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

#### Abhishek Halder

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

Joint work with S. Haddad, K.F. Caluya (UC Santa Cruz) B. Singh (Ford), T.T. Georgiou (UC Irvine), W. Krichene (Google)

> Yahoo! Research March 30, 2022

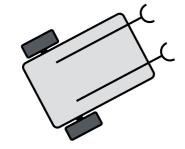


### **Overarching Theme**

## Systems-control theory and algorithms for densities

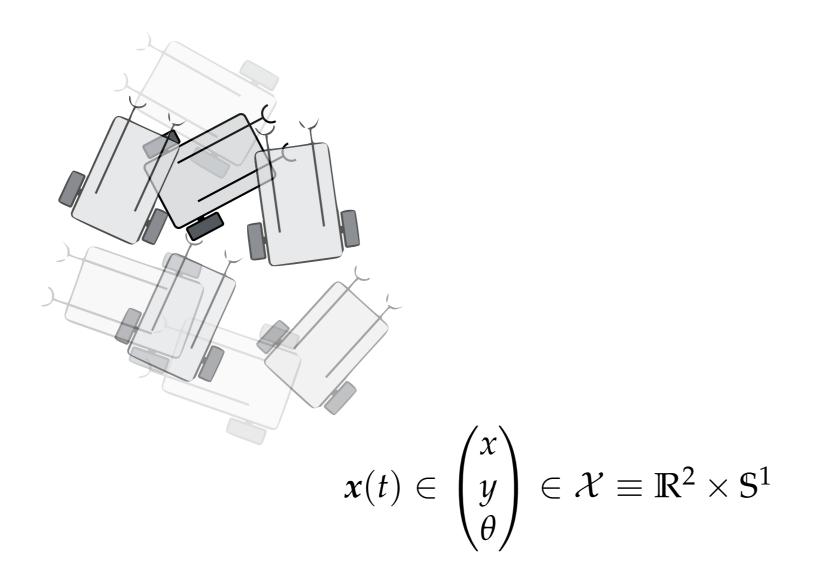
## What is density?

#### **Probability Density Fn.**



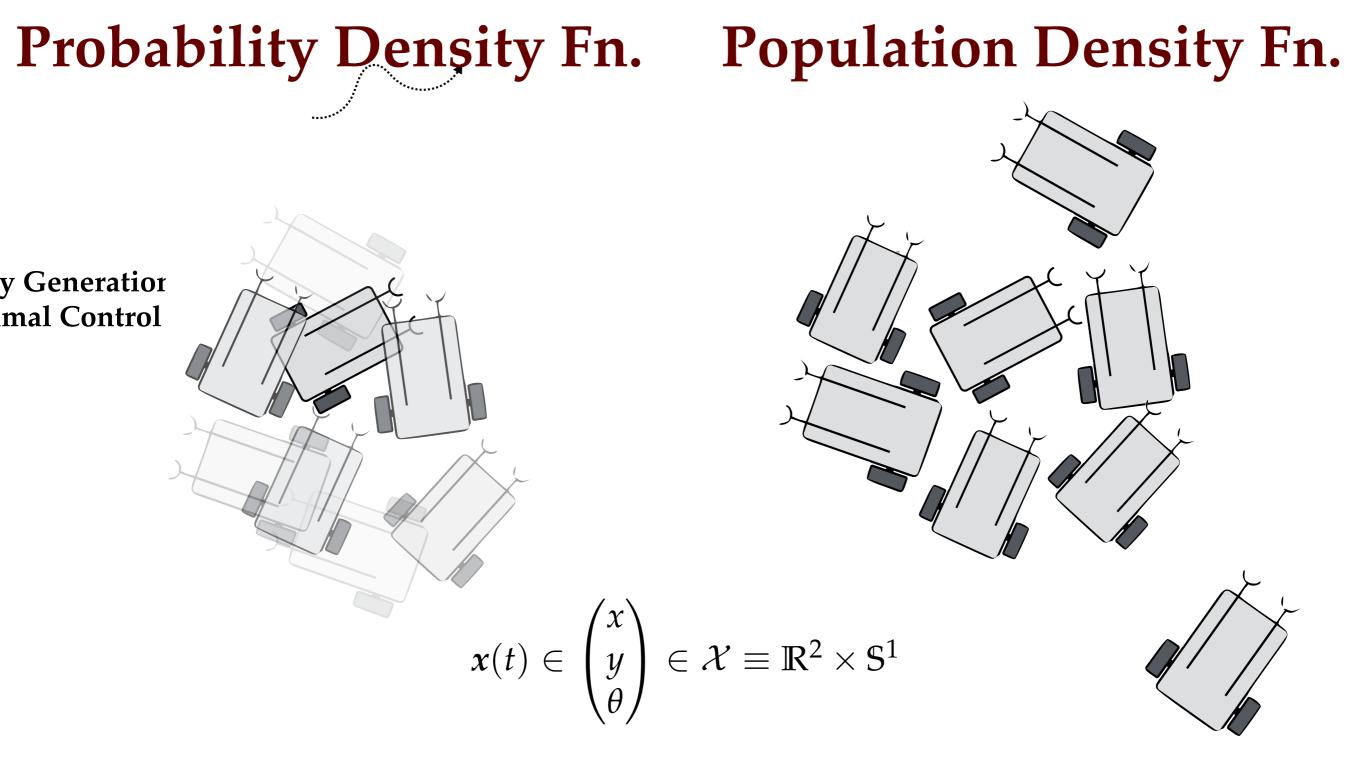
$$\mathbf{x}(t) \in \begin{pmatrix} x \\ y \\ \theta \end{pmatrix} \in \mathcal{X} \equiv \mathbb{R}^2 \times \mathbb{S}^1$$

### **Probability Density Fn.**



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

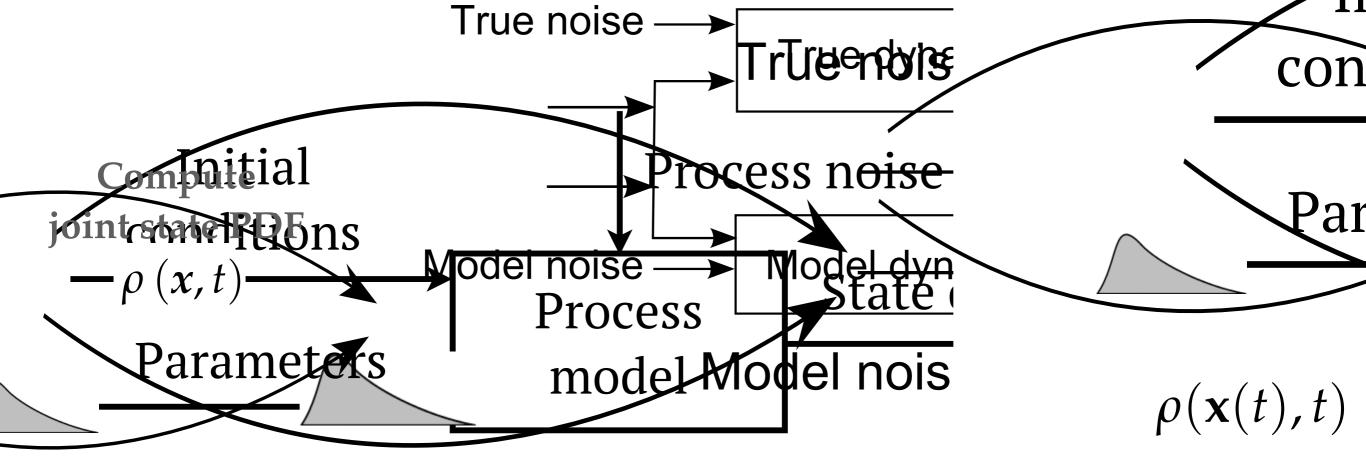
$$\int_{\mathcal{X}} \rho \, \mathrm{d}x = 1 \quad \text{for all } t \in [0, \infty)$$



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

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

## Why do we care about densities in systems-control problems?

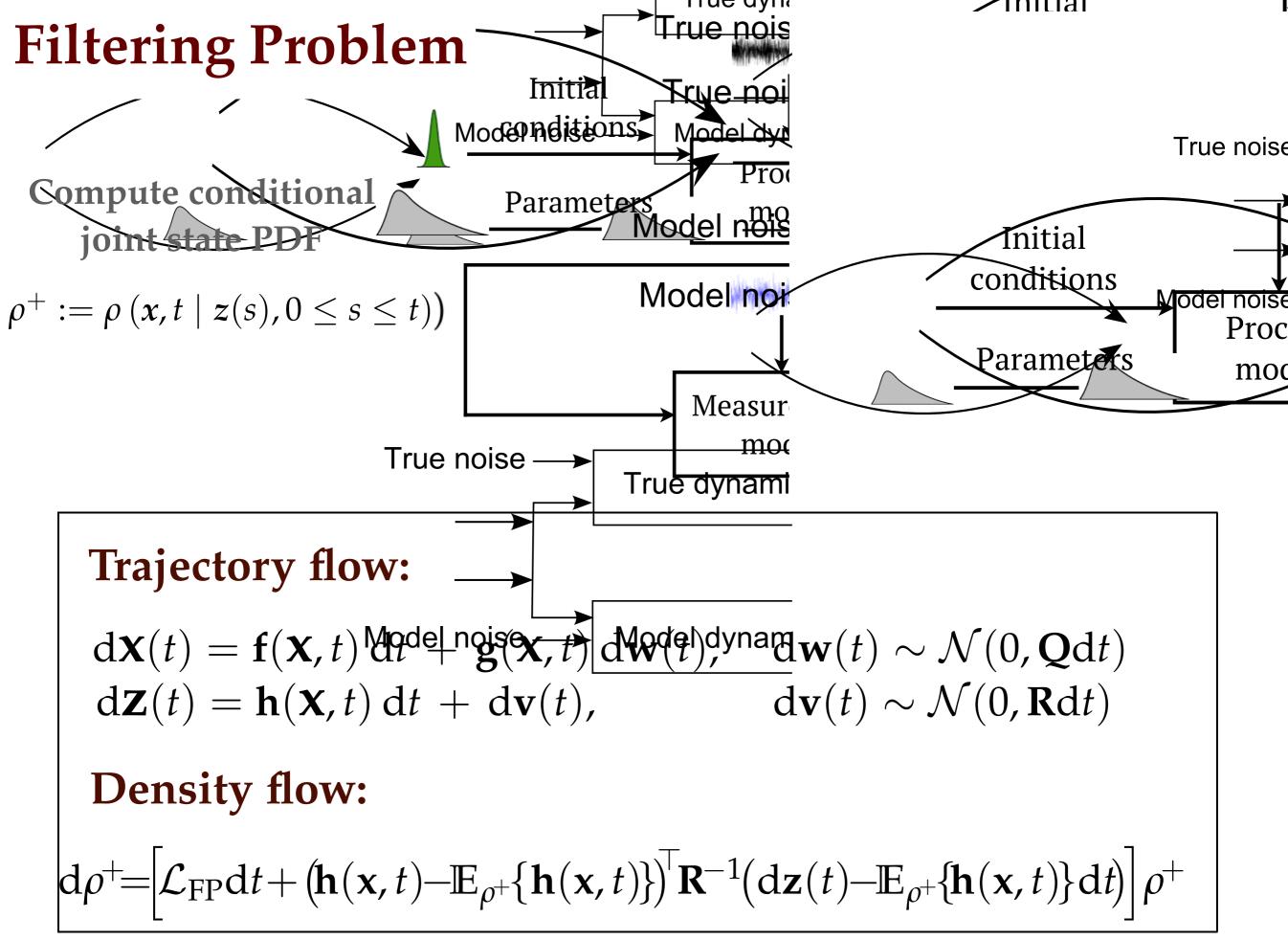


#### **Trajectory flow:**

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

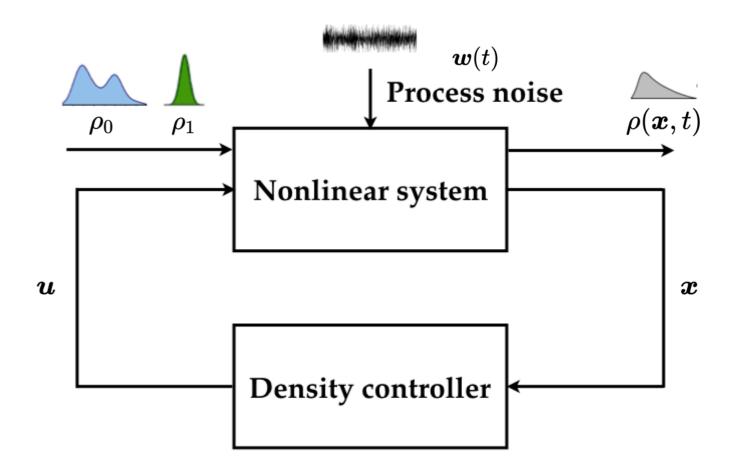
**Density flow:** 

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



## **Control Problem**

**Steer joint state PDF via feedback control over finite time horizon** 



$$\begin{array}{ll} \underset{u \in \mathcal{U}}{\text{minimize}} & \mathbb{E}\left[\int_{0}^{1} \|u\|_{2}^{2} \, \mathrm{d}t\right] \\ \text{subject to} \\ \mathrm{d}x = f\left(x, u, t\right) \, \mathrm{d}t + g\left(x, t\right) \, \mathrm{d}w, \\ x\left(t = 0\right) \sim \rho_{0}, \quad x\left(t = 1\right) \sim \rho_{1} \end{array}$$

