Debiaser Beware: Pitfalls of Centering Regularized Transport maps

Aram-Alexandre Pooladian
New York University
ICML 2022

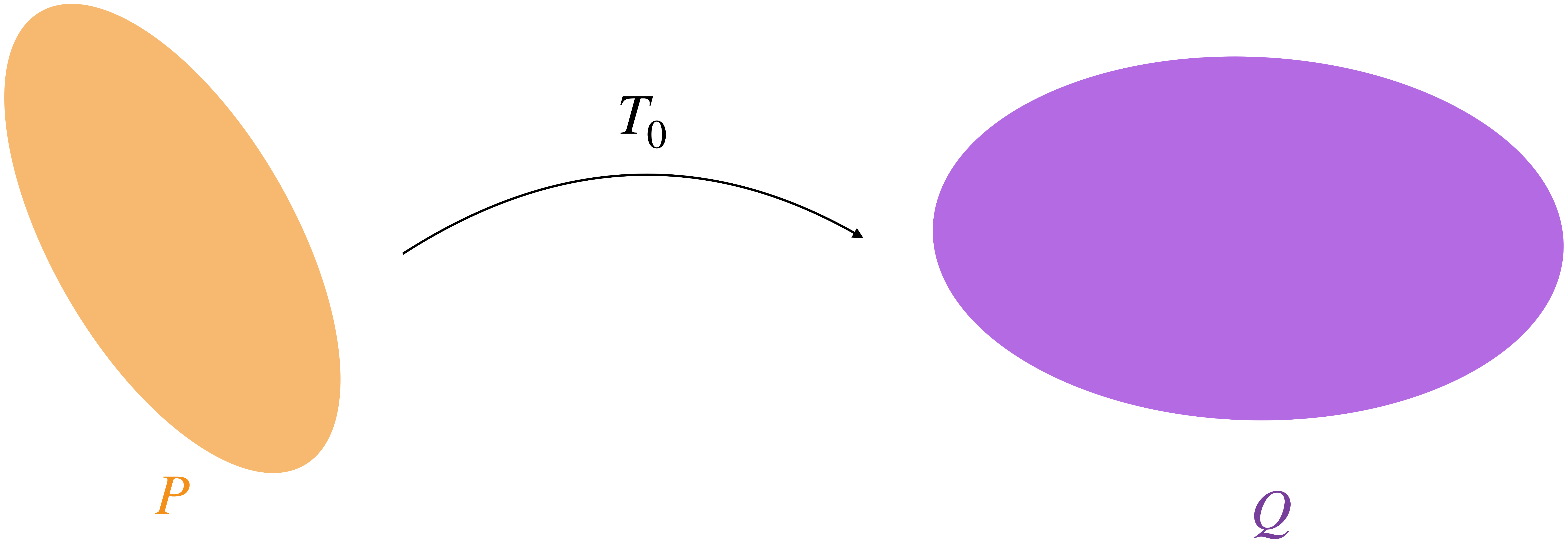
joint with

Jonathan Niles-Weed (NYU)

Marco Cuturi (Apple/ENSAE)

