Debiaser Beware: Pitfalls of Centering Regularized Transport maps

Aram-Alexandre Pooladian
New York University
ICML 2022

Debiaser Beware: Pitfalls of Centering Regularized Transport maps

Aram-Alexandre Pooladian
New York University
ICML 2022

Joint work with Jonathan Niles-Weed, and Marco Cuturi

1. Regularized Transport maps

2. Centering/Debiasing

3. Beware of Pitfalls

Estimating optimal transport maps

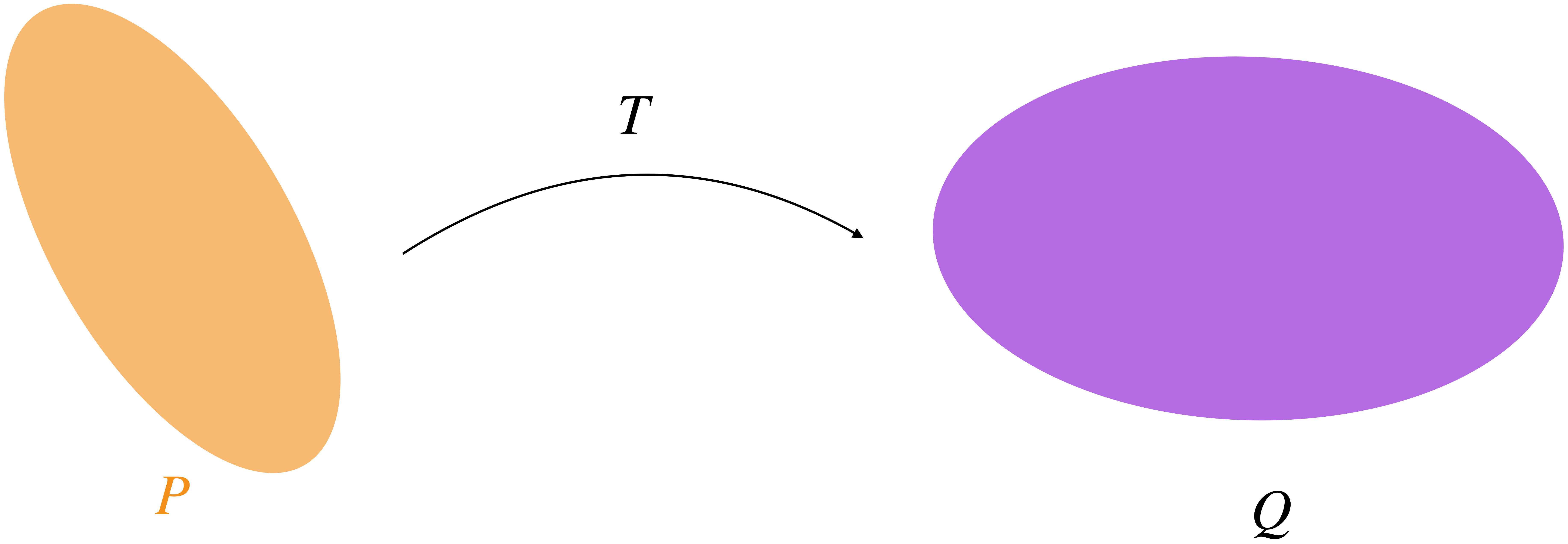
P and Q are two (nice) probability measures on $\mathbb{R}^d$:

Estimating optimal transport maps

P and Q are two (nice) probability measures on $\mathbb{R}^d$:

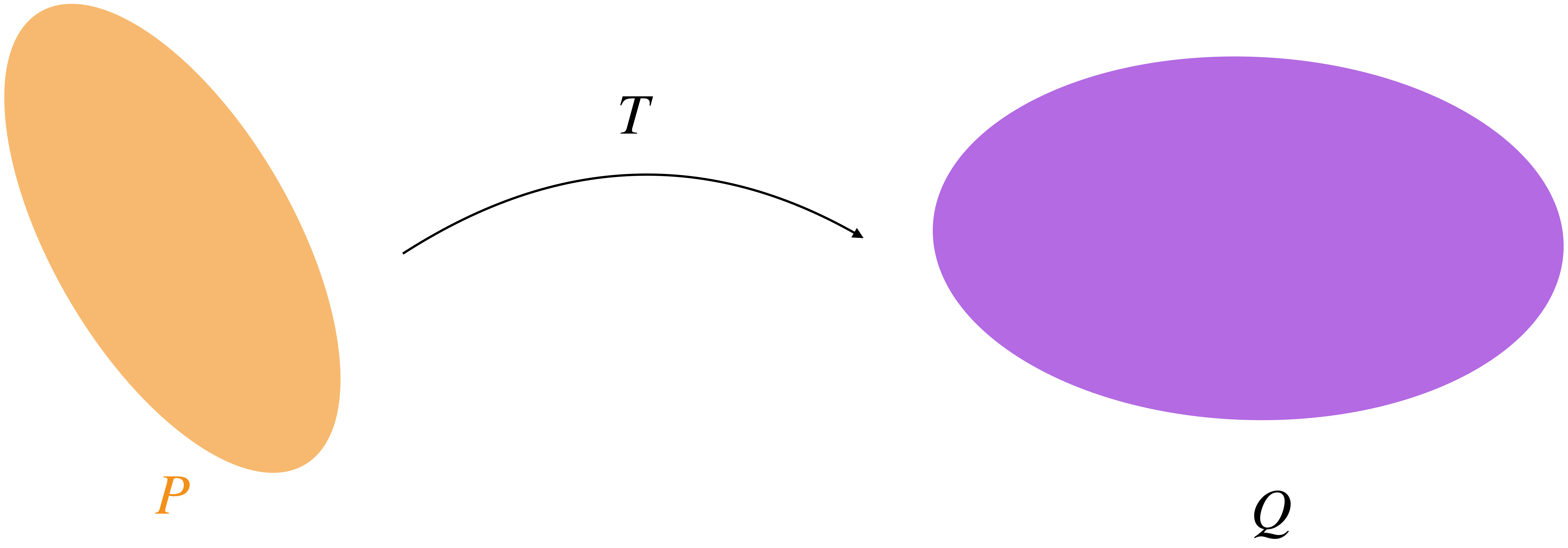
- have densities
- bounded domain

Estimating optimal transport maps



T is a *transport map* from P to Q if: for $X \sim P, T(X) \sim Q$

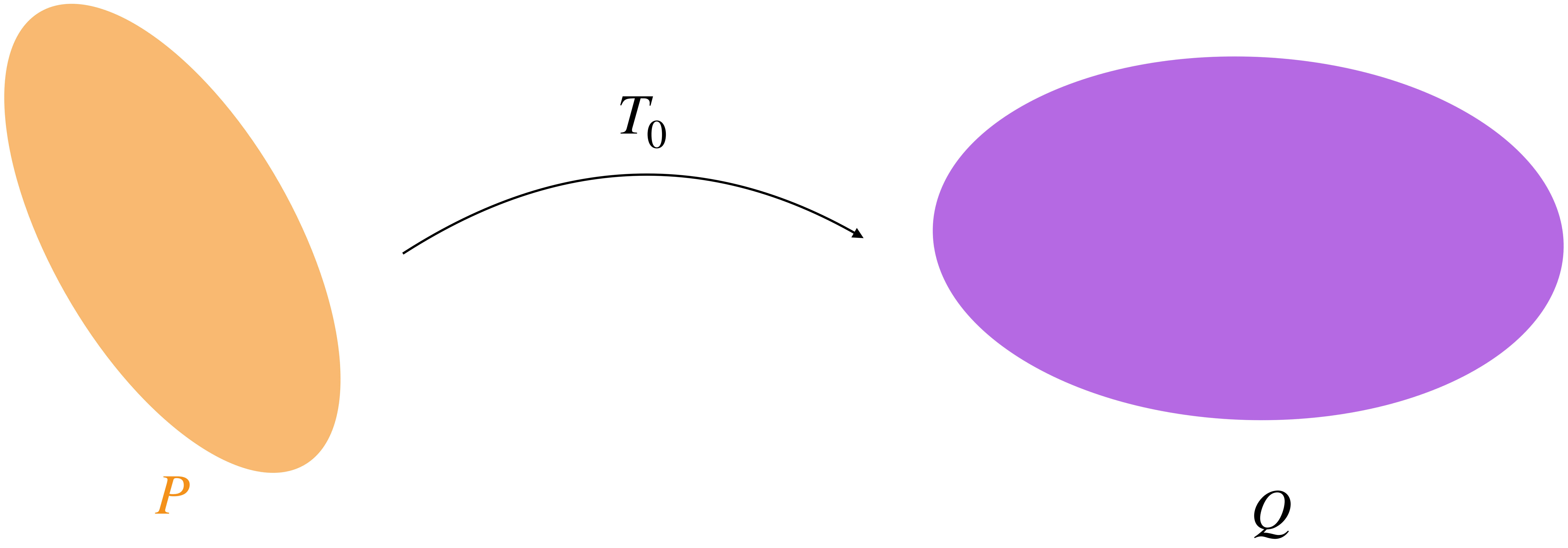
