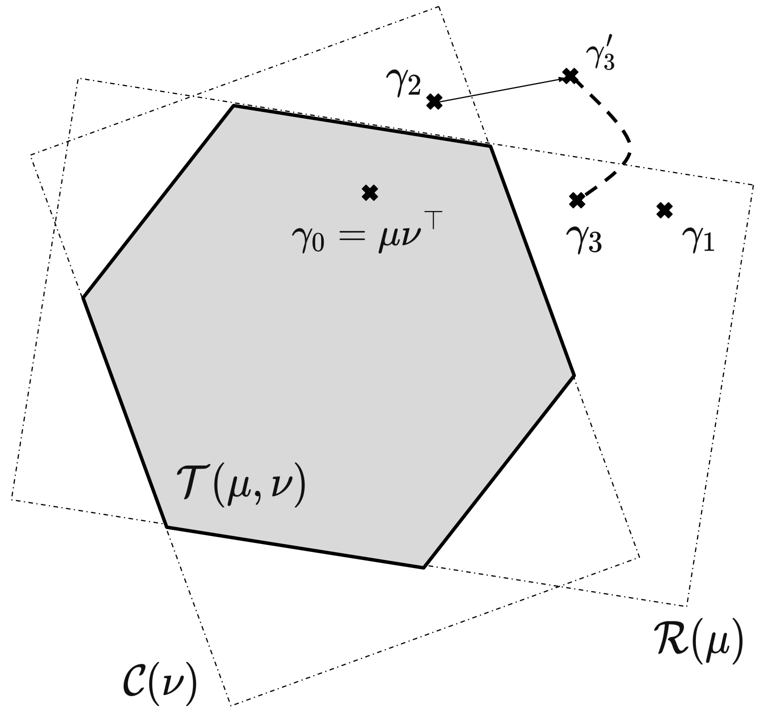
Mirror Sinkhorn:

Fast Online Optimization on Transport Polytopes



M. Ballu
(U. of Cambridge)

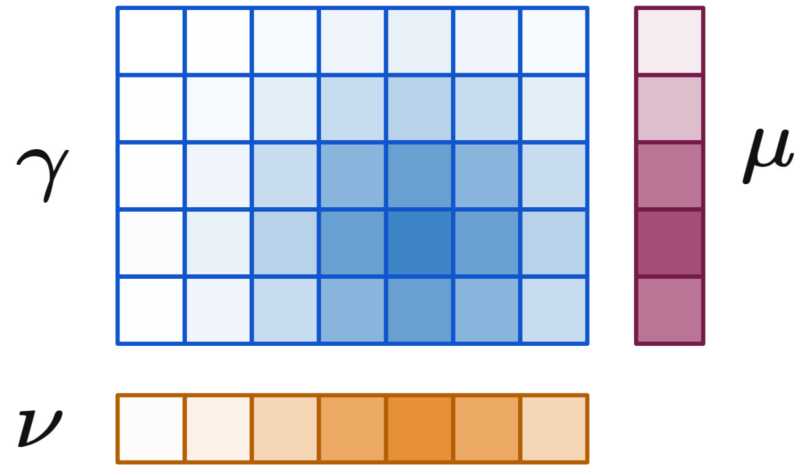
Q. Berthet
(Google DeepMind)

International Conference on Machine Learning 2023

Optimization over the transport polytope

Optimization of **convex** function f over **transport poytope** $\mathcal{T}(\mu, \nu)$.

$$\min_{\gamma \in \mathcal{T}(\mu,\nu)} f(\gamma)$$



- Variable γ: $m \times n$ probability matrix.
- Marginal constraints:
  probability vectors μ and ν

Domain of **optimal transport** (OT) problem, and entropic-regularized (Cuturi, 13).

