

A Distributed Algorithm for Measure-valued Optimization with Additive Objective

Abhishek Halder

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

Joint work with I. Nodoozi (UC Santa Cruz)

Invited Session: Optimal Transport in Networks and Systems, MTNS 2022

September 15, 2022



Measure-valued Optimization Problems

$$\arg \inf_{\mu \in \mathcal{P}_2(\mathbb{R}^d)} F(\mu)$$

Space of Borel probability measures
on \mathbb{R}^d with finite second moments

2-Wasserstein geodescially
convex functional

In many applications, we have additive structure:

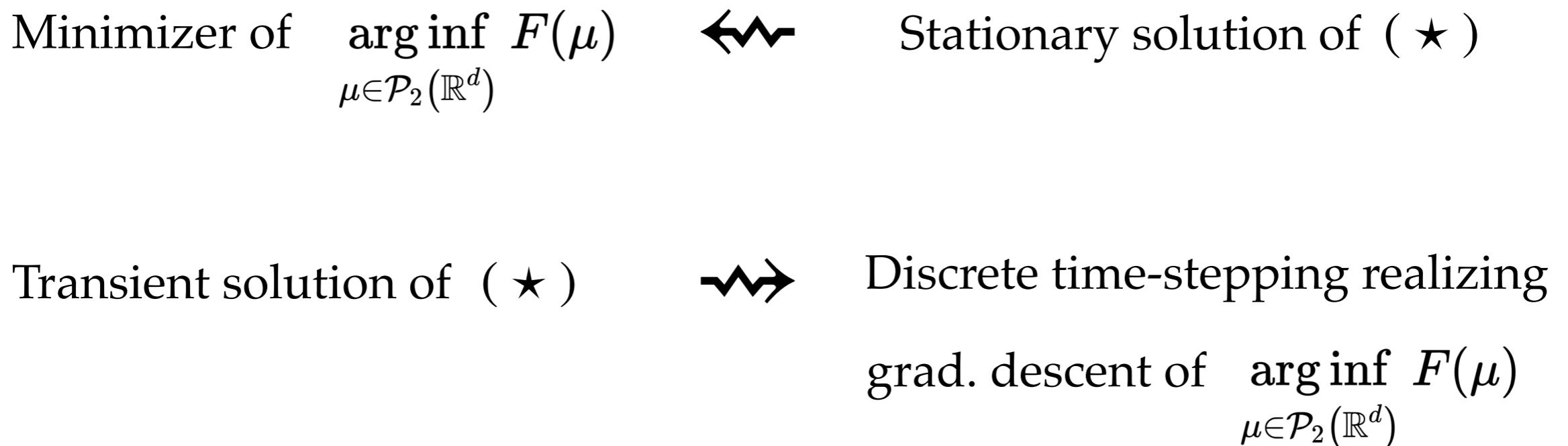
$$F(\mu) = F_1(\mu) + F_2(\mu) + \dots + F_n(\mu)$$

where each $F_i : \mathcal{P}_2(\mathbb{R}^d) \mapsto (-\infty, +\infty]$ is proper, lsc,
and 2-Wasserstein geodescially convex

Connection with Wasserstein Gradient Flows

$$\frac{\partial \mu}{\partial t} = -\nabla^{W_2} F(\mu) := \nabla \cdot \left(\mu \frac{\delta F}{\delta \mu} \right) \quad (\star)$$

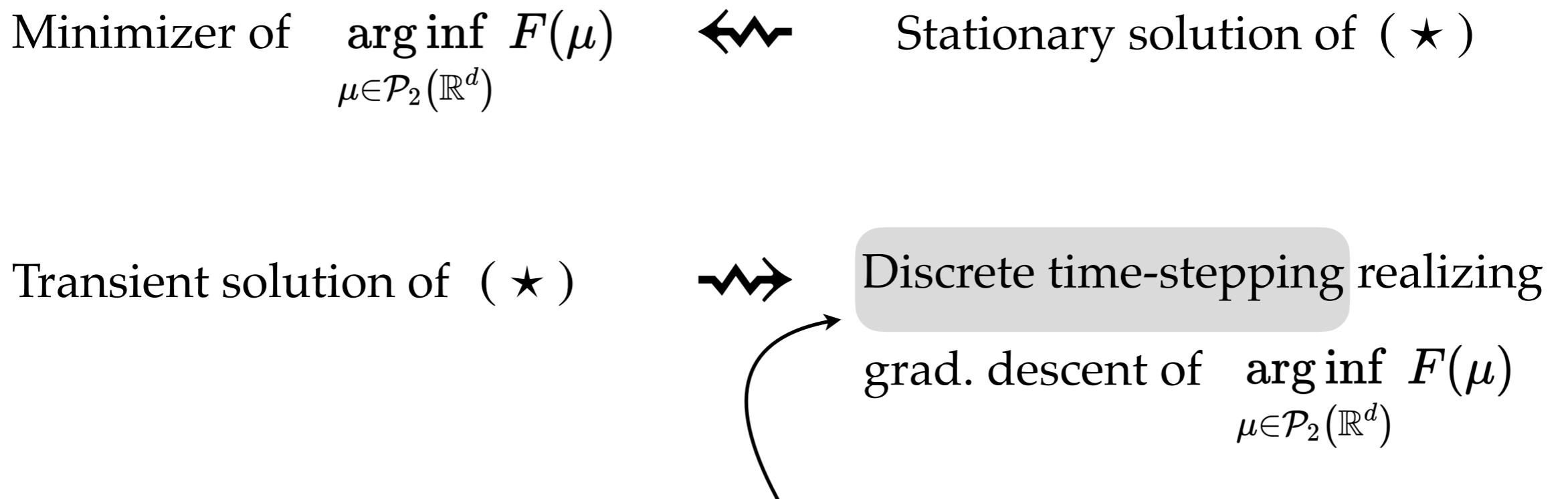
Wasserstein gradient



Connection with Wasserstein Gradient Flows

$$\frac{\partial \mu}{\partial t} = -\nabla^{W_2} F(\mu) := \nabla \cdot \left(\mu \frac{\delta F}{\delta \mu} \right) \quad (\star)$$

Wasserstein gradient



Wasserstein proximal recursion à la Jordan-Kinderlehrer-Otto (JKO) scheme

Gradient Flows

Gradient Flow in \mathcal{X}

$$\frac{d\mathbf{x}}{dt} = -\nabla f(\mathbf{x}), \quad \mathbf{x}(0) = \mathbf{x}_0$$

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

$$\frac{\partial \mu}{\partial t} = -\nabla^W F(\mu), \quad \mu(\mathbf{x}, 0) = \mu_0$$

Recursion:

$$\begin{aligned}\mathbf{x}_k &= \mathbf{x}_{k-1} - h \nabla f(\mathbf{x}_k) \\ &= \arg \min_{\mathbf{x} \in \mathcal{X}} \left\{ \frac{1}{2} \|\mathbf{x} - \mathbf{x}_{k-1}\|_2^2 + h f(\mathbf{x}) \right\} \\ &=: \text{prox}_{hf}^{\|\cdot\|_2}(\mathbf{x}_{k-1})\end{aligned}$$

Recursion:

$$\begin{aligned}\mu_k &= \mu(\cdot, t = kh) \\ &= \arg \min_{\mu \in \mathcal{P}_2(\mathcal{X})} \left\{ \frac{1}{2} W^2(\mu, \mu_{k-1}) + h F(\mu) \right\} \\ &=: \text{prox}_{hF}^W(\mu_{k-1})\end{aligned}$$

Convergence:

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

Convergence:

$$\mu_k \rightarrow \mu(\cdot, t = kh) \quad \text{as} \quad h \downarrow 0$$

f as Lyapunov function:

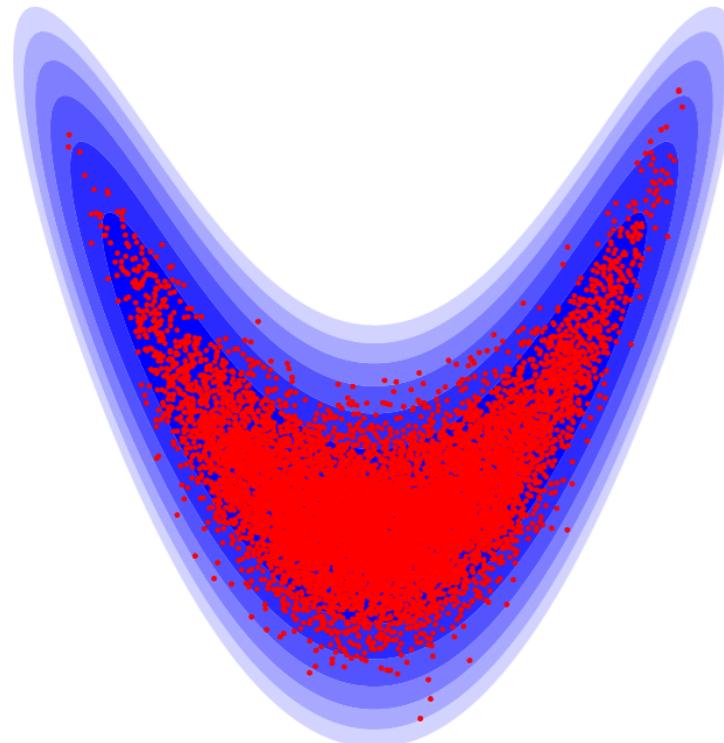
$$\frac{d}{dt} f = - \|\nabla f\|_2^2 \leq 0$$