### Neural Network Learning Problem

Consider fully connected NN

Think "layers" as interacting population of neurons

Mean field learning problem:

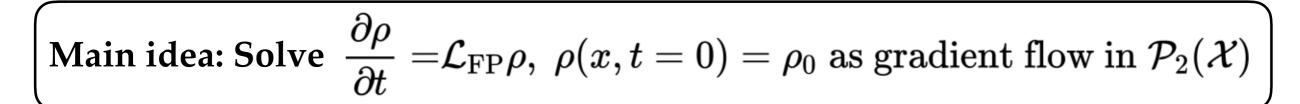
$$\inf_{
ho \in \mathcal{P}_2(\mathbb{R}^p)} \; Rigg(\int \Phi(oldsymbol{x},oldsymbol{ heta})
ho(oldsymbol{ heta})\mathrm{d}oldsymbol{ heta}igg)$$

PDF dynamics:

$$rac{\partial 
ho}{\partial t} = -
abla^W Rigg(\int \Phi 
hoigg) = 
abla \cdot igg(
ho 
abla rac{\delta}{\delta 
ho} Rigg(\int \Phi 
hoigg)igg)$$

# Solving prediction problem as generalized gradient flow

#### What's New?



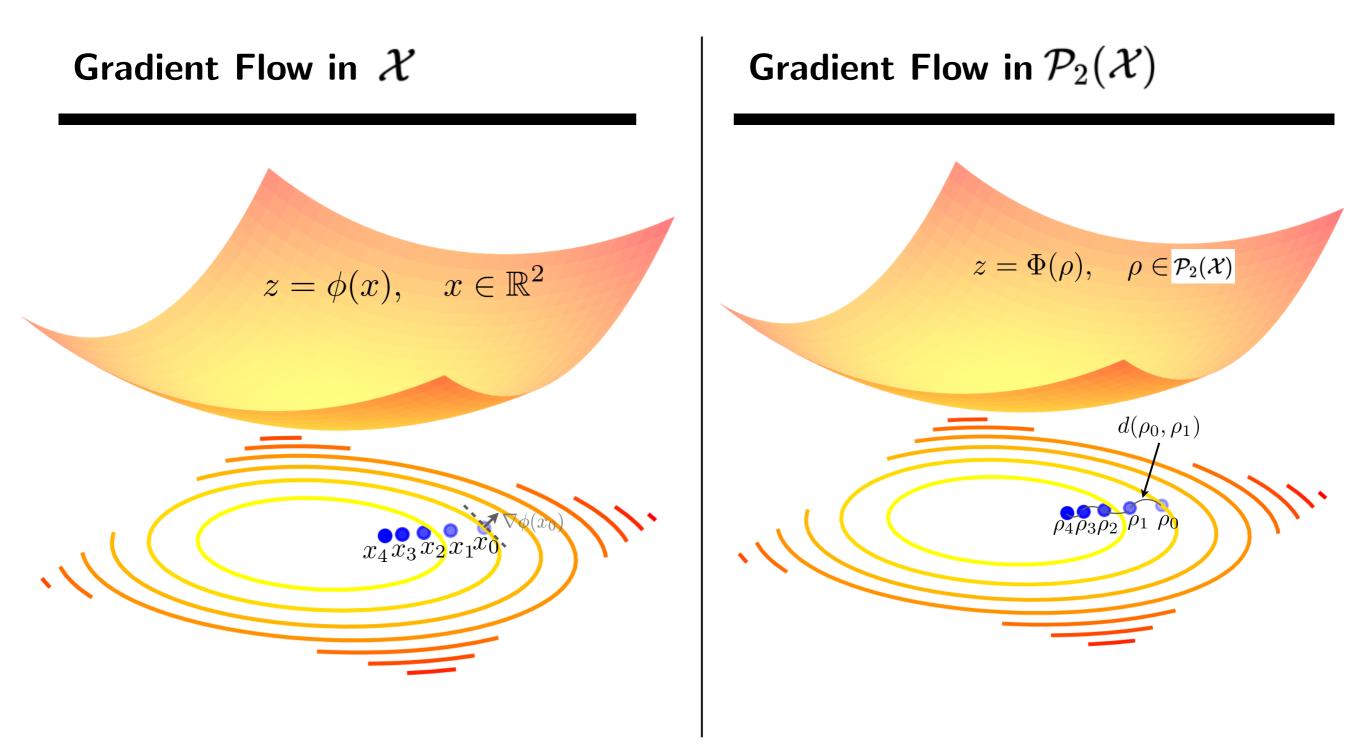
#### Infinite dimensional variational recursion:

$$\begin{array}{c} & & & & & \\ & & & \\ & & & \\ & & \\ & & \\ & & \\ \end{array} \end{array}$$
Proximal operator:  $\varrho_k = \operatorname{prox}_{h\Phi}^{W^2}(\varrho_{k-1}) := \operatorname*{arg\,inf}_{\varrho \in \mathcal{P}_2(\mathcal{X})} \left\{ \frac{1}{2} W^2(\varrho, \varrho_{k-1}) + h\Phi(\varrho) \right\}$ 
Optimal transport cost:  $W^2(\varrho, \varrho_{k-1}) := \underset{\pi \in \Pi(\varrho, \varrho_{k-1})}{\inf} \int_{\mathcal{X} \times \mathcal{X}} c(x, y) \, \mathrm{d}\pi(x, y)$ 
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**

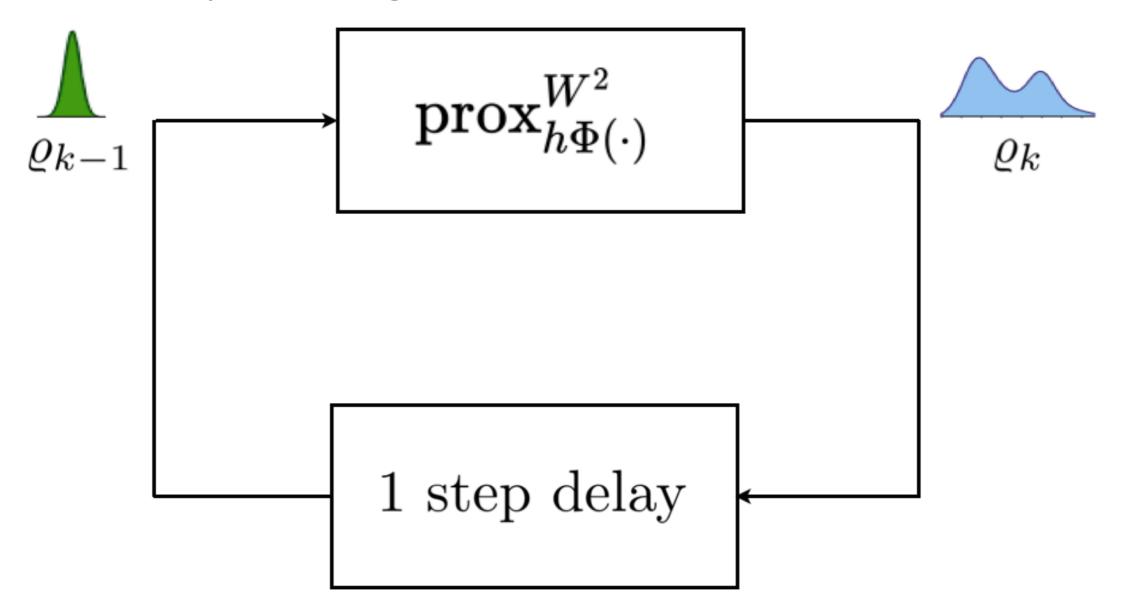
Gradient Flow in  $\mathcal{P}_2(\mathcal{X})$ Gradient Flow in  $\mathcal{X}$  $\frac{\partial \rho}{\partial t} = -\nabla^W \Phi(\rho), \quad \rho(\mathbf{x}, 0) = \rho_0$  $\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}\boldsymbol{t}} = -\nabla\varphi(\boldsymbol{x}), \quad \boldsymbol{x}(0) = \boldsymbol{x}_0$ **Recursion: Recursion:**  $\rho_k = \rho(\cdot, t = kh)$  $\mathbf{x}_k = \mathbf{x}_{k-1} - h \nabla \varphi(\mathbf{x}_k)$  $= \arg\min_{\boldsymbol{x}\in\mathcal{X}} \left\{ \frac{1}{2} \|\boldsymbol{x}-\boldsymbol{x}_{k-1}\|_{2}^{2} + h\varphi(\boldsymbol{x}) \right\}$  $= \arg \min_{\rho \in \mathcal{P}_2(\mathcal{X})} \left\{ \frac{1}{2} W^2(\rho, \rho_{k-1}) + h \Phi(\rho) \right\}$  $=: \operatorname{prox}_{h_{\mathcal{O}}}^{\|\cdot\|_2}(\mathbf{x}_{k-1})$  $=: \operatorname{prox}_{h\Phi}^{W^2}(\rho_{k-1})$ **Convergence: Convergence:**  $\mathbf{x}_k \to \mathbf{x}(t = kh)$  as  $h \downarrow 0$  $\rho_k \to \rho(\cdot, t = kh)$  as  $h \downarrow 0$  $\varphi$  as Lyapunov function:  $\Phi$  as Lyapunov functional:  $rac{\mathrm{d}}{\mathrm{d}t}\Phi = -\mathbb{E}_
ho \left[ \left\| 
abla rac{\delta \Phi}{\delta 
ho} \right\|_2^2 
ight] \le 0$  $rac{\mathrm{d}}{\mathrm{d}t}arphi = - \parallel 
abla arphi \parallel_2^2 \ \le \ 0$ 

### **Geometric Meaning of Gradient Flow**



## **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

#### **Recursion on the Cone**

$$\mathbf{y} = e^{\frac{\boldsymbol{\lambda}_0^*}{\epsilon}h} \left| \quad \mathbf{z} = e^{\frac{\boldsymbol{\lambda}_1^*}{\epsilon}h}$$

Coupled Transcendental Equations in y and z

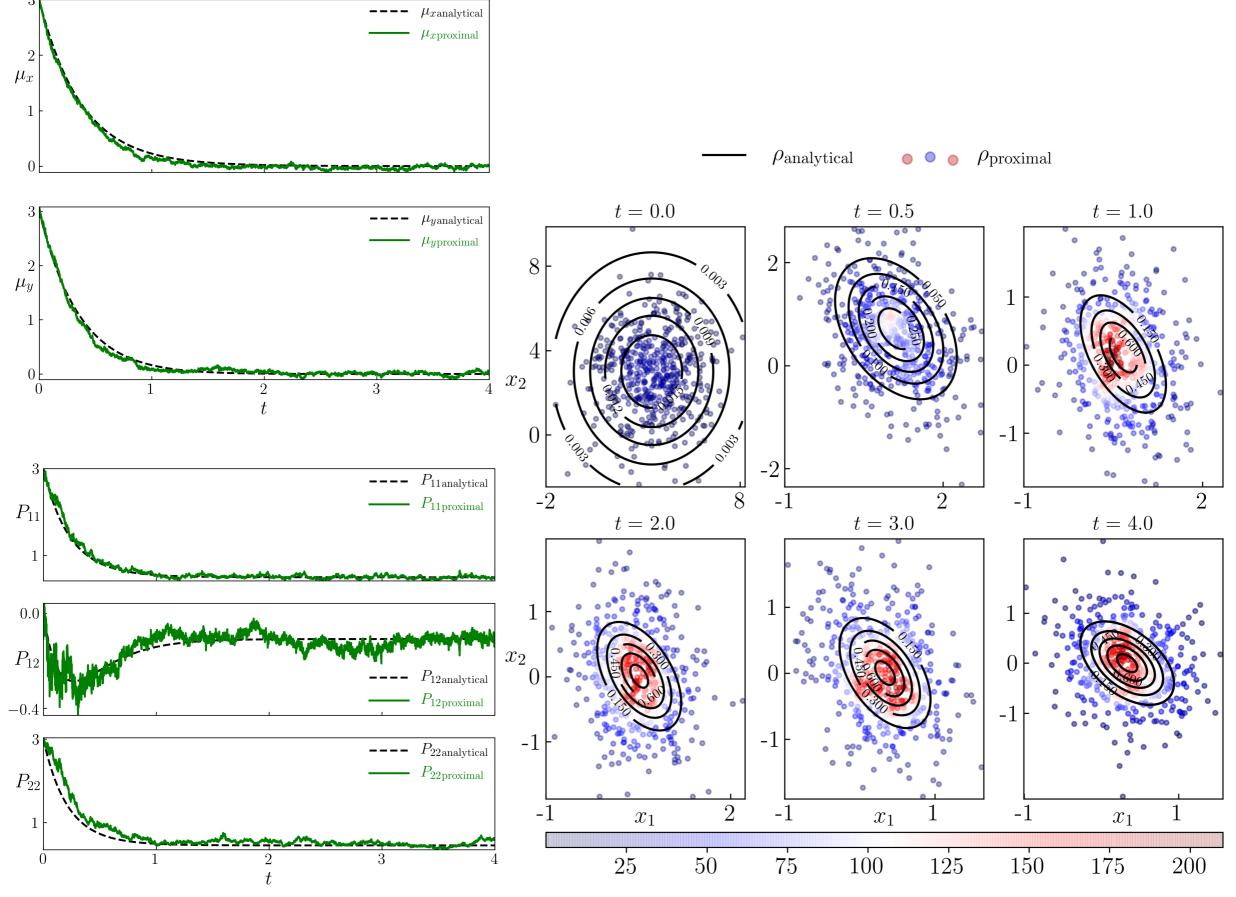
$$\begin{split} \mathbf{\Gamma}_{k} &= e^{\frac{-\mathbf{C}_{k}}{2\epsilon}} \longrightarrow \\ \mathbf{\mathcal{Q}}_{k-1} \longrightarrow \\ \mathbf{\mathcal{Q}}_{k-1} \longrightarrow \\ \mathbf{\mathcal{Z}}_{k-1} &\stackrel{\mathbf{\mathcal{Q}}_{k-1}}{\longrightarrow} \end{split} \qquad \begin{array}{c} \mathbf{y} \odot \mathbf{\Gamma}_{k}^{\mathsf{T}} \mathbf{z} = \mathbf{\mathcal{Q}}_{k-1} \\ \mathbf{z} \odot \mathbf{\Gamma}_{k}^{\mathsf{T}} \mathbf{y} = \mathbf{\mathcal{Z}}_{k-1} \xrightarrow{\mathbf{\mathcal{Z}}^{\mathsf{F}} \epsilon/2h} \end{array} \end{array} \longrightarrow \mathbf{\mathcal{Q}}_{k} = \mathbf{z} \odot \mathbf{\Gamma}_{k}^{\mathsf{T}} \mathbf{y}$$