Optimal transport maps

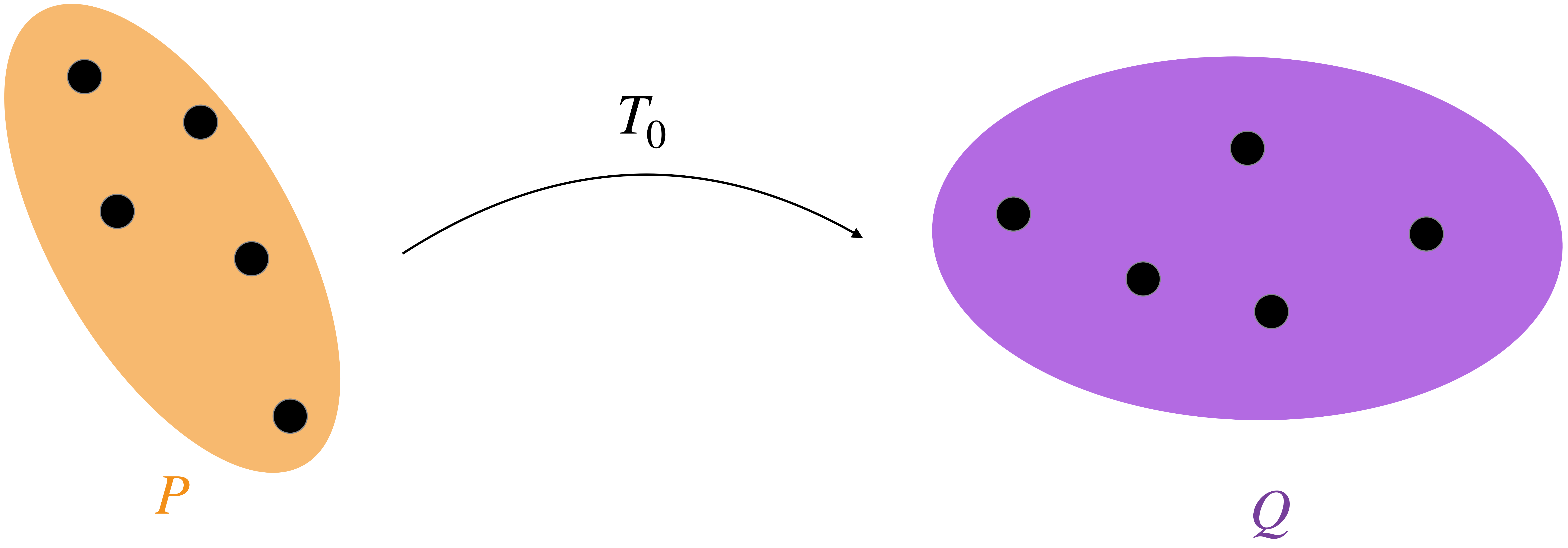
Optimal transport maps



Monge Problem

$$T_0 := \operatorname{argmin}_T \int \frac{1}{2} \|x - T(x)\|_2^2 \, \mathrm{d}P(x) \quad \text{s.t.} \quad X \sim P, T(X) \sim Q$$

Estimating optimal transport maps



Given i.i.d samples $X_1, \ldots, X_n \sim P$ and $Y_1, \ldots, Y_n \sim Q$

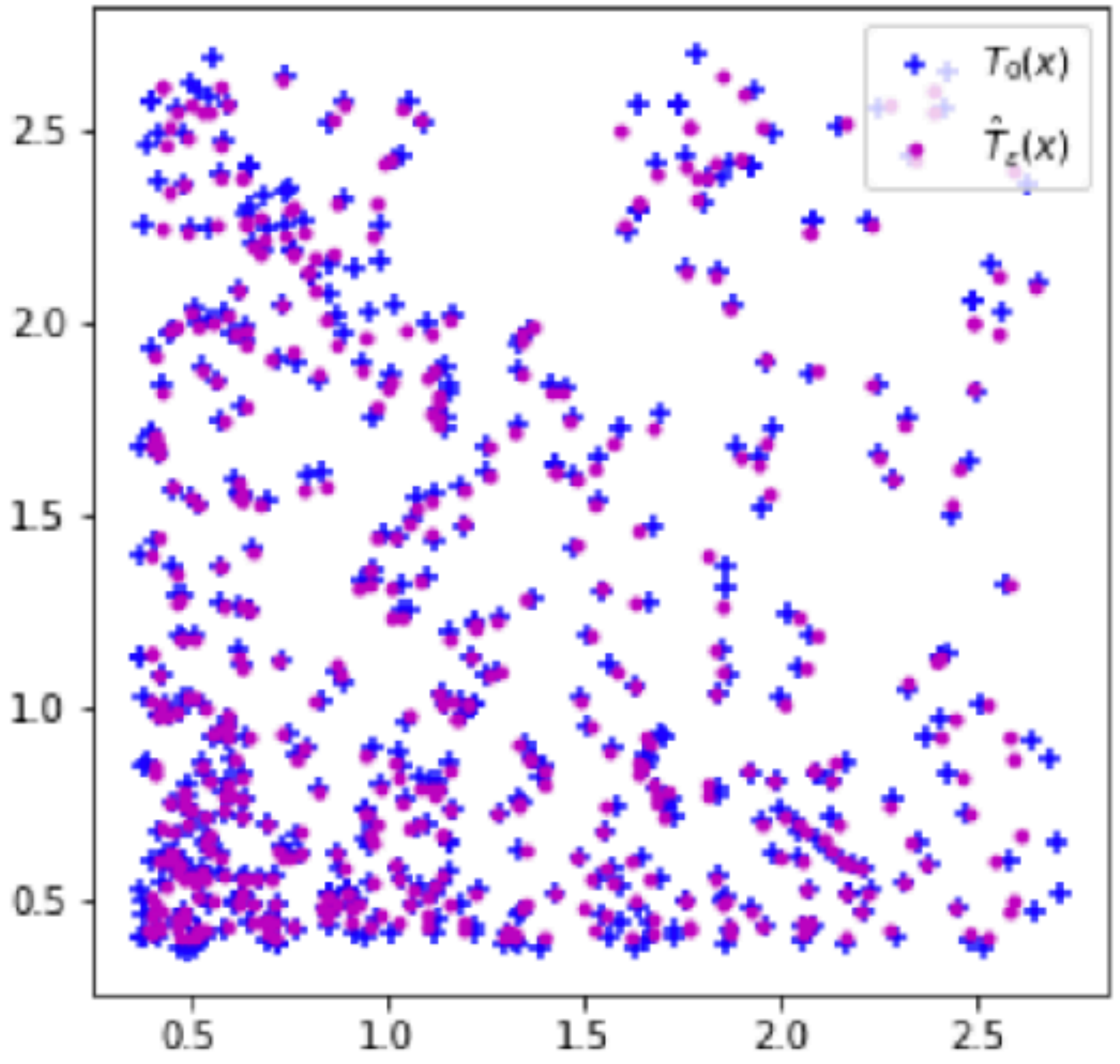
Question: How to estimate T_0 on the basis of samples?