F as Lyapunov functional:

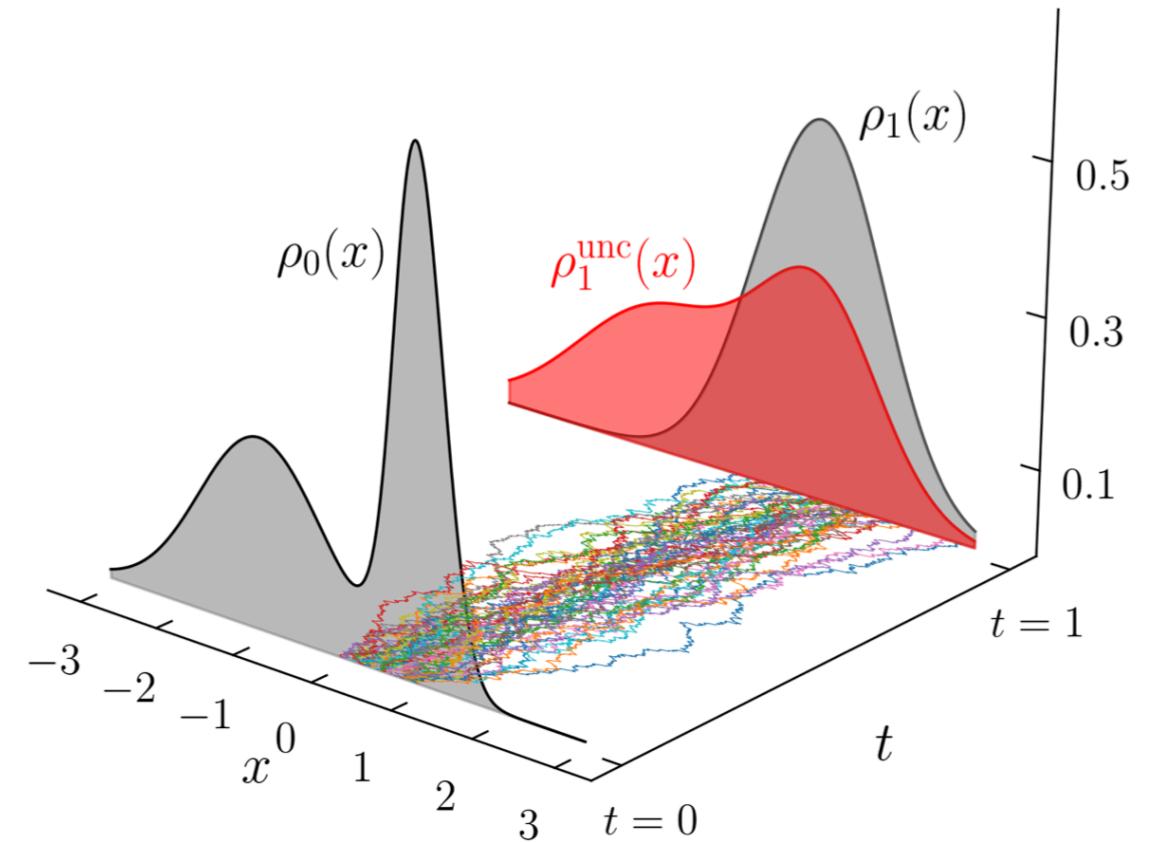
$$\frac{d}{dt} F = -\mathbb{E}_\mu \left[\left\| \nabla \frac{\delta F}{\delta \mu} \right\|_2^2 \right] \leq 0$$

Motivating Applications

Langevin sampling from
an unnormalized prior



Optimal control of distributions
a.k.a. Schrödinger bridge problems



Stramer and Tweedie, *Methodology and Computing in Applied Probability*, 1999

Jarner and Hansen, *Stochastic Processes and their Applications*, 2000

Roberts and Stramer, *Methodology and Computing in Applied Probability*, 2002

Vempala and Wibisino, *NeurIPS*, 2019

Chen, Georgiou and Pavon, *SIAM Review*, 2021

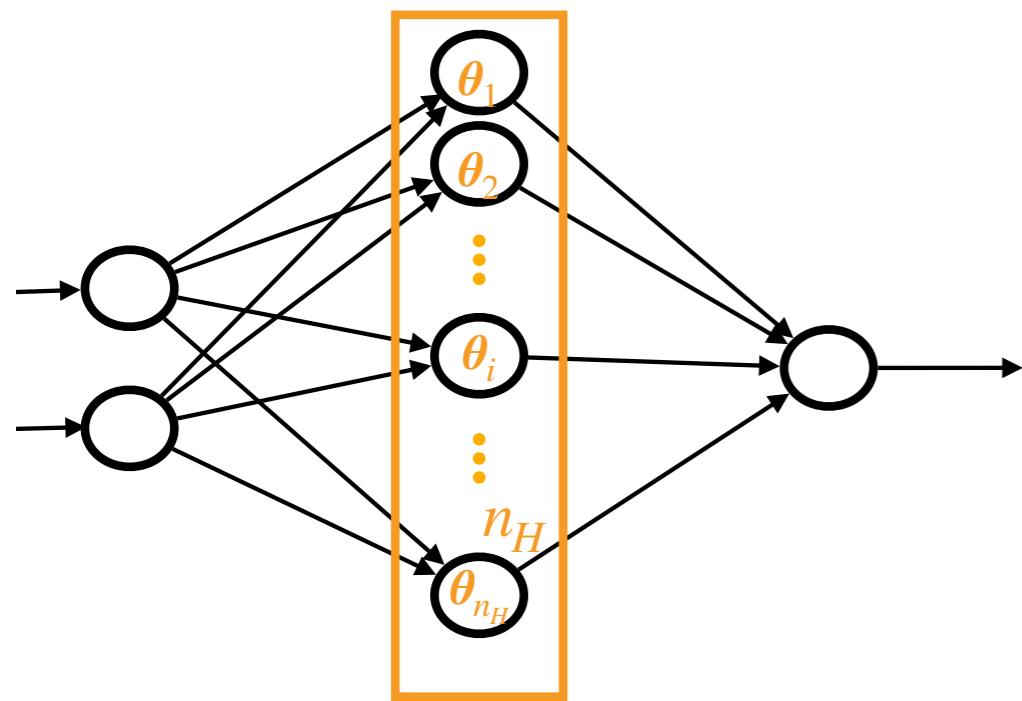
Chen, Georgiou and Pavon, *SIAM Journal on Applied Mathematics*, 2016

Chen, Georgiou and Pavon, *Journal on Optimization Theory and Applications*, 2016

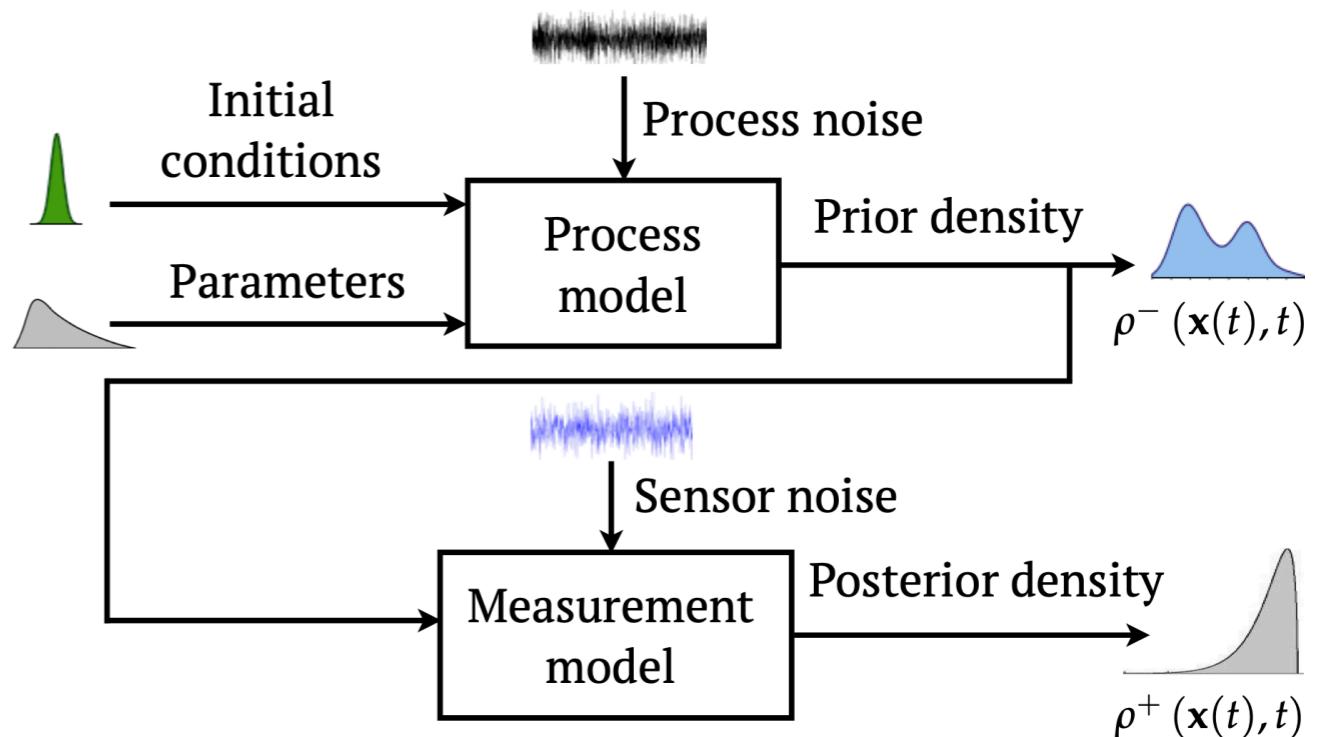
Caluya and Halder, *IEEE Transactions on Automatic Control*, 2021