**Theorem:** Consider the recursion on the cone  $\mathbb{R}_{\geq 0}^n \times \mathbb{R}_{\geq 0}^n$  $\mathbf{y} \odot (\mathbf{\Gamma}_k \mathbf{z}) = \mathbf{\varrho}_{k-1}, \quad \mathbf{z} \odot (\mathbf{\Gamma}_k^{\top} \mathbf{y}) = \mathbf{\xi}_{k-1} \odot \mathbf{z}^{-\frac{\beta\epsilon}{h}},$ Then the solution  $(\mathbf{y}^*, \mathbf{z}^*)$  gives the proximal update  $\mathbf{\varrho}_k = \mathbf{z}^* \odot (\mathbf{\Gamma}_k^{\top} \mathbf{y}^*)$ 

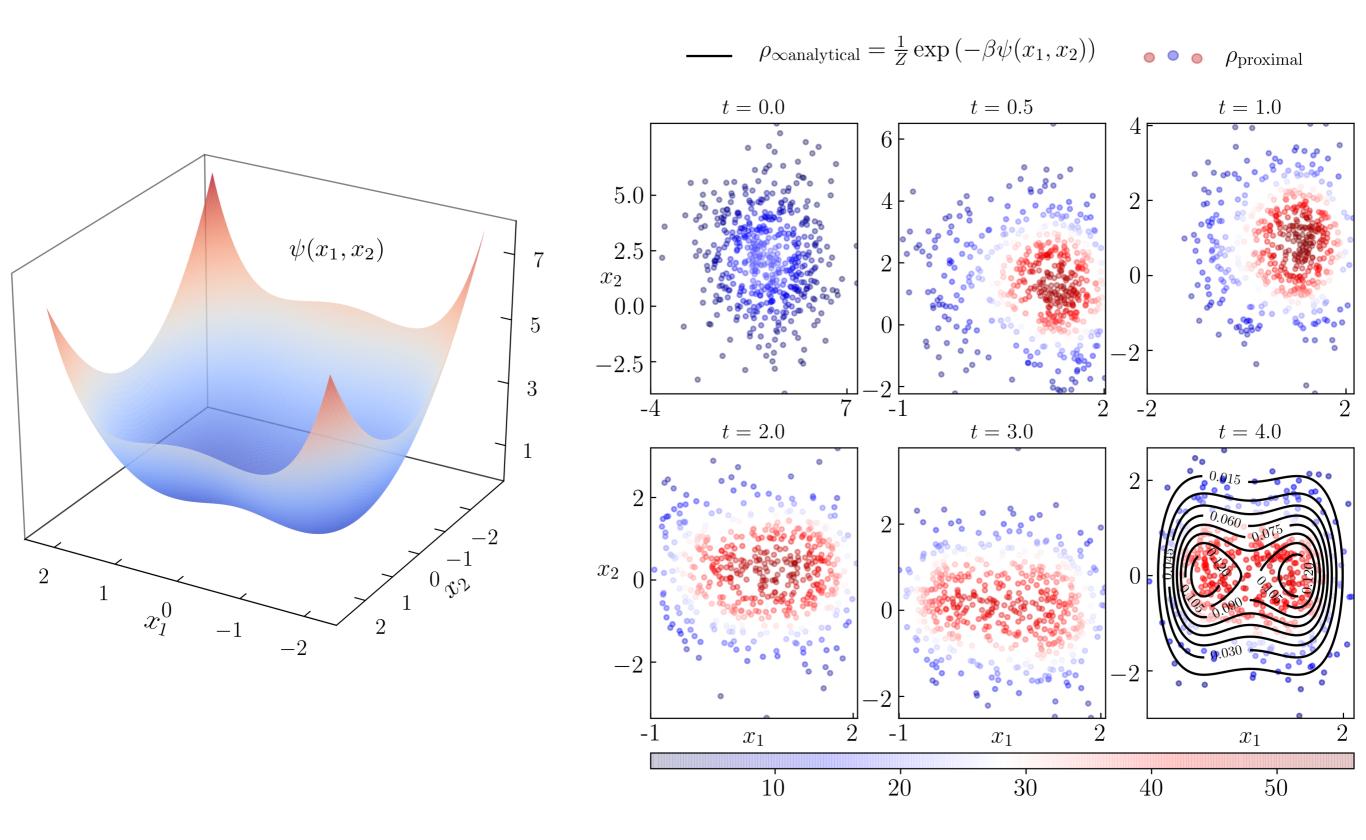
— K.F. Caluya and A.H., Gradient flow algorithms for density propagation in stochastic systems, *IEEE TAC* 2020.

 $\begin{array}{c} \operatorname{arg} \min \left[ \frac{c_{an}}{2} \right] \xrightarrow{\mathbf{r}}_{2} = \operatorname{arg} \mathbf{r} \\ x & 2 \end{array} \right] \xrightarrow{\mathbf{r}}_{2} = \operatorname{arg} \mathbf{r} \\ x & 2 \end{array} \right] \xrightarrow{\mathbf{r}}_{2} \xrightarrow{\mathbf{r}}_{2} = \operatorname{arg} \mathbf{r} \\ x & 2 \end{array} \right] \xrightarrow{\mathbf{r}}_{2} \xrightarrow{\mathbf{r}}_{$  $\begin{array}{c} \mathbf{x} \quad \mathbf{z} \\ \text{primization literature, the initial prime literature, the trapping lay <math>k_{k-1} \quad \mathbf{x}_{k}, \\ \mathbf{D}_{\mathrm{KI}} \quad (\mathrm{d}\pi_1 \parallel \mathrm{d}\pi_2) \quad = \quad \mathbf{y}_{\mathrm{I}} \left( \mathbf{x} \right) \quad \mathbf{y}_{\mathrm{I}} \left$ given by given by  $\mathbf{x}_{k-1}$  := arg minrox  $\lim_{k \to \infty} \frac{1}{2} \lim_{k \to \infty}$ the "proximal operator the "proximal preverse fulled NI and the suffice the oranges in settisty the triangle in neisated of the proximation of the proximal recursion Definition 2: The 2-Wasteristern Amethe Between the  $\{x_k\}$  generate  $\begin{array}{c} \{\boldsymbol{x}_{k}\} \text{ generation} \\ \boldsymbol{x}_{k} = \mathrm{prox}_{h\varphi}^{\|\cdot\|}(\boldsymbol{x}_{k-1}), \quad \boldsymbol{x}_{k} = 0, 1, 2, \dots, k_{(\mathfrak{F})}^{\|\cdot\|}(\boldsymbol{x}_{k-1}), \quad \boldsymbol{x}_{(\mathfrak{F})}^{\mathrm{definity}} \text{ measures } \boldsymbol{y}_{1}(\boldsymbol{x}_{k}) = \boldsymbol{y}_{1}(\boldsymbol{x}_{k}) =$  $x_k = x_k = x_k$  is denoted the flow of the flow of the ode  $(4)_{\mu}$  is the ode  $(4)_{\mu}$  is the ode  $(4)_{\mu}$  is the ode  $(4)_{\mu}$  is the sequence  $(4)_{\mu}$  is the sequen converges to satisfies  $x_k \rightarrow x(t = kh)$  as the step-size how in the step size to the step size how in the step sis size how in the st  $\begin{array}{c} \text{Inite dimensiv} \\ (\varrho_{k \text{prbx}})_{h \Phi}^{d^{2}}(\underset{\varrho \in \mathscr{D}_{2}}{\operatorname{arg inf}} \inf \frac{1}{2} d^{2}(\varrho_{k-1}) := \underset{\varphi \in \mathscr{D}_{2}}{\operatorname{arg inf}} \inf \frac{1}{2} d^{2}(\underset{\varphi, \varrho_{k-1}}{\operatorname{arg inf}}) := \underset{\varphi \in \mathscr{D}_{2}}{\operatorname{arg inf}} \inf \frac{1}{2} d^{2}(\underset{\varphi, \varrho_{k-1}}{\operatorname{arg inf}}) := \underset{W(\pi_{1}, \pi_{2}) := }{\operatorname{arg inf}} d^{2}(\underset{\varphi, \varrho_{k-1}}{\operatorname{arg inf}}) := \underset{W(\pi_{1}, \pi_{2}) := }{\operatorname{arg inf}} d^{2}(\underset{\varphi, \varrho_{k-1}}{\operatorname{arg inf}}) := \underset{W(\pi_{1}, \pi_{2}) := }{\operatorname{arg inf}} d^{2}(\underset{\varphi, \varrho_{k-1}}{\operatorname{arg inf}}) := \underset{W(\pi_{1}, \pi_{2}) := }{\operatorname{arg inf}} d^{2}(\underset{\varphi, \varrho_{k-1}}{\operatorname{arg inf}}) := \underset{W(\pi_{1}, \pi_{2}) := }{\operatorname{arg inf}} d^{2}(\underset{\varphi, \varrho_{k-1}}{\operatorname{arg inf}}) := \underset{W(\pi_{1}, \pi_{2}) := }{\operatorname{arg inf}} d^{2}(\underset{\varphi, \varrho_{k-1}}{\operatorname{arg inf}}) := \underset{W(\pi_{1}, \pi_{2}) := }{\operatorname{arg inf}} d^{2}(\underset{\varphi, \varrho_{k-1}}{\operatorname{arg inf}}) := \underset{W(\pi_{1}, \pi_{2}) := }{\operatorname{arg inf}} d^{2}(\underset{\varphi, \varrho_{k-1}}{\operatorname{arg inf}}) := \underset{W(\pi_{1}, \pi_{2}) := }{\operatorname{arg inf}} d^{2}(\underset{\varphi, \varrho_{k-1}}{\operatorname{arg inf}}) := \underset{W(\pi_{1}, \pi_{2}) := }{\operatorname{arg inf}} d^{2}(\underset{\varphi, \varrho_{k-1}}{\operatorname{arg inf}}) := \underset{W(\pi_{1}, \pi_{2}) := }{\operatorname{arg inf}} d^{2}(\underset{\varphi, \varrho_{k-1}}{\operatorname{arg inf}}) := \underset{W(\pi_{1}, \pi_{2}) := }{\operatorname{arg inf}} d^{2}(\underset{\varphi, \varrho_{k-1}}{\operatorname{arg inf}}) := \underset{W(\pi_{1}, \pi_{2}) := }{\operatorname{arg inf}} d^{2}(\underset{\varphi, \varrho_{k-1}}{\operatorname{arg inf}}) := \underset{W(\pi_{1}, \pi_{2}) := }{\operatorname{arg inf}} d^{2}(\underset{\varphi, \varrho_{k-1}}{\operatorname{arg inf}}) := \underset{W(\pi_{1}, \pi_{2}) := }{\operatorname{arg inf}} d^{2}(\underset{\varphi, \varrho_{k-1}}{\operatorname{arg inf}}) := \underset{W(\pi_{1}, \pi_{2}) := }{\operatorname{arg inf}} d^{2}(\underset{\varphi, \varrho_{k-1}}{\operatorname{arg inf}}) := \underset{W(\pi_{1}, \pi_{2}) := }{\operatorname{arg inf}} d^{2}(\underset{\varphi, \varrho_{k-1}}{\operatorname{arg inf}}) := \underset{W(\pi_{1}, \pi_{2}) := }{\operatorname{arg inf}} d^{2}(\underset{\varphi, \varrho_{k-1}}{\operatorname{arg inf}}) := \underset{W(\pi_{1}, \pi_{2}) := }{\operatorname{arg inf}} d^{2}(\underset{\varphi, \varrho_{k-1}}{\operatorname{arg inf}}) := \underset{W(\pi_{1}, \pi_{2}) := }{\operatorname{arg inf}} d^{2}(\underset{\varphi, \varrho_{k-1}}{\operatorname{arg inf}}) := \underset{W(\pi_{1}, \pi_{2}) := }{\operatorname{arg inf}} d^{2}(\underset{\varphi, \varrho_{k-1}}{\operatorname{arg inf}}) := \underset{W(\pi_{1}, \pi_{2}) := }{\operatorname{arg inf}} d^{2}(\underset{\varphi, \varrho_{k-1}}{\operatorname{arg inf}}) := \underset{W(\pi_{1}, \pi_{2}) := }{\operatorname{arg inf}} d^{2}(\underset{\varphi, \varrho_{k-1}}{\operatorname{arg inf}}) := \underset{W(\pi_{1}, \pi_{2}) := }{\operatorname{arg inf}} d^{2}(\underset{\varphi, \varrho_{k-1}}{\operatorname{arg inf}}) := \underset{W(\pi_{1}, \pi_{2}) := }{\operatorname{arg inf}} d^{2}(\underset{\varphi, \varrho_{k-1}}{\operatorname{arg inf}}) := \underset{W(\pi_{1}$  $W(\pi_1,\pi_2):=$  $\begin{array}{c} (\underline{x},\underline{y}) \mapsto \underline{x}_{h\Phi}(\underline{y}_{\underline{\rho}} \in \mathcal{D}_{2} = 2 \\ \text{as an infinite dimensional proximal operator. As mentioned} \\ (\underline{y}_{\underline{\rho}} \in \underline{y}_{\underline{\rho}} \in \underline{y}_{\underline{\rho}} = 1 \\ (\underline{y}_{\underline{\rho}} = 1 \\ (\underline{y}_{\underline{\rho} = 1$ 3) unoinverge scto sequence satisfies  $PDE_{k}(4)$  i.e., the <u>kh</u> as the step-size where  $II(\pi_1, \pi_2)$  denotes the constraint sequence satisfies  $Q_k(x)$  where  $II(\pi_1, \pi_2)$  denotes the sequence satisfies  $Q_k(x)$  and  $PDE_{k}(4)$  i.e., the <u>kh</u> as the step-size denotes the solution of all approximations of the step-size denotes the solution of the step-size denotes the solution of the step-size denotes the solution of the step-size denotes the step-size denotes the solution of the step-size denotes denotes the solution of the step-size denotes denotes the solution of the step-size denotes denote ma finited e halsour by Slason **Theorem:** Block co-ordinate iteration of (**y**, **z**) recur- $\frac{\mathrm{d}}{\mathrm{d}t}\varphi =$ t-20 topk sion is contractive on  $\mathbb{R}^n_{>0} \times \mathbb{R}^n_{>0}$ . natt on the plies s a me trom the fact that the generative of a direct non-second can be appropriate to choos in a provinte dependent of the paper of the paper

