Measure-valued Proximal Recursions for Learning and Control

Iman Nodozi

inodozi@ucsc.edu

Department of Electrical and Computer Engineering
University of California, Santa Cruz

Committee: Abhishek Halder (Advisor), Ali Mesbah, Dejan Milutinovic (Chair), Yu Zhang

Advancement to Candidacy, February 09, 2022

Convex optimization over the space of probability measures

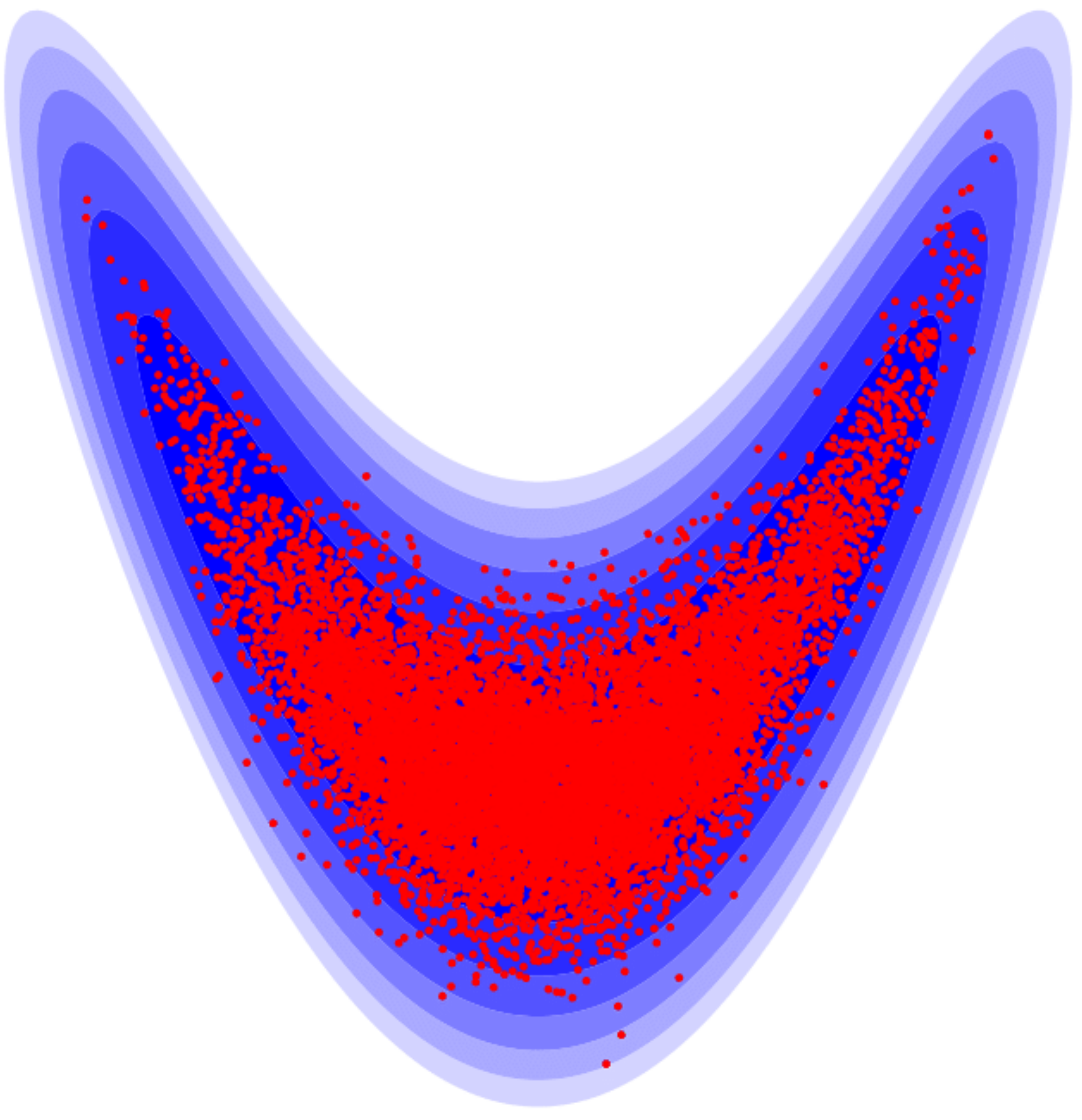
Convex functional

$$\mu^{\text{opt}} = \arg\inf_{\mu \in \mathscr{P}_2(\mathbb{R}^d)} F(\mu)$$

Manifold of probability measures supported on $\mathbb{R}^d$ with finite second moments

Motivating Applications

Langevin sampling from given unnormalized prior



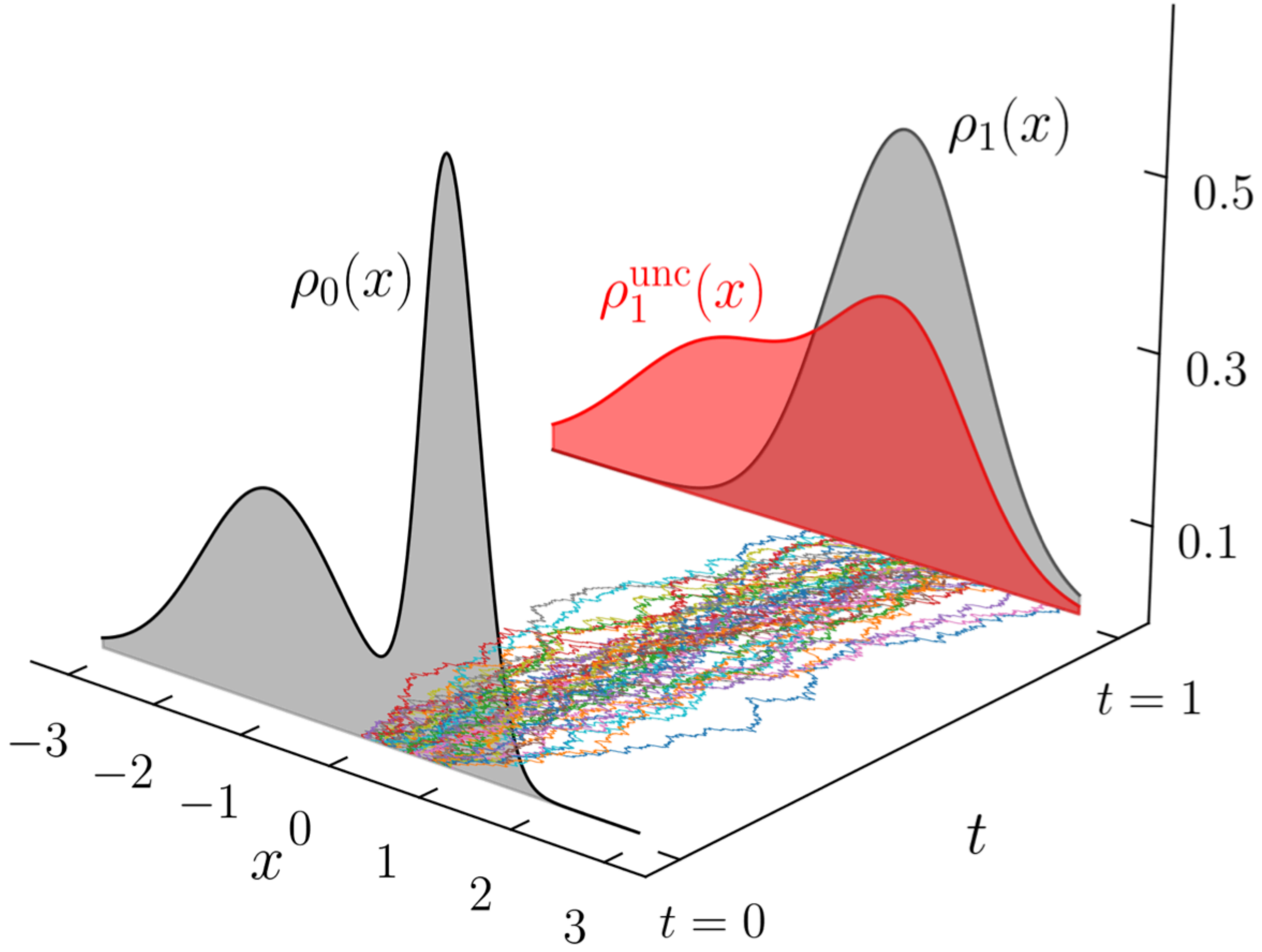
[Stramer and Tweedie, 1999]

[Jarner and Hansen, 2000]

[Roberts and Stramer, 2002]

[Vempala and Wibisono, 2019]

Density control

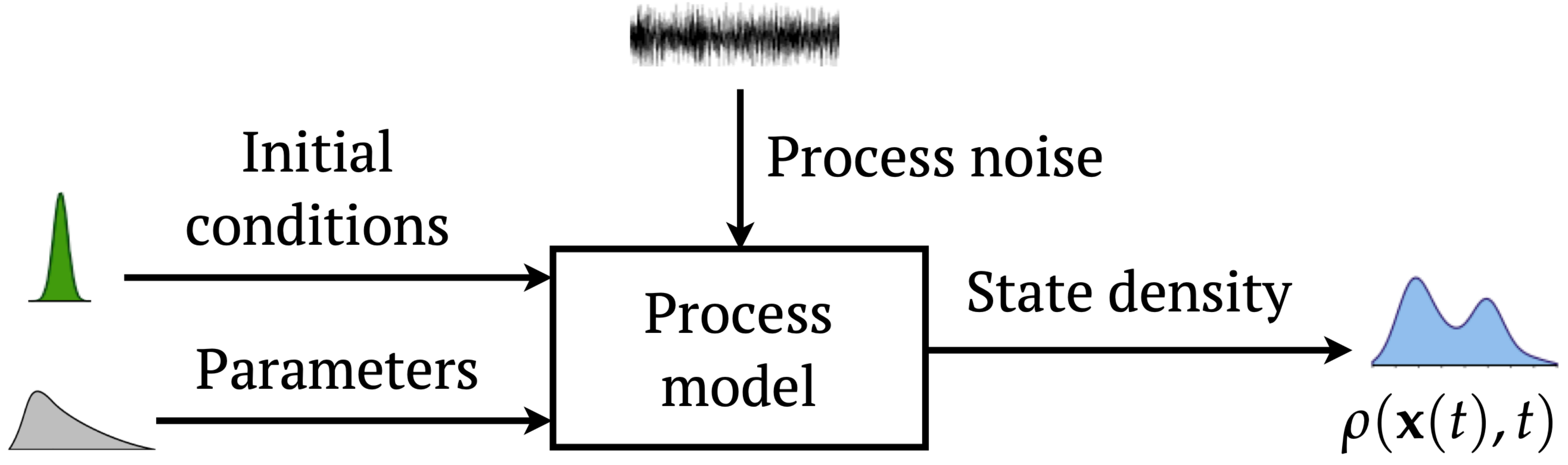


[Caluya and Halder,., 2021]

[Y. Chen et al., 2021]

Motivating Applications

Stochastic prediction

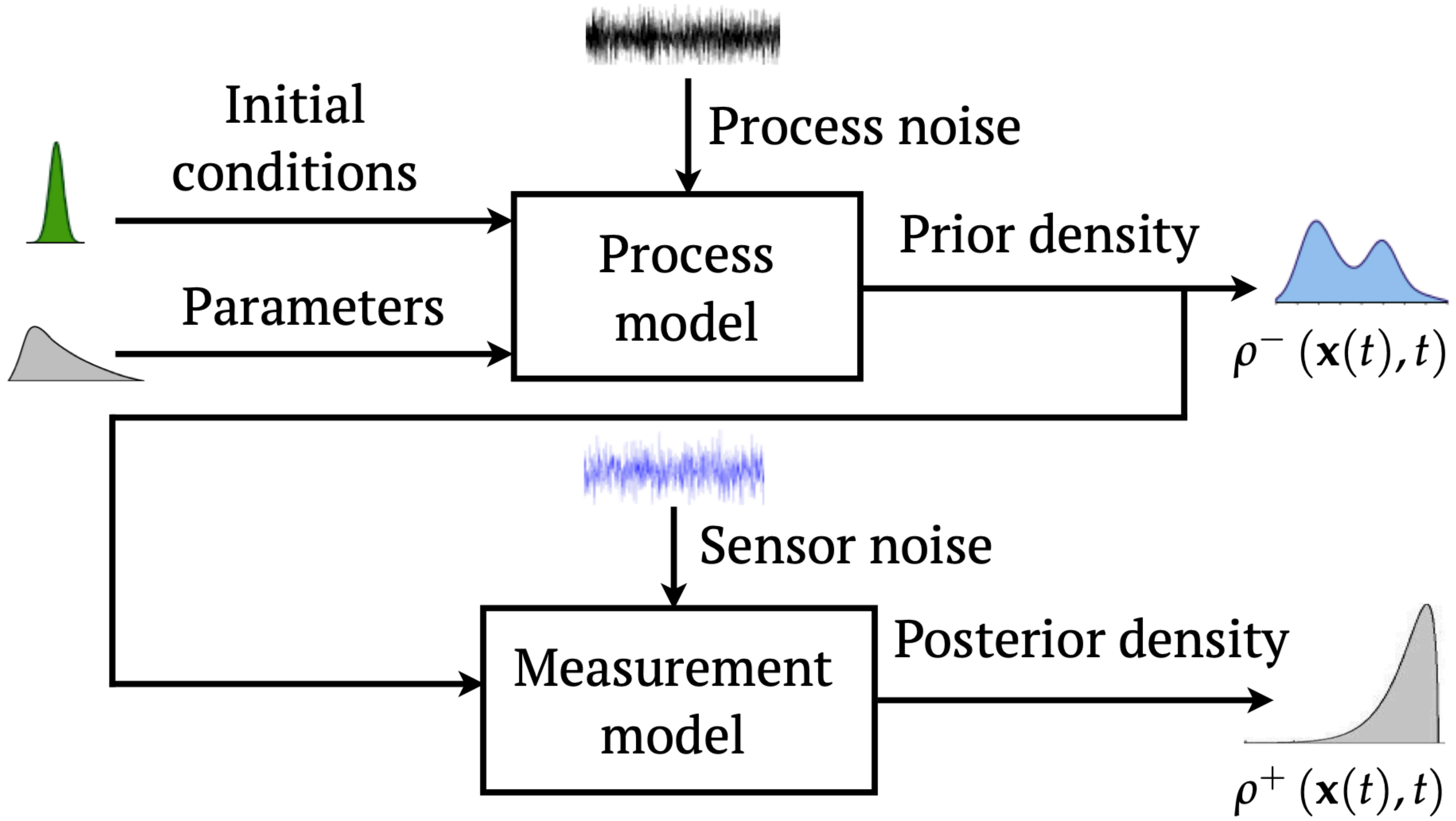


[Jordan et al., 1998]

[Ambrosio et al., 2005]

[Caluya and Halder, 2019]

Stochastic estimation



[Kushner, 1964]

[Stratonovich, 1965]

[Bucy, 1965]