Prior work: entropic map

Inspired by **entropic optimal transport** [Cut13], prior work [PNW21] studied the **entropic map** between two distributions

Prior work: entropic map

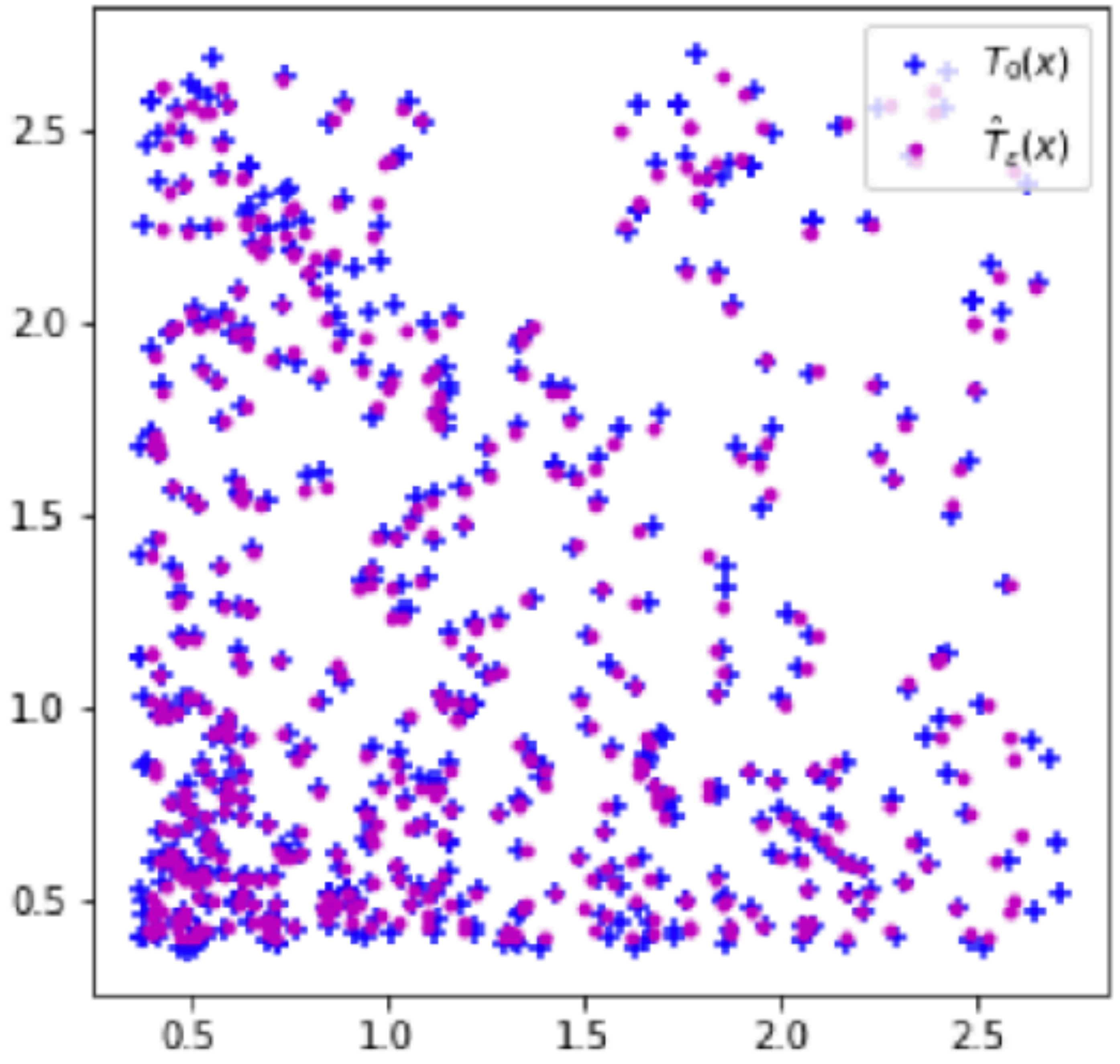
Inspired by **entropic optimal transport** [Cut13], prior work [PNW21] studied the **entropic map** between two distributions