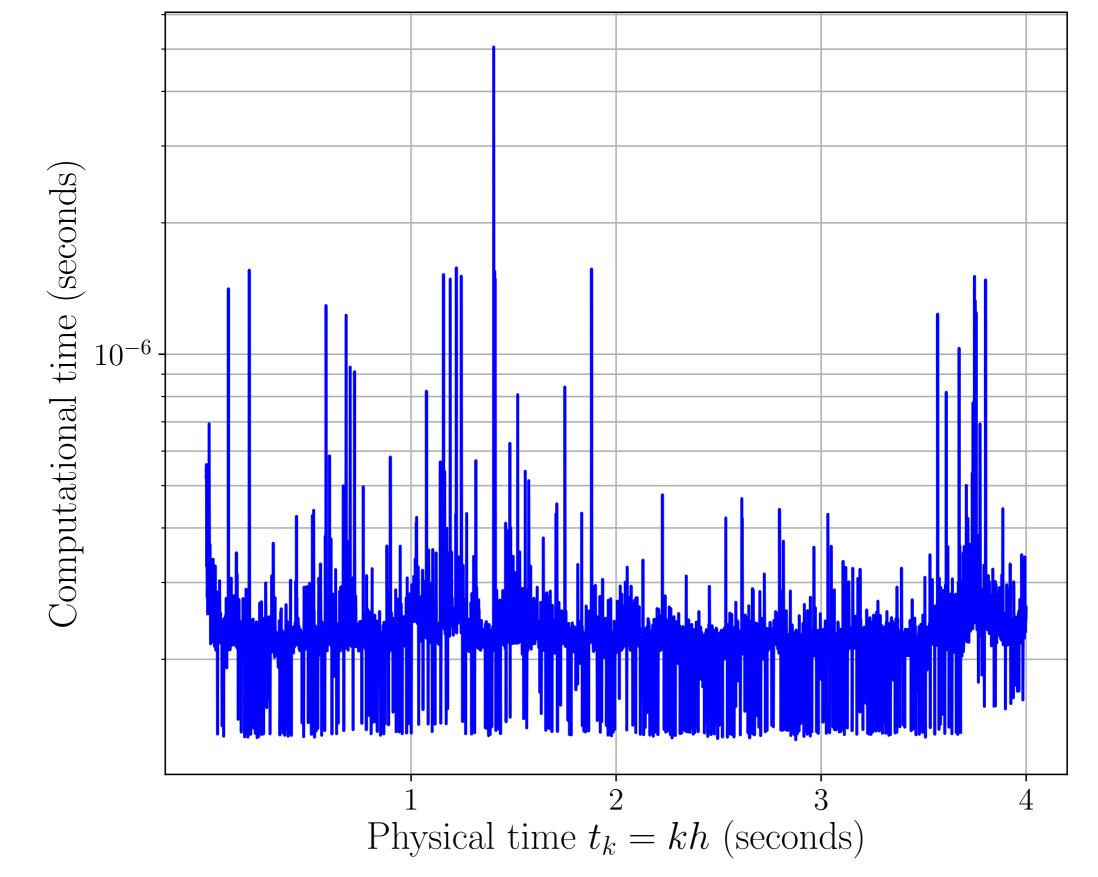
#### **Proximal Prediction: 2D Linear Gaussian**



#### **Proximal Prediction: Nonlinear Non-Gaussian**



#### **Computational Time: Nonlinear Non-Gaussian**



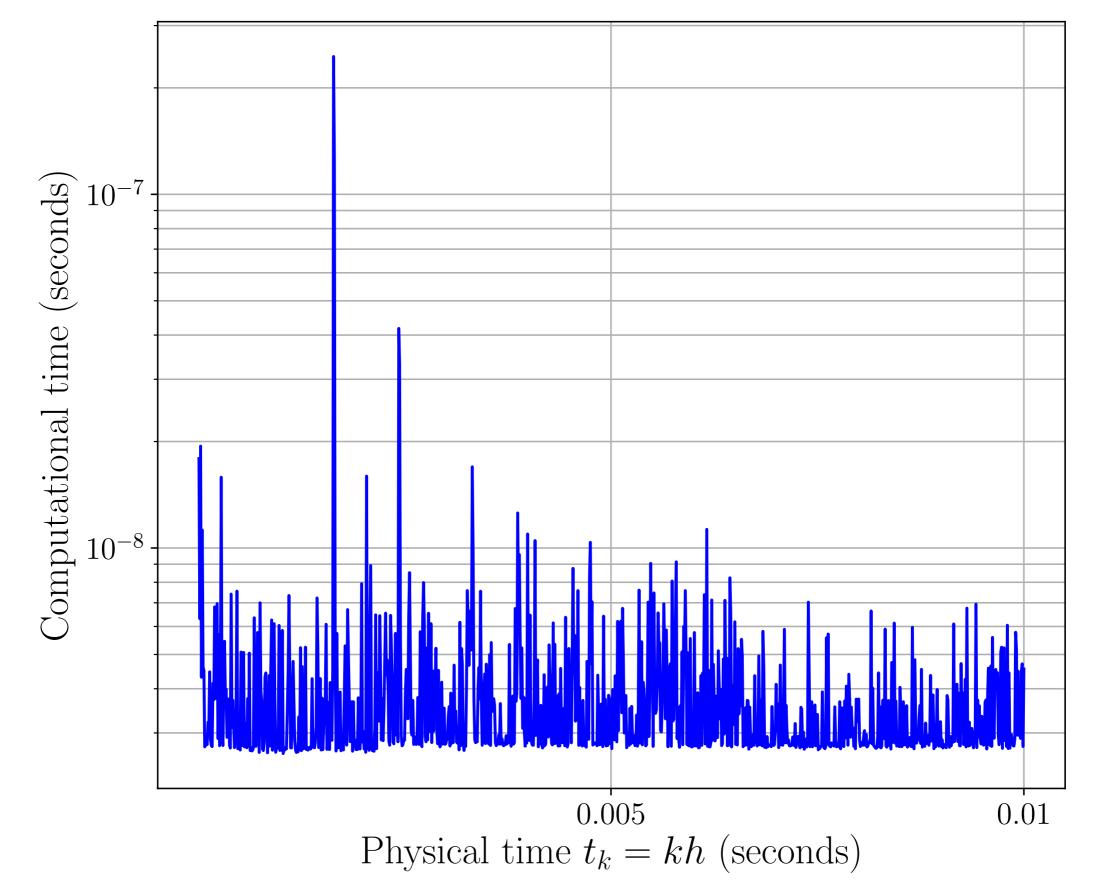
#### **Proximal Prediction: Satellite in Geocentric Orbit**

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

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

$$\begin{pmatrix} f_x \\ f_y \\ f_z \end{pmatrix}_{\text{pert}} = \begin{pmatrix} s\theta \ c\phi \ c\theta \ c\phi \ -s\phi \\ s\theta \ s\phi \ c\theta \ s\phi \ c\phi \\ c\theta \ -s\theta \ 0 \end{pmatrix} \begin{pmatrix} \frac{k}{2r^4} \left(3(s\theta)^2 - 1\right) \\ -\frac{k}{r^5}s\theta \ c\theta \\ 0 \end{pmatrix}, k := 3J_2 R_{\text{E}}^2, \mu = \text{constant}$$

#### **Computational Time: Satellite in Geocentric Orbit**



#### **Extensions: Nonlocal Interactions**

**PDF dependent sample path dynamics:**  $d\mathbf{x} = -\left(\nabla U\left(\mathbf{x}\right) + \nabla \rho * V\right) dt + \sqrt{2\beta^{-1}} d\mathbf{w}$ 

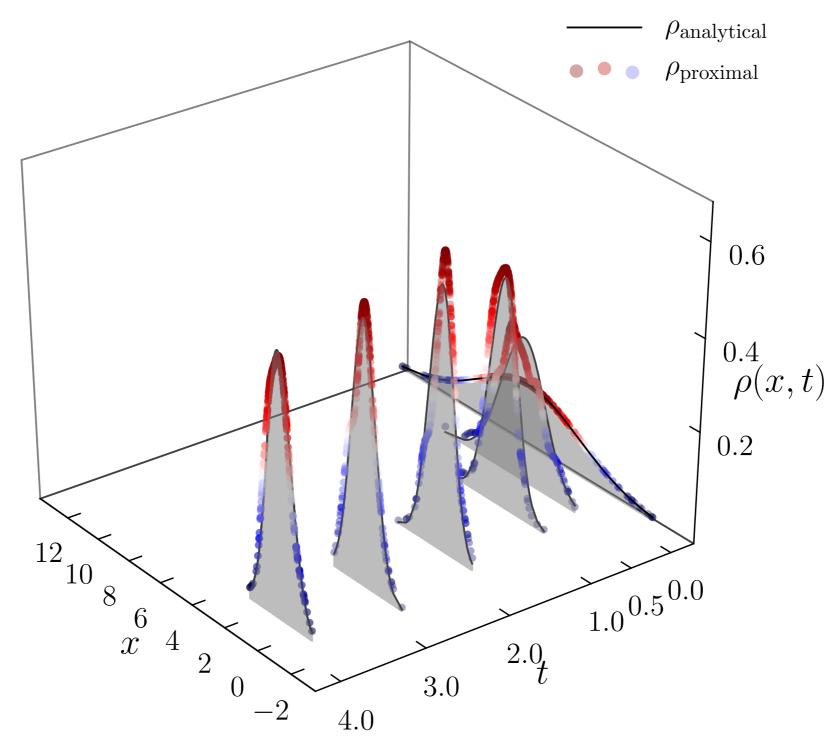
Mckean-Vlasov-Fokker-Planck-  
Kolmogorov integro PDE:  
$$\frac{\partial \rho}{\partial t} = \nabla \cdot \left(\rho \nabla \left(U + \rho * V\right)\right) + \beta^{-1} \Delta \rho$$

Free energy:  

$$F(\rho) := \mathbb{E}_{\rho} \left[ U + \beta^{-1} \rho \log \rho + \rho * V \right]$$