Motivating Applications (contd.)

Mean field learning dynamics
in neural networks



Prediction and estimation of time-varying
joint state probability densities



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

Caluya and Halder, *IEEE Transactions on Automatic Control*, 2019

Halder and Georgiou, *CDC*, 2019

Halder and Georgiou, *ACC*, 2018

Halder and Georgiou, *CDC*, 2017

Many Recently Proposed Algorithms to Solve Measure-valued Optimization Problems

Peyré, *SIAM Journal on Imaging Sciences*, 2015

Benamou, Carlier and Laborde, *ESAIM: Proceedings and Surveys*, 2016

Carlier, Duval, Peyré and Schimtzer, *SIAM Journal on Mathematical Analysis*, 2017

Karlsson and Ringh, *SIAM Journal on Imaging Sciences*, 2017

Caluya and Halder, *IEEE Transactions on Automatic Control*, 2019

Carrillo, Craig, Wang and Wei, *Foundations of Computational Mathematics*, 2021

Mokrov, Korotin, Li, Gnevay, Solomon, and Burnaev, *NeurIPS*, 2021

Alvarez-Melis, Schiff, and Mroueh, *NeurIPS*, 2021

Many Recently Proposed Algorithms to Solve Measure-valued Optimization Problems

Peyré, *SIAM Journal on Imaging Sciences*, 2015

Benamou, Carlier and Laborde, *ESAIM: Proceedings and Surveys*, 2016

Carlier, Duval, Peyré and Schimtzer, *SIAM Journal on Mathematical Analysis*, 2017

Karlsson and Ringh, *SIAM Journal on Imaging Sciences*, 2017

Caluya and Halder, *IEEE Transactions on Automatic Control*, 2019

Carrillo, Craig, Wang and Wei, *Foundations of Computational Mathematics*, 2021

Mokrov, Korotin, Li, Gnevay, Solomon, and Burnaev, *NeurIPS*, 2021

Alvarez-Melis, Schiff, and Mroueh, *NeurIPS*, 2021

But all require centralized computing

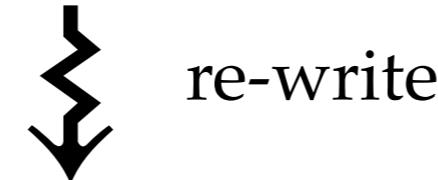
Present Work: Distributed Algorithm

$$\arg \inf_{\mu \in \mathcal{P}_2(\mathbb{R}^d)} F_1(\mu) + F_2(\mu) + \dots + F_n(\mu)$$

Present Work: Distributed Algorithm

$$\arg \inf_{\mu \in \mathcal{P}_2(\mathbb{R}^d)} F_1(\mu) + F_2(\mu) + \dots + F_n(\mu)$$

Main idea:



$$\begin{aligned} & \arg \inf_{(\mu_1, \dots, \mu_n, \zeta) \in \mathcal{P}_2^{n+1}(\mathbb{R}^d)} F_1(\mu_1) + F_2(\mu_2) + \dots + F_n(\mu_n) \\ & \text{subject to} \quad \mu_i = \zeta \quad \text{for all } i \in [n] \end{aligned}$$

Present Work: Distributed Algorithm

$$\arg \inf_{\mu \in \mathcal{P}_2(\mathbb{R}^d)} F_1(\mu) + F_2(\mu) + \dots + F_n(\mu)$$

Main idea:

↓ re-write

$$\begin{aligned} & \arg \inf_{(\mu_1, \dots, \mu_n, \zeta) \in \mathcal{P}_2^{n+1}(\mathbb{R}^d)} F_1(\mu_1) + F_2(\mu_2) + \dots + F_n(\mu_n) \\ & \text{subject to } \mu_i = \zeta \quad \text{for all } i \in [n] \end{aligned}$$

Define Wasserstein augmented Lagrangian:

$$L_\alpha(\mu_1, \dots, \mu_n, \zeta, \nu_1, \dots, \nu_n) := \sum_{i=1}^n \left\{ F_i(\mu_i) + \frac{\alpha}{2} W^2(\mu_i, \zeta) + \int_{\mathbb{R}^d} \nu_i(\theta) (\mathrm{d}\mu_i - \mathrm{d}\zeta) \right\}$$

↑ regularization > 0 ↑ (Lagrange multipliers

Proposed Consensus ADMM

$$\begin{aligned}\mu_i^{k+1} &= \arg \inf_{\mu_i \in \mathcal{P}_2(\mathbb{R}^d)} L_\alpha(\mu_1, \dots, \mu_n, \zeta^k, \nu_1^k, \dots, \nu_n^k) \\ \zeta^{k+1} &= \arg \inf_{\zeta \in \mathcal{P}_2(\mathbb{R}^d)} L_\alpha(\mu_1^{k+1}, \dots, \mu_n^{k+1}, \zeta, \nu_1^k, \dots, \nu_n^k) \\ \nu_i^{k+1} &= \nu_i^k + \alpha(\mu_i^{k+1} - \zeta^{k+1})\end{aligned}\quad \text{where } i \in [n], k \in \mathbb{N}_0$$

Proposed Consensus ADMM

$$\begin{aligned}
\mu_i^{k+1} &= \arg \inf_{\mu_i \in \mathcal{P}_2(\mathbb{R}^d)} L_\alpha(\mu_1, \dots, \mu_n, \zeta^k, \nu_1^k, \dots, \nu_n^k) \\
\zeta^{k+1} &= \arg \inf_{\zeta \in \mathcal{P}_2(\mathbb{R}^d)} L_\alpha(\mu_1^{k+1}, \dots, \mu_n^{k+1}, \zeta, \nu_1^k, \dots, \nu_n^k) \\
\nu_i^{k+1} &= \nu_i^k + \alpha(\mu_i^{k+1} - \zeta^{k+1})
\end{aligned}
\quad \text{where } i \in [n], k \in \mathbb{N}_0$$

Define

$$\nu_{\text{sum}}^k(\boldsymbol{\theta}) := \sum_{i=1}^n \nu_i^k(\boldsymbol{\theta}), \quad k \in \mathbb{N}_0$$

and simplify the recursions to

$$\begin{aligned}
\mu_i^{k+1} &= \text{prox}_{\frac{1}{\alpha}(F_i(\cdot) + \int \nu_i^k d(\cdot))}^W(\zeta^k) \\
\zeta^{k+1} &= \arg \inf_{\zeta \in \mathcal{P}_2(\mathbb{R}^d)} \left\{ \left(\sum_{i=1}^n W^2(\mu_i^{k+1}, \zeta) \right) - \frac{2}{\alpha} \int_{\mathbb{R}^d} \nu_{\text{sum}}^k(\boldsymbol{\theta}) d\zeta \right\} \\
\nu_i^{k+1} &= \nu_i^k + \alpha(\mu_i^{k+1} - \zeta^{k+1})
\end{aligned}$$

Proposed Consensus ADMM (contd.)

$$\mu_i^{k+1} = \text{prox}_{\frac{1}{\alpha}(F_i(\cdot) + \int \nu_i^k d(\cdot))}^W(\zeta^k)$$

$$\zeta^{k+1} = \arg \inf_{\zeta \in \mathcal{P}_2(\mathbb{R}^d)} \left\{ \left(\sum_{i=1}^n W^2(\mu_i^{k+1}, \zeta) \right) - \frac{2}{\alpha} \int_{\mathbb{R}^d} \nu_{\text{sum}}^k(\boldsymbol{\theta}) d\zeta \right\}$$