$$T_\varepsilon := \mathbb{E}_{\pi_\varepsilon}[Y \mid X = x]$$

Prior work: entropic map

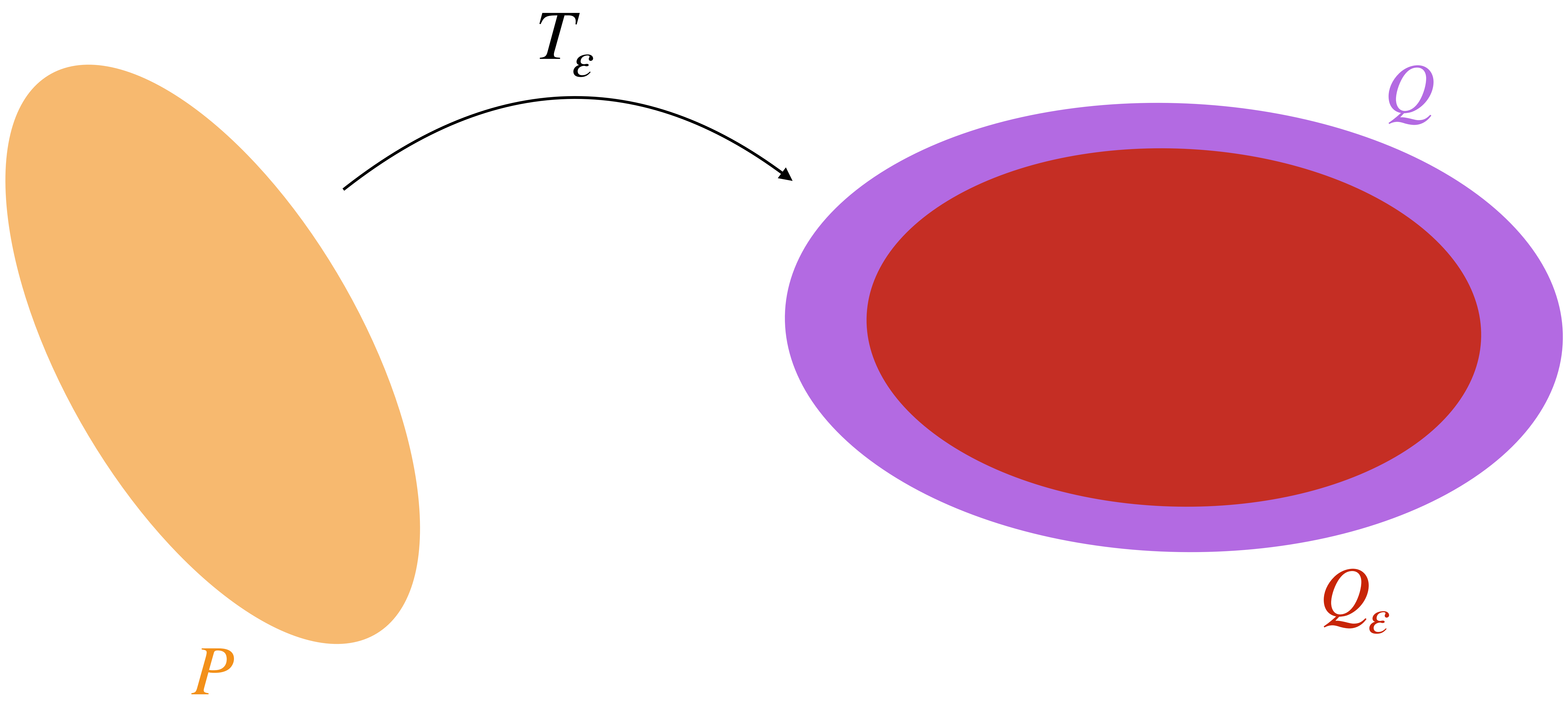
Inspired by **entropic optimal transport** [Cut13], prior work [PNW21] studied the **entropic map** between two distributions



$$T_\varepsilon := \mathbb{E}_{\pi_\varepsilon}[Y|X = x]$$

- GPU-friendly implementations
- Complexity: $O(n^2\varepsilon^{-2})$
- Provably approximates Monge map