[Halder and Georgiou, 2017, 2018, 2019]

Mean field neural network learning

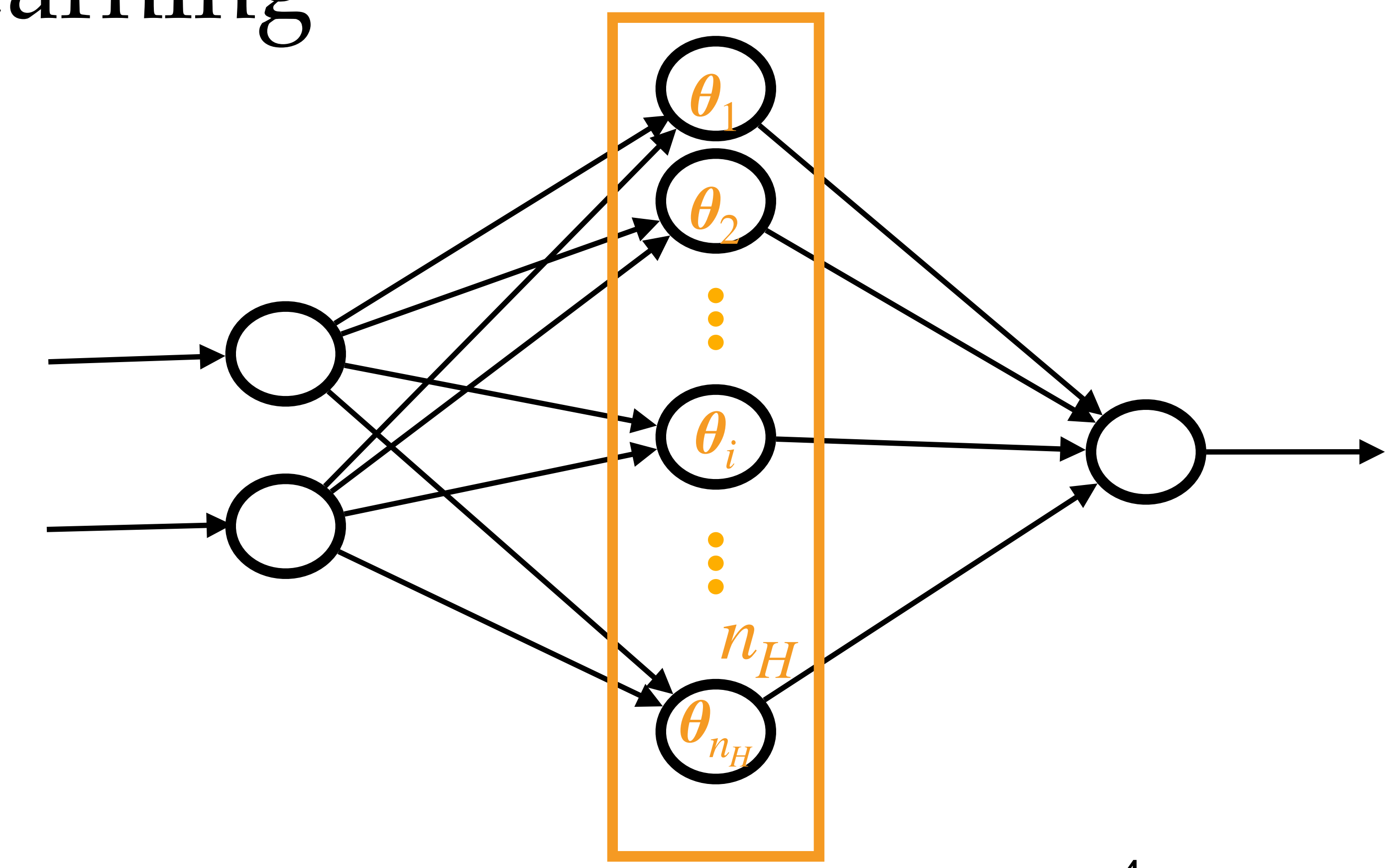
[Rotskoff and Vanden-Eijnden, 2018]

[Sirignano and Spiliopoulos, 2020]

[Domingo-Enrich et al., 2020]

[Krichene, et al., 2020]

[Halder et al., 2020]



Measure valued proximal operator

$$\mu^{\text{opt}} = \text{prox}_{hF}^{\text{dist}}(\nu) := \underset{\mu \in \mathscr{P}_2(\mathbb{R}^d)}{\arg\inf} \; \frac{1}{2}\Big(\text{dist}\,(\mu, \nu)\Big)^2 + hF(\mu)$$

Distance

Step size

Convex functional

Outline of this talk

1. **Measure-valued Proximal Recursions for Mean Field Neural Network Learning**

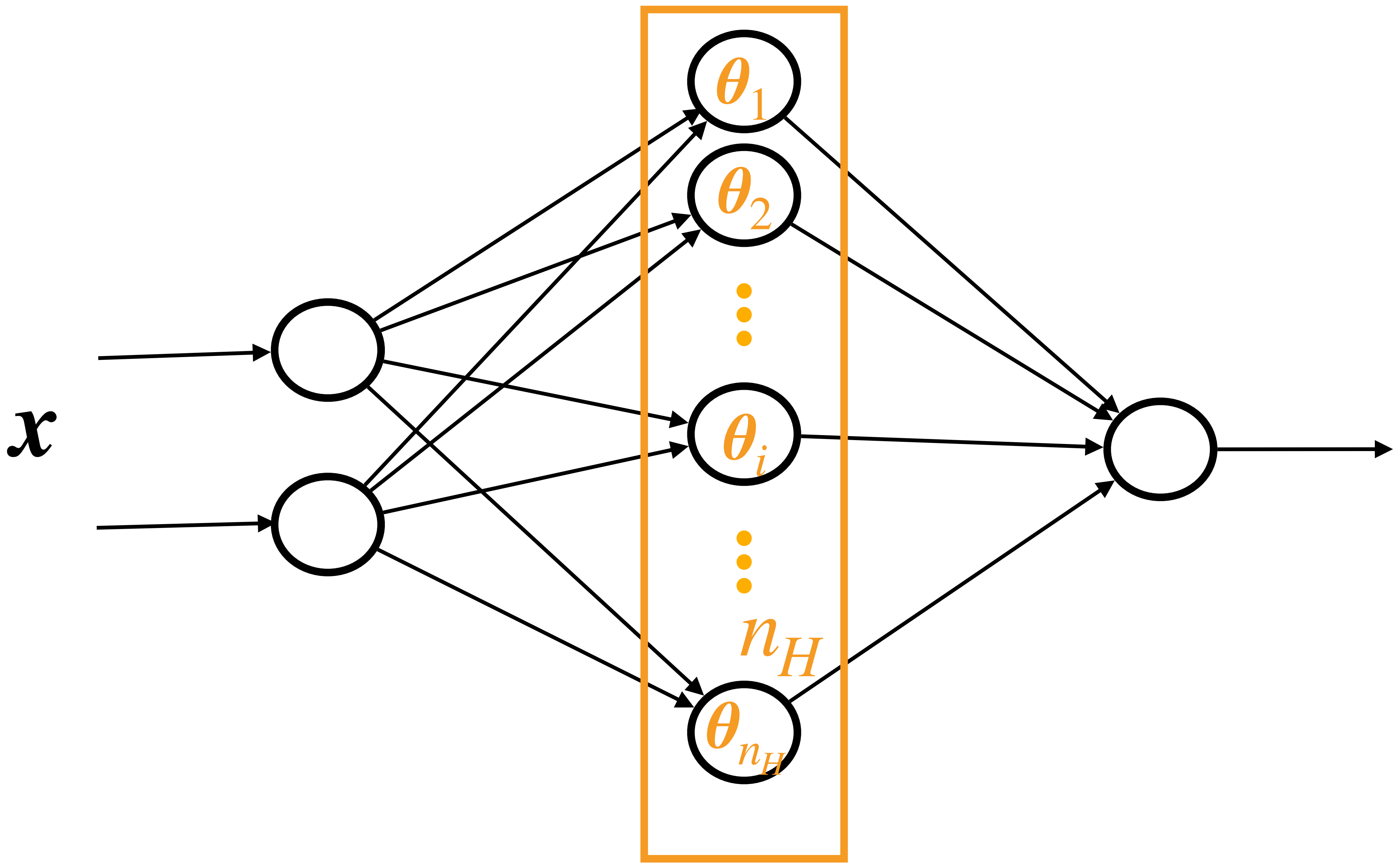
2. **Measure-valued Proximal Recursions for Optimal Steering of Distributions via Feedback Control**

3. **Distributed Algorithms**

4. **Future Plans**

Measure-valued Proximal Recursions for Mean Field Neural Network Learning

Empirical Risk Minimization for Supervised Learning



Hidden layer

Given Data: Feature $\boldsymbol{x}$ Labels $\boldsymbol{y}$

$$(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{X} \times \mathcal{Y} \subseteq \mathbb{R}^{n_x} \times \mathbb{R}^{n_y}$$

$$f(\boldsymbol{x}, \boldsymbol{\theta}) := \frac{1}{n_H} \sum_{i=1}^{n_H} \Phi\left(\boldsymbol{x}, \boldsymbol{\theta}_i\right)$$

Activation function

$$\Phi\left(\boldsymbol{x}, \boldsymbol{\theta}_i\right) = a_i \sigma\left(\left\langle \boldsymbol{w}_i, \boldsymbol{x} \right\rangle + \boldsymbol{b}_i\right)$$

Scaling

Weight vector

Bias

$$\boldsymbol{\theta}_i := \begin{pmatrix} a_i \\ b_i \\ \boldsymbol{w}_i \end{pmatrix} \in \mathbb{R}^p$$

$$\boldsymbol{y} \approx f(\boldsymbol{x}, \boldsymbol{\theta})$$

$$\boldsymbol{\theta} = \begin{pmatrix} \boldsymbol{\theta}_1 \\ \vdots \\ \boldsymbol{\theta}_{n_H} \end{pmatrix}$$

Empirical Risk Minimization for Supervised Learning

$$l(\boldsymbol{y},\boldsymbol{x},\boldsymbol{\theta}) \equiv l(\boldsymbol{y},f(\boldsymbol{x},\boldsymbol{\theta})) = \underbrace{\frac{1}{2}\|\boldsymbol{y}-f(\boldsymbol{x},\boldsymbol{\theta})\|_2^2}_{\text{Quadratic loss}}$$

$$R(f) := \mathbb{E}_{\Delta}[l(\boldsymbol{y},\boldsymbol{x},\boldsymbol{\theta})] \xrightarrow{\Delta \text{ unknown}} R(f) \approx \underbrace{\frac{1}{n}\sum_{j=1}^{n} l\left(\boldsymbol{y}_j,\boldsymbol{x}_j,\boldsymbol{\theta}\right)}_{\text{Empirical risk}}$$

$$\min_{\boldsymbol{\theta}\in\mathbb{R}^{pn_H}} R(f)$$

Finite dimensional nonconvex problem

State-of-the-art: search optimal $\boldsymbol{\theta}$ using variants of SGD

Learning Algorithm Dynamics: the Mean Field Limit

Absolutely continuous

$$f = \int_{\mathbb{R}^p} \Phi(\boldsymbol{x}, \boldsymbol{\theta}) \underbrace{\mathrm{d}\mu(\boldsymbol{\theta})} = \int_{\mathbb{R}^p} \Phi(\boldsymbol{x}, \boldsymbol{\theta}) \rho(\boldsymbol{\theta}) \mathrm{d}\boldsymbol{\theta} = \mathbb{E}_{\boldsymbol{\theta}}[\Phi(\boldsymbol{x}, \boldsymbol{\theta})]$$

Hidden neuronal population mass

$$F(\rho) := R(f(\boldsymbol{x}, \rho)) = \mathbb{E}_\Delta \left[\frac{1}{2} \left\| \boldsymbol{y} - \int_{\mathbb{R}^p} \Phi(\boldsymbol{x}, \boldsymbol{\theta}) \rho(\boldsymbol{\theta}) \mathrm{d}\boldsymbol{\theta} \right\|_2^2 \right]$$

$$= \boxed{F_0} + \int_{\mathbb{R}^p} \boxed{V(\boldsymbol{\theta})} \rho(\boldsymbol{\theta}) \mathrm{d}\theta + \int_{\mathbb{R}^{2p}} \boxed{U(\boldsymbol{\theta}, \tilde{\boldsymbol{\theta}})} \rho(\boldsymbol{\theta}) \rho(\tilde{\boldsymbol{\theta}}) \mathrm{d}\theta \mathrm{d}\tilde{\theta}$$

$$\mathbb{E}_\Delta \left[\|\boldsymbol{y}\|_2^2 \right]$$

$$\mathbb{E}_\Delta[-2\boldsymbol{y}\Phi(\boldsymbol{x}, \boldsymbol{\theta})]$$

$$\mathbb{E}_\Delta[\Phi(\boldsymbol{x}, \boldsymbol{\theta})\Phi(\boldsymbol{x}, \widetilde{\boldsymbol{\theta}})]$$

$$\min_\rho F(\rho)$$

Infinite dimensional optimization

Regularized Ensemble Risk Minimization

Entropy regularized risk functional

$$F_\beta(\rho) := F(\rho) + \overbrace{\beta^{-1}\int_{\mathbb{R}^p}\rho\log\rho\,\mathrm{d}\boldsymbol{\theta}}^{\text{strictly convex regularizer}}, \quad \beta > 0$$

Sample path dynamics: noisy SGD

$$\mathrm{d}\boldsymbol{\theta} = -\nabla_{\boldsymbol{\theta}}\left(V(\boldsymbol{\theta}) + \int_{\mathbb{R}^p} U(\boldsymbol{\theta},\tilde{\boldsymbol{\theta}})\rho(\tilde{\boldsymbol{\theta}})\mathrm{d}\tilde{\boldsymbol{\theta}}\right)\mathrm{d}t + \sqrt{2\beta^{-1}}\,\mathrm{d}\boldsymbol{w}$$

Ensemble dynamics: mean field PDE IVP

$$\frac{\partial\rho}{\partial t} = \nabla_{\boldsymbol{\theta}}\cdot\left(\rho\left(V(\boldsymbol{\theta}) + \int_{\mathbb{R}^p} U(\boldsymbol{\theta},\tilde{\boldsymbol{\theta}})\rho(\tilde{\boldsymbol{\theta}})\mathrm{d}\tilde{\boldsymbol{\theta}}\right)\right) + \beta^{-1}\Delta_{\boldsymbol{\theta}}\rho \quad \rho(\boldsymbol{\theta},0) = \rho_0(\boldsymbol{\theta}))$$

Regularized Ensemble Risk Minimization:

Static variational problem:

$$\min_{\rho} F_{\beta}(\rho)$$

$$\Updownarrow \text{ Wasserstein gradient flow}$$

Mean field PDE:

$$\frac{\partial \rho}{\partial t} = -\nabla^{W_2} F_{\beta}(\rho) = -\nabla \cdot \left(\rho \nabla \frac{\delta F_{\beta}}{\partial \rho}\right)$$

$$\Updownarrow \text{ Gradient descent}$$

Gradient descent time-stepping:

$$\varrho_k = \operatorname{prox}^{d}_{hF_{\beta}}\left(\varrho_{k-1}\right) := \underset{\varrho \in \mathscr{P}_2\left(\mathbb{R}^p\right)}{\arg\inf} \; \frac{1}{2}\left(d\left(\varrho, \varrho_{k-1}\right)\right)^2 + hF_{\beta}(\varrho)$$

Convergence guarantee:

$$\varrho_k(h, \boldsymbol{\theta}) \xrightarrow{h \downarrow 0} \rho(t = kh, \boldsymbol{\theta}) \quad \text{in } L^1(\mathbb{R}^p), \quad k \in \mathbb{N}$$

Proximal Algorithm

$$\boldsymbol{V}_{k-1} \equiv V\left(\boldsymbol{\theta}_{k-1}\right) := \mathbb{E}_{\Delta}\left[-2\boldsymbol{y}\Phi\left(\boldsymbol{x}, \boldsymbol{\theta}_{k-1}\right)\right]$$

$$\boldsymbol{U}_{k-1} \equiv U\left(\boldsymbol{\theta}_{k-1}, \tilde{\boldsymbol{\theta}}_{k-1}\right) := \mathbb{E}_{\Delta}\left[\Phi\left(\boldsymbol{x}, \boldsymbol{\theta}_{k-1}\right)\Phi\left(\boldsymbol{x}, \tilde{\boldsymbol{\theta}}_{k-1}\right)\right]$$

$$\boldsymbol{C}_k(i,j) := \left\| \boldsymbol{\theta}_k^i - \boldsymbol{\theta}_{k-1}^j \right\|_2^2$$

$$\text{PROXRECUR}\left(\varrho_{k-1}, \boldsymbol{V}_{k-1}, \boldsymbol{U}_{k-1}, \boldsymbol{C}_k, \beta, h, \varepsilon, N, \delta, L\right)$$

$$\boldsymbol{\Gamma}_k \leftarrow \exp\left(-\boldsymbol{C}_k / 2\varepsilon\right)$$

$$\boldsymbol{\xi}_{k-1} \leftarrow \exp\left(-\beta \boldsymbol{V}_{k-1} - \beta \boldsymbol{U}_{k-1}\boldsymbol{\varrho}_{k-1} - \mathbf{1}\right)$$

While converge:

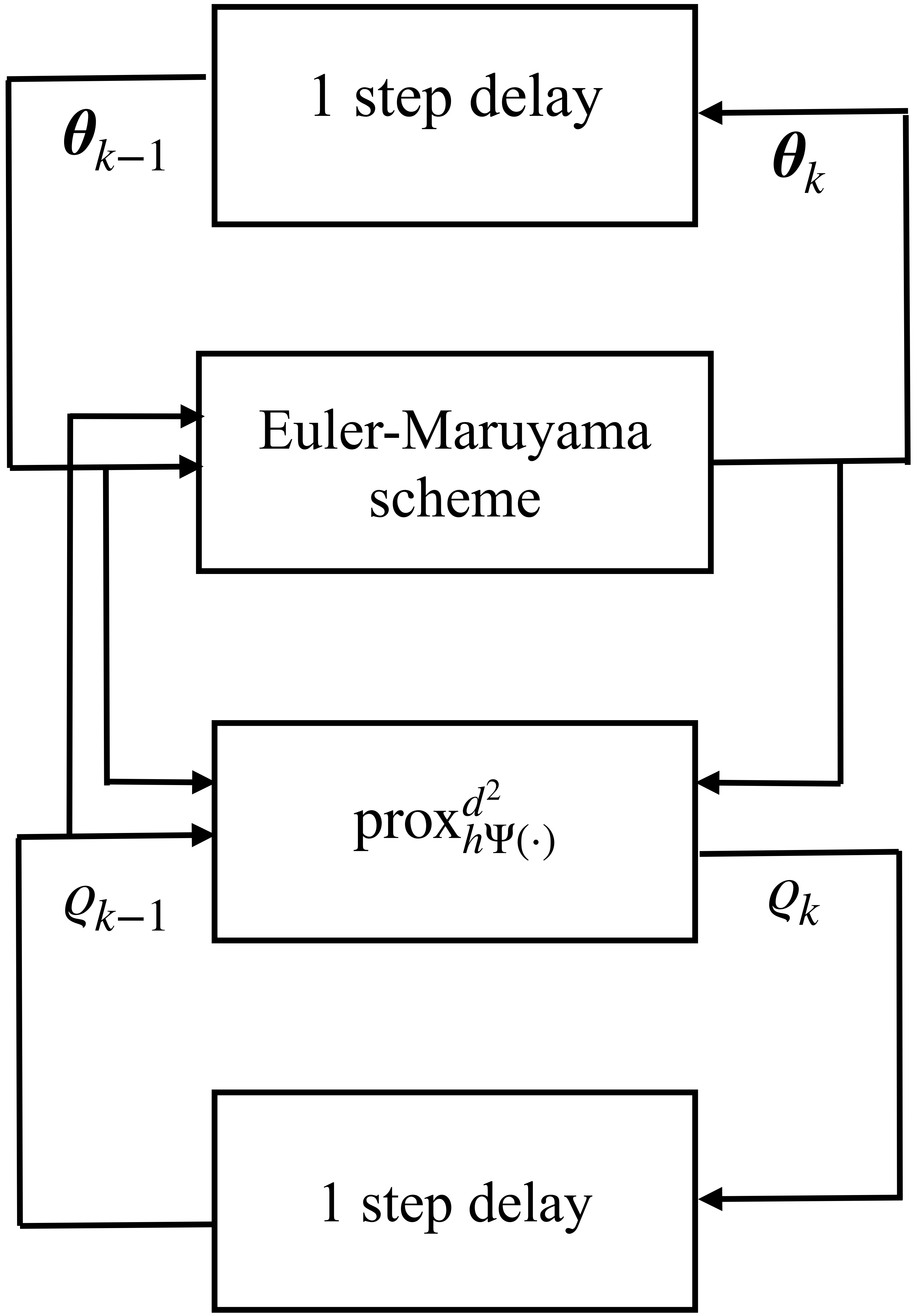
$$\boldsymbol{y} \odot \left(\boldsymbol{\Gamma}_k \boldsymbol{z}\right) = \boldsymbol{\varrho}_{k-1}$$

$$\boldsymbol{z} \odot \left(\boldsymbol{\Gamma}_k^\top \boldsymbol{y}\right) = \boldsymbol{\xi}_{k-1} \odot \boldsymbol{z}^{-\frac{\beta\epsilon}{h}}$$

$$\boldsymbol{\varrho}_k = \boldsymbol{z}^{\text{opt}} \odot \left(\boldsymbol{\Gamma}_k^\top \boldsymbol{y}^{\text{opt}}\right)$$

Convergence guarantee: [Caluya and Halder, 2019]

Schematic of the Proximal Algorithm



$$\boldsymbol{\theta}_k^i = \boldsymbol{\theta}_{k-1}^i - h\nabla\left(\boldsymbol{V}\left(\boldsymbol{\theta}_{k-1}^i\right) + \omega\left(\boldsymbol{\theta}_{k-1}^i\right)\right) + \sqrt{2\beta^{-1}}\left(\boldsymbol{w}_k^i - \boldsymbol{w}_{k-1}^i\right)$$

$$\text{PROXRECUR}\left(\varrho_{k-1}, \boldsymbol{V}_{k-1}, \boldsymbol{U}_{k-1}, \boldsymbol{C}_k, \beta, h, \varepsilon, N, \delta, L\right)$$

Case study: Classification for Breast Cancer Wisconsin (Diagnostic) Data Set

Number of features: $n_x = 30$

Dimension of the neuronal population ensemble support: $p = n_x + 2 = 32$

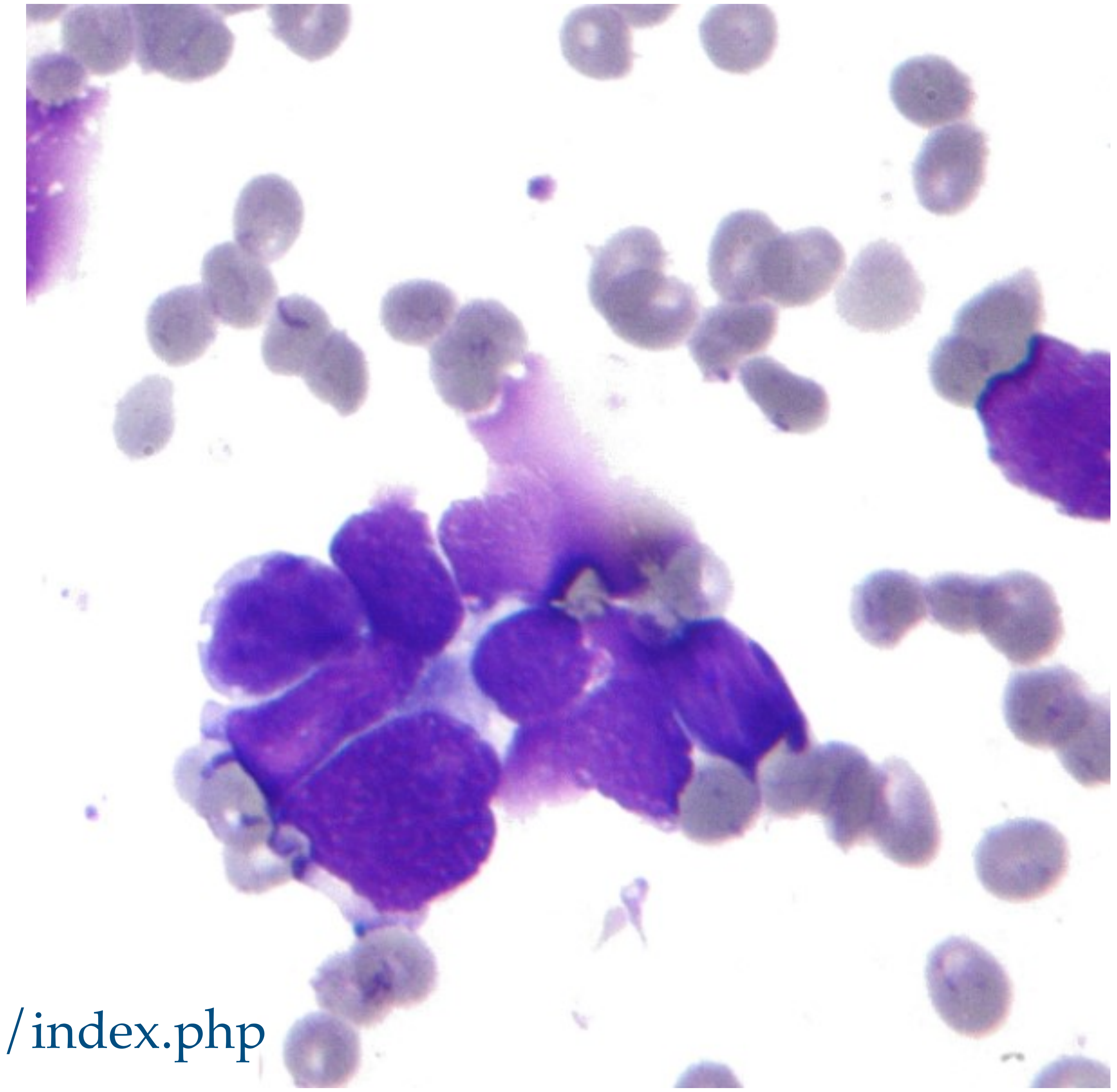
Number of data points: $n = 569$

357 instances

The label 0 denotes "benign"

The label 1 denotes “malignant".

212 instances



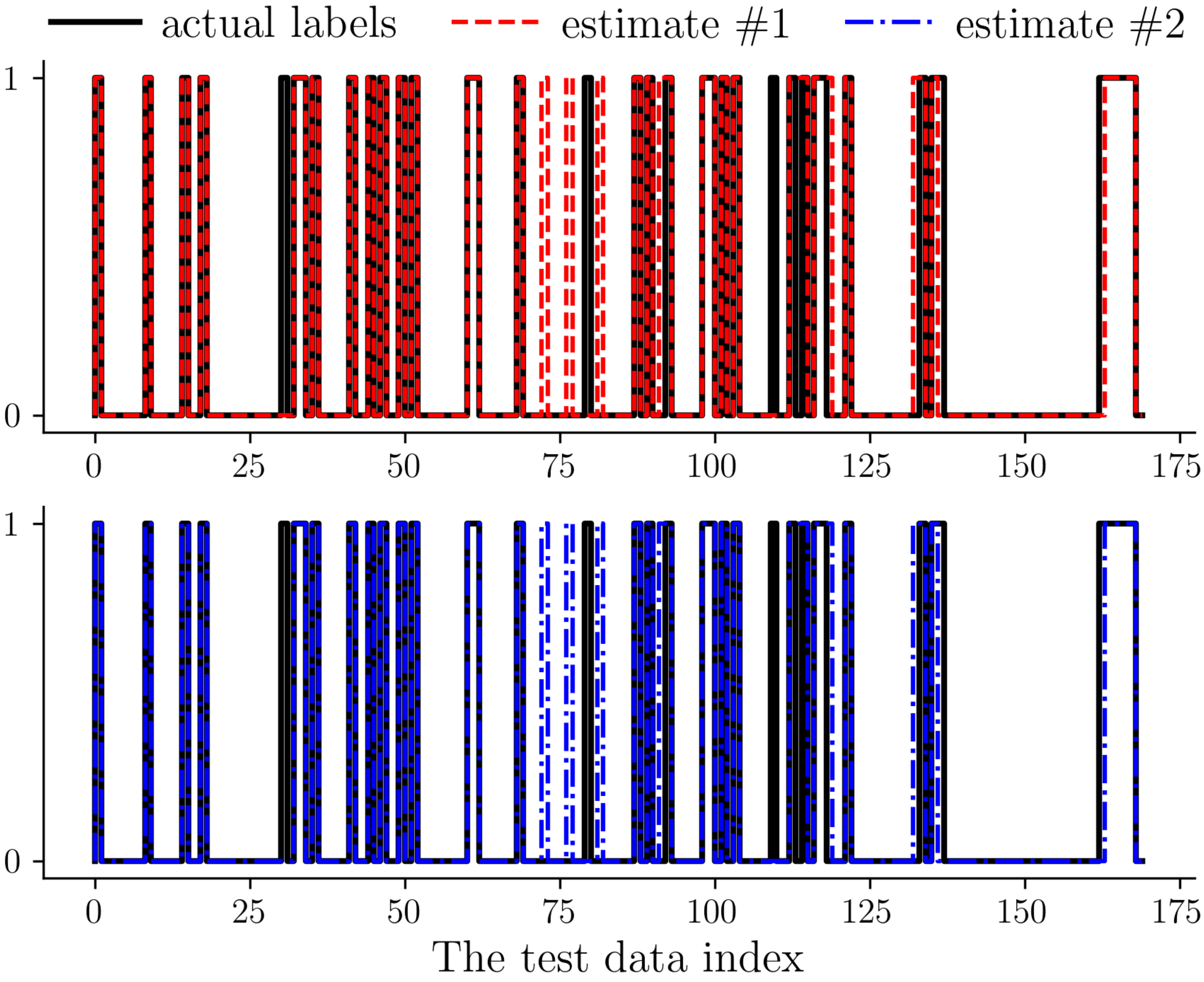
Source: UCI machine learning repository, 2017, Available: http://archive.ics.uci.edu/ml/index.php