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

Split free energy functionals: $\Phi_i(\mu_i) := F_i(\mu_i) + \int_{\mathbb{R}^d} \nu_i^k d\mu_i$

\therefore Distributed Wasserstein prox \approx time updates of $\frac{\partial \tilde{\mu}_i}{\partial t} = -\nabla^W \Phi_i(\tilde{\mu}_i)$

Proposed Consensus ADMM (contd.)

$$\mu_i^{k+1} = \text{prox}_{\frac{1}{\alpha}(F_i(\cdot) + \int \nu_i^k d(\cdot))}^W(\zeta^k)$$

$$\zeta^{k+1} = \arg \inf_{\zeta \in \mathcal{P}_2(\mathbb{R}^d)} \left\{ \left(\sum_{i=1}^n W^2(\mu_i^{k+1}, \zeta) \right) - \frac{2}{\alpha} \int_{\mathbb{R}^d} \nu_{\text{sum}}^k(\boldsymbol{\theta}) d\zeta \right\}$$

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

Split free energy functionals: $\Phi_i(\mu_i) := F_i(\mu_i) + \int_{\mathbb{R}^d} \nu_i^k d\mu_i$

\therefore Distributed Wasserstein prox \approx time updates of $\frac{\partial \tilde{\mu}_i}{\partial t} = -\nabla^W \Phi_i(\tilde{\mu}_i)$

Examples:

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

Discrete Version of the Proposed ADMM

$$\begin{aligned}
 \boldsymbol{\mu}_i^{k+1} &= \text{prox}_{\frac{1}{\alpha}(F_i(\boldsymbol{\mu}_i) + \langle \boldsymbol{\nu}_i^k, \boldsymbol{\mu}_i \rangle)}^W(\boldsymbol{\zeta}^k) \\
 &= \arg \inf_{\boldsymbol{\mu}_i \in \Delta^{N-1}} \left\{ \min_{\boldsymbol{M} \in \Pi_N(\boldsymbol{\mu}_i, \boldsymbol{\zeta}^k)} \frac{1}{2} \langle \boldsymbol{C}, \boldsymbol{M} \rangle + \frac{1}{\alpha} (F_i(\boldsymbol{\mu}_i) + \langle \boldsymbol{\nu}_i^k, \boldsymbol{\mu}_i \rangle) \right\} \\
 \boldsymbol{\zeta}^{k+1} &= \arg \inf_{\boldsymbol{\zeta} \in \Delta^{N-1}} \left\{ \left(\sum_{i=1}^n \min_{\boldsymbol{M}_i \in \Pi_N(\boldsymbol{\mu}_i^{k+1}, \boldsymbol{\zeta})} \frac{1}{2} \langle \boldsymbol{C}, \boldsymbol{M}_i \rangle \right) - \frac{2}{\alpha} \langle \boldsymbol{\nu}_{\text{sum}}^k, \boldsymbol{\zeta} \rangle \right\} \\
 \boldsymbol{\nu}_i^{k+1} &= \boldsymbol{\nu}_i^k + \alpha (\boldsymbol{\mu}_i^{k+1} - \boldsymbol{\zeta}^{k+1})
 \end{aligned}$$

Euclidean distance matrix
where N is the number of samples

Discrete Version of the Proposed ADMM

$$\begin{aligned}
\boldsymbol{\mu}_i^{k+1} &= \text{prox}_{\frac{1}{\alpha}(F_i(\boldsymbol{\mu}_i) + \langle \boldsymbol{\nu}_i^k, \boldsymbol{\mu}_i \rangle)}^W(\boldsymbol{\zeta}^k) \\
&= \arg \inf_{\boldsymbol{\mu}_i \in \Delta^{N-1}} \left\{ \min_{\boldsymbol{M} \in \Pi_N(\boldsymbol{\mu}_i, \boldsymbol{\zeta}^k)} \frac{1}{2} \langle \boldsymbol{C}, \boldsymbol{M} \rangle + \frac{1}{\alpha} (F_i(\boldsymbol{\mu}_i) + \langle \boldsymbol{\nu}_i^k, \boldsymbol{\mu}_i \rangle) \right\} \\
\boldsymbol{\zeta}^{k+1} &= \arg \inf_{\boldsymbol{\zeta} \in \Delta^{N-1}} \left\{ \left(\sum_{i=1}^n \min_{\boldsymbol{M}_i \in \Pi_N(\boldsymbol{\mu}_i^{k+1}, \boldsymbol{\zeta})} \frac{1}{2} \langle \boldsymbol{C}, \boldsymbol{M}_i \rangle \right) - \frac{2}{\alpha} \langle \boldsymbol{\nu}_{\text{sum}}^k, \boldsymbol{\zeta} \rangle \right\} \\
\boldsymbol{\nu}_i^{k+1} &= \boldsymbol{\nu}_i^k + \alpha (\boldsymbol{\mu}_i^{k+1} - \boldsymbol{\zeta}^{k+1})
\end{aligned}$$

With Sinkhorn regularization:

Discrete Sinkhorn divergence

$$\begin{aligned}
\boldsymbol{\mu}_i^{k+1} &= \text{prox}_{\frac{1}{\alpha}(F_i(\boldsymbol{\mu}_i) + \langle \boldsymbol{\nu}_i^k, \boldsymbol{\mu}_i \rangle)}^{W_\varepsilon}(\boldsymbol{\zeta}^k) \\
&= \arg \inf_{\boldsymbol{\mu}_i \in \Delta^{N-1}} \left\{ \min_{\boldsymbol{M} \in \Pi_N(\boldsymbol{\mu}_i, \boldsymbol{\zeta}^k)} \left\langle \frac{1}{2} \boldsymbol{C} + \varepsilon \log \boldsymbol{M}, \boldsymbol{M} \right\rangle + \frac{1}{\alpha} (F_i(\boldsymbol{\mu}_i) + \langle \boldsymbol{\nu}_i^k, \boldsymbol{\mu}_i \rangle) \right\} \\
\boldsymbol{\zeta}^{k+1} &= \arg \inf_{\boldsymbol{\zeta} \in \Delta^{N-1}} \left\{ \left(\sum_{i=1}^n \min_{\boldsymbol{M}_i \in \Pi_N(\boldsymbol{\mu}_i^{k+1}, \boldsymbol{\zeta})} \left\langle \frac{1}{2} \boldsymbol{C} + \varepsilon \log \boldsymbol{M}_i, \boldsymbol{M}_i \right\rangle \right) - \frac{2}{\alpha} \langle \boldsymbol{\nu}_{\text{sum}}^k, \boldsymbol{\zeta} \rangle \right\} \\
\boldsymbol{\nu}_i^{k+1} &= \boldsymbol{\nu}_i^k + \alpha (\boldsymbol{\mu}_i^{k+1} - \boldsymbol{\zeta}^{k+1})
\end{aligned}$$

Discrete Version of the Proposed ADMM

$$\begin{aligned}
\boldsymbol{\mu}_i^{k+1} &= \text{prox}_{\frac{1}{\alpha}(F_i(\boldsymbol{\mu}_i) + \langle \boldsymbol{\nu}_i^k, \boldsymbol{\mu}_i \rangle)}^W(\boldsymbol{\zeta}^k) \\
&= \arg \inf_{\boldsymbol{\mu}_i \in \Delta^{N-1}} \left\{ \min_{\boldsymbol{M} \in \Pi_N(\boldsymbol{\mu}_i, \boldsymbol{\zeta}^k)} \frac{1}{2} \langle \boldsymbol{C}, \boldsymbol{M} \rangle + \frac{1}{\alpha} (F_i(\boldsymbol{\mu}_i) + \langle \boldsymbol{\nu}_i^k, \boldsymbol{\mu}_i \rangle) \right\} \\
\boldsymbol{\zeta}^{k+1} &= \arg \inf_{\boldsymbol{\zeta} \in \Delta^{N-1}} \left\{ \left(\sum_{i=1}^n \min_{\boldsymbol{M}_i \in \Pi_N(\boldsymbol{\mu}_i^{k+1}, \boldsymbol{\zeta})} \frac{1}{2} \langle \boldsymbol{C}, \boldsymbol{M}_i \rangle \right) - \frac{2}{\alpha} \langle \boldsymbol{\nu}_{\text{sum}}^k, \boldsymbol{\zeta} \rangle \right\} \\
\boldsymbol{\nu}_i^{k+1} &= \boldsymbol{\nu}_i^k + \alpha (\boldsymbol{\mu}_i^{k+1} - \boldsymbol{\zeta}^{k+1})
\end{aligned}$$

With Sinkhorn regularization:

Outer layer ADMM

$$\begin{aligned}
\boldsymbol{\mu}_i^{k+1} &= \text{prox}_{\frac{1}{\alpha}(F_i(\boldsymbol{\mu}_i) + \langle \boldsymbol{\nu}_i^k, \boldsymbol{\mu}_i \rangle)}^{W_\varepsilon}(\boldsymbol{\zeta}^k)
\end{aligned}$$