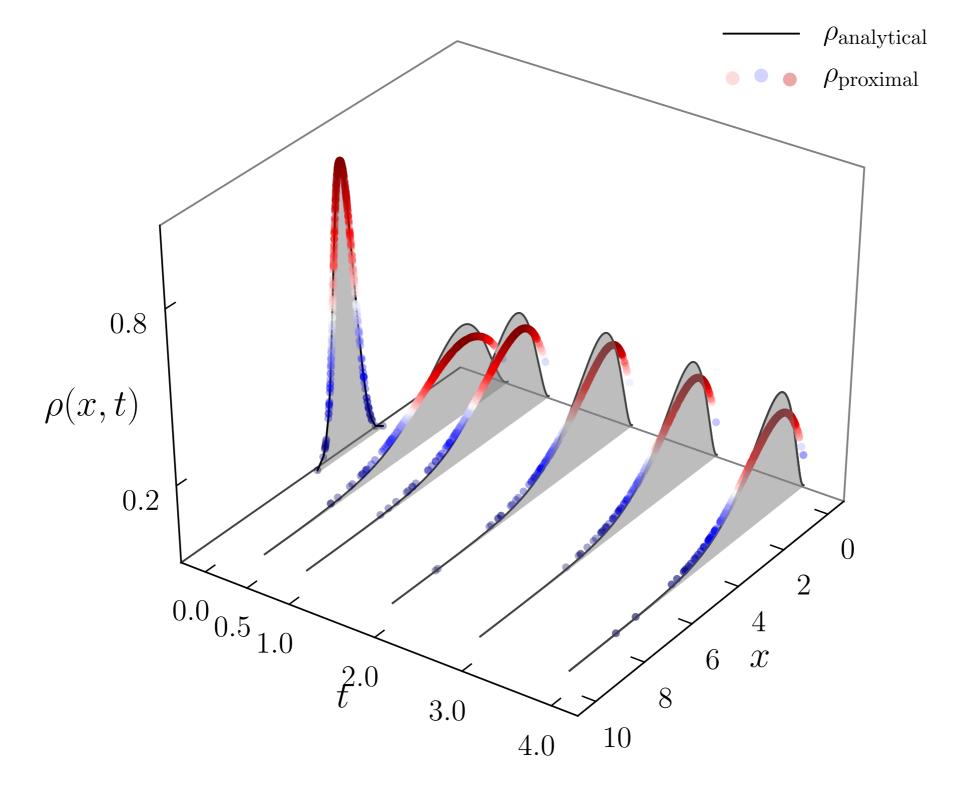
#### **Extensions: Nonlocal Interactions**

$$U(\cdot) = V(\cdot) = \|\cdot\|_2^2$$



#### **Extensions: Multiplicative Noise**

**Cox-Ingersoll-Ross:**  $dx = a(\theta - x) dt + b\sqrt{x} dw$ ,  $2a > b^2$ ,  $\theta > 0$ 



## 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 • stochastic} proximal operators:** 

$$\begin{array}{c}
\varrho_{k-1}^{+} & \rho_{k}^{-} & \rho_{k}^{-} & \rho_{k}^{+} & \rho_{k}^{+} & \rho_{k}^{+} & \rho_{k}^{+} & \rho_{k}^{+} & \rho_{k}^{+} & \rho_{k}^{-} & \rho_{k}^{-$$

 $\textbf{Convergence:} \hspace{0.2cm} \varrho_{k}^{+}(h) \rightarrow \rho^{+}(x,t=kh) \hspace{0.2cm} \text{as} \hspace{0.2cm} h \downarrow 0$ 

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

For posterior:  $d^+ \equiv d_{ ext{FR}}^2$  or  $D_{ ext{KL}}, \quad \Phi^+ \underset{_{29}}{\equiv} \; rac{1}{2} \mathbb{E}_{arrho^+} \Big[ (y_k - h(x))^ op R^{-1} (y_k - h(x)) \Big]$ 

### **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:

$$\begin{aligned} \mathbf{P}^{+}\mathbf{C}\mathbf{R}^{-1} \\ \mathbf{I} \\ \mathbf{d}\mu^{+}(t) &= \mathbf{A}\mu^{+}(t)\mathbf{d}t + \mathbf{K}(t) \quad (\mathbf{d}\mathbf{z}(t) - \mathbf{C}\mu^{+}(t)\mathbf{d}t), \\ \dot{\mathbf{P}}^{+}(t) &= \mathbf{A}\mathbf{P}^{+}(t) + \mathbf{P}^{+}(t)\mathbf{A}^{\top} + \mathbf{B}\mathbf{Q}\mathbf{B}^{\top} - \mathbf{K}(t)\mathbf{R}\mathbf{K}(t)^{\top}. \end{aligned}$$

— 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:

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

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:** 

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

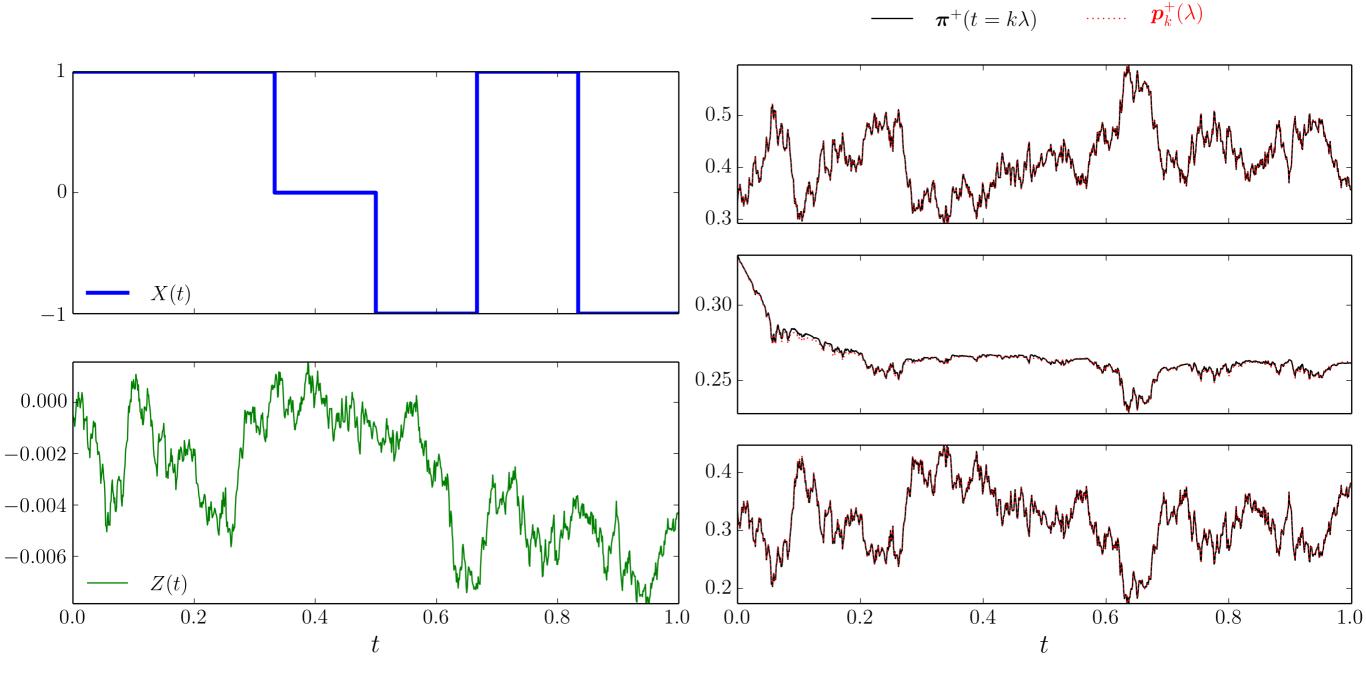
where  $H := ext{diag}(h(a_1), \dots, h(a_m)), \quad \widehat{h}(t) := \sum_{i=1}^m h(a_i) \pi_i^+(t),$ 

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

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

— 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**

**Process noise**  $\rho_0$  $\rho_1$ **Steer joint state PDF via feedback** Nonlinear system control over finite time horizon  $\boldsymbol{u}$ Density controller Common scenario:  $G \equiv B$  $\underset{u \in \mathcal{U}}{\text{minimize}} \quad \mathbb{E} \left[ \int_0^1 \left( \frac{1}{2} \| u(t, x_t^u) \|_2^2 + q(t, x_t^u) \right) dt \right]$ subject to  $\mathrm{d} x_t^u = \{f(t, x_t^u) + B(t, x_t^u)u\}\mathrm{d} t + \sqrt{2}G(t, x_t^u)\mathrm{d} w_t$  $x_0^u := x_t^u(t=0) \sim \rho_0, \quad x_1^u := x_t^u(t=1) \sim \rho_1$ 

 $\boldsymbol{w}(t)$ 

 $ho(oldsymbol{x},t)$ 

 $oldsymbol{x}$ 

#### **Optimal Control Problem over PDFs**

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

$$P_{o1} := \left\{ \begin{array}{c} & & \\ P_{o1} & & \\ P_{o} & & \\ \end{array} \right\}$$

Hessian operator w.r.t. state: Hess

$$\inf_{\substack{(\rho,u)\in\mathcal{P}_{01}\times\mathcal{U}\\ \forall t \in \mathcal{P}_{01}\times\mathcal{U}}} \int_{\mathbb{R}^{n}} \int_{0}^{1} \left(\frac{1}{2} \|u(t, x_{t}^{u})\|_{2}^{2} + q(t, x_{t}^{u})\right) \rho(t, x_{t}^{u}) dt dx_{t}^{u} dt dx_{t}^{u}$$
subject to
$$\frac{\partial\rho}{\partial t} + \nabla \cdot \left((f + Bu)\rho\right) = \langle \text{Hess}, D\rho \rangle$$

$$\rho(t = 0, x_{0}^{u}) = \rho_{0}, \quad \rho(t = 1, x_{1}^{u}) = \rho_{1}$$

### **Necessary Conditions of Optimality** (Assuming $G \equiv B$ )

#### **Coupled nonlinear PDEs + linear boundary conditions**

**Controlled Fokker-Planck or Kolmogorov's forward PDE** 

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

#### Hamilton-Jacobi-Bellman-like PDE

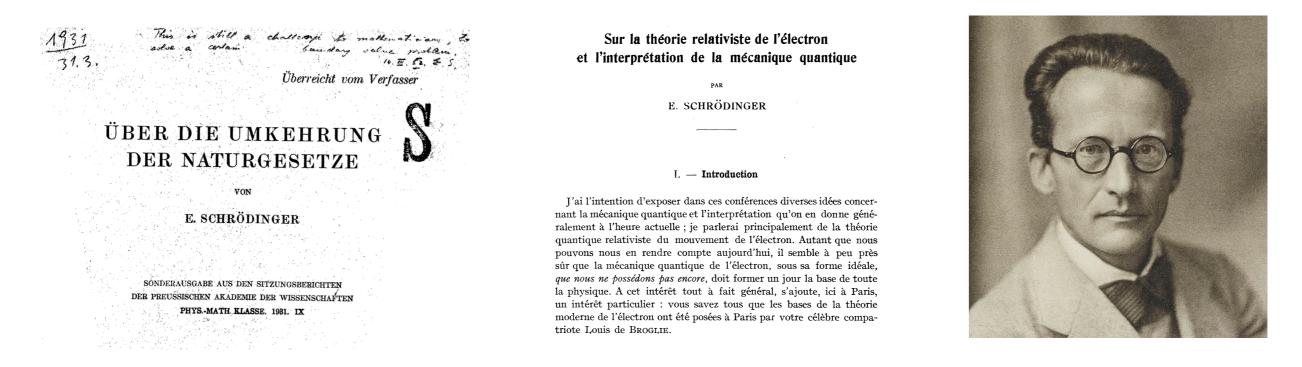
$$\frac{\partial \psi}{\partial t} + \langle \nabla \psi, f \rangle + \langle D, \text{Hess}(\psi) \rangle + \frac{1}{2} \langle \nabla \psi, D \nabla \psi \rangle = q$$

**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



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

 $(\rho^{\text{opt}}, \psi) \mapsto (\widehat{\varphi}, \varphi) -$ Schrödinger factors

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

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

### Feedback Synthesis via the Schrödinger System

2 coupled nonlinear PDEs  $\rightarrow$  boundary-coupled linear PDEs!!