Case study: Classification for Breast Cancer Wisconsin (Diagnostic) Data Set

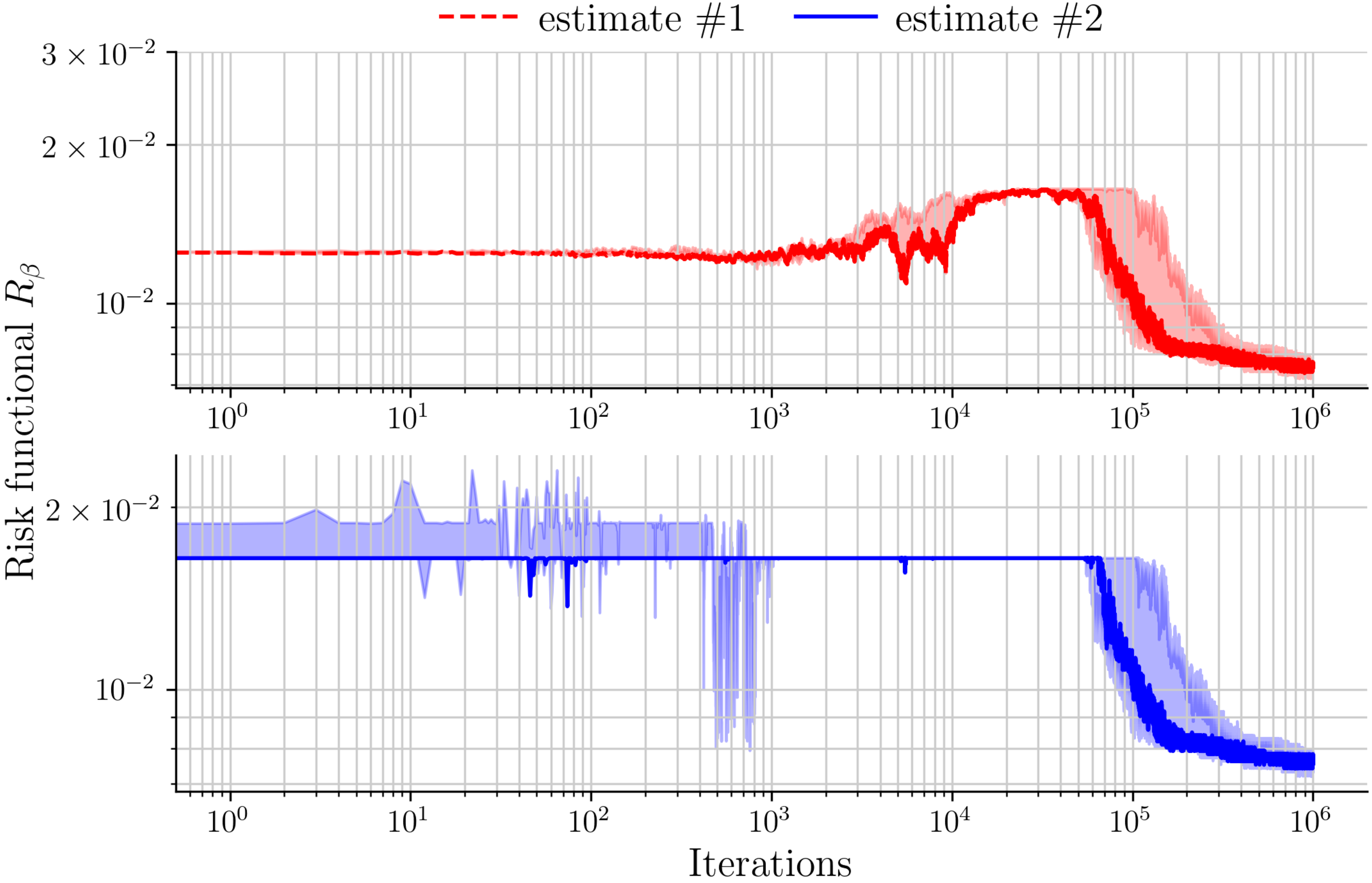
Classification accuracy

β	$Estimate\#1$	$Estimate\#2$
0.03	91.17 %	92.35 %
0.05	92.94 %	92.94 %
0.07	78.23 %	92.94 %

For each fixed β, computational time $\approx$ 33 hours

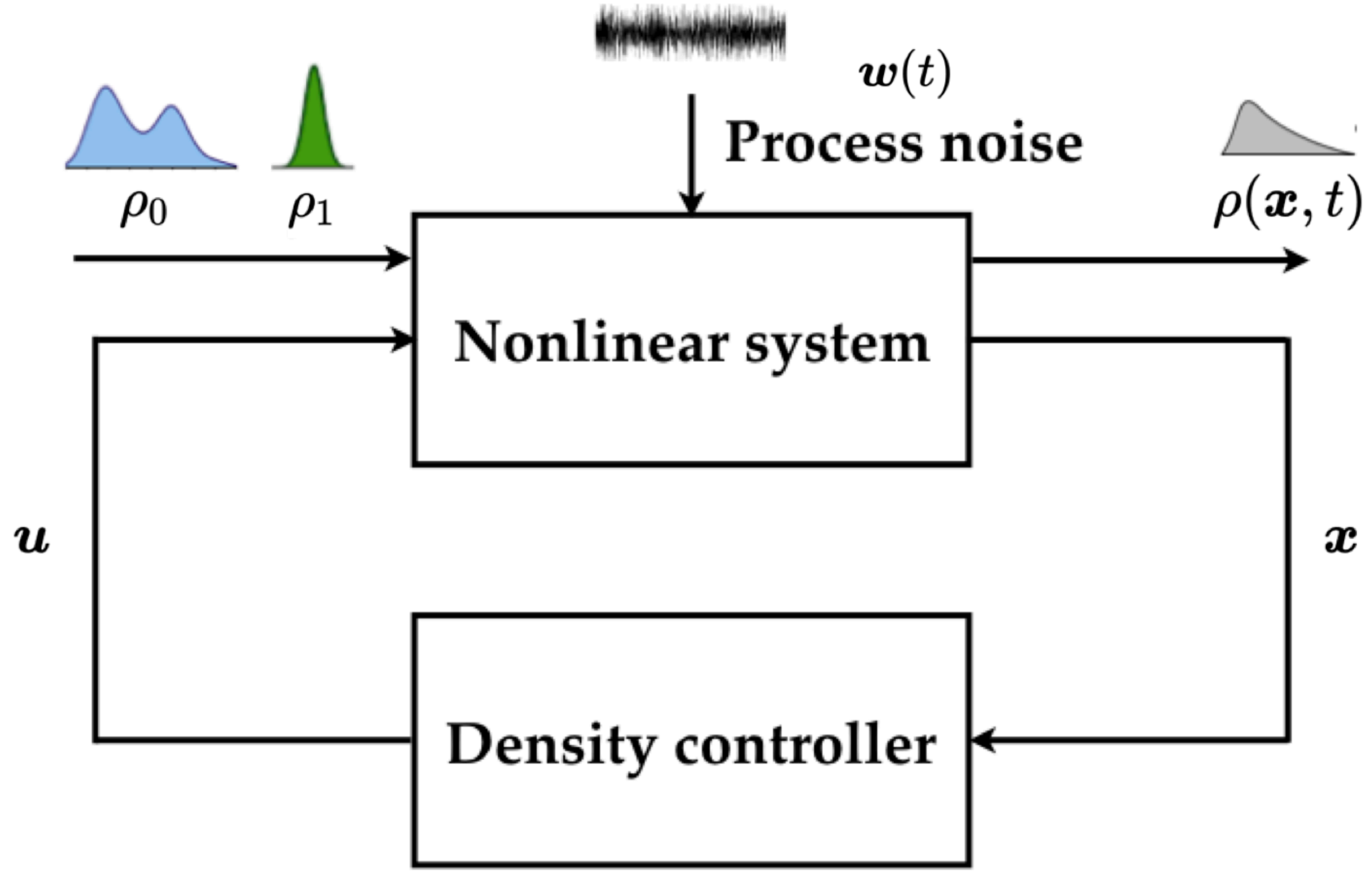


actual labels
estimate #1
estimate #2
1
0
0
25
50
75
100
125
150
175
The test data index


estimate #1
estimate #2
3×10^{-2}
2×10^{-2}
10^{-2}
Risk functional R_β
10^0
10^1
10^2
10^3
10^4
10^5
10^6
2×10^{-2}
10^{-2}
10^0
10^1
10^2
10^3
10^4
10^5
10^6
Iterations

Measure-valued Proximal Recursions for Optimal Steering of Distributions via Feedback Control

State Feedback Density Steering

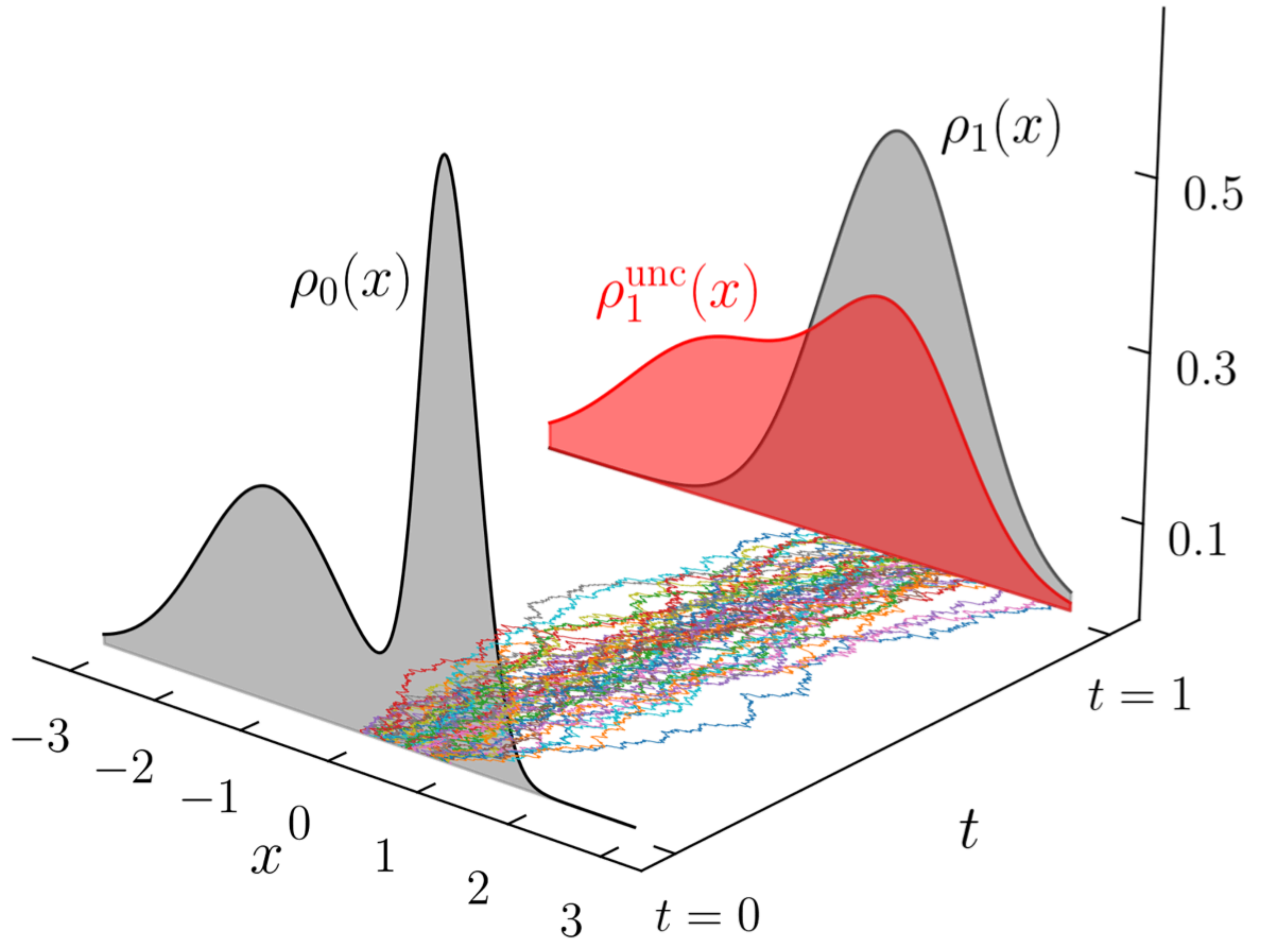


$$\inf_{\boldsymbol{u}\in\mathcal{U}} \quad \mathbb{E}_{\mu^u}\left\{ \int_0^T \frac{1}{2}\|\boldsymbol{u}(\boldsymbol{x}, t)\|_2^2 \, \mathrm{d}t \right\}$$

subject to $\quad \mathrm{d}\boldsymbol{x} = \boldsymbol{f}(\boldsymbol{x}, t)\mathrm{d}t + \boldsymbol{B}(t)\boldsymbol{u}(\boldsymbol{x}, t)\mathrm{d}t + \sqrt{2}\epsilon\boldsymbol{B}(t)\mathrm{d}\boldsymbol{w}(t)$

$$\boldsymbol{x}(t=0) \sim \mu_0(\boldsymbol{x}), \quad \boldsymbol{x}(t=T) \sim \mu_T(\boldsymbol{x})$$

Optimal Control Problem over PDFs



$$\inf_{(\rho^u, \boldsymbol{u})} \frac{1}{2}\int_0^T \int_{\mathbb{R}^n} \|\boldsymbol{u}(\boldsymbol{x},t)\|_2^2 \rho^u(\boldsymbol{x},t) \mathrm{d}\boldsymbol{x}\mathrm{d}t$$

subject to $\dfrac{\partial \rho^u}{\partial t} + \nabla \cdot \left(\rho^u(\boldsymbol{f} + \boldsymbol{B}(t)\boldsymbol{u})\right) = \epsilon \left\langle \boldsymbol{D}(t),\ \text{Hess}\ \left(\rho^u\right)\right\rangle$ $\quad BB^\top$

$$\rho^u(\boldsymbol{x},0) = \rho_0(\boldsymbol{x}) \text{ (given)}, \quad \rho^u(\boldsymbol{x},T) = \rho_T(\boldsymbol{x}) \text{ (given)}.$$

Necessary Conditions for Optimality

Controlled Fokker-Planck or Kolmogorov's forward PDE

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

Hamilton-Jacobi-Bellman PDE:

Value function

$$\frac{\partial\psi}{\partial t} + \frac{1}{2}\left\| \boldsymbol{B}(t)^\top \nabla\psi \right\|_2^2 + \langle \nabla\psi, \boldsymbol{f}\rangle = -\epsilon\langle \boldsymbol{D}(t),\ \text{Hess}\ (\psi)\rangle$$

Boundary conditions: $\rho^{\text{opt}}(\boldsymbol{x},0) = \rho_0(\boldsymbol{x}), \quad \rho^{\text{opt}}(\boldsymbol{x},T) = \rho_T(\boldsymbol{x})$

Optimal control: $\boldsymbol{u}^{\text{opt}}(\boldsymbol{x},t) = \boldsymbol{B}(t)^\top \nabla\psi(\boldsymbol{x},t)$

Feedback Synthesis via the Schrödinger System

Hopf-Cole a.k.a. Fleming's logarithmic transform: $\left(\rho^{\text{opt}}, \psi\right) \mapsto \underbrace{(\widehat{\varphi}, \varphi)}_{\text{Schrödinger factors}}$

$$\varphi(\boldsymbol{x}, t) = \exp\left(\frac{\psi(\boldsymbol{x}, t)}{2\epsilon}\right)$$

$$\hat{\varphi}(\boldsymbol{x}, t) = \rho^{\text{opt}}(\boldsymbol{x}, t)\exp\left(-\frac{\psi(\boldsymbol{x}, t)}{2\epsilon}\right)$$

Feedback Synthesis via the Schrödinger System

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

Uncontrolled forward-backward Kolmogrov PDEs

$$\frac{\partial \varphi}{\partial t} = -\langle \nabla \varphi, \boldsymbol{f} \rangle - \epsilon \langle \boldsymbol{D}(t), \mathrm{Hess}(\varphi) \rangle$$

$$\frac{\partial \hat{\varphi}}{\partial t} = -\nabla \cdot (\hat{\varphi} \boldsymbol{f}) + \epsilon \langle \boldsymbol{D}(t), \mathrm{Hess}(\hat{\varphi}) \rangle$$

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

Optimal control: $\boldsymbol{u}^{\mathrm{opt}}(\boldsymbol{x}, t) = 2\epsilon \boldsymbol{B}(t)^{\top} \nabla \log \varphi$

Fixed Point Recursion over Pair $\left(\varphi_1, \hat{\varphi}_0\right)$

$$
\begin{tikzcd}
\hat{\varphi}_0(\boldsymbol{x}) \arrow[rr, "\int"] & & \hat{\varphi}_1(\boldsymbol{x}) \arrow[d, "\rho_1(\boldsymbol{x}) \oslash \hat{\varphi}_1(\boldsymbol{x})"] \\
\varphi_0(\boldsymbol{x}) \arrow[u, "\rho_0(\boldsymbol{x}) \oslash \varphi_0(\boldsymbol{x})"] & & \varphi_1(\boldsymbol{x}) \arrow[ll, "\int"]
\end{tikzcd}
$$

This recursion is contractive in the Hilbert metric

Case study: Optimal Steering of Distributions for the Nonuniform Noisy Kuramoto Oscillators

Potential $V(\boldsymbol{\theta}) := \sum_{i<j} k_{ij}(1 - \cos(\theta_i - \theta_j - \varphi_{ij})) - \sum_{i=1}^{n} P_i \theta_i$

Coupling > 0 Phase difference $\in [0, \pi/2)$ Linear coeff. > 0

Positive diag matrices $\boldsymbol{M}, \boldsymbol{\Gamma}, \boldsymbol{S}$

Case study: Optimal Steering of Distributions for the Nonuniform Noisy Kuramoto Oscillators

$$\inf_{(\rho^u, u)} \int_0^T \int_{\mathcal{X}} \|\boldsymbol{u}(\boldsymbol{x}, t)\|_2^2 \rho^u(\boldsymbol{x}, t)\mathrm{d}\boldsymbol{x}\mathrm{d}t$$

First order, $\mathcal{X} \equiv \mathbb{T}^n$

$$\frac{\partial \rho^u}{\partial t} = -\nabla_{\boldsymbol{\theta}} \cdot \left(\rho^u \left(\boldsymbol{S}\boldsymbol{u} - \nabla_{\boldsymbol{\theta}} V\right)\right) + \left\langle \boldsymbol{D}, \mathbf{Hess}\left(\rho^u\right)\right\rangle$$

$\boldsymbol{S}\boldsymbol{S}^\top$

Second order, $\mathcal{X} \equiv \mathbb{T}^n \times \mathbb{R}^n$

$$\frac{\partial \rho^u}{\partial t} = \nabla_\omega \cdot \Big(\rho^u \big(\boldsymbol{M}^{-1} \nabla_{\boldsymbol{\theta}} V(\boldsymbol{\theta}) + \boldsymbol{M}^{-1}\boldsymbol{\Gamma}\boldsymbol{\omega} - \boldsymbol{M}^{-1}\boldsymbol{S}\boldsymbol{u} + \boldsymbol{M}^{-1}\boldsymbol{D}\boldsymbol{M}^{-1}\nabla_{\boldsymbol{\omega}} \log \rho^u\big)\Big) - \left\langle \boldsymbol{\omega}, \nabla_{\boldsymbol{\theta}} \rho^u \right\rangle$$

Boundary conditions

$$\rho^u(\boldsymbol{x}, t = 0) = \rho_0$$

$$\rho^u(\boldsymbol{x}, t = T) = \rho_T$$

From Anisotropic to Isotropic Degenerate Diffusion

The First Order Case

$$\boldsymbol{\theta} \mapsto \boldsymbol{\xi} := \boldsymbol{S}^{-1}\boldsymbol{\theta}$$

$$\mathrm{d}\boldsymbol{\xi} = \left(\boldsymbol{u} - \boldsymbol{\Upsilon}\,\nabla_{\boldsymbol{\xi}}\tilde{V}(\boldsymbol{\xi})\right)\mathrm{d}t + \sqrt{2}\;\mathrm{d}\boldsymbol{w}$$

$$\boldsymbol{\Upsilon} := \left(\prod_{i=1}^{n}\sigma_i^2\right)\boldsymbol{S}^{-2} = \mathrm{diag}\left(\prod_{j\neq i}\sigma_j^2\right) \succ \boldsymbol{0}$$

$$\tilde{V}(\boldsymbol{\xi}) := \left(\frac{1}{2}\sum_{i<j}k_{ij}\left(1-\cos\left(\sigma_i\xi_i-\sigma_j\xi_j-\varphi_{ij}\right)\right)-\sum_{i=1}^{n}\sigma_i P_i\xi_i\right)/\left(\prod_{i=1}^{n}\sigma_i^2\right)$$

Isotropic Degenerate Diffusion For The First Order Case

The Second Order Case

$$\begin{pmatrix}\boldsymbol{\theta}\\ \boldsymbol{\omega}\end{pmatrix} \mapsto \begin{pmatrix}\boldsymbol{\xi}\\ \boldsymbol{\eta}\end{pmatrix} := \left(\boldsymbol{I}_2 \otimes \left(\boldsymbol{M}\boldsymbol{S}^{-1}\right)\right)\begin{pmatrix}\boldsymbol{\theta}\\ \boldsymbol{\omega}\end{pmatrix}$$

$$\begin{pmatrix}\mathrm{d}\boldsymbol{\xi}\\ \mathrm{d}\boldsymbol{\eta}\end{pmatrix} = \begin{pmatrix}\boldsymbol{\eta}\\ \boldsymbol{u} - \widetilde{\boldsymbol{\Upsilon}}\,\nabla_{\boldsymbol{\xi}} U(\boldsymbol{\xi}) - \nabla_{\boldsymbol{\eta}} F(\boldsymbol{\eta})\end{pmatrix}\mathrm{d}t + \begin{pmatrix}\boldsymbol{0}_{n\times n}\\ \boldsymbol{I}_n\end{pmatrix}\mathrm{d}\boldsymbol{w}$$

$$\widetilde{\Upsilon} := \left(\prod_{i=1}^{n} \sigma_i^2 m_i^{-2}\right) MS^{-2}$$

$$U(\boldsymbol{\xi}) := \left[\frac{1}{2}\sum_{i<j} k_{ij}\left(1 - \cos\left(\frac{\sigma_i}{m_i}\xi_i - \frac{\sigma_j}{m_j}\xi_j - \varphi_{ij}\right)\right) - \sum_{i=1}^{n}\frac{\sigma_i}{m_i}P_i\xi_i\right]\left[\prod_{i=1}^{n}\left(\frac{m_i}{\sigma_i}\right)^2\right]$$

$$F(\boldsymbol{\eta}) := \frac{1}{2}\left\langle \boldsymbol{\eta}, S^{-1}\boldsymbol{\Gamma}\boldsymbol{\eta}\right\rangle$$

Feedback Synthesis via the Schrödinger System: First Order Case

Uncontrolled forward-backward Kolmogrov PDEs

$$\frac{\partial \hat{\varphi}}{\partial t} = \nabla_{\xi} \cdot \left(\hat{\varphi} \Upsilon \nabla_{\xi} \tilde{V} \right) + \Delta_{\xi} \hat{\varphi}$$

$$\frac{\partial \varphi}{\partial t} = \left\langle \nabla_{\xi} \varphi, \Upsilon \nabla_{\xi} \tilde{V} \right\rangle - \Delta_{\xi} \varphi$$

Boundary conditions

$$\hat{\varphi}_0(\boldsymbol{\xi}) \varphi_0(\boldsymbol{\xi}) = \rho_0(\boldsymbol{S}\boldsymbol{\xi}) \left(\prod_{i=1}^{n} \sigma_i \right)$$

$$\hat{\varphi}_T(\boldsymbol{\xi}) \varphi_T(\boldsymbol{\xi}) = \rho_T(\boldsymbol{S}\boldsymbol{\xi}) \left(\prod_{i=1}^{n} \sigma_i \right)$$

Optimal controlled joint state PDF: $\rho^{\text{opt}}(\boldsymbol{\theta}, t) = \hat{\varphi}(\boldsymbol{S}^{-1}\boldsymbol{\theta}, t) \varphi(\boldsymbol{S}^{-1}\boldsymbol{\theta}, t) / \left(\prod_{i=1}^{n} \sigma_i \right)$

Optimal control: $\boldsymbol{u}^{\text{opt}}(\boldsymbol{\theta}, t) = \boldsymbol{S} \nabla_{\boldsymbol{\theta}} \log \varphi \left(\boldsymbol{S}^{-1}\boldsymbol{\theta}, t \right)$

Feedback Synthesis via the Schrödinger System: Second Order Case

Uncontrolled forward-backward Kolmogrov PDEs

$$\frac{\partial \hat{\varphi}}{\partial t} = -\left\langle \boldsymbol{\eta}, \nabla_{\boldsymbol{\xi}} \hat{\varphi} \right\rangle + \nabla_{\boldsymbol{\eta}} \cdot \left(\hat{\varphi} \left(\widetilde{\boldsymbol{\Upsilon}}\, \nabla_{\boldsymbol{\xi}} \boldsymbol{U}(\boldsymbol{\xi}) + \nabla_{\boldsymbol{\eta}} \boldsymbol{F}(\boldsymbol{\eta}) \right) \right) + \Delta_{\boldsymbol{\eta}} \hat{\varphi}$$

$$\frac{\partial \varphi}{\partial t} = -\left\langle \boldsymbol{\eta}, \nabla_{\boldsymbol{\xi}} \varphi \right\rangle + \left\langle \widetilde{\boldsymbol{\Upsilon}}\, \nabla_{\boldsymbol{\xi}} \boldsymbol{U}(\boldsymbol{\xi}) + \nabla_{\boldsymbol{\eta}} \boldsymbol{F}(\boldsymbol{\eta}), \nabla_{\boldsymbol{\eta}} \varphi \right\rangle - \Delta_{\boldsymbol{\eta}} \varphi$$

Boundary conditions

$$\hat{\varphi}_0(\boldsymbol{\xi})\varphi_0(\boldsymbol{\xi}) = \rho_0 \left(\left(\boldsymbol{I}_2 \otimes \boldsymbol{S}\boldsymbol{M}^{-1} \right) \begin{pmatrix} \boldsymbol{\xi} \\ \boldsymbol{\eta} \end{pmatrix} \right) \left(\prod_{i=1}^{n} \frac{\sigma_i^2}{m_i^2} \right)$$

$$\hat{\varphi}_T(\boldsymbol{\xi})\varphi_T(\boldsymbol{\xi}) = \rho_T \left(\left(\boldsymbol{I}_2 \otimes \boldsymbol{S}\boldsymbol{M}^{-1} \right) \begin{pmatrix} \boldsymbol{\xi} \\ \boldsymbol{\eta} \end{pmatrix} \right) \left(\prod_{i=1}^{n} \frac{\sigma_i^2}{m_i^2} \right)$$

Optimal controlled joint state PDF: $$\rho^{\mathrm{opt}}(\boldsymbol{\theta}, \boldsymbol{\omega}, t) = \hat{\varphi} \left(\left(\boldsymbol{I}_2 \otimes \boldsymbol{M}\boldsymbol{S}^{-1} \right) \begin{pmatrix} \boldsymbol{\theta} \\ \boldsymbol{\omega} \end{pmatrix}, t \right) \varphi \left(\left(\boldsymbol{I}_2 \otimes \boldsymbol{M}\boldsymbol{S}^{-1} \right) \begin{pmatrix} \boldsymbol{\theta} \\ \boldsymbol{\omega} \end{pmatrix}, t \right) \left(\prod_{i=1}^{n} \frac{m_i^2}{\sigma_i^2} \right)$$