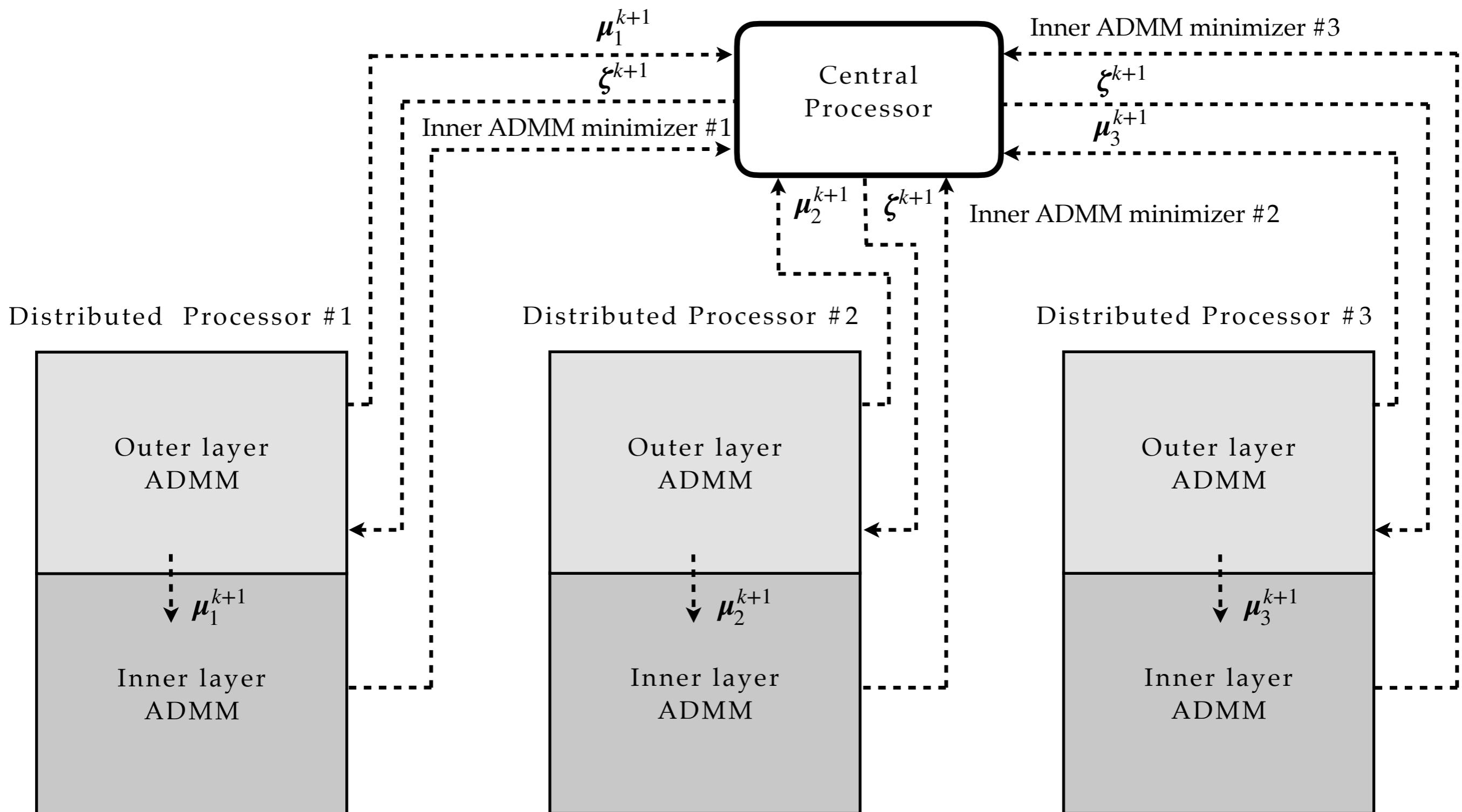
Discrete Sinkhorn divergence

$$\begin{aligned}
&= \arg \inf_{\boldsymbol{\mu}_i \in \Delta^{N-1}} \left\{ \min_{\boldsymbol{M} \in \Pi_N(\boldsymbol{\mu}_i, \boldsymbol{\zeta}^k)} \left\langle \frac{1}{2} \boldsymbol{C} + \varepsilon \log \boldsymbol{M}, \boldsymbol{M} \right\rangle + \frac{1}{\alpha} (F_i(\boldsymbol{\mu}_i) + \langle \boldsymbol{\nu}_i^k, \boldsymbol{\mu}_i \rangle) \right\}
\end{aligned}$$

Inner layer ADMM

$$\begin{aligned}
\boldsymbol{\zeta}^{k+1} &= \arg \inf_{\boldsymbol{\zeta} \in \Delta^{N-1}} \left\{ \left(\sum_{i=1}^n \min_{\boldsymbol{M}_i \in \Pi_N(\boldsymbol{\mu}_i^{k+1}, \boldsymbol{\zeta})} \left\langle \frac{1}{2} \boldsymbol{C} + \varepsilon \log \boldsymbol{M}_i, \boldsymbol{M}_i \right\rangle \right) - \frac{2}{\alpha} \langle \boldsymbol{\nu}_{\text{sum}}^k, \boldsymbol{\zeta} \rangle \right\} \\
\boldsymbol{\nu}_i^{k+1} &= \boldsymbol{\nu}_i^k + \alpha (\boldsymbol{\mu}_i^{k+1} - \boldsymbol{\zeta}^{k+1})
\end{aligned}$$

Overall Schematic



μ_i update \rightsquigarrow Outer Consensus (Sinkhorn) ADMM

Example. $\Phi(\mu) := \langle a, \mu \rangle$, $a \in \mathbb{R}^N \setminus \{\mathbf{0}\}$, $\mu, \zeta \in \Delta^{N-1}$, $\Gamma := \exp(-C/2\epsilon)$, $\epsilon > 0$

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

μ_i update \rightsquigarrow Outer Consensus (Sinkhorn) ADMM

Example. $\Phi(\boldsymbol{\mu}) := \langle \mathbf{a}, \boldsymbol{\mu} \rangle$, $\mathbf{a} \in \mathbb{R}^N \setminus \{\mathbf{0}\}$, $\boldsymbol{\mu}, \boldsymbol{\zeta} \in \Delta^{N-1}$, $\Gamma := \exp(-C/2\varepsilon)$, $\varepsilon > 0$

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

Example. $G_i(\boldsymbol{\mu}_i) := F_i(\boldsymbol{\mu}_i) + \langle \boldsymbol{\nu}_i^k, \boldsymbol{\mu}_i \rangle$, $\boldsymbol{\zeta}^k \in \Delta^{N-1}$, $k \in \mathbb{N}_0$.

\uparrow
 Convex

$$\boldsymbol{\mu}_i^{k+1} = \text{prox}_{\frac{1}{\alpha}(F_i(\boldsymbol{\mu}_i) + \langle \boldsymbol{\nu}_i^k, \boldsymbol{\mu}_i \rangle)}^{W_\varepsilon}(\boldsymbol{\zeta}^k) = \exp\left(\frac{\boldsymbol{\lambda}_{1i}^{\text{opt}}}{\alpha\varepsilon}\right) \odot \left(\exp\left(-\frac{C^\top}{2\varepsilon}\right) \exp\left(\frac{\boldsymbol{\lambda}_{0i}^{\text{opt}}}{\alpha\varepsilon}\right) \right)$$

where $\boldsymbol{\lambda}_{0i}^{\text{opt}}, \boldsymbol{\lambda}_{1i}^{\text{opt}} \in \mathbb{R}^N$ solve

$$\exp\left(\frac{\boldsymbol{\lambda}_{0i}^{\text{opt}}}{\alpha\varepsilon}\right) \odot \left(\exp\left(-\frac{C}{2\varepsilon}\right) \exp\left(\frac{\boldsymbol{\lambda}_{1i}^{\text{opt}}}{\alpha\varepsilon}\right) \right) = \boldsymbol{\zeta}_k,$$

$$\mathbf{0} \in \partial_{\boldsymbol{\lambda}_{1i}^{\text{opt}}} G_i^*(-\boldsymbol{\lambda}_{1i}^{\text{opt}}) - \exp\left(\frac{\boldsymbol{\lambda}_{1i}^{\text{opt}}}{\alpha\varepsilon}\right) \odot \left(\exp\left(-\frac{C^\top}{2\varepsilon}\right) \exp\left(\frac{\boldsymbol{\lambda}_{0i}^{\text{opt}}}{\alpha\varepsilon}\right) \right).$$

ζ update \rightsquigarrow Inner (Euclidean) ADMM

Theorem.