Estimating optimal transport maps



T is a *transport map* from P to Q if: for $X \sim P, T(X) \sim Q$

(write $T \in \mathcal{T}(P, Q)$)

Estimating optimal transport maps

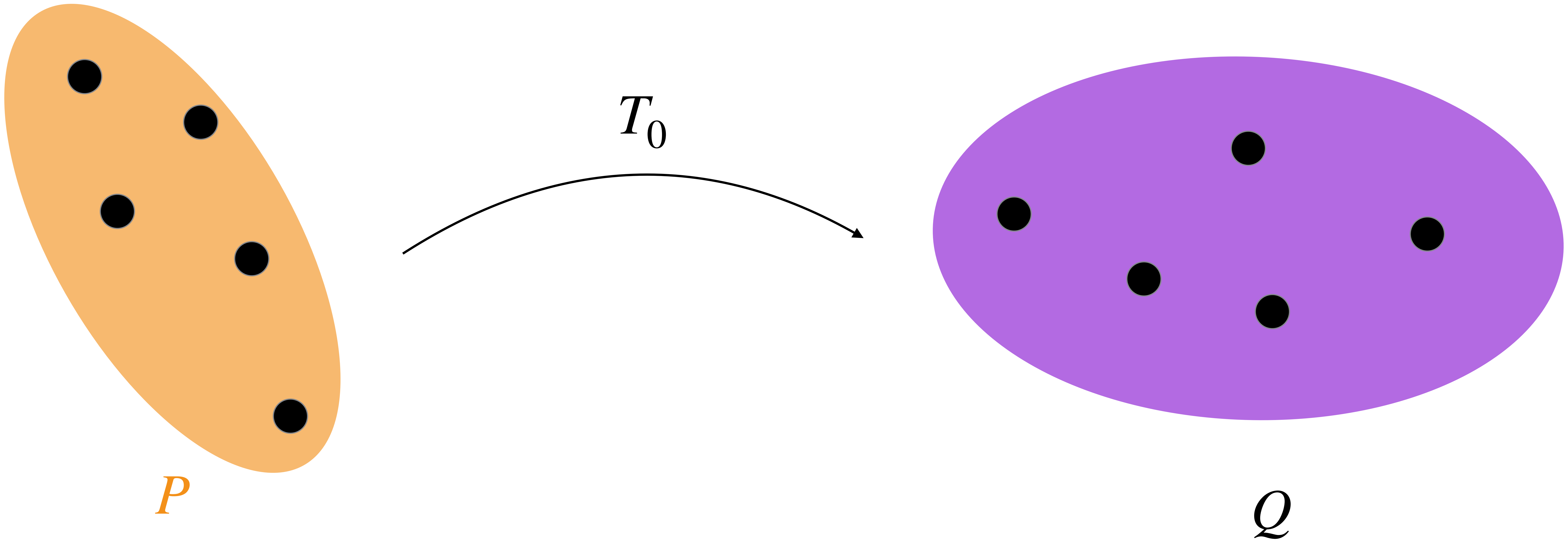


Monge Problem

optimal transport map:

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

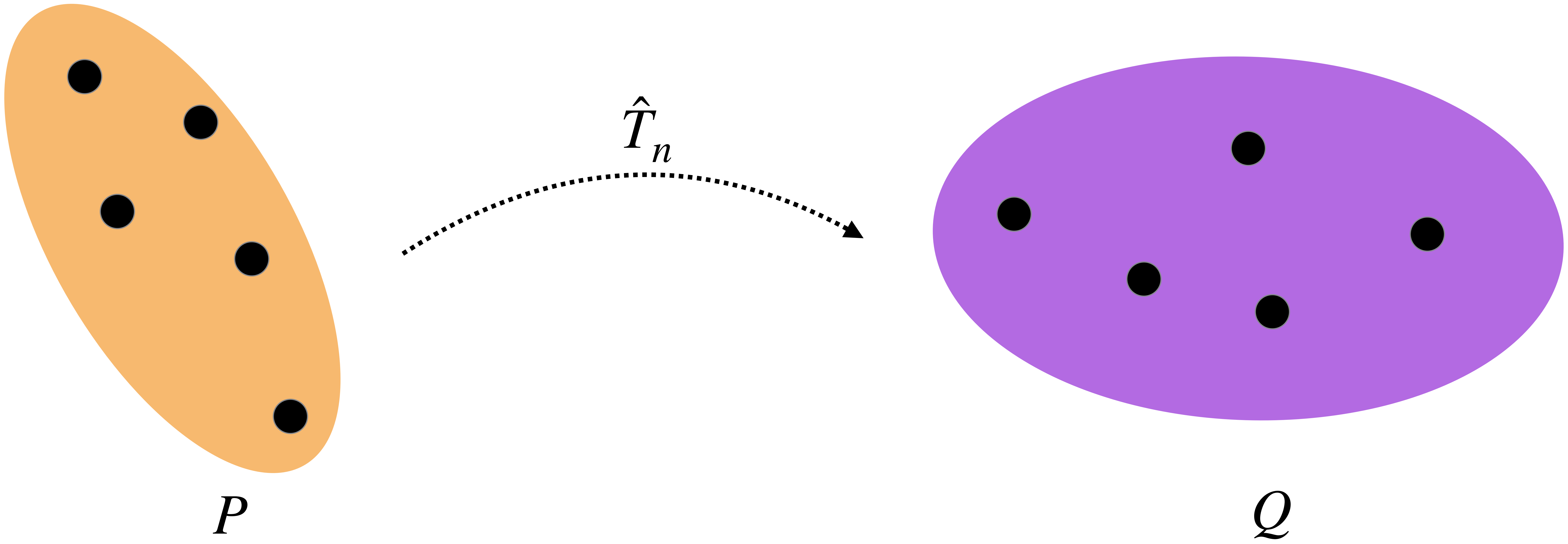
Estimating optimal transport maps



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?

Estimating optimal transport maps



Goal: Construct estimator $\hat{T}_n$ with "good" computational and statistical properties

Prior work

[HR21]: Wavelet-based estimator that achieves the following estimation rate

$$\mathbb{E}\|\hat{T}_n - T_0\|^2_{L^2(P)} \lesssim n^{-\frac{2\alpha}{2\alpha - 2 + d}} \log^3(n) \qquad (T_0 \in C^\alpha)$$

Prior work

[HR21]: Wavelet-based estimator that achieves the following estimation rate

$$\mathbb{E}\|\hat{T}_n - T_0\|^2_{L^2(P)} \lesssim n^{-\frac{2\alpha}{2\alpha - 2 + d}} \log^3(n) \qquad (T_0 \in C^\alpha)$$

- (Near) minimax optimal
- numerically intractable in $d > 3$; complexity $O(N^d)$

Prior work

[HR21]: Wavelet-based estimator that achieves the following estimation rate

$$\mathbb{E}\|\hat{T}_n - T_0\|^2_{L^2(P)} \lesssim n^{-\frac{2\alpha}{2\alpha - 2 + d}} \log^3(n) \qquad (T_0 \in C^\alpha)$$

- (Near) minimax optimal
- numerically intractable in $d > 3$; complexity $O(N^d)$

Estimators from [DGS21] and [MB+21] also achieve minimax rate

Prior work

[HR21]: Wavelet-based estimator that achieves the following estimation rate

$$\mathbb{E}\|\hat{T}_n - T_0\|^2_{L^2(P)} \lesssim n^{-\frac{2\alpha}{2\alpha - 2 + d}} \log^3(n) \qquad (T_0 \in C^\alpha)$$