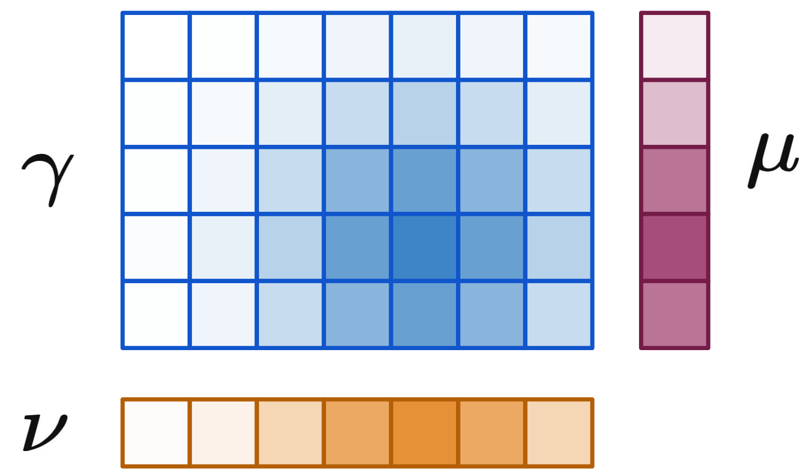
$$\min_{\gamma \in \mathcal{T}(\mu,\nu)} \langle C, \gamma \rangle \,.$$

- Special case, linear objective, solved with Sinkhorn algorithm. (Sinkhorn, 64)

Optimization over the transport polytope

Optimization of **convex** function f over **transport poytope** $\mathcal{T}(\mu,\nu)$.

$$\min_{\gamma \in \mathcal{T}(\mu,\nu)} f(\gamma)$$



- Variable γ: $m \times n$ probability matrix.
- Marginal constraints:
  probability vectors μ and ν

Domain of **optimal transport** (OT) problem, and entropic-regularized (Cuturi, 13).

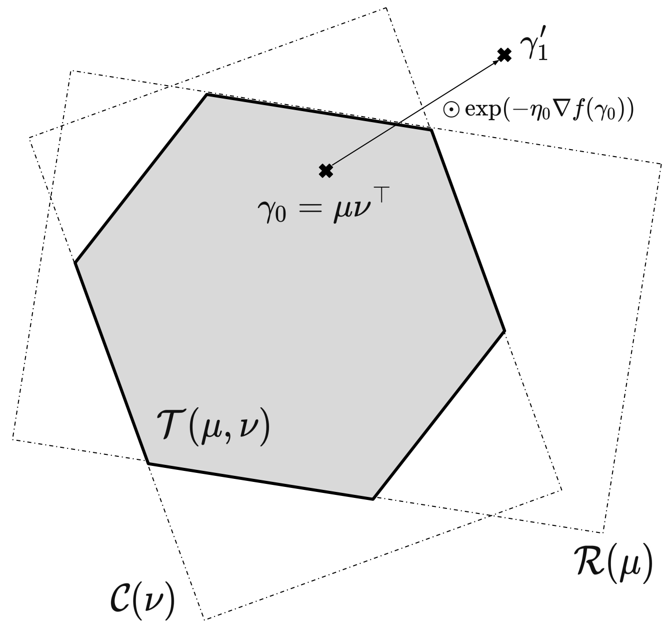
$$\min_{\gamma \in \mathcal{T}(\mu,\nu)} \langle C, \gamma \rangle - \alpha H(\gamma)\,.$$

- Special case, linear objective, solved with Sinkhorn algorithm. (Sinkhorn, 64)

Mirror Sinkhorn algorithm

Mirror descent steps equivalent to regularized OT problem

$$\gamma_{t+1} = \operatorname*{argmin}_{\gamma \in \mathcal{T}(\mu,\nu)} \left\{ \langle \gamma, \nabla f(\gamma_t) \rangle + \frac{D_{\mathsf{KL}}(\gamma, \gamma_t)}{\eta_t} \right\}$$



- Updates require several steps of Sinkhorn algorithm (**nested loop**).
- Replaced by **closed form**, alternating normalization (**single loop**).