Uncontrolled forward-backward Kolmogorov PDEs:  

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

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

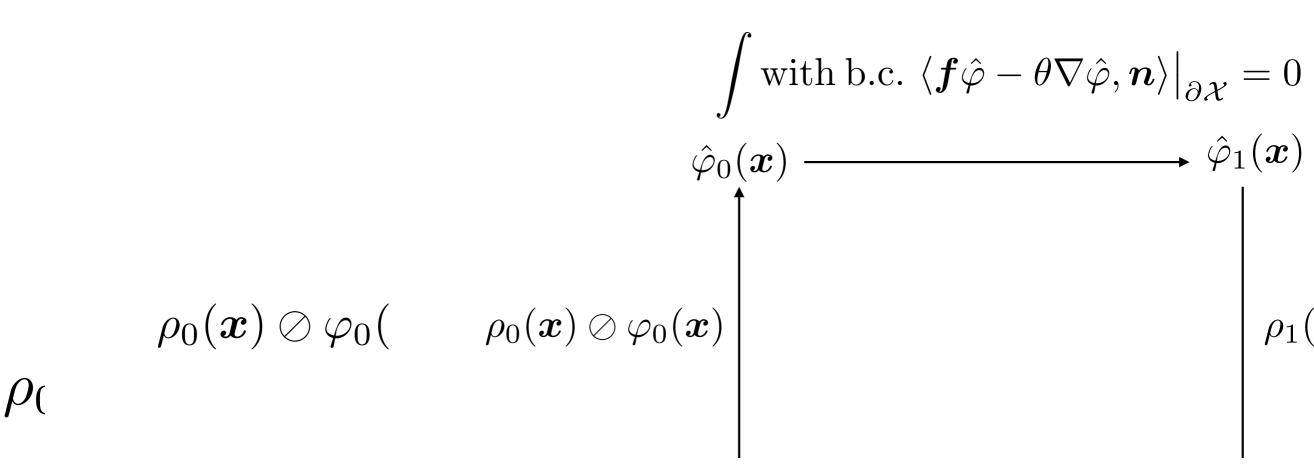
**Optimal controlled joint state PDF:** 

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

**Optimal control:** 

$$\boldsymbol{u}^{\text{opt}}(\boldsymbol{x},t) = 2\boldsymbol{B}^{\top}\nabla_{\boldsymbol{x}}\log\varphi(\boldsymbol{x},t)$$

# Fized Point Recursion (kgr, khe pair

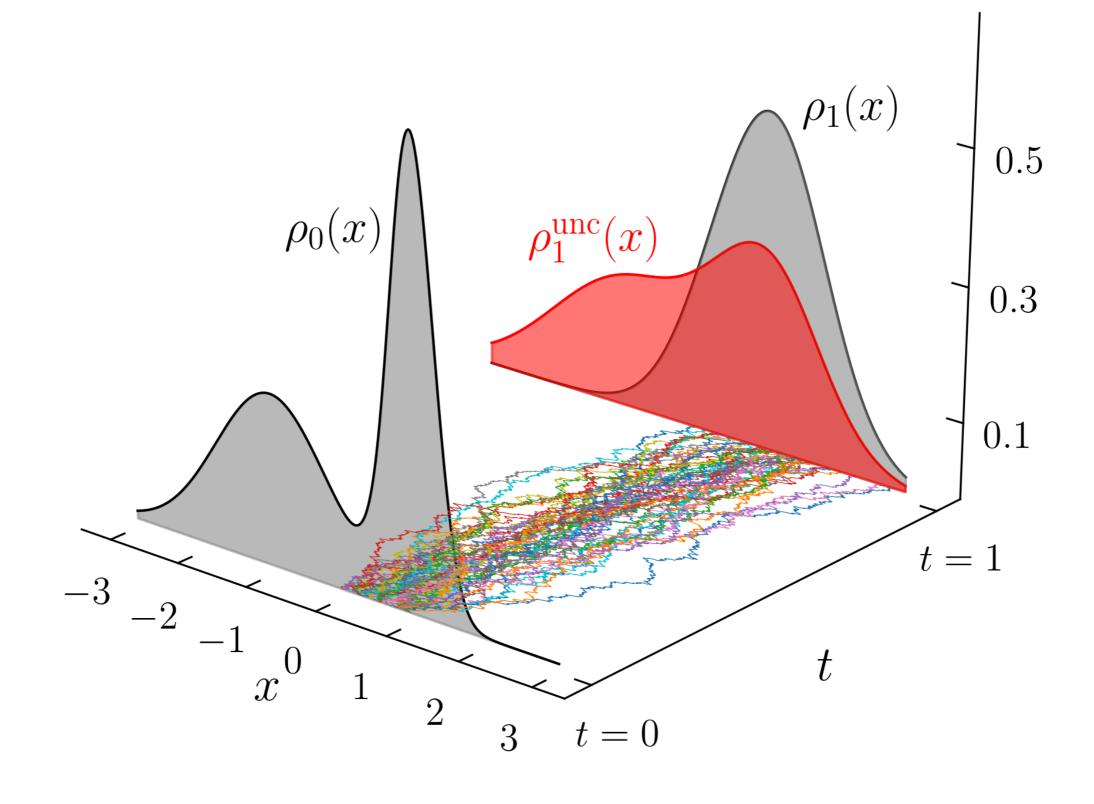


### Fixed Point Recursion for the pair

$$\int \text{with b.c. } \left\langle \boldsymbol{f}\hat{\varphi} - \theta \nabla \hat{\varphi}, \boldsymbol{n} \right\rangle \Big|_{\partial \mathcal{X}} = 0$$

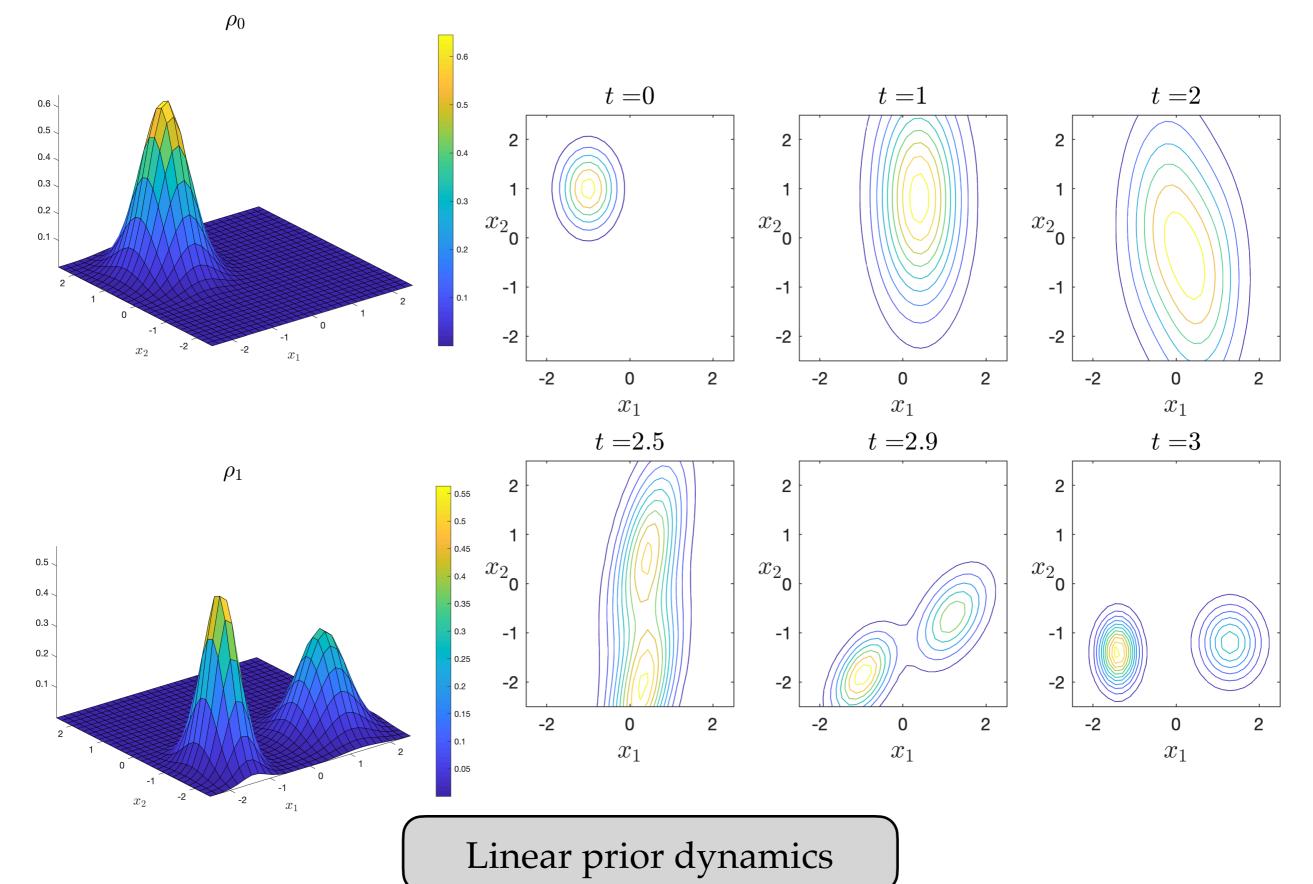
 $\hat{\varphi}_0(\boldsymbol{x}) \longrightarrow \hat{\varphi}_1(\boldsymbol{x})$ This recursion is contractive in the Hilbert metric!!

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



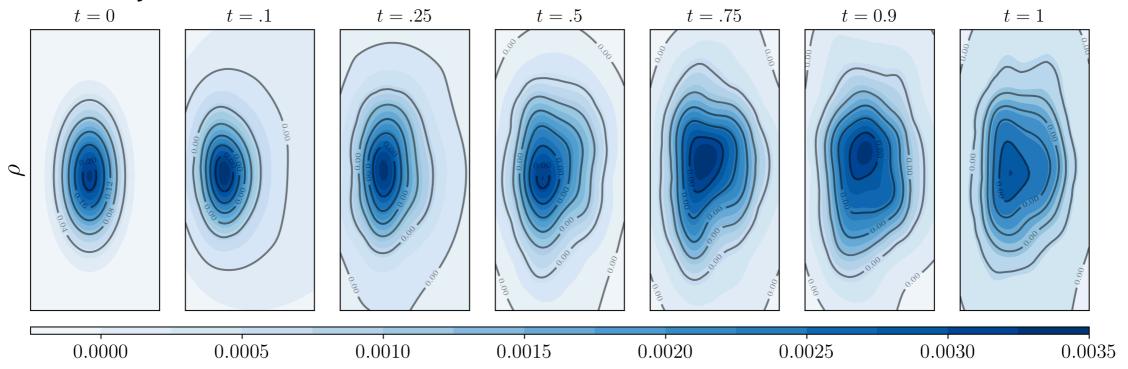
Zero prior dynamics

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

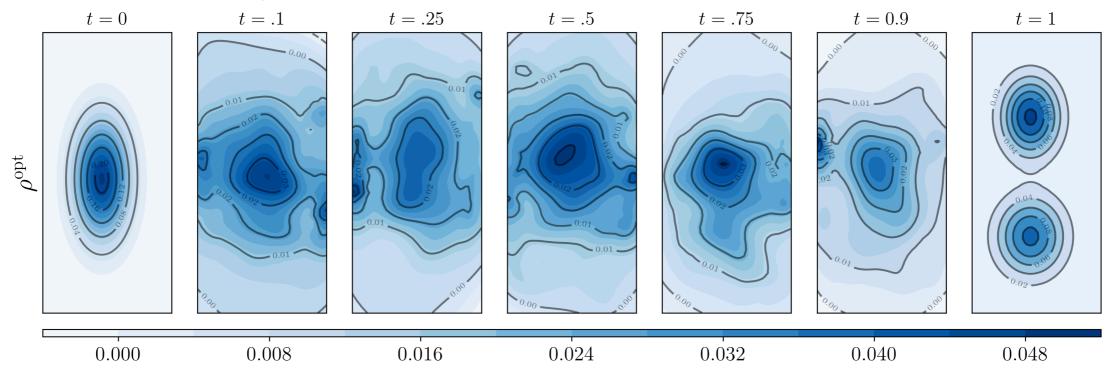


## Feedback Density Control: Nonlinear Grad. Drift

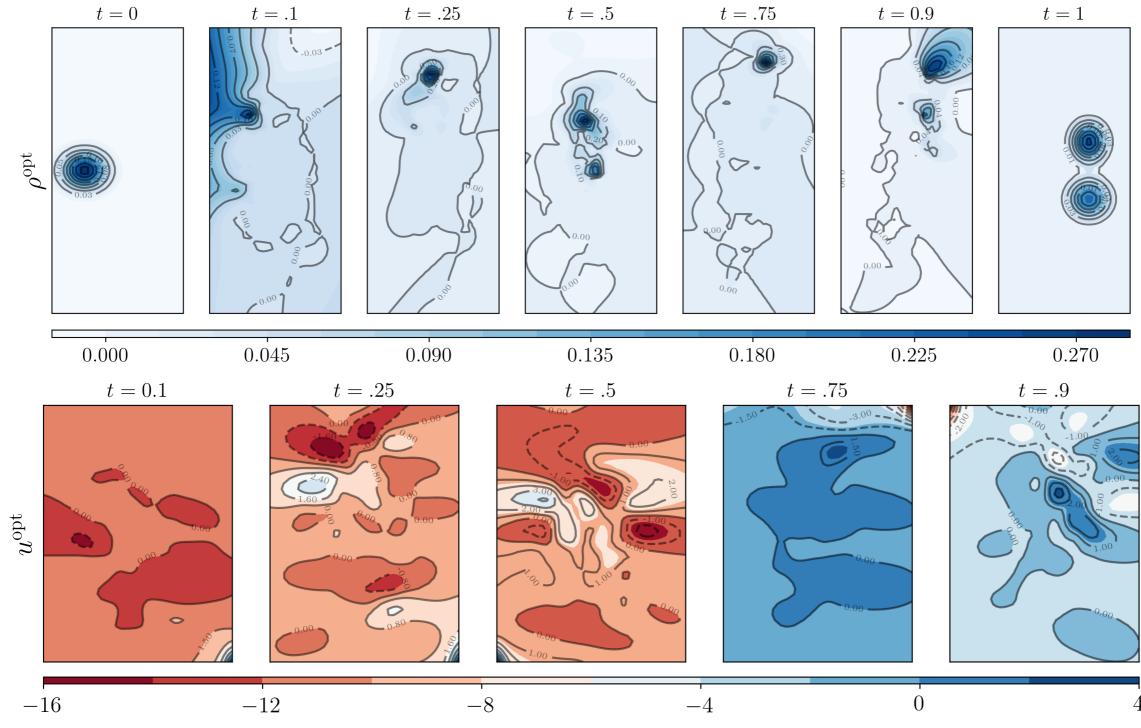
### **Uncontrolled joint PDF evolution:**



