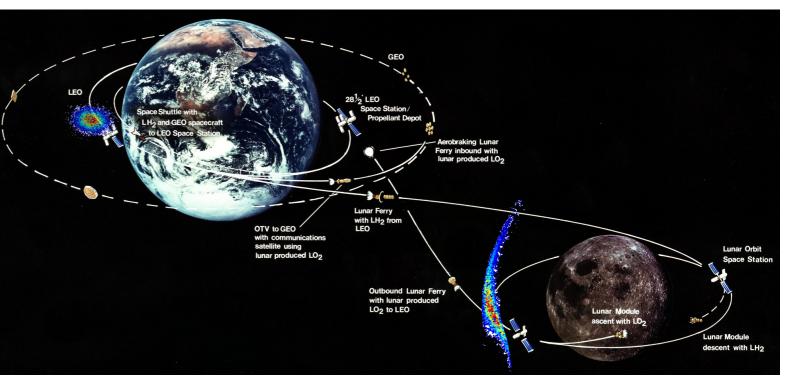


### Stochastic guidance of spacecraft

Provably correct guidance law synthesis

Nonparametric statistical computation

Orbit transfer case studies



### Take Home Message



# Thank You