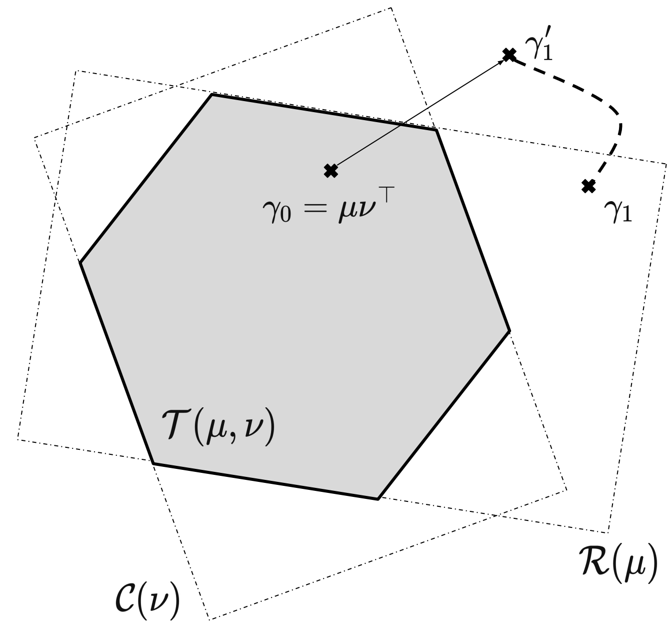
Mirror Sinkhorn

- Gradient update: $\gamma'_{t+1} = \gamma_t \odot \exp\left(-\eta_t \nabla f(\gamma_t)\right)$
- Alternating **row**/column normalization: $\gamma_{t+1} = \gamma'_{t+1} \mathrm{Diag}\left(\frac{\nu}{(\gamma'_{t+1})^\top 1_m}\right)$

Mirror Sinkhorn algorithm

Mirror descent steps equivalent to regularized OT problem

$$\gamma_{t+1} = \underset{\gamma \in \mathcal{T}(\mu,\nu)}{\text{argmin}} \left\{ \langle \gamma, \nabla f(\gamma_t) \rangle + \frac{D_{\mathsf{KL}}(\gamma, \gamma_t)}{\eta_t} \right\}$$



- Updates require several steps of Sinkhorn algorithm (**nested loop**).
- Replaced by **closed form**, alternating normalization (**single loop**).

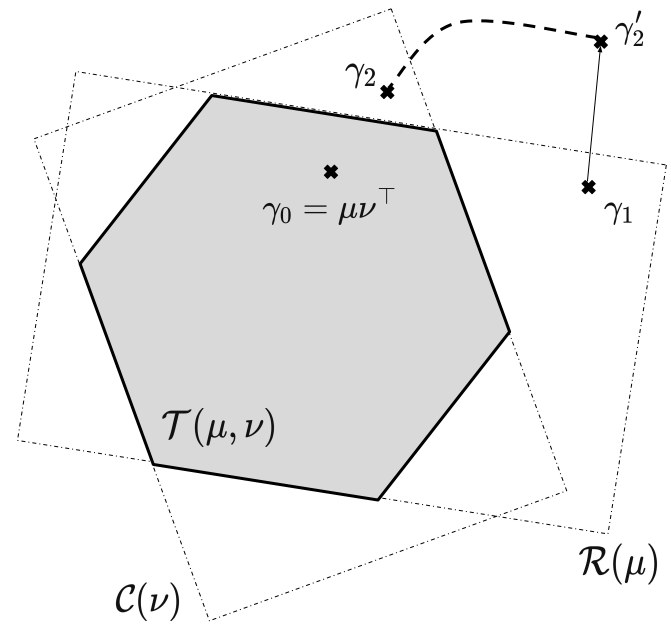
Mirror Sinkhorn

- Gradient update: $\gamma'_{t+1} = \gamma_t \odot \exp\left(-\eta_t \nabla f(\gamma_t)\right)$
- Alternating **row**/column normalization: $\gamma_{t+1} = \gamma'_{t+1} \mathrm{Diag}\left(\frac{\nu}{(\gamma'_{t+1})^\top 1_m}\right)$

Mirror Sinkhorn algorithm

Mirror descent steps equivalent to regularized OT problem

$$\gamma_{t+1} = \operatorname*{argmin}_{\gamma \in \mathcal{T}(\mu,\nu)} \left\{ \langle \gamma, \nabla f(\gamma_t) \rangle + \frac{D_{\mathsf{KL}}(\gamma, \gamma_t)}{\eta_t} \right\}$$



- Updates require several steps of Sinkhorn algorithm (**nested loop**).
- Replaced by **closed form**, alternating normalization (**single loop**).