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

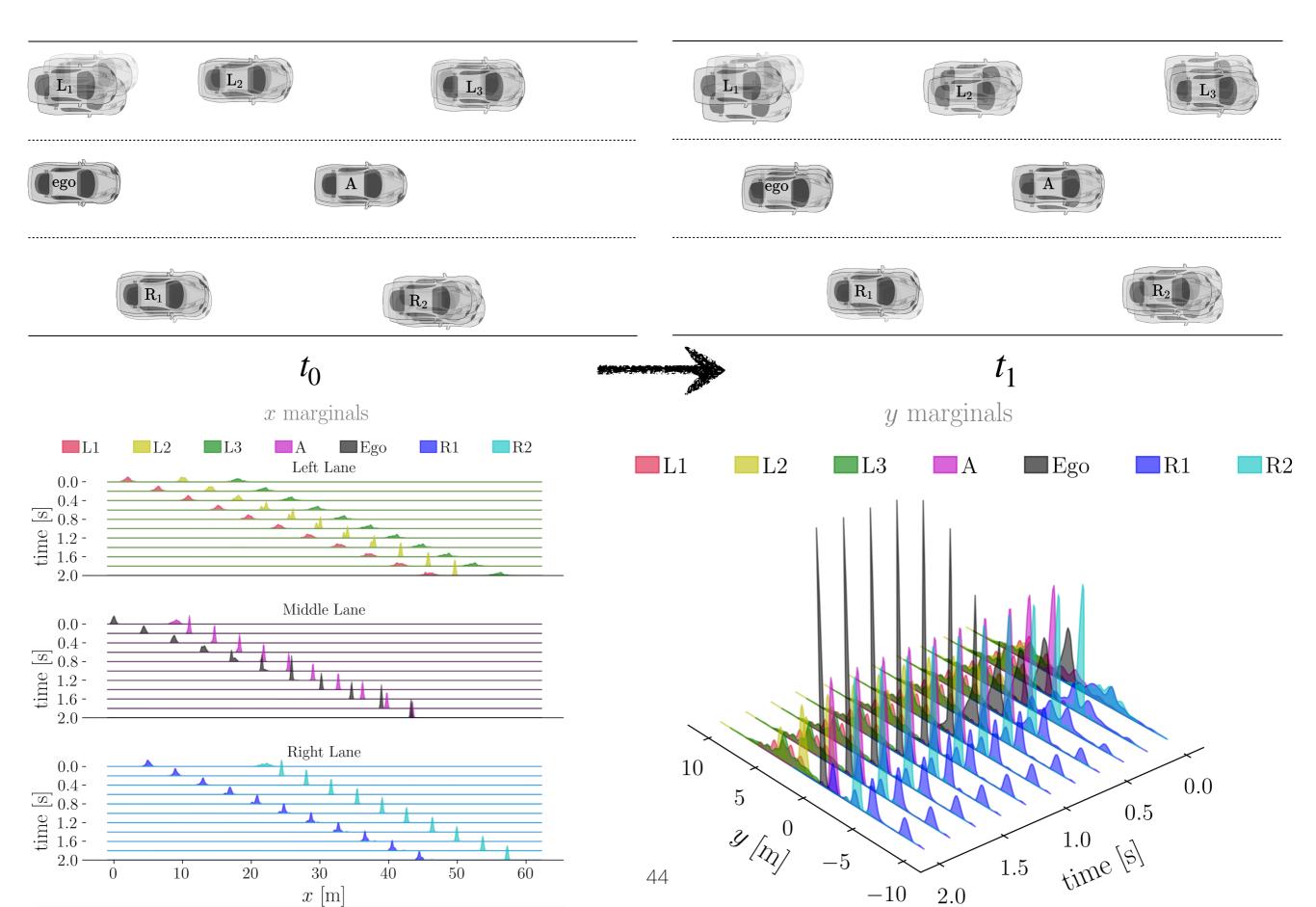


## Feedback Density Control: Mixed Conservative-Dissipative Drift

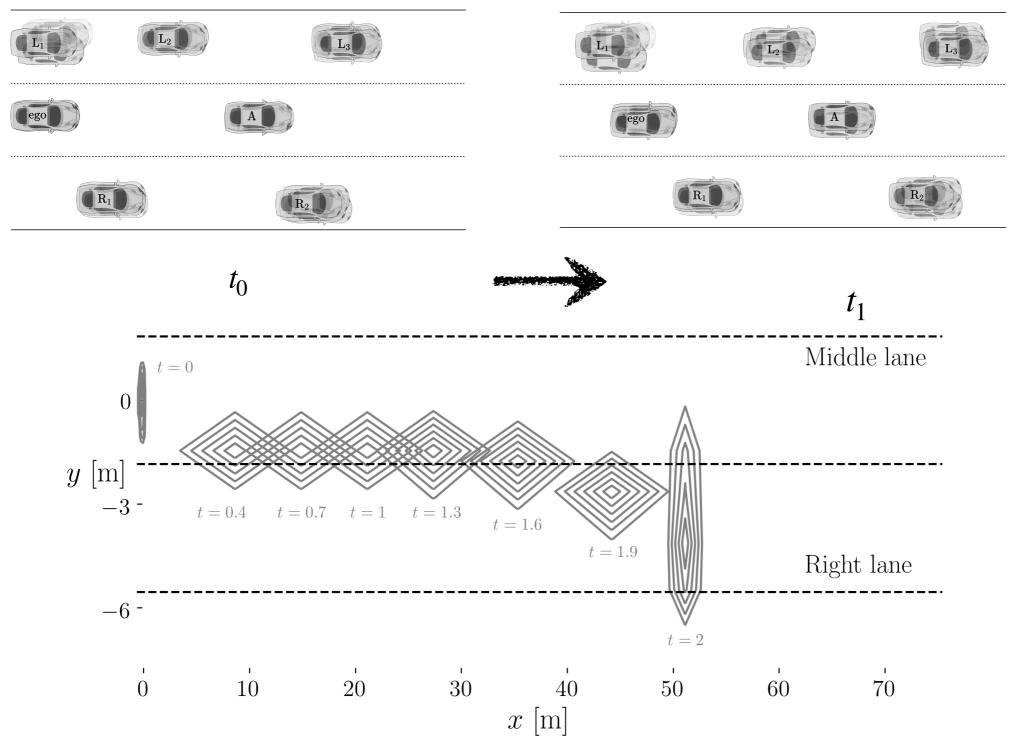


— K.F. Caluya and A.H., Wasserstein proximal algorithms for the Schrodinger bridge problem: density control with nonlinear drift, *IEEE TAC* 2021.

## **Density Prediction for Safe Automated Driving**



### **Density Control for Safe Automated Driving**



— S. Haddad, A.H., and B. Singh, Density-based stochastic reachability computation for occupancy prediction in automated driving, *IEEE TCST* 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 LCSS* 2021.

# Learning a neural network as generalized gradient flow

### Learning Neural Network from Data

 $( ext{feature vector, label}) = (oldsymbol{x}_i, y_i) \in \mathbb{R}^d imes \mathbb{R}, \quad i = 1, \dots, n$ 

Consider shallow NN: 1 hidden layer with  $n_{\rm H}$  neurons

NN parameter vector  $\boldsymbol{\theta} := (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \dots, \boldsymbol{\theta}_{n_{\mathrm{H}}})^\top \in \mathbb{R}^{pn_{\mathrm{H}}}$ 

Approximating function:

$$\hat{f}\left(oldsymbol{x},oldsymbol{ heta}
ight) = rac{1}{n_{ ext{H}}}\sum_{i=1}^{n_{ ext{H}}} \Phi(oldsymbol{x},oldsymbol{ heta}_i), ext{ example: } \Phi(oldsymbol{x},oldsymbol{ heta}_i) = a_i \sigmaig(oldsymbol{w}_i^ opoldsymbol{x}+b_iig)$$

Population risk functional:

### **Mean Field Density Dynamics of SGD**

Free energy functional:  $F(\rho) := R(\hat{f}(\boldsymbol{x}, \rho))$ 

For quadratic loss:

$$F(\rho) = \underbrace{F_0}_{\text{independent of } \rho} + \underbrace{\int_{\mathbb{R}^p} V(\theta) \rho(\theta) d\theta}_{\text{advection potential energy, linear in } \rho} + \underbrace{\int_{\mathbb{R}^p} \int_{\mathbb{R}^p} U(\theta, \tilde{\theta}) \rho(\theta) \rho(\tilde{\theta}) d\theta d\tilde{\theta}}_{\text{interaction potential energy, nonlinear in } \rho}$$

### where

$$F_0 := \mathbb{E}_{(\boldsymbol{x},y)} [y^2], \quad V(\boldsymbol{\theta}) := \mathbb{E}_{(\boldsymbol{x},y)} [-2y\Phi(\boldsymbol{x},\boldsymbol{\theta})],$$

PDF dynamics for SGD:

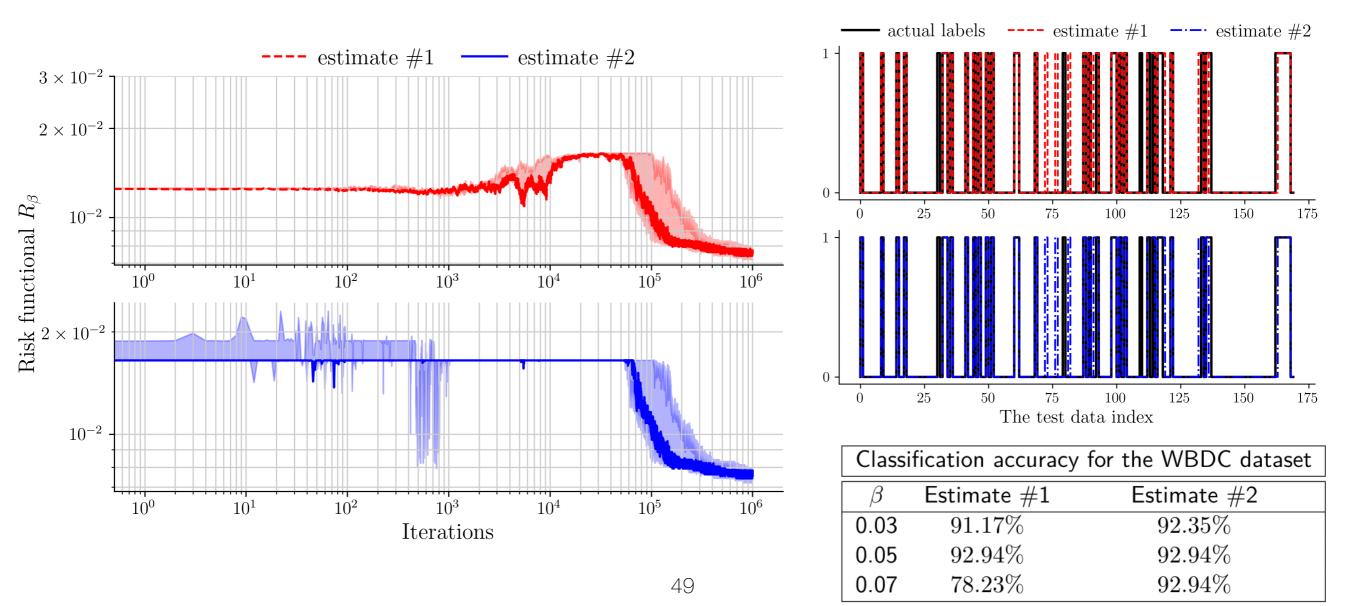
$$\frac{\partial \rho}{\partial t} = \nabla \cdot (\rho \nabla (V + U \circledast \rho)), \text{ where } (U \circledast \rho)(\boldsymbol{\theta}) := \int_{\mathbb{R}^p} U(\boldsymbol{\theta}, \tilde{\boldsymbol{\theta}}) \rho(\tilde{\boldsymbol{\theta}}) \mathrm{d}\tilde{\boldsymbol{\theta}}$$
$$\frac{\delta F}{\delta \rho}$$