Drawbacks: underdispersed



Approximation of the target distribution is <u>*underdispersed*</u> for large ε

Fix: Debiasing/Centering

- Conventional wisdom in optimal transport: *debias* the entropic problem

- Seen in several works [GPC18, GC+19, FS+19, CR+20]

- Idea: add a ***correction*** *term* so that when $P = Q$, we recover the identity map

- The correction term $\xi_\varepsilon : \mathbb{R}^d \to \mathbb{R}^d$ is obtained by solving the entropic transport problem from the source measure onto itself

Main findings

Main findings

- Asymptotic guarantees

Main findings

- Asymptotic guarantees

- For **judicious** choice of ε, debiasing *corrects* underdispersion

Main findings

- Asymptotic guarantees
- For **judicious** choice of ε, debiasing *corrects* underdispersion
- For **wrong** choice of ε, debiasing leads to unnecessary overdispersion

Main findings

- Asymptotic guarantees
- For **judicious** choice of ε, debiasing *corrects* underdispersion
- For **wrong** choice of ε, debiasing leads to unnecessary overdispersion
- Whether debiasing is better or worse is sensitive to P and Q

Main findings

- Asymptotic guarantees

- For **judicious** choice of ε, debiasing *corrects* underdispersion

- For **wrong** choice of ε, debiasing leads to unnecessary overdispersion

- Whether debiasing is better or worse is sensitive to P and Q