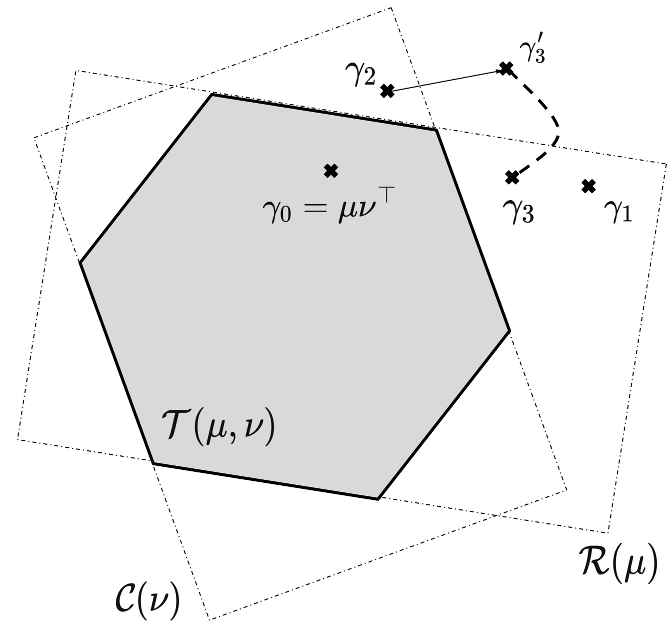
Mirror Sinkhorn

- Gradient update: $$\gamma'_{t+1} = \gamma_t \odot \exp\left(-\eta_t \nabla f(\gamma_t)\right)$$
- Alternating row/**column** normalization: $$\gamma_{t+1} = \mathsf{Diag}\left(\frac{\mu}{\gamma'_{t+1} 1_n}\right) \gamma'_{t+1}$$

Mirror Sinkhorn algorithm

Mirror descent steps equivalent to regularized OT problem

$$\gamma_{t+1} = \underset{\gamma \in \mathcal{T}(\mu,\nu)}{\text{argmin}} \left\{ \langle \gamma, \nabla f(\gamma_t) \rangle + \frac{D_{\mathsf{KL}}(\gamma, \gamma_t)}{\eta_t} \right\}$$



- Updates require several steps of Sinkhorn algorithm (**nested loop**).
- Replaced by **closed form**, alternating normalization (**single loop**).

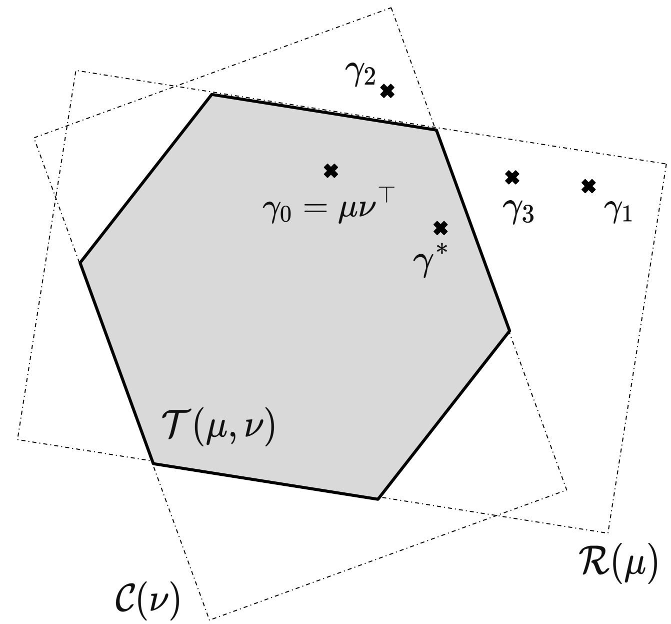
Mirror Sinkhorn

- Gradient update: $\gamma'_{t+1} = \gamma_t \odot \exp\left(-\eta_t \nabla f(\gamma_t)\right)$
- Alternating **row**/column normalization: $\gamma_{t+1} = \gamma'_{t+1} \mathsf{Diag}\Big(\frac{\nu}{(\gamma'_{t+1})^\top 1_m}\Big)$