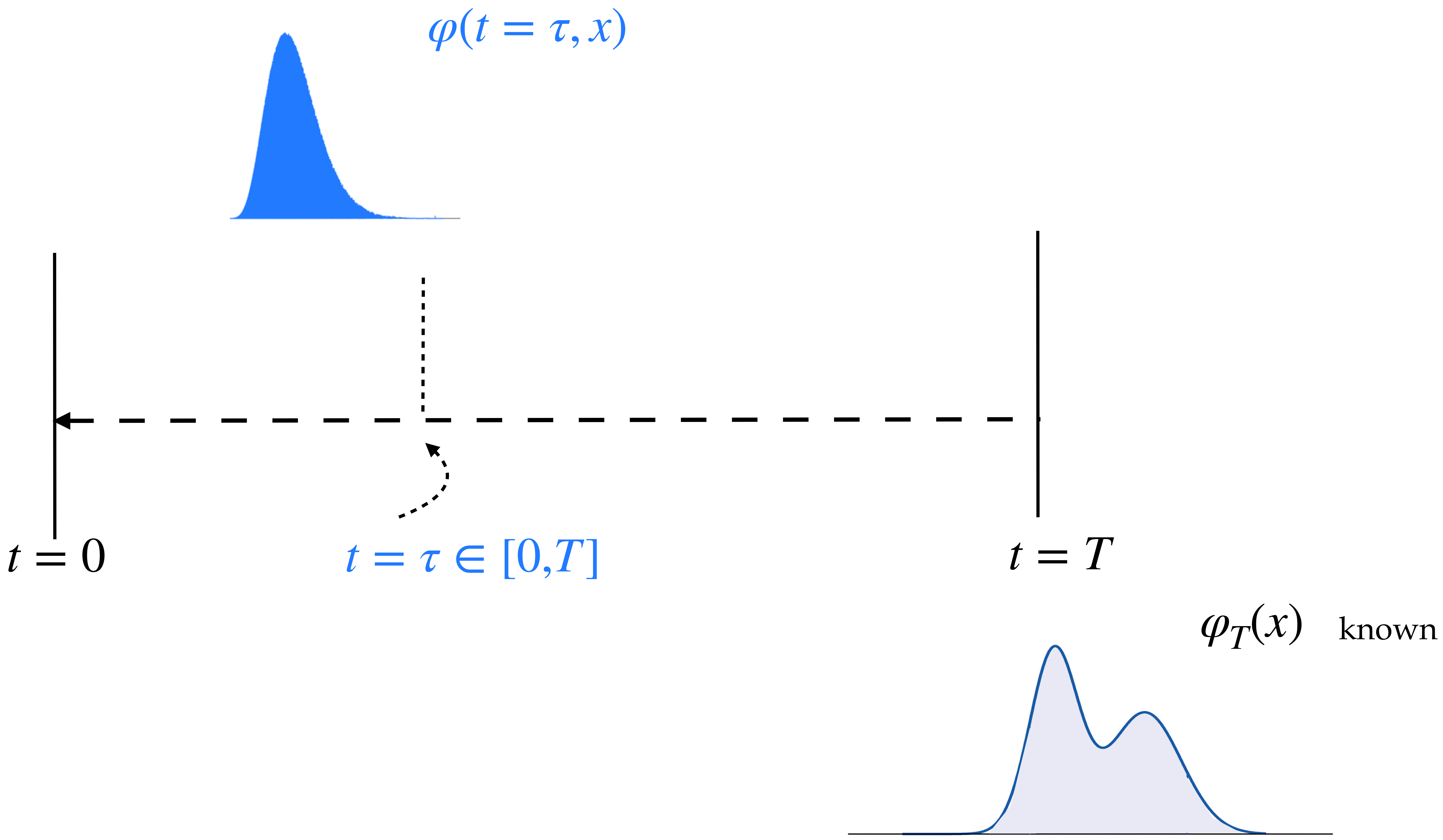
Optimal control: $$\boldsymbol{u}^{\mathrm{opt}} \left(\left(\boldsymbol{I}_2 \otimes \boldsymbol{M}\boldsymbol{S}^{-1} \right) \begin{pmatrix} \boldsymbol{\theta} \\ \boldsymbol{\omega} \end{pmatrix}, t \right) = \left(\boldsymbol{I}_2 \otimes \boldsymbol{S}\boldsymbol{M}^{-1} \right) \nabla_{\boldsymbol{\theta}} \log \varphi \left(\left(\boldsymbol{I}_2 \otimes \boldsymbol{M}\boldsymbol{S}^{-1} \right) \begin{pmatrix} \boldsymbol{\theta} \\ \boldsymbol{\omega} \end{pmatrix}, t \right)$$

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

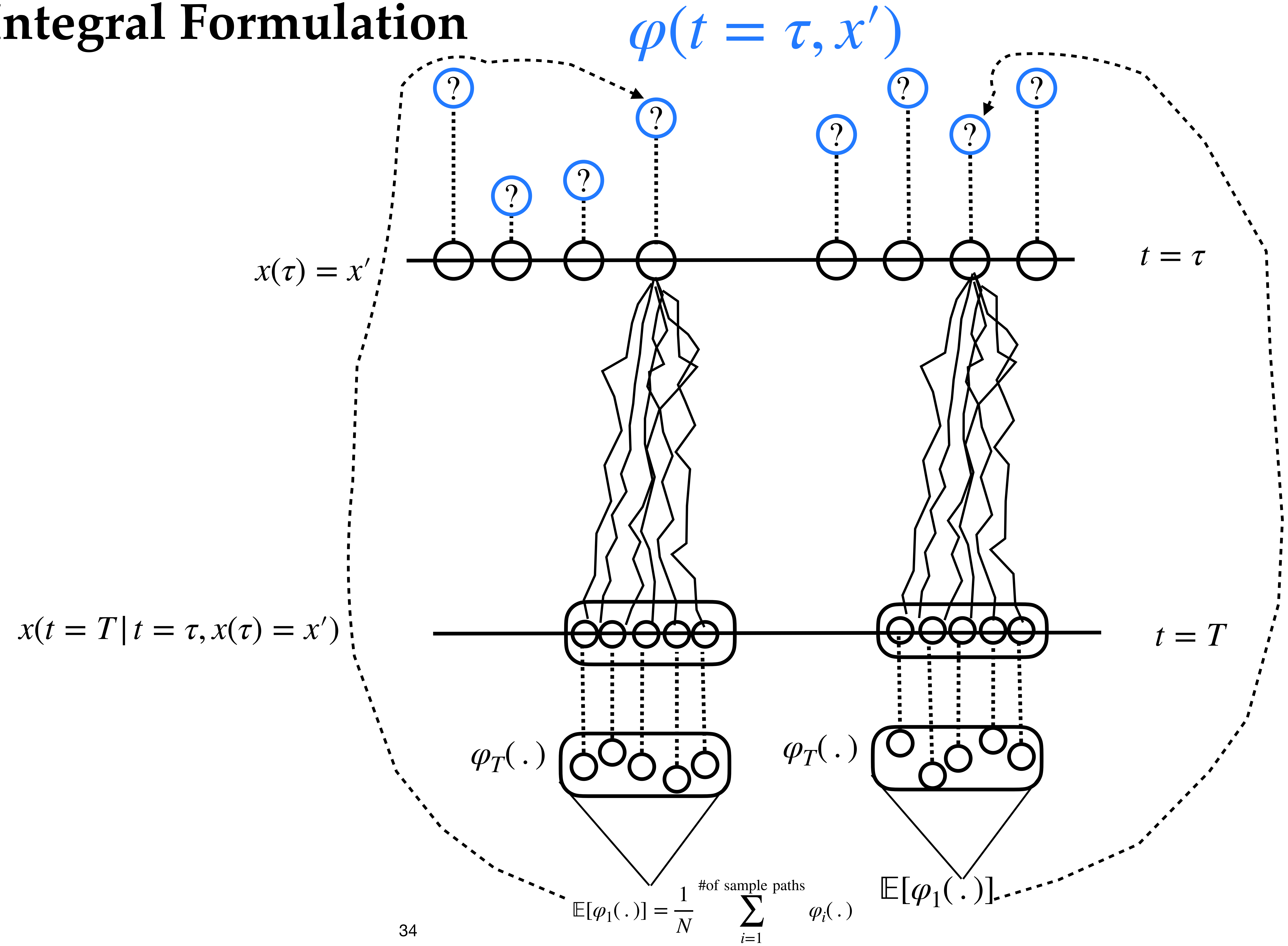
$$
\begin{tikzcd}
\hat{\varphi}_0 \arrow[r, "\text{Proximal recursion}"] & \hat{\varphi}_T \arrow[d, "\tilde{\rho}_T \oslash \hat{\varphi}_T"] \\
\varphi_0 \arrow[u, "\tilde{\rho}_0 \oslash \varphi_0"] & \varphi_T \arrow[l, "\text{Feynman-Kac algorithm}"]
\end{tikzcd}
$$

Feynman-Kac Path Integral Formulation

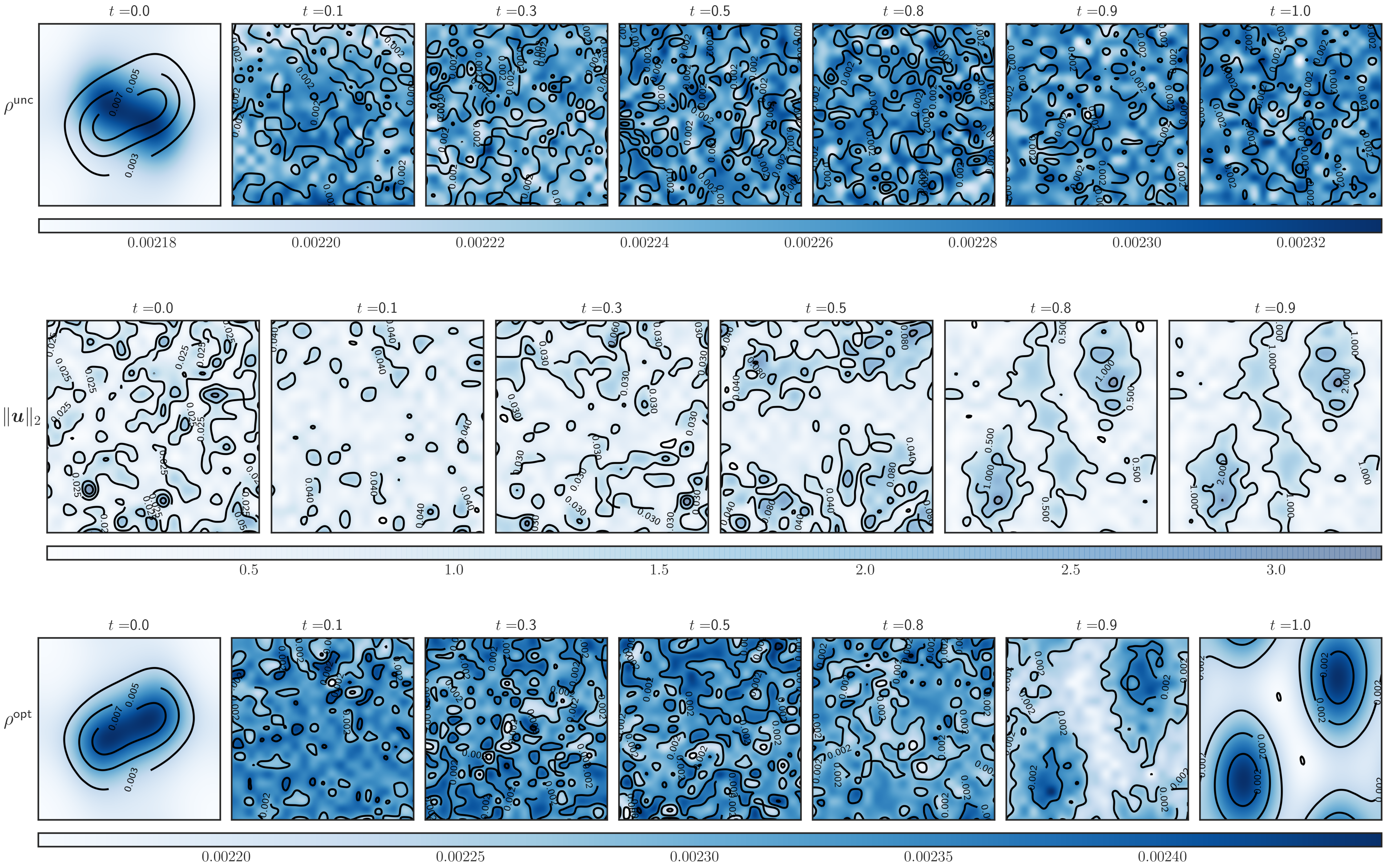
$$\frac{\partial \varphi}{\partial t} = L_{\text{Backward}} \varphi$$