- Debiasing seems to be much more sensitive to statistical errors

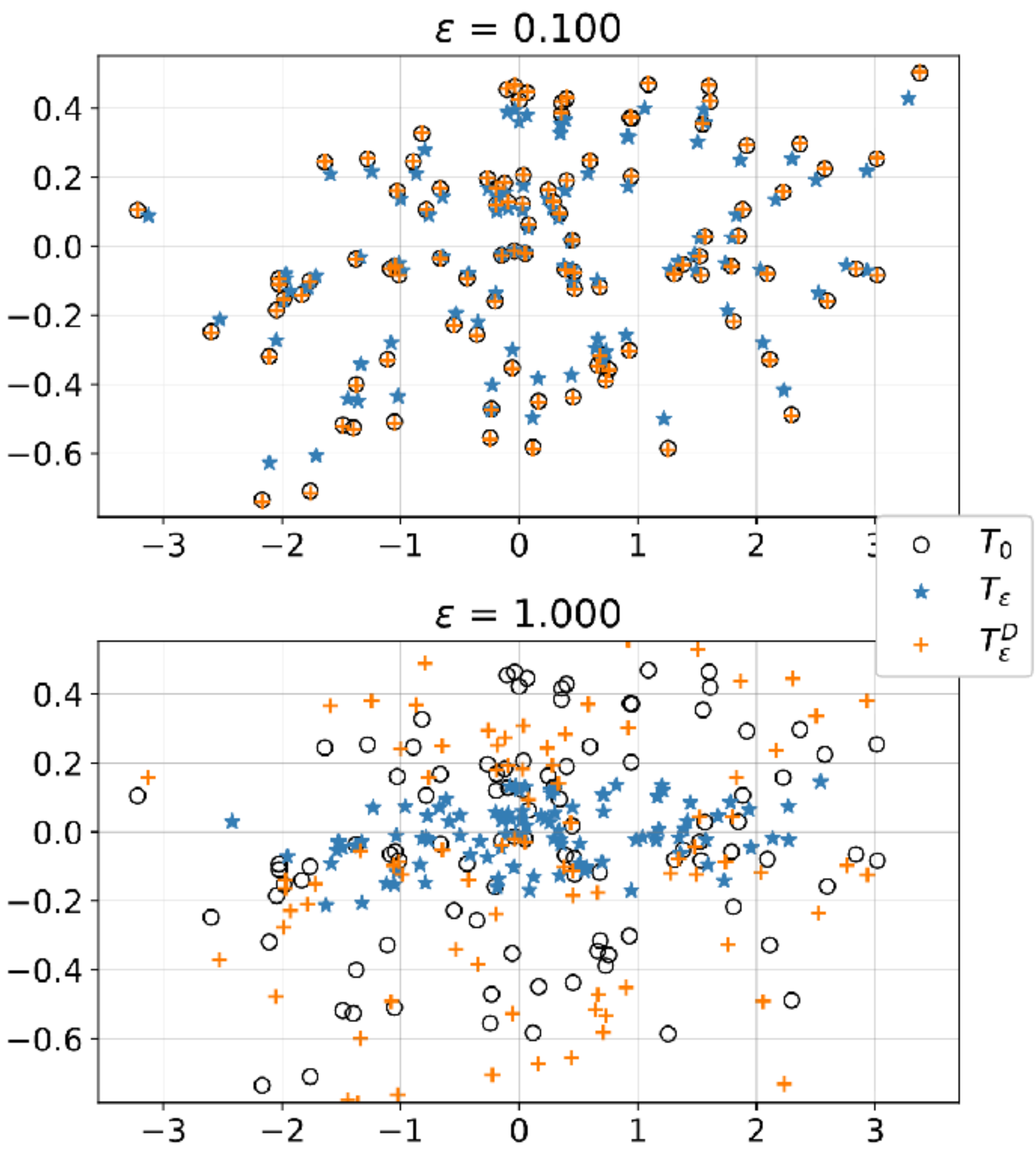
Asymptotic behavior in ε

Debiased entropic map

$$T_\varepsilon^D := T_\varepsilon + \xi_\varepsilon$$

versus (biased) entropic map T_ε

Asymptotic behavior in ε



Debiased entropic map

$$T_\varepsilon^D := T_\varepsilon + \xi_\varepsilon$$

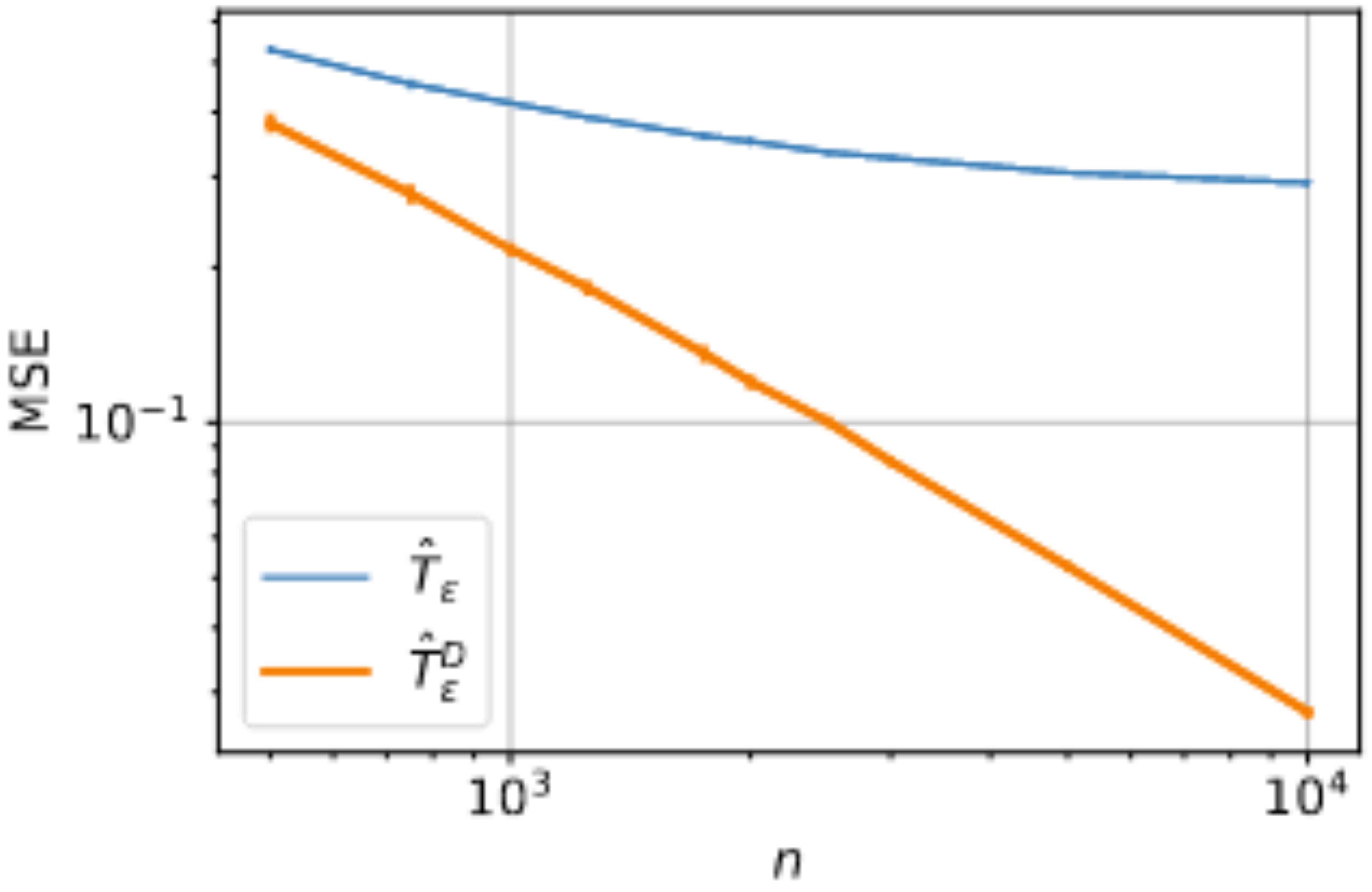
versus (biased) entropic map T_ε

For large ε, the entropic map concentrates around the mean of Q

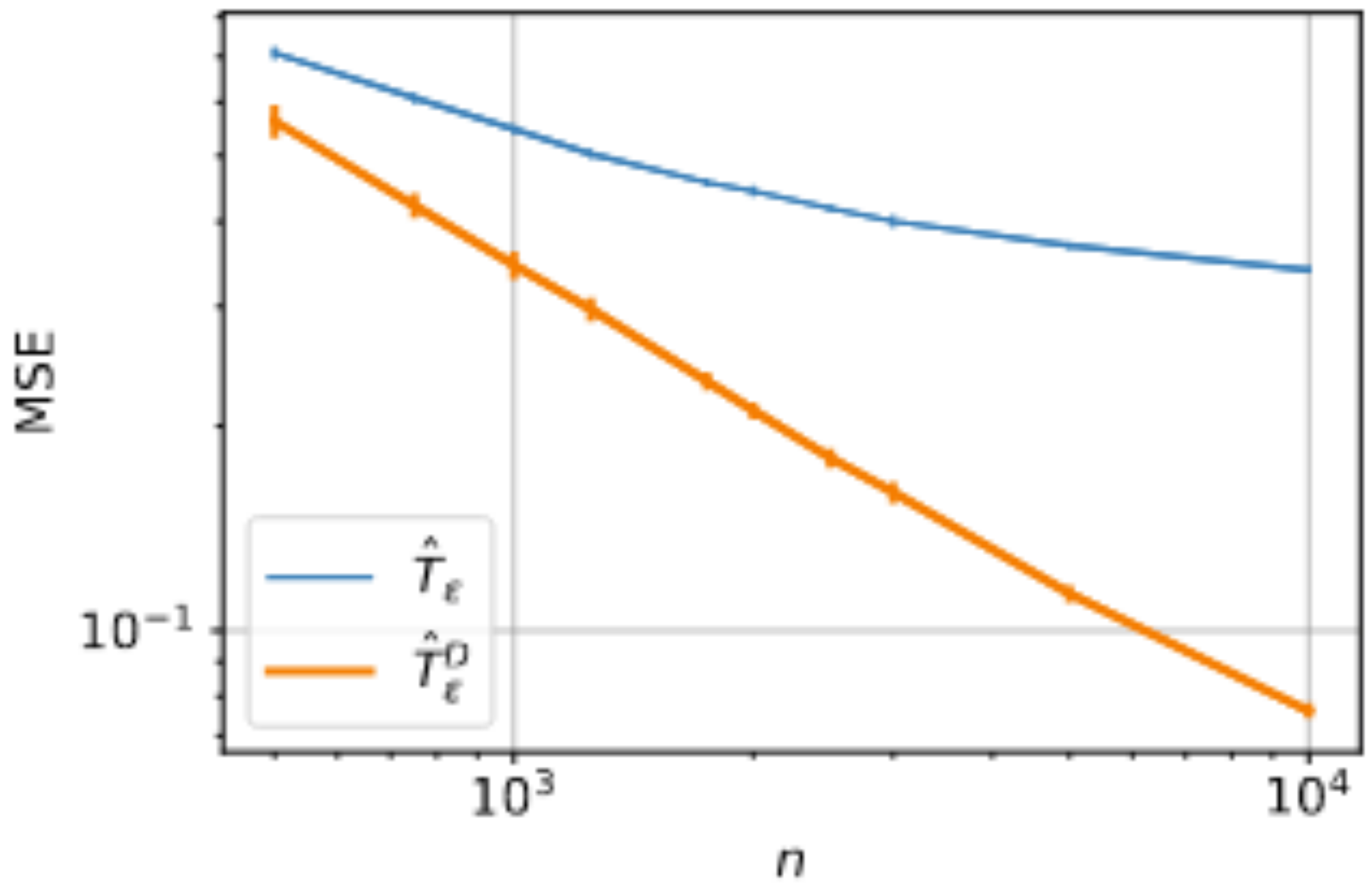
Judicious choice of ε

Synthetic examples: estimating optimal transport maps (plots are in $d = 10$)