Mirror Sinkhorn algorithm

Mirror descent steps equivalent to regularized OT problem

$$\gamma_{t+1} = \operatorname*{argmin}_{\gamma \in \mathcal{T}(\mu,\nu)} \left\{ \langle \gamma, \nabla f(\gamma_t) \rangle + \frac{D_{\mathsf{KL}}(\gamma, \gamma_t)}{\eta_t} \right\}$$



- Updates require several steps of Sinkhorn algorithm (**nested loop**).
- Replaced by **closed form**, alternating normalization (**single loop**).

Mirror Sinkhorn

- Gradient update: $\gamma'_{t+1} = \gamma_t \odot \exp\left(-\eta_t \nabla f(\gamma_t)\right)$
- Alternating **row**/column normalization: $\gamma_{t+1} = \gamma'_{t+1} \mathrm{Diag}\left(\frac{\nu}{(\gamma'_{t+1})^\top 1_m}\right)$

Results - convex optimization

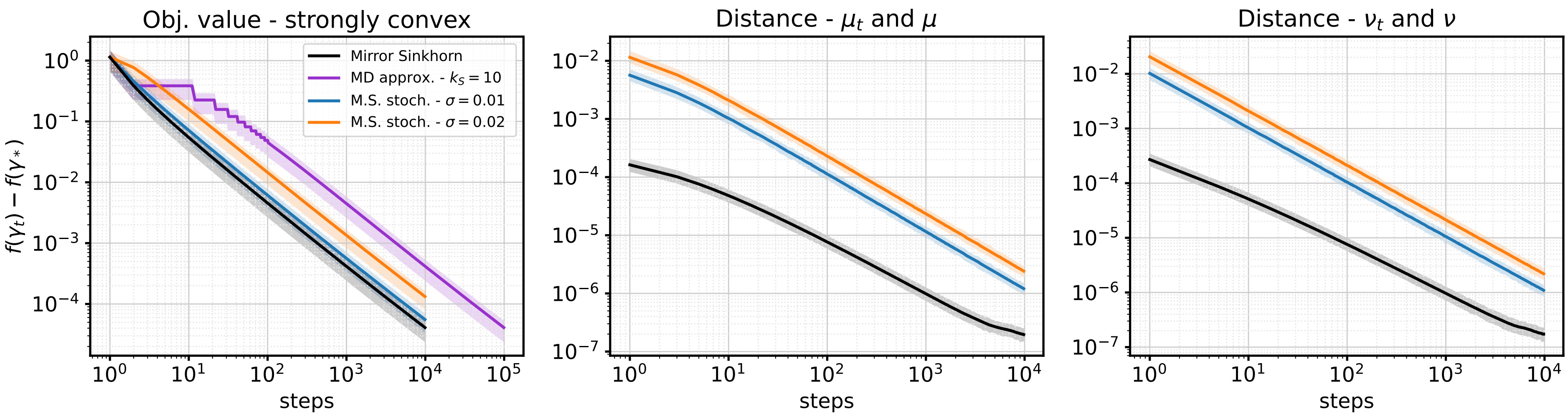
- Smooth f, σ-stochastic gradients:

$$\mathbf{E}[f(\tilde{\gamma}_T) - f(\gamma^*)] \leq \frac{\kappa_\sigma \log(T)}{\sqrt{T}}$$