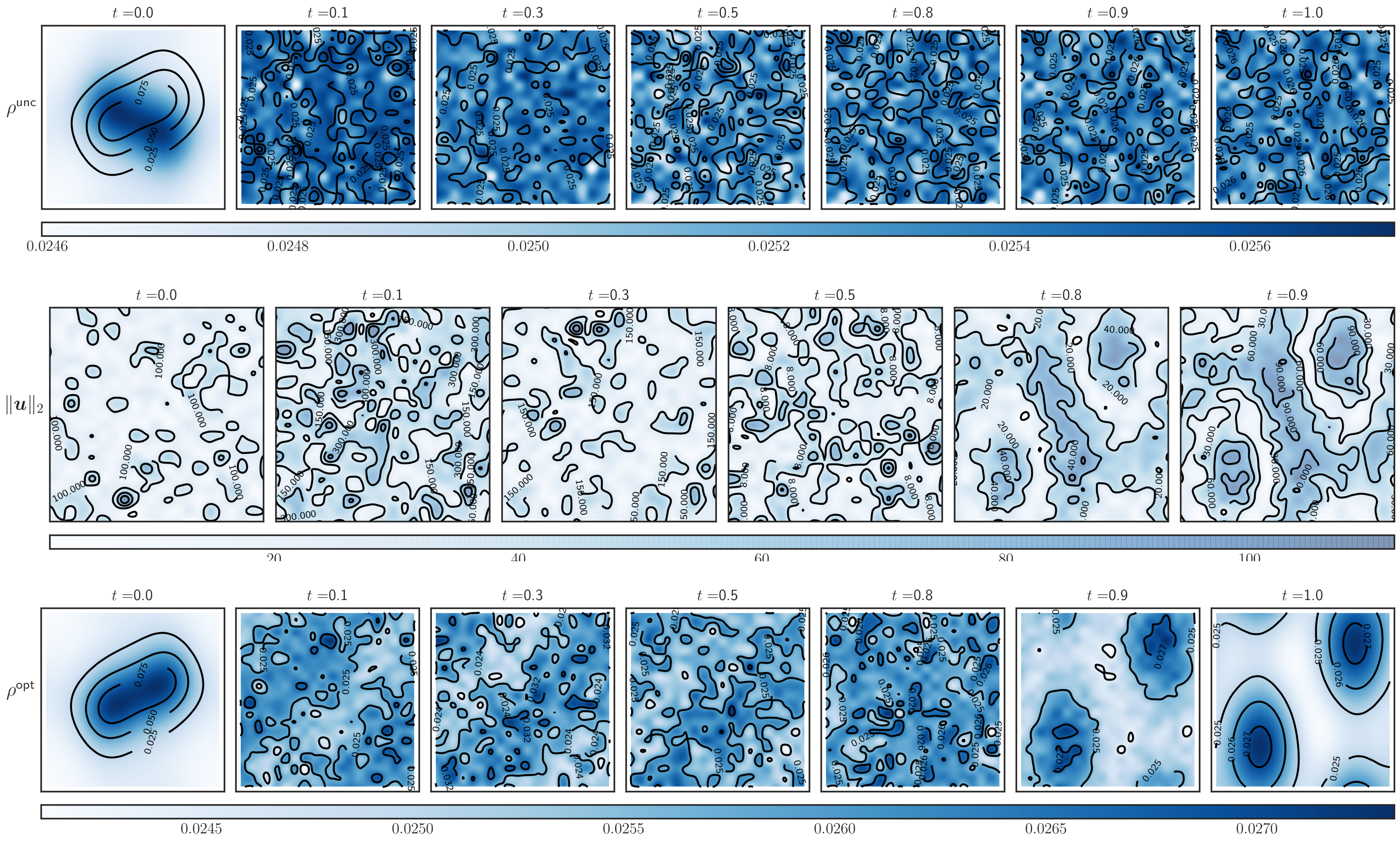
Feynman-Kac Path Integral Formulation



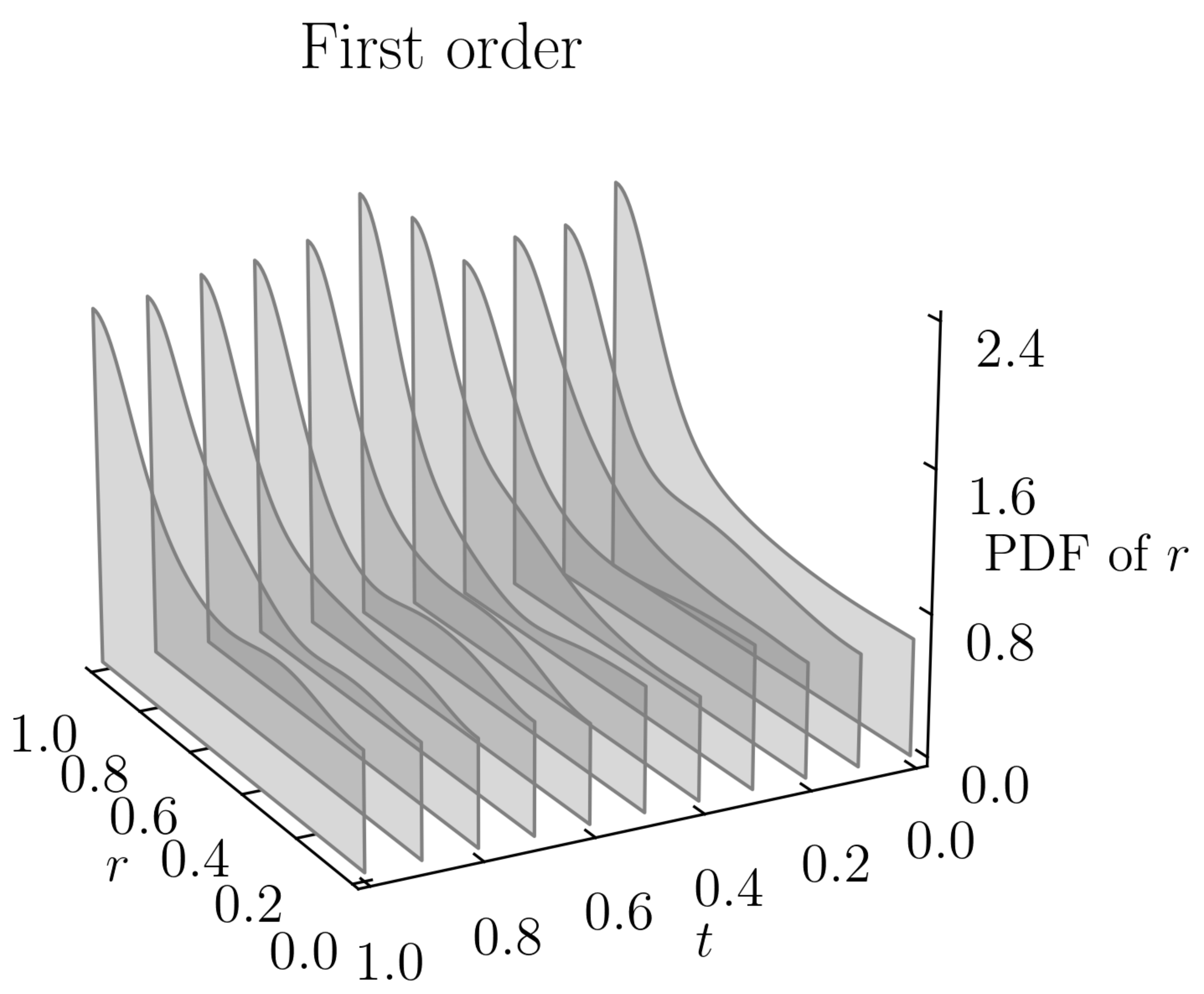
Numerical Example: First Order Case

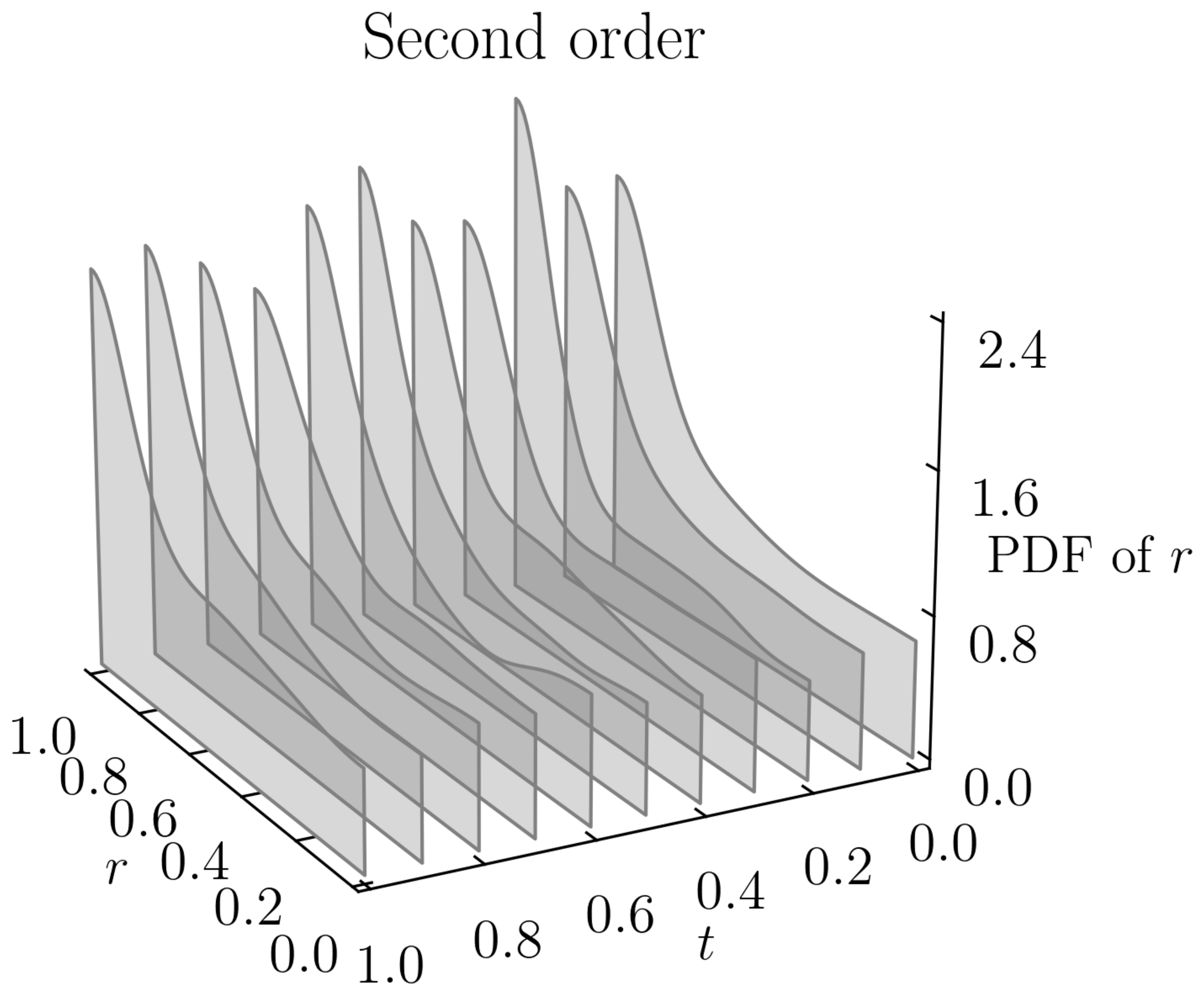


Numerical Example: Second Order Case



Numerical Example: Controlled Order Parameter PDFs





PDF of order parameter $r := \frac{1}{n}\sqrt{\left(\sum_{i=1}^{n}\cos\theta_i\right)^2 + \left(\sum_{i=1}^{n}\sin\theta_i\right)^2}$

Distributed Algorithms

Minimizing Convex Additive Measure-valued Objective

$$\underset{\mu}{\operatorname{arginf}}\, F_1(\mu) + F_2(\mu) + \ldots + F_n(\mu)$$

$\Phi_i(\,\cdot\,) = F_i(\,\cdot\,) + \int \nu_i^k \,\mathrm{d}(\,\cdot\,)$	PDE	Name
$\int_{\mathbb{R}^d} \left(V(\boldsymbol{\theta}) + \nu_i^k(\boldsymbol{\theta})\right) \mathrm{d}\mu_i(\boldsymbol{\theta})$	$\frac{\partial \widetilde{\mu}_i}{\partial t} = \nabla \cdot \left(\widetilde{\mu}_i \left(\nabla V + \nabla \nu_i^k\right)\right)$	Liouville PDE
$\int_{\mathbb{R}^d} \left(\nu_i^k(\boldsymbol{\theta}) + \beta^{-1} \log \mu_i(\boldsymbol{\theta})\right) \mathrm{d}\mu_i(\boldsymbol{\theta})$	$\frac{\partial \widetilde{\mu}_i}{\partial t} = \nabla \cdot \left(\widetilde{\mu}_i \nabla \nu_i^k\right) + \beta^{-1} \Delta \widetilde{\mu}_i$	Fokker-Planck PDE
$\int_{\mathbb{R}^d} \nu_i^k(\boldsymbol{\theta}) \mathrm{d}\mu_i(\boldsymbol{\theta}) + \int_{\mathbb{R}^{2d}} U(\boldsymbol{\theta}, \boldsymbol{\sigma}) \mathrm{d}\mu_i(\boldsymbol{\theta}) \mathrm{d}\mu_i(\boldsymbol{\sigma})$	$\frac{\partial \widetilde{\mu}_i}{\partial t} = \nabla \cdot \left(\widetilde{\mu}_i \left(\nabla \nu_i^k + \nabla \left(U \circledast \widetilde{\mu}_i\right)\right)\right)$	Propagation of chaos PDE

Measure-valued Consensus ADMM

$$\operatorname*{arginf}_{\mu} F_1(\mu) + F_2(\mu) + \ldots + F_n(\mu)$$

$$\rightsquigarrow$$

$$\operatorname*{arg\,inf}_{(\mu_1,\ldots,\mu_n,\zeta)\in\mathscr{P}_2^{n+1}(\mathbb{R}^d)} F_1\left(\mu_1\right) + F_2\left(\mu_2\right) + \ldots + F_n\left(\mu_n\right)$$

$$\mu_1 = \mu_2 = \ldots = \mu_n = \zeta$$

Primal updates

$$\begin{cases} \mu_i^{k+1} = \operatorname*{arginf}_{\mu_i\in\mathscr{P}_2(\mathbb{R}^d)} \frac{1}{2}W^2\left(\mu_i,\zeta^k\right) + \frac{1}{\alpha}\left\{F_i\left(\mu_i\right) + \int_{\mathbb{R}^d}\nu_i^k(\boldsymbol{\theta})\mathrm{d}\mu_i\right\} = \operatorname{prox}_{\frac{1}{\alpha}\left(F_i(\,\cdot\,)+\int\nu_i^k\,\mathrm{d}(\,\cdot\,)\right)}\left(\zeta^k\right) \\ \zeta^{k+1} = \operatorname*{arg\,inf}_{\zeta\in\mathscr{P}_2(\mathbb{R}^d)} \sum_{i=1}^n \left\{\frac{1}{2}W^2\left(\mu_i^{k+1},\zeta\right) - \frac{1}{\alpha}\int_{\mathbb{R}^d}\nu_i^k(\boldsymbol{\theta})\mathrm{d}\zeta\right\} = \operatorname*{arg\,inf}_{\zeta\in\mathscr{P}_2(\mathbb{R}^d)} \left\{\left(\sum_{i=1}^n W^2\left(\mu_i^{k+1},\zeta\right)\right) - \frac{2}{\alpha}\int_{\mathbb{R}^d}\nu_{\mathrm{sum}}^k(\boldsymbol{\theta})\mathrm{d}\zeta\right\} \end{cases}$$

$$\nu_i^{k+1} = \nu_i^k + \alpha\left(\mu_i^{k+1} - \zeta^{k+1}\right)$$

Dual ascent

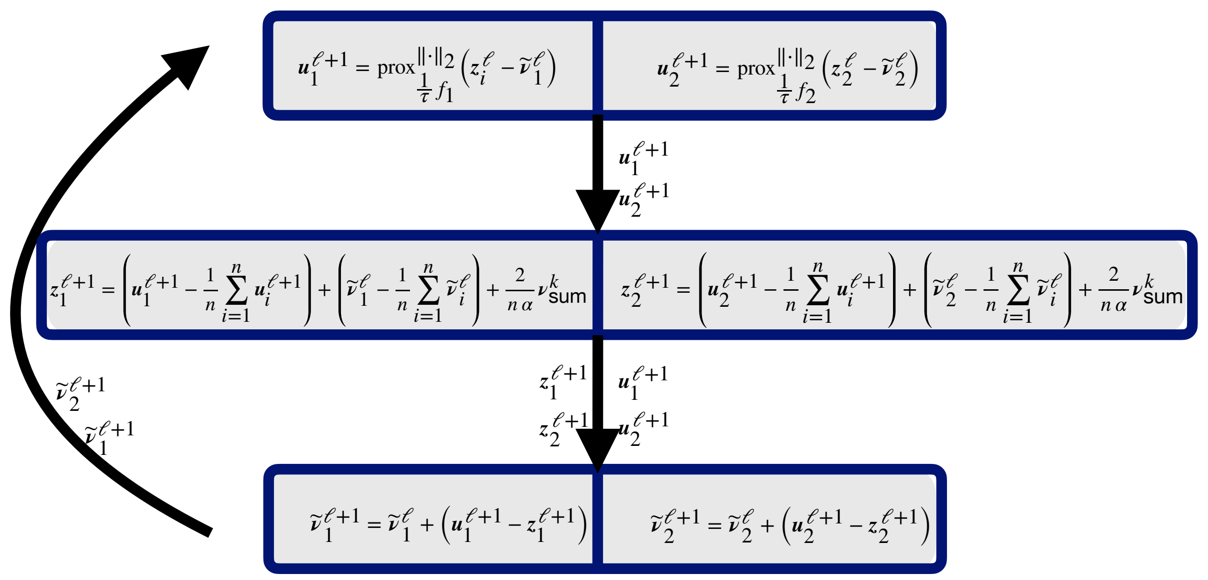
Measure-valued Consensus ADMM Structure for the case that $n = 2$

$$\mu_1^{k+1} = \operatorname{prox}^{W_\epsilon}_{\frac{1}{\alpha}\left(F_1(\boldsymbol{\mu}_1) + \langle \boldsymbol{\nu}_1^k, \boldsymbol{\mu}_1 \rangle\right)}\left(\boldsymbol{\zeta}^k\right)$$

$$\mu_2^{k+1} = \operatorname{prox}^{W_\epsilon}_{\frac{1}{\alpha}\left(F_2(\boldsymbol{\mu}_2) + \langle \boldsymbol{\nu}_2^k, \boldsymbol{\mu}_2 \rangle\right)}\left(\boldsymbol{\zeta}^k\right)$$