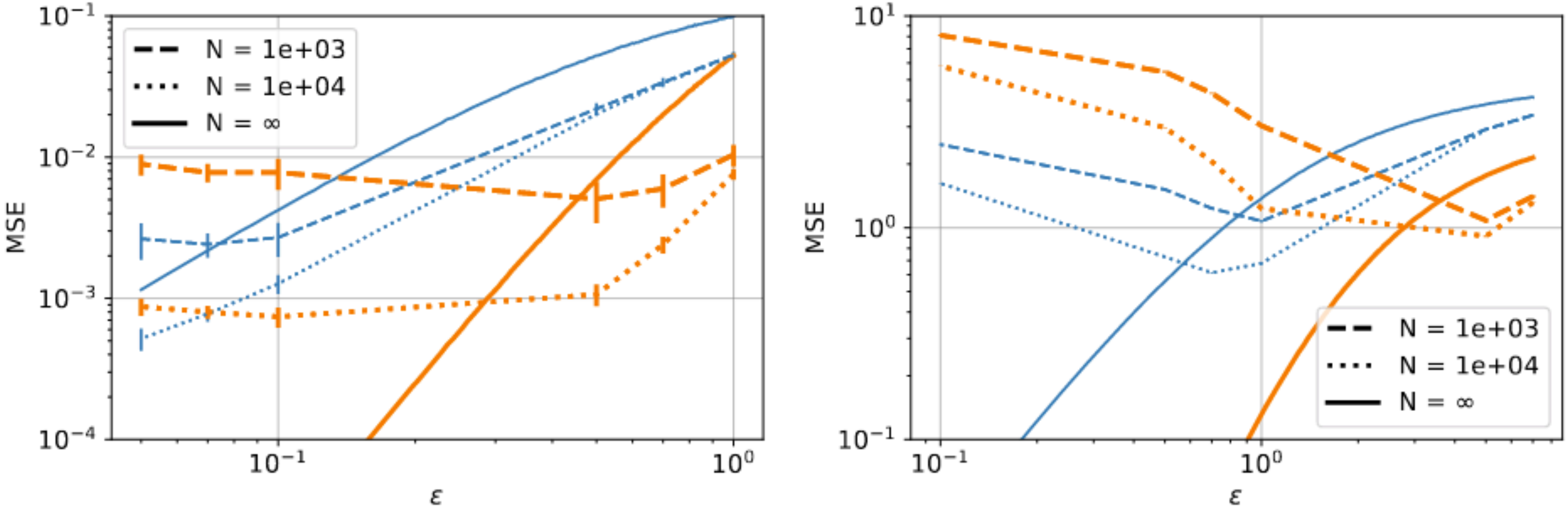
$T_0(x) = Ax$



$T_0(x) = (\exp(x_i))_{i=1}^d$



Beware of pitfalls



(a) $\hat{T}_\varepsilon$ vs. $\hat{T}_\varepsilon^D$ with Σ concentrated in $d = 2$ (b) $\hat{T}_\varepsilon$ vs. $\hat{T}_\varepsilon^D$ with Σ concentrated in $d = 15$

What isn't covered in this presentation:

- Theorems (asymptotic behavior of T_ε^D and T_ε)
- Gaussian-to-Gaussian case: *rates* of convergence showing that debiasing is asymptotically better
- Counter-results showing that debiasing does *not* always lead to better estimation in MSE

Thanks!

Bibliography

- [PNW21] A-A. Pooladian, and J. Niles-Weed. *Entropic estimation of optimal transport maps.* (submitted) 2021
- [CT+20] M. Cuturi, O. Teboule, J. Niles-Weed, and JP. Vert. *Supervised Quantile Normalization for Low Rank Matrix Factorization.* ICML 2020
- [CP+20] L. Chizat, P. Roussillon, F. Léger, F-X. Vialard, and G. Peyré. *Faster Wasserstein Distance Estimation with the Sinkhorn divergence*. NeurIPS 2020
- [GC+19] A. Genevay, L. Chizat, F. Bach, M. Cuturi, and G. Peyré. *Sample complexity of Sinkhorn divergences*. AISTATS 2019
- [GPC18] A. Genevay, G. Peyré, and M. Cuturi. *Learning generative models with the Sinkhorn divergences.* AISTATS 2018
- [F+19] J. Feydy, T. Séjourné, F-X. Vialard, S-i. Amari, A. Trouvé, and G. Peyré. *Interpolating between optimal transport and MMD using Sinkhorn Divergences*. AISTATS 2019
- [C13] M. Cuturi. *Sinkhorn distances: Lightspeed computation of optimal transport*. NIPS, 2013
- [SS+19] G. Schiebinger, J. Shu, M. Tabaka, B. Cleary, et al. *Optimal Transport Analysis of Single-Cell Gene Expression Identifies Developmental Trajectories in Reprogramming* Cell, 2019