- Strongly convex setting, averaging:

$$f(\overline{\gamma}_T) - f(\gamma^*) \leq \frac{\kappa \log(T)}{T}$$

Experiments - simulated data



Results - convex optimization

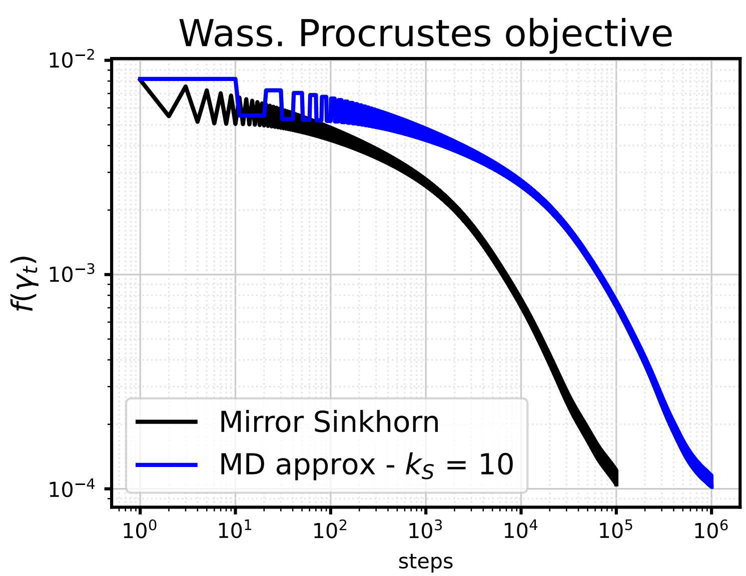
- Smooth f, σ-stochastic gradients:

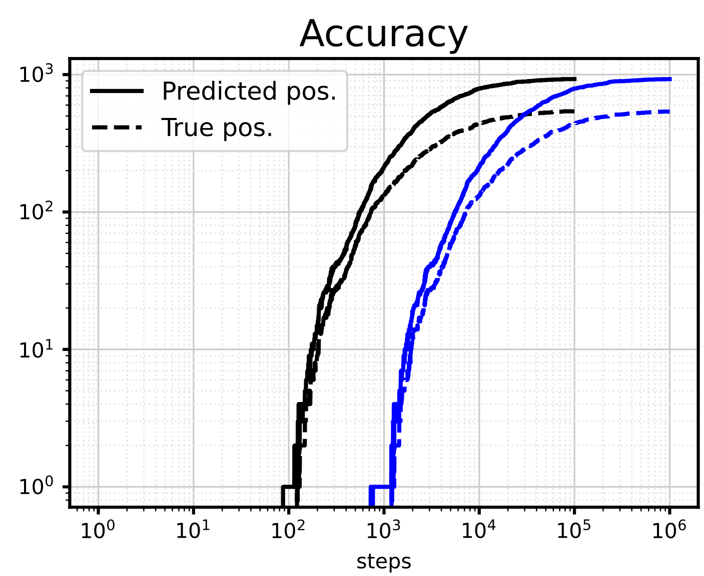
$$\mathbf{E}[f(\tilde{\gamma}_T) - f(\gamma^*)] \leq \frac{\kappa_\sigma \log(T)}{\sqrt{T}}$$

- Strongly convex setting, averaging:

$$f(\overline{\gamma}_T) - f(\gamma^*) \leq \frac{\kappa \log(T)}{T}$$

Experiments - SNARE-seq data, WP objective (Demetci et al., 20; Grave et al., 19)