μ_2^{k+1} μ_1^{k+1}

μ_2^{k+1} μ_1^{k+1}



$\boldsymbol{u}_i^{\mathrm{opt}}$

$\boldsymbol{\nu}_1^{k+1}$ $\boldsymbol{\nu}_2^{k+1}$

$$\boldsymbol{\zeta}^{k+1} = \exp\left(\boldsymbol{u}_i^{\mathrm{opt}}/\varepsilon\right) \odot \left(\boldsymbol{\Gamma}\left(\boldsymbol{\mu}_i^{k+1} \oslash \left(\boldsymbol{\Gamma} \exp\left(\boldsymbol{u}_i^{\mathrm{opt}}/\varepsilon\right)\right)\right)\right) \in \Delta^{N-1}, \quad \text{for all } i \in \{1,2\}$$

μ_1^{k+1} μ_2^{k+1} $\boldsymbol{\zeta}^{k+1}$

$$\boldsymbol{\nu}_1^{k+1} = \boldsymbol{\nu}_1^k + \alpha\left(\boldsymbol{\mu}_1^{k+1} - \boldsymbol{\zeta}^{k+1}\right)$$

$$\boldsymbol{\nu}_2^{k+1} = \boldsymbol{\nu}_2^k + \alpha\left(\boldsymbol{\mu}_2^{k+1} - \boldsymbol{\zeta}^{k+1}\right)$$

Inner Layer ADMM

$$\boldsymbol{u}_1^{\ell+1} = \mathrm{prox}_{\frac{1}{\tau}f_1}^{\|\cdot\|_2}\left(\boldsymbol{z}_i^{\ell} - \widetilde{\boldsymbol{\nu}}_1^{\ell}\right)$$

$$\boldsymbol{u}_2^{\ell+1} = \mathrm{prox}_{\frac{1}{\tau}f_2}^{\|\cdot\|_2}\left(\boldsymbol{z}_2^{\ell} - \widetilde{\boldsymbol{\nu}}_2^{\ell}\right)$$

$\boldsymbol{u}_1^{\ell+1}$

$\boldsymbol{u}_2^{\ell+1}$

$$\boldsymbol{z}_1^{\ell+1} = \left(\boldsymbol{u}_1^{\ell+1} - \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{u}_i^{\ell+1}\right) + \left(\widetilde{\boldsymbol{\nu}}_1^{\ell} - \frac{1}{n}\sum_{i=1}^{n}\widetilde{\boldsymbol{\nu}}_i^{\ell}\right) + \frac{2}{n\alpha}\boldsymbol{\nu}_{\text{sum}}^{k}$$

$$\boldsymbol{z}_2^{\ell+1} = \left(\boldsymbol{u}_2^{\ell+1} - \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{u}_i^{\ell+1}\right) + \left(\widetilde{\boldsymbol{\nu}}_2^{\ell} - \frac{1}{n}\sum_{i=1}^{n}\widetilde{\boldsymbol{\nu}}_i^{\ell}\right) + \frac{2}{n\alpha}\boldsymbol{\nu}_{\text{sum}}^{k}$$

$\widetilde{\boldsymbol{\nu}}_2^{\ell+1}$

$\widetilde{\boldsymbol{\nu}}_1^{\ell+1}$

$\boldsymbol{z}_1^{\ell+1}$ $\boldsymbol{u}_1^{\ell+1}$

$\boldsymbol{z}_2^{\ell+1}$ $\boldsymbol{u}_2^{\ell+1}$

$$\widetilde{\boldsymbol{\nu}}_1^{\ell+1} = \widetilde{\boldsymbol{\nu}}_1^{\ell} + \left(\boldsymbol{u}_1^{\ell+1} - \boldsymbol{z}_1^{\ell+1}\right)$$

$$\widetilde{\boldsymbol{\nu}}_2^{\ell+1} = \widetilde{\boldsymbol{\nu}}_2^{\ell} + \left(\boldsymbol{u}_2^{\ell+1} - \boldsymbol{z}_2^{\ell+1}\right)$$

The μ Update

$$\boldsymbol{\mu}_i^{k+1} = \mathrm{prox}_{\frac{1}{\alpha}\left(F_i(\boldsymbol{\mu}_i) + \langle \boldsymbol{\nu}_i^k, \boldsymbol{\mu}_i \rangle\right)}\left(\boldsymbol{\zeta}^k\right) = \mathrm{arginf}_{\boldsymbol{\mu}_i \in \Delta^{N-1}} \left\{ \min_{\boldsymbol{M} \in \Pi_N(\boldsymbol{\mu}_i, \boldsymbol{\zeta}^k)} \frac{1}{2}\langle \boldsymbol{C}, \boldsymbol{M} \rangle + \frac{1}{\alpha}\left(F_i\left(\boldsymbol{\mu}_i\right) + \left\langle \boldsymbol{\nu}_i^k, \boldsymbol{\mu}_i \right\rangle\right) \right\}$$

Theorem

Given $\boldsymbol{a} \in \mathbb{R}^N \backslash \{\mathbf{0}\}$

Let $\Phi(\boldsymbol{\mu}) := \langle \boldsymbol{a}, \boldsymbol{\mu} \rangle$ for $\boldsymbol{\mu} \in \Delta^{N-1}$ and $\Gamma := \exp(-\boldsymbol{C}/2\varepsilon)$

Then for any $\zeta \in \Delta^{N-1}, \alpha > 0$

$$\mathrm{prox}_{\frac{1}{\alpha}\Phi}^{W_\varepsilon}(\boldsymbol{\zeta}) = \exp\left(-\tfrac{1}{\alpha\varepsilon}\boldsymbol{a}\right) \odot \left(\boldsymbol{\Gamma}^\top \left(\boldsymbol{\zeta} \oslash \left(\boldsymbol{\Gamma} \exp\left(-\tfrac{1}{\alpha\varepsilon}\boldsymbol{a}\right)\right)\right)\right)$$

The ζ Update

$$\boldsymbol{\zeta}^{k+1} = \underset{\boldsymbol{\zeta}\in\Delta^{N-1}}{\arg\inf}\left\{\left(\sum_{i=1}^{n}\min_{\boldsymbol{M}_i\in\Pi_N(\boldsymbol{\mu}_i^{k+1},\boldsymbol{\zeta})}\left\langle\tfrac{1}{2}\boldsymbol{C}+\varepsilon\log\boldsymbol{M}_i,\boldsymbol{M}_i\right\rangle\right)-\tfrac{2}{\alpha}\left\langle\boldsymbol{\nu}_{\mathrm{sum}}^{k},\boldsymbol{\zeta}\right\rangle\right\}$$

Theorem Given $\alpha, \varepsilon > 0$

Let $\boldsymbol{\Gamma} := \exp(-\boldsymbol{C}/2\varepsilon)$

Then

$$\zeta^{k+1} = \exp\left(\boldsymbol{u}_i^{\mathrm{opt}}/\varepsilon\right)\odot\left(\boldsymbol{\Gamma}\left(\boldsymbol{\mu}_i^{k+1}\oslash\left(\boldsymbol{\Gamma}\exp\left(\boldsymbol{u}_i^{\mathrm{opt}}/\varepsilon\right)\right)\right)\right)\in\Delta^{N-1}, \quad \text{for all } i\in[n]$$

Where

$$\left(\boldsymbol{u}_1^{\mathrm{opt}},\ldots,\boldsymbol{u}_n^{\mathrm{opt}}\right) = \underset{(\boldsymbol{u}_1,\ldots,\boldsymbol{u}_n)\in\mathbb{R}^{nN}}{\arg\min}\sum_{i=1}^{n}\left\langle\boldsymbol{\mu}_i^{k+1},\log\left(\boldsymbol{\Gamma}\exp\left(\boldsymbol{u}_i/\varepsilon\right)\right)\right\rangle$$

$$\text{Subject to} \quad \sum_{i=1}^{n}\boldsymbol{u}_i = \frac{2}{\alpha}\boldsymbol{\nu}_{\mathrm{sum}}^{k}$$

The ζ Update: Inner Layer ADMM

Newton's method

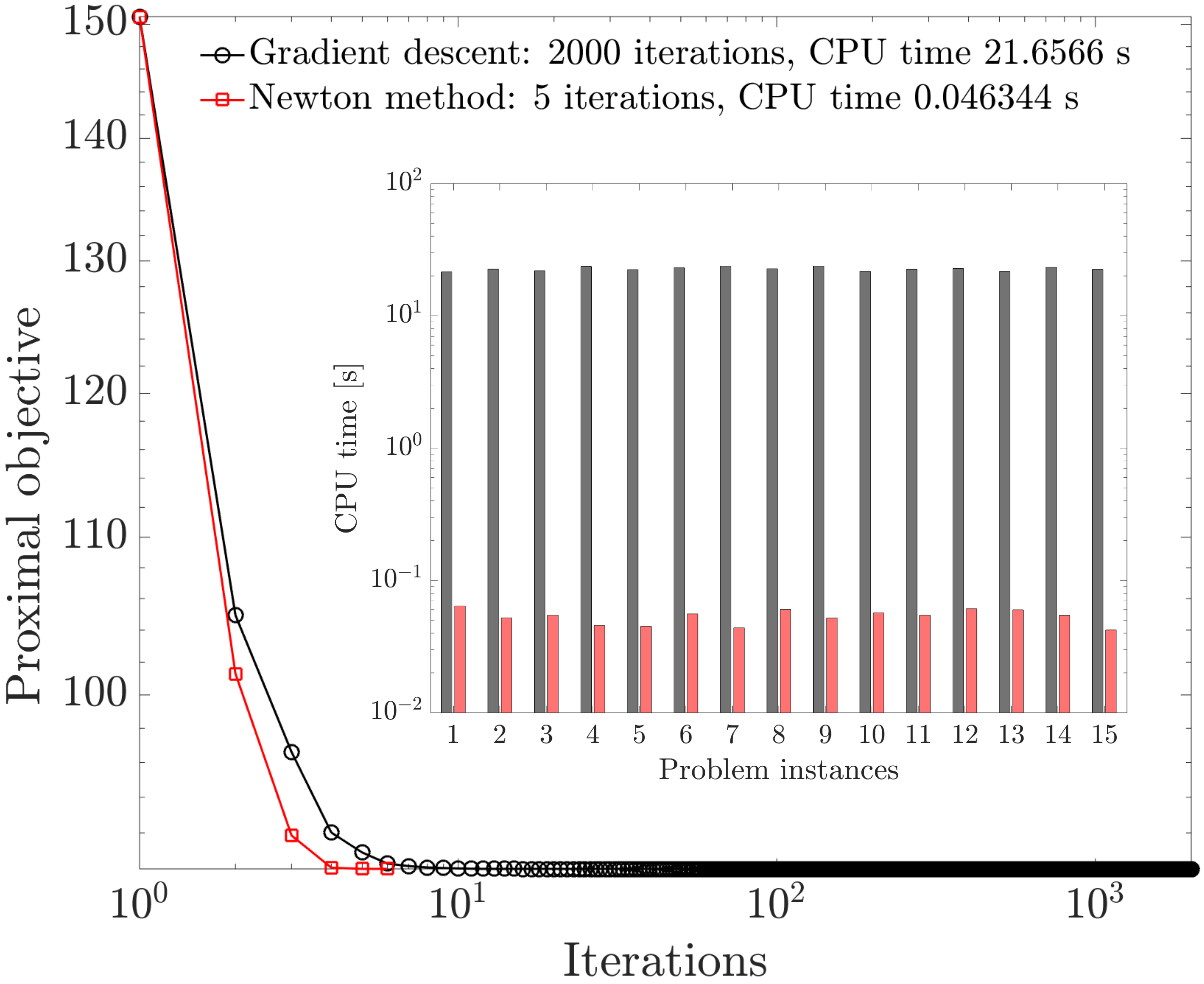
$$\boldsymbol{u}_i^{\ell+1} = \operatorname{prox}_{\frac{1}{\tau}f_i}^{\|\cdot\|_2}\left(\boldsymbol{z}_i^{\ell} - \widetilde{\boldsymbol{\nu}}_i^{\ell}\right), \quad i \in [n],$$

$$\boldsymbol{z}^{\ell+1} = \operatorname{proj}_{\mathscr{C}}\left(\boldsymbol{u}^{\ell+1} + \widetilde{\boldsymbol{\nu}}^{\ell}\right),$$

$$\widetilde{\boldsymbol{\nu}}_i^{\ell+1} = \widetilde{\boldsymbol{\nu}}_i^{\ell} + \left(\boldsymbol{u}_i^{\ell+1} - \boldsymbol{z}_i^{\ell+1}\right), \quad i \in [n],$$

$$\boldsymbol{z}_i^{\ell+1} = \left(\boldsymbol{u}_i^{\ell+1} - \frac{1}{n}\sum_{i=1}^n \boldsymbol{u}_i^{\ell+1}\right) + \left(\widetilde{\boldsymbol{\nu}}_i^{\ell} - \frac{1}{n}\sum_{i=1}^n \widetilde{\boldsymbol{\nu}}_i^{\ell}\right) + \frac{2}{n\alpha}\boldsymbol{\nu}_{\mathrm{sum}}^k, \quad i \in [n]$$

$$\operatorname{proj}_{\mathscr{C}}(\boldsymbol{v}) = \left(\boldsymbol{v}_1 - \overline{\boldsymbol{v}} + \frac{2}{n\alpha}\boldsymbol{\nu}_{\mathrm{sum}}^k, \ldots, \boldsymbol{v}_n - \overline{\boldsymbol{v}} + \frac{2}{n\alpha}\boldsymbol{\nu}_{\mathrm{sum}}^k\right) \in \mathbb{R}^{nN}$$



Gradient descent: 2000 iterations, CPU time 21.6566 s
Newton method: 5 iterations, CPU time 0.046344 s
Proximal objective
Iterations
150
140
130
120
110
100
10^0
10^1
10^2
10^3
CPU time [s]
Problem instances
10^2
10^1
10^0
10^{-1}
10^{-2}
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15

Near term Publications Plan

I. N., and A. Halder. Schrödinger Meets Kuramoto via Feynman-Kac: Minimum Effort Distribution Steering for Noisy Nonuniform Kuramoto Oscillators.

I. N., and A. Halder, Wasserstein Consensus ADMM.

Future Timeline

Numerical case studies for the distributed algorithms **(Winter - Spring 2022)**

Optimal distribution steering algorithms for molecular self-assembly **(Summer - Fall 2022)**

Adaptive distributional learning and control **(Fall 2022 - Spring 2023)**

Application to policy optimization for reinforcement learning **(Spring - Summer 2023)**

Write dissertation and graduate **(Fall 2023 - Winter 2024)**

Thank You