Consider the convex problem

$$(\mathbf{u}_1^{\text{opt}}, \dots, \mathbf{u}_n^{\text{opt}}) = \arg \min_{(\mathbf{u}_1, \dots, \mathbf{u}_n) \in \mathbb{R}^{nN}} \sum_{i=1}^n \langle \boldsymbol{\mu}_i^{k+1}, \log(\Gamma \exp(\mathbf{u}_i/\varepsilon)) \rangle$$

(♥)

subject to $\sum_{i=1}^n \mathbf{u}_i = \frac{2}{\alpha} \boldsymbol{\nu}_{\text{sum}}^k$.

Then

$$\boldsymbol{\zeta}^{k+1} = \exp(\mathbf{u}_i^{\text{opt}}/\varepsilon) \odot (\Gamma(\boldsymbol{\mu}_i^{k+1} \oslash (\Gamma \exp(\mathbf{u}_i^{\text{opt}}/\varepsilon)))) \in \Delta^{N-1} \quad \forall i \in [n].$$

ζ update \rightsquigarrow Inner (Euclidean) ADMM

Theorem.

Let $f_i(\mathbf{u}_i) := \langle \boldsymbol{\mu}_i^{k+1}, \log(\Gamma \exp(\mathbf{u}_i/\varepsilon)) \rangle$, $\mathbf{u}_i \in \mathbb{R}^N$, for all $i \in [n]$,

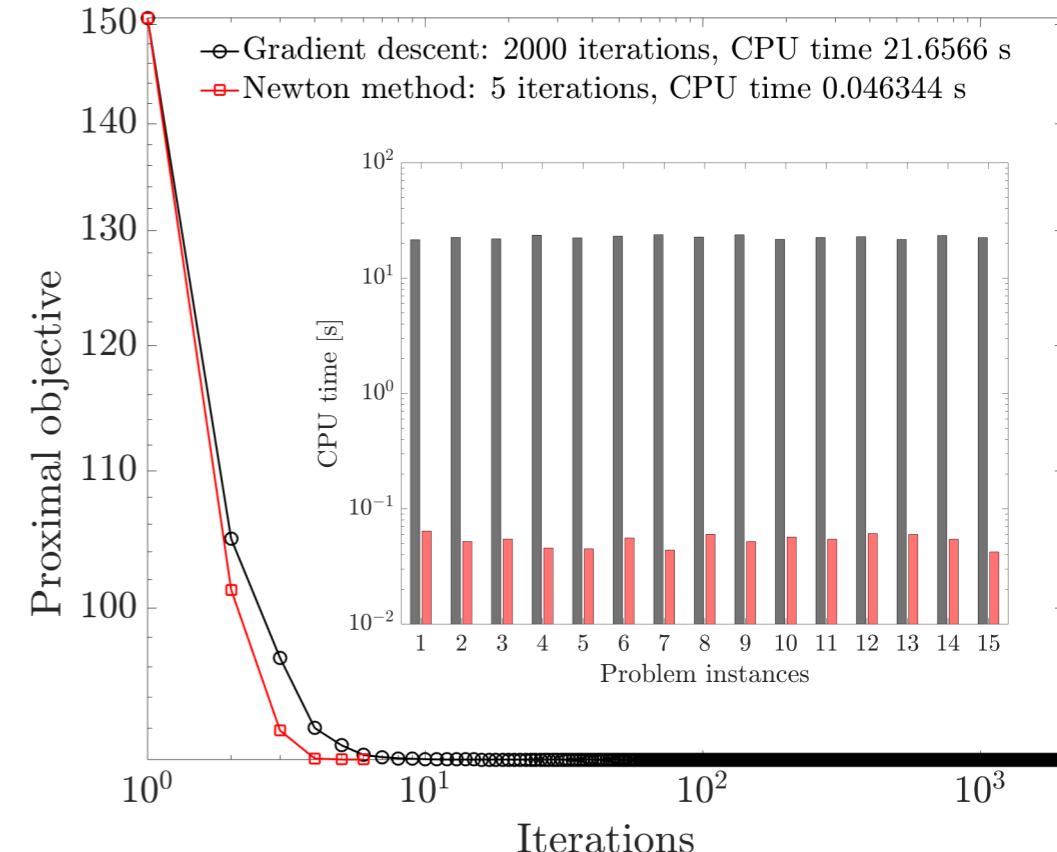
Then the following Euclidean ADMM solves (♥)

$$\mathbf{u}_i^{\ell+1} = \text{prox}_{\frac{1}{\tau} f_i}^{\|\cdot\|_2} (\mathbf{z}_i^\ell - \tilde{\boldsymbol{\nu}}_i^\ell)$$

No analytical solution, use e.g.,
Newton's method (has structured Hess)

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

$$\tilde{\boldsymbol{\nu}}_i^{\ell+1} = \tilde{\boldsymbol{\nu}}_i^\ell + (\mathbf{u}_i^{\ell+1} - \mathbf{z}_i^{\ell+1})$$



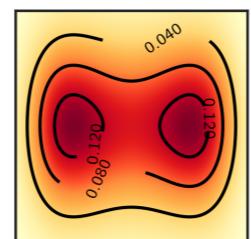
Experiment #1

Linear Fokker-Planck-Kolmogorov PDE

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

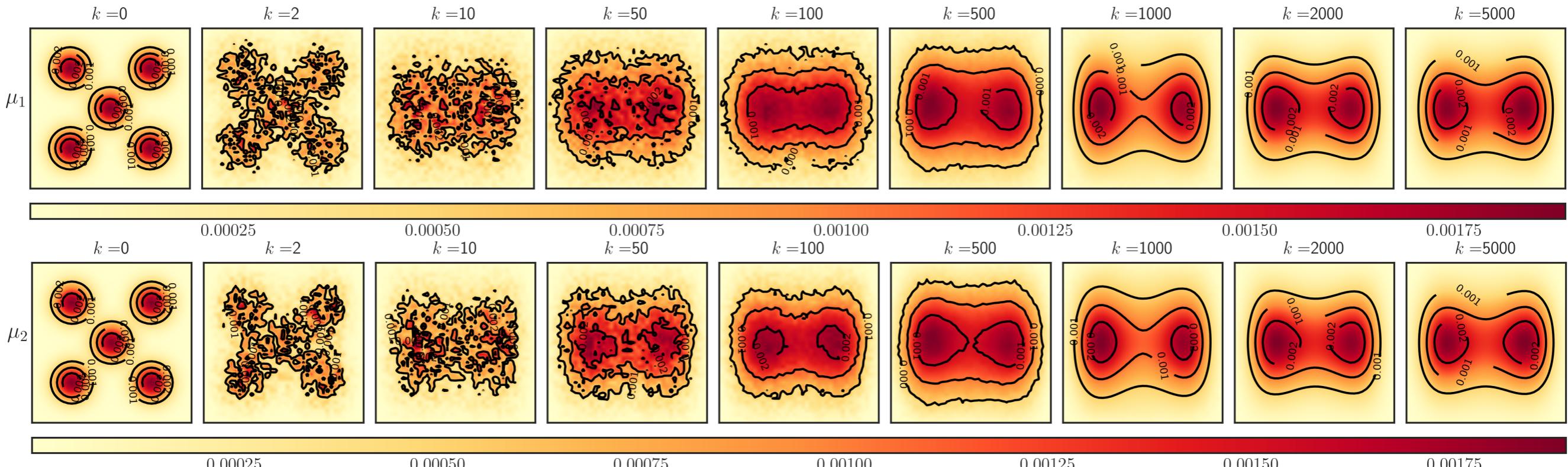
$$V(x_1, x_2) = \frac{1}{4} (1 + x_1^4) + \frac{1}{2} (x_2^2 - x_1^2)$$

$$\mu_\infty \propto \exp(-\beta V(x_1, x_2)) dx_1 dx_2$$



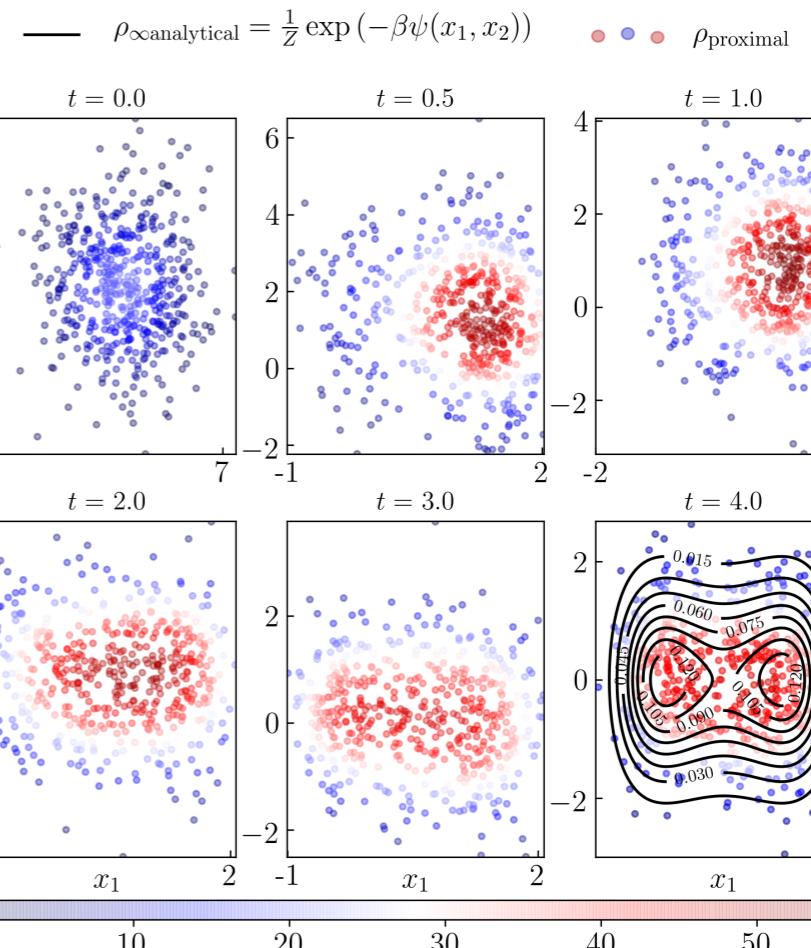
Distributed computation:

$$F_1(\mu) = \langle V_k, \mu \rangle \quad F_2(\mu) = \langle \beta^{-1} \log \mu, \mu \rangle$$



Centralized computation:

Caluya and Halder, IEEE Trans. Automatic Control, 2019



Experiment #2

Aggregation-drift-diffusion nonlinear PDE

$$\frac{\partial \mu}{\partial t} = \underbrace{\nabla \cdot (\mu \nabla (U * \mu))}_{i=1} + \underbrace{\nabla \cdot (\mu \nabla V) + \beta^{-1} \Delta \mu^2}_{i=2}$$

$$U(\mathbf{x}) = \frac{1}{2} \|\mathbf{x}\|_2^2 - \ln \|\mathbf{x}\|_2$$

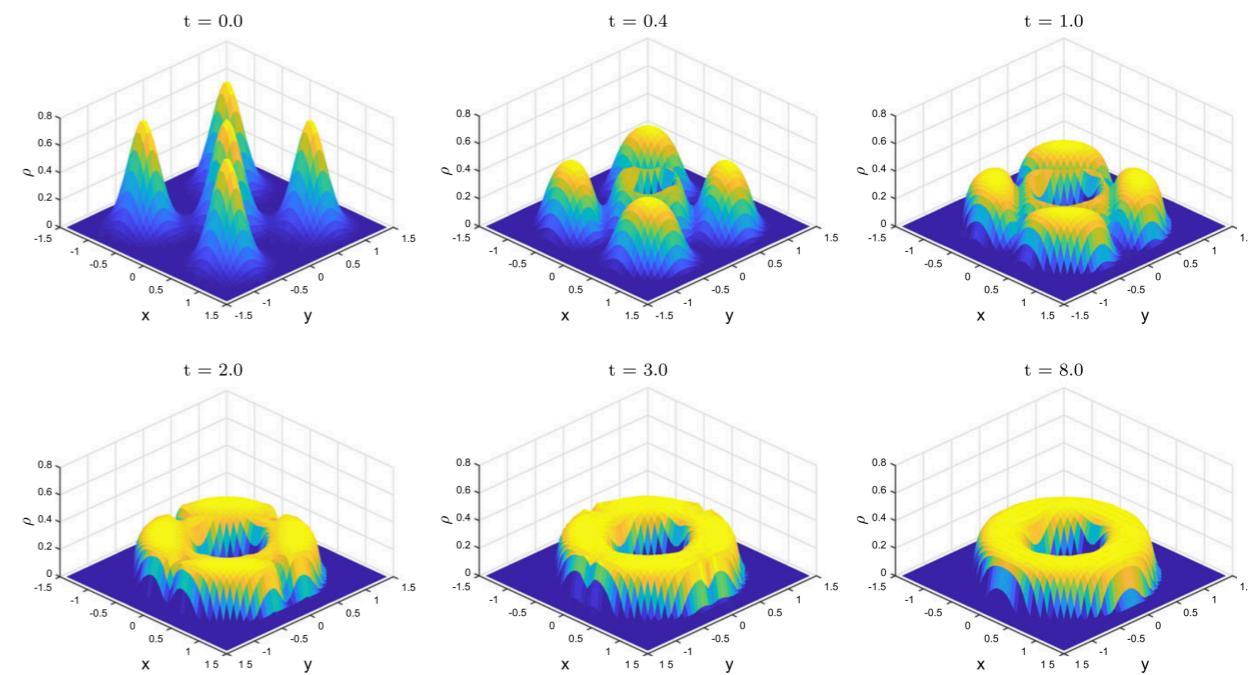
$$V(\mathbf{x}) = -\frac{1}{4} \ln \|\mathbf{x}\|_2$$

Distributed computation:

$$F_1(\mu) = \langle U_k \mu, \mu \rangle \quad F_2(\mu) = \langle V_k + \beta^{-1} \log \mu, \mu \rangle$$

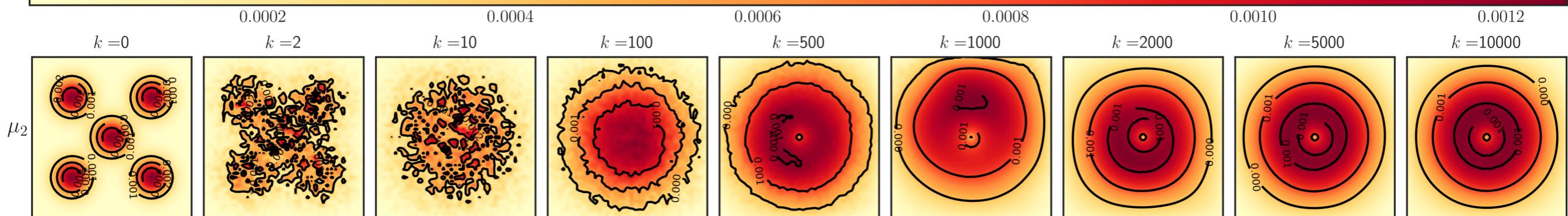
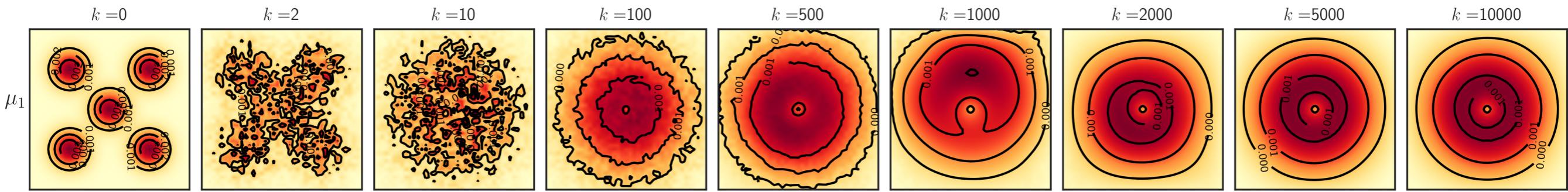
Centralized computation:

Carrillo, Craig, Wang and Wei, FOCM, 2021



$$\lim_{\beta^{-1} \downarrow 0} \mu_\infty = \text{Unif}(\mathcal{A})$$

Annulus with inner radius 1/2 and outer radius $\sqrt{5}/2$



Experiment #2 (contd.)

Aggregation-drift-diffusion nonlinear PDE

$$\frac{\partial \mu}{\partial t} = \underbrace{\nabla \cdot (\mu \nabla (U * \mu))}_{i=1} + \underbrace{\nabla \cdot (\mu \nabla V) + \beta^{-1} \Delta \mu^2}_{i=2}$$