This PDE is the gradient flow of functional F w.r.t. the Wasserstein metric W

## **Proximal Recursion for SGD Training of NN**

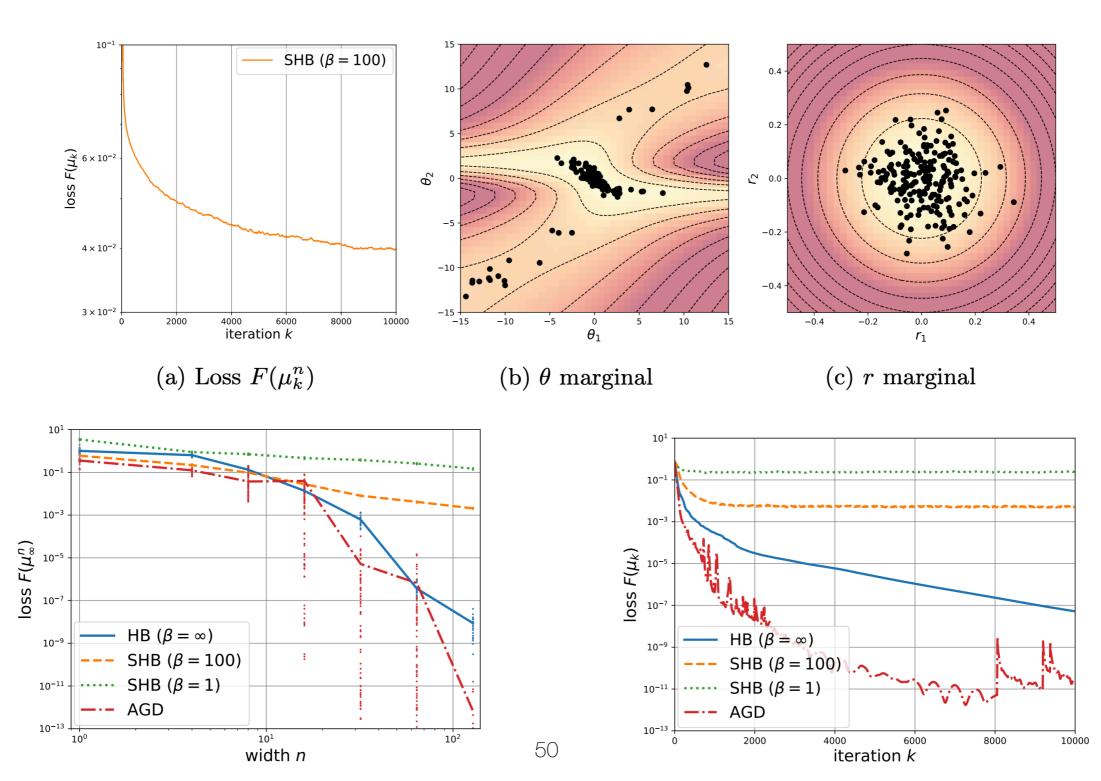
$$\varrho_{k}(\tau, \boldsymbol{\theta}) = \underset{\varrho \in \mathcal{P}(\mathbb{R}^{p})}{\operatorname{arg\,min}} \frac{1}{2} \left( W \left( \varrho(\boldsymbol{\theta}), \varrho_{k-1}(\tau, \boldsymbol{\theta}) \right) \right)^{2} + \tau F(\varrho(\boldsymbol{\theta}))$$
$$= \operatorname{prox}_{\tau F}^{W} \left( \varrho_{k-1} \right)$$

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



### Mean Field Density Dynamics of Stoc. Heavy Ball

$$\partial_t \mu_t = -\nabla \cdot \left[ \mu_t \cdot \left( \begin{array}{c} r \\ -\nabla F'([\mu_t]^{\theta}) - \gamma r \end{array} \right) \right] + \gamma \beta^{-1} \Delta_r \mu_t$$



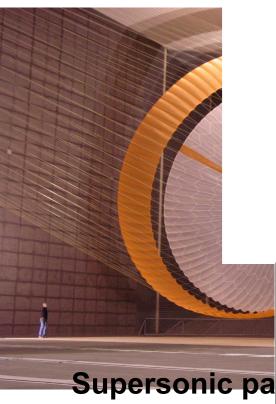
## Filtering F

Navigational

uncertainty

Heating uncertain

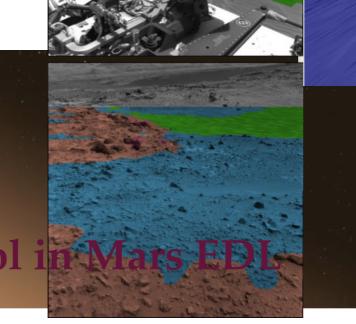
edit: NASA IPI



**Figure 2**. Mars Science Laborato joing full-scale wind-tunnel testing PDF control

Heliver such a large and capable r cally compelling site, which is rich and preserve biomarkers, presents chalEigeureotTonlyExample terrain requirements are Mso more 2012, MSL will enter the Martian at aeroshell ever flown to Mars, fly th at Mars, generate a higher hypersc

any previous Marsing and decelerate behind the largest and SITATI TOCKS subvy cluder ent Classes supersonic parachute ever deployed at Mars. The MSL EDL supersonic parachule ever deployed at Wars. The WSL EDL pose of proposing and selecting of proposing a 300ectly and exprove the state of the state Parachile Belief didingingars we found that in practices while belief and the second and a second and the secon convergence. This was especially type given the small size of for this, we developed a pre-processing scheme for scale-norm



of classifier output on terrain images from MSL Navcam. On the left are some of the thanindiwidura Maraning same high affect heating class if a tight in Qn

nute GB parachute under-

safely to a scientifi-

cally compelling site, which is rich in milieral Blikely to trap and preserve biomarkers, presents a myriad of engineering 2thallenges. Not only is the payload mass significantly larger than all previous Mars missions, the delivery accuracy and than all previous Mars missions, the delivery accuracy and Smooterraw requirements for Mso more stringent. In August of 1.9012, MSL will enter the Martian atmosphere with the largest aeroshell ever flown to Mars, fly the first guided lifting entry aeroshell ever flown to Mars, fly the first guided lifting entry ROCKSt 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 fyes Ant Max 1451fth

**30** Dectly

at Mars. However, Cr range somewhere betw Gale Crater This presents a challe Eberswalde Crater must then weigh the s sochaldenifiragiavi altitudes and Mach nu quantified, probability

mgner Mach numbers of parachute structure and (3) at Mach numb

exhibit an instability, sult in multiple partia

The chief concern with

parachute deployment driving factor, is there

The Viking parachute s Mach 1.4 and 2.1, and 700 Pa [1]. However cessfully operating DB little flight test data al the amount of increas

the relevant flight tests

the plann The Sh pave agree that higher Mac ity of failure, they have should be blaced. For

per boundanfe Mach 2 f

oscillations.

deploying a DGB at M ose of proposing and increase in tisk over a many sites were initial

Loose Sand Parachut testrighte side the full classified panoramic mosaic

Bedrock

algorithm parachurany

quantified, probability of parachanger to avoid scheres of the source of the second schere the second schere the second schere the second schere to avoid sche will often be identical in appearance learn to understand projective geomet

Learn surface feature from data



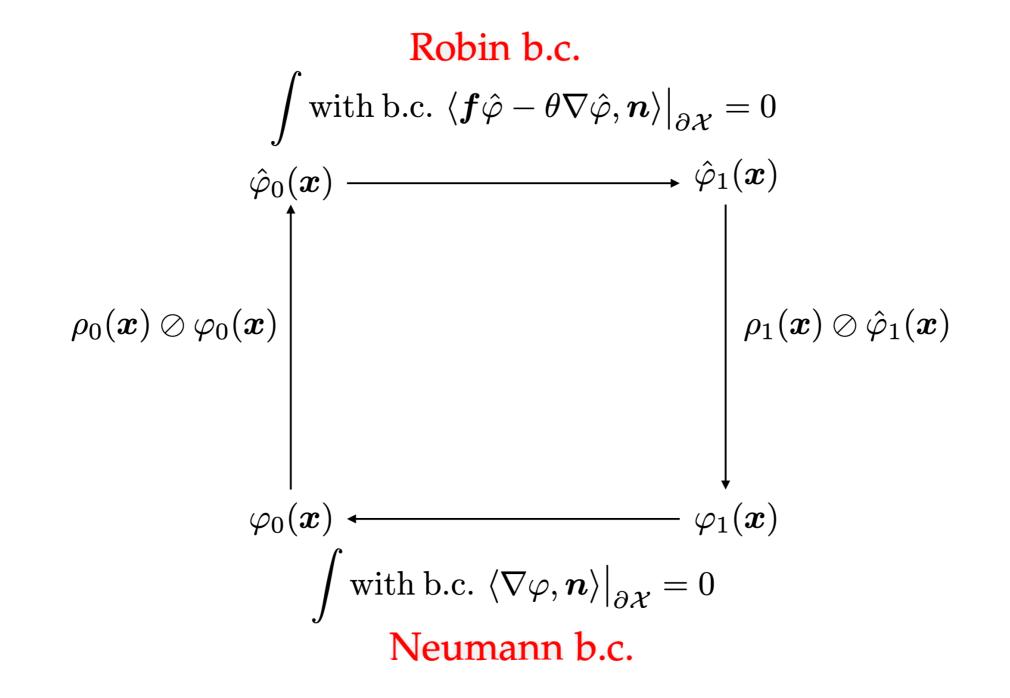
## Thank You



## **Back up slides**

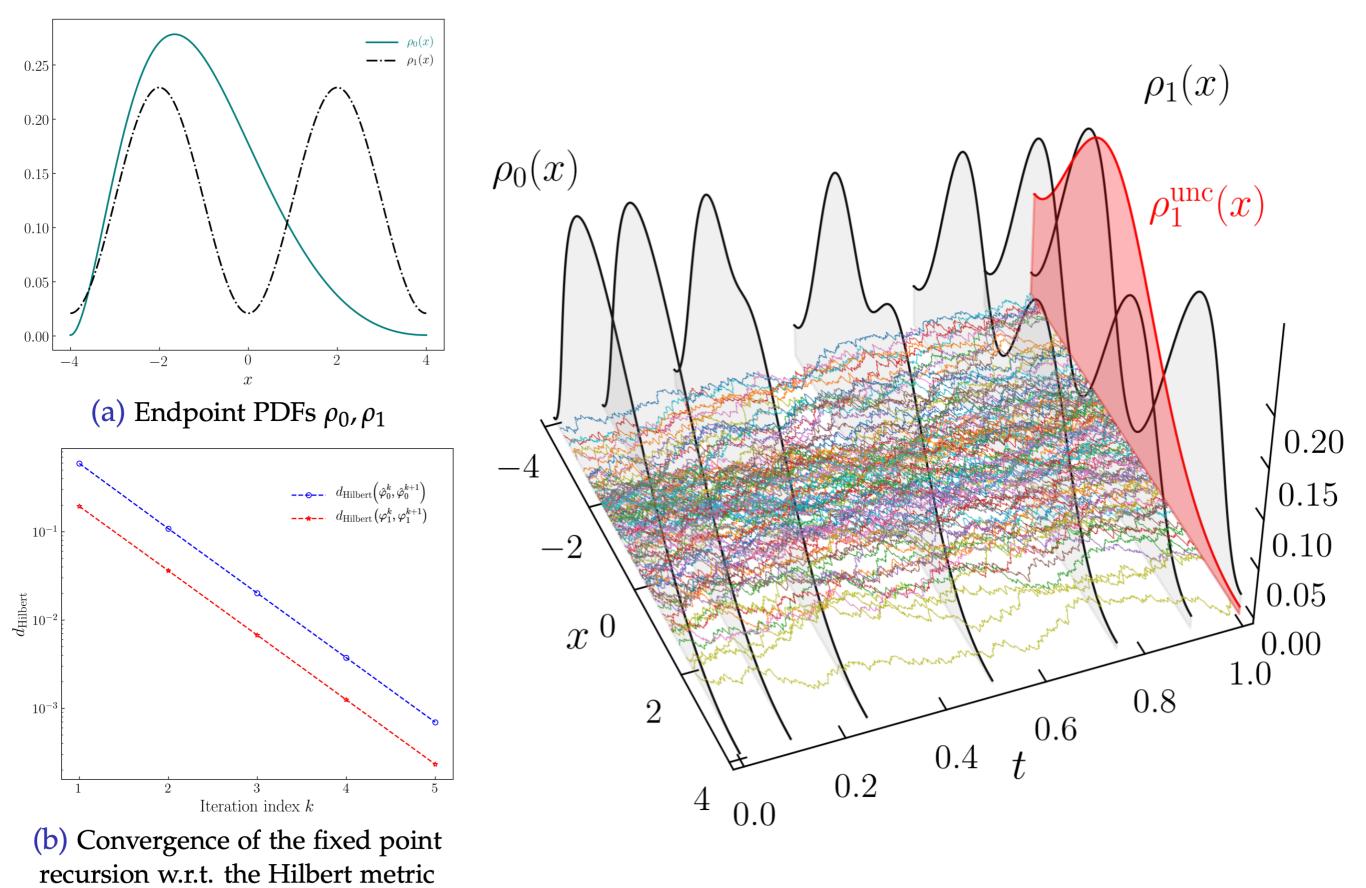
### **Density Control with Hard State Constraints**

**SDE** sample paths reflecting from a given boundary



- K.F. Caluya and A.H., Reflecting Schrodinger bridge: density control with path constraints, ACC 2021.

### **Density Control with Hard State Constraints**



— K.F. Caluya and A.H., Reflecting Schrodinger bridge: density control with path constraints, ACC 2021.