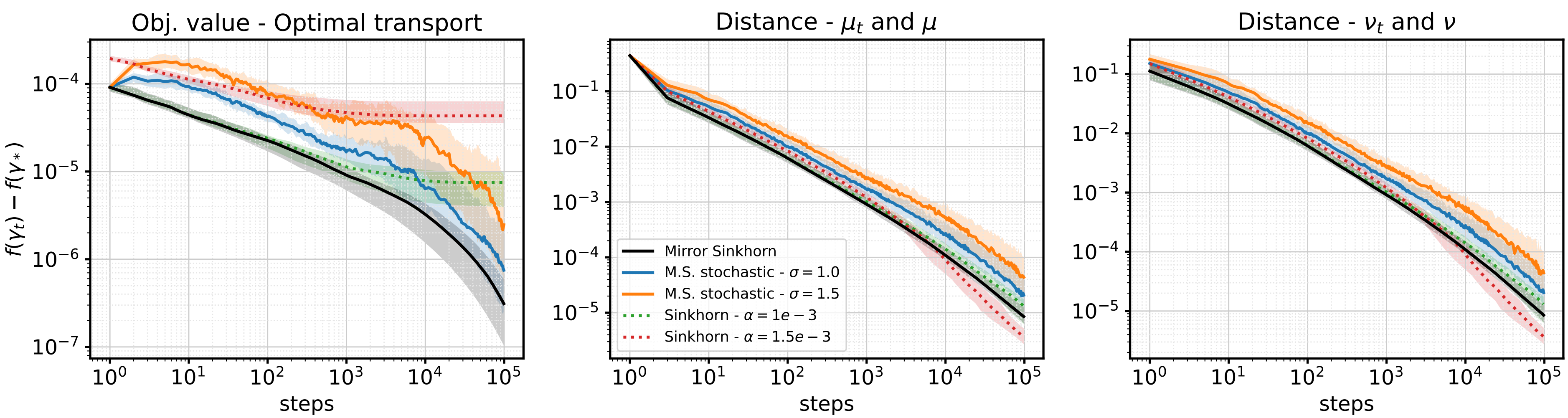
Results - Optimal transport

- OT - linear objective:

$$\mathbf{E}[\langle C, \tilde{\gamma}_T \rangle] - \langle C, \gamma^* \rangle \leq \frac{\kappa_\sigma \log(T)}{\sqrt{T}}$$

- No regularization bias, online algorithm, stochastic access to cost matrix C.

Experiments - simulated data



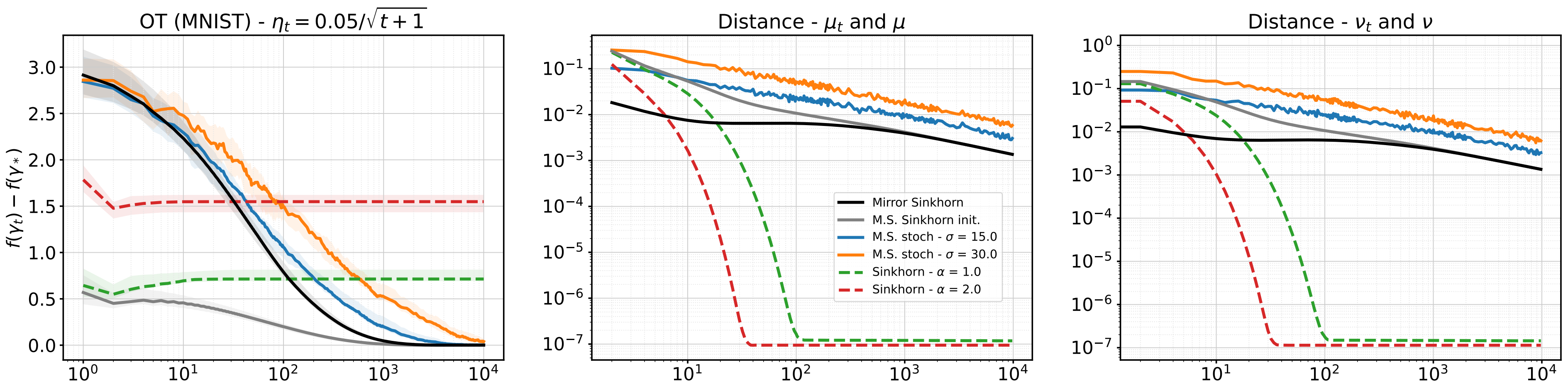
Results - Optimal transport

- OT - linear objective:

$$\mathbf{E}[\langle C, \tilde{\gamma}_T \rangle] - \langle C, \gamma^* \rangle \leq \frac{\kappa_\sigma \log(T)}{\sqrt{T}}$$

- No regularization bias, online algorithm, stochastic access to cost matrix C.

Experiments - MNIST data (Altschuler et al., 17)



Wednesday July 26th, Session 4

Poster #131