$$U(\mathbf{x}) = \frac{1}{2} \|\mathbf{x}\|_2^2 - \ln \|\mathbf{x}\|_2$$

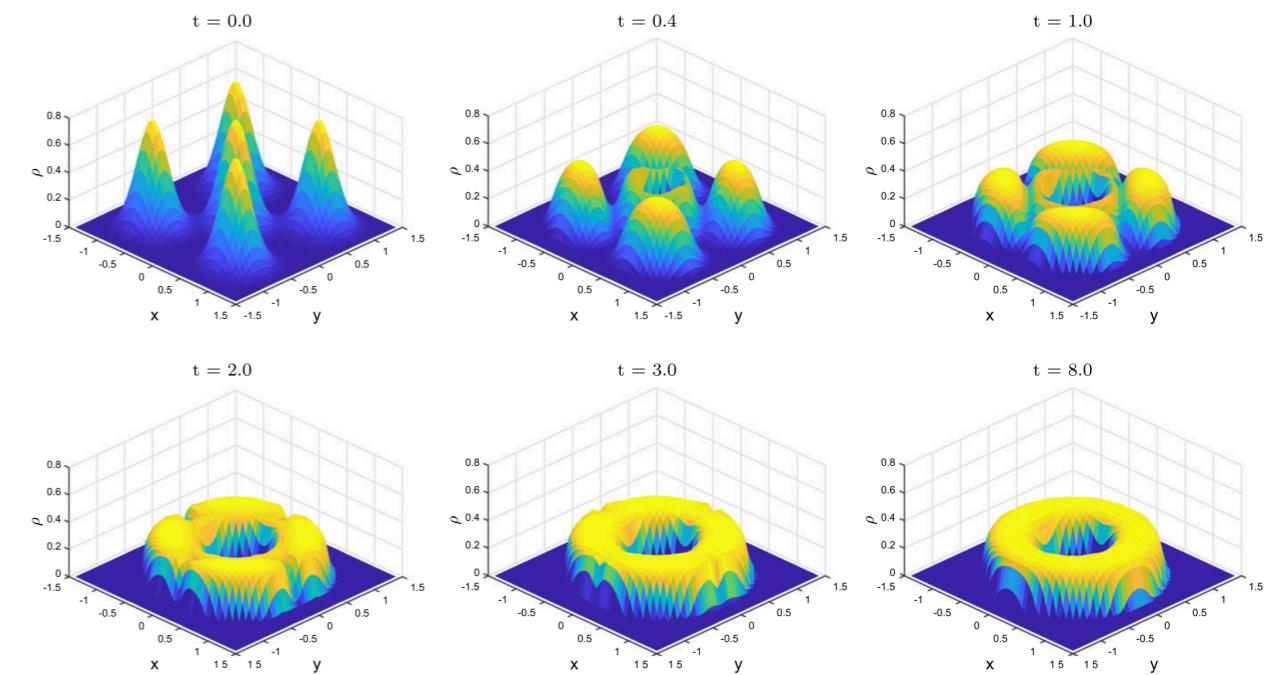
$$V(\mathbf{x}) = -\frac{1}{4} \ln \|\mathbf{x}\|_2$$

Distributed computation:

$$F_1(\boldsymbol{\mu}) = \langle \mathbf{U}_k \boldsymbol{\mu}, \boldsymbol{\mu} \rangle \quad F_2(\boldsymbol{\mu}) = \langle \mathbf{V}_k + \beta^{-1} \log \boldsymbol{\mu}, \boldsymbol{\mu} \rangle$$

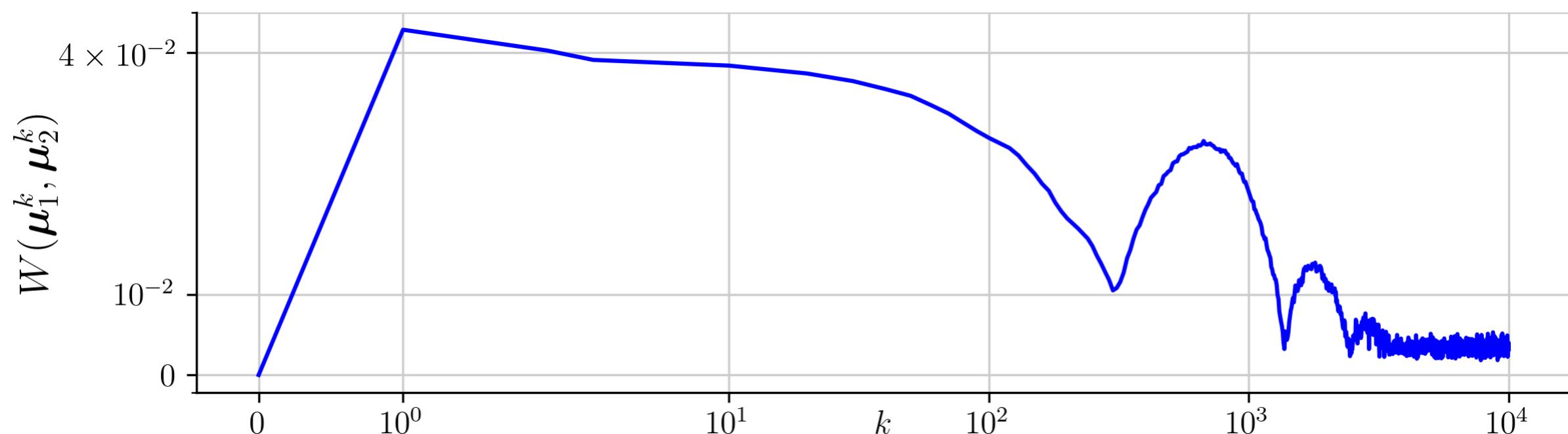
Centralized computation:

Carrillo, Craig, Wang and Wei, FOCM, 2021



$$\lim_{\beta^{-1} \downarrow 0} \boldsymbol{\mu}_\infty = \text{Unif}(\mathcal{A})$$

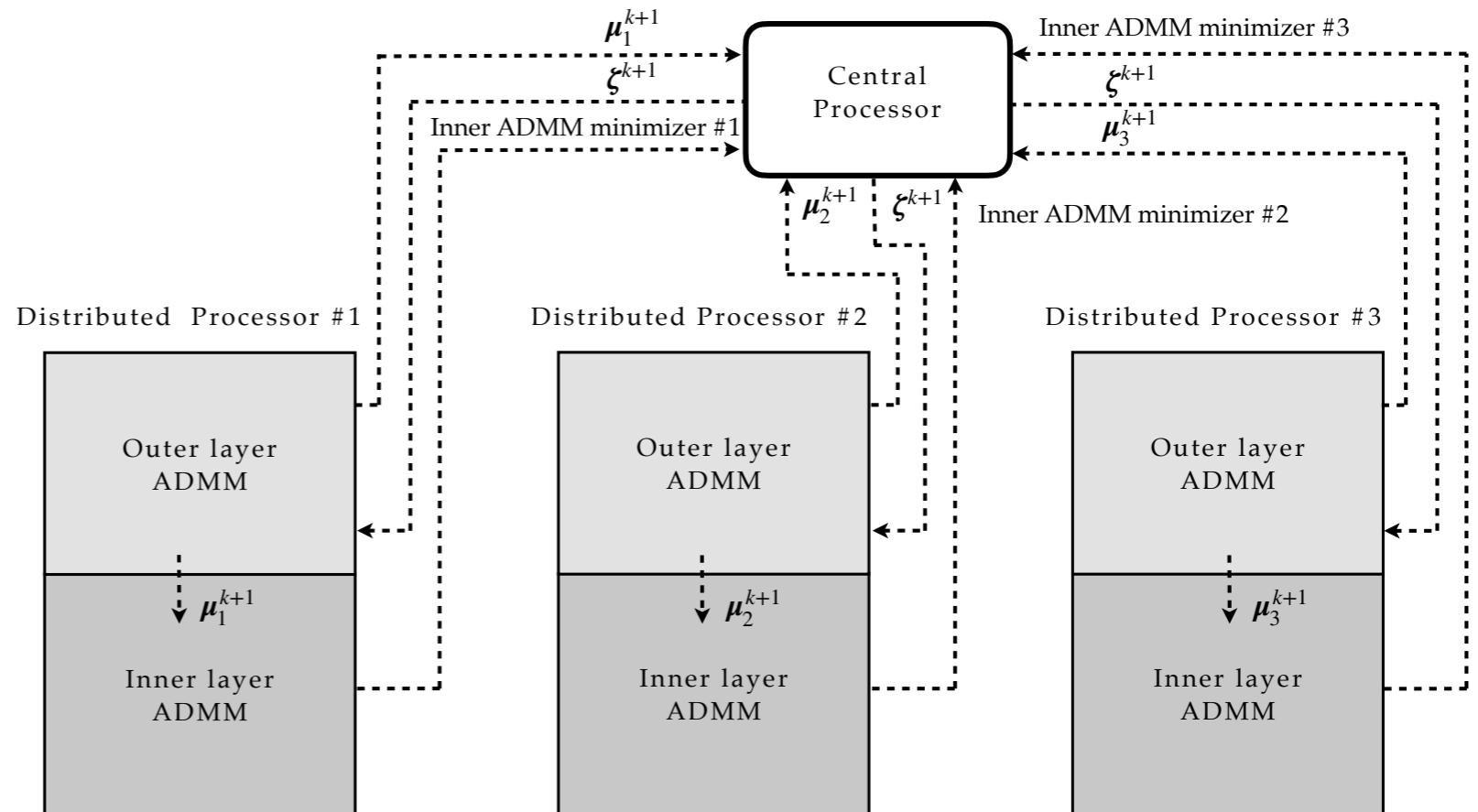
Annulus with inner radius $1/2$ and outer radius $\sqrt{5}/2$



Summary

Distributed computation for measure-valued optimization

Realizes measure-valued operator splitting



Takes advantage of the existing proximal and JKO type algorithms

Ongoing

Convergence guarantees for the overall scheme

High dimensional case studies

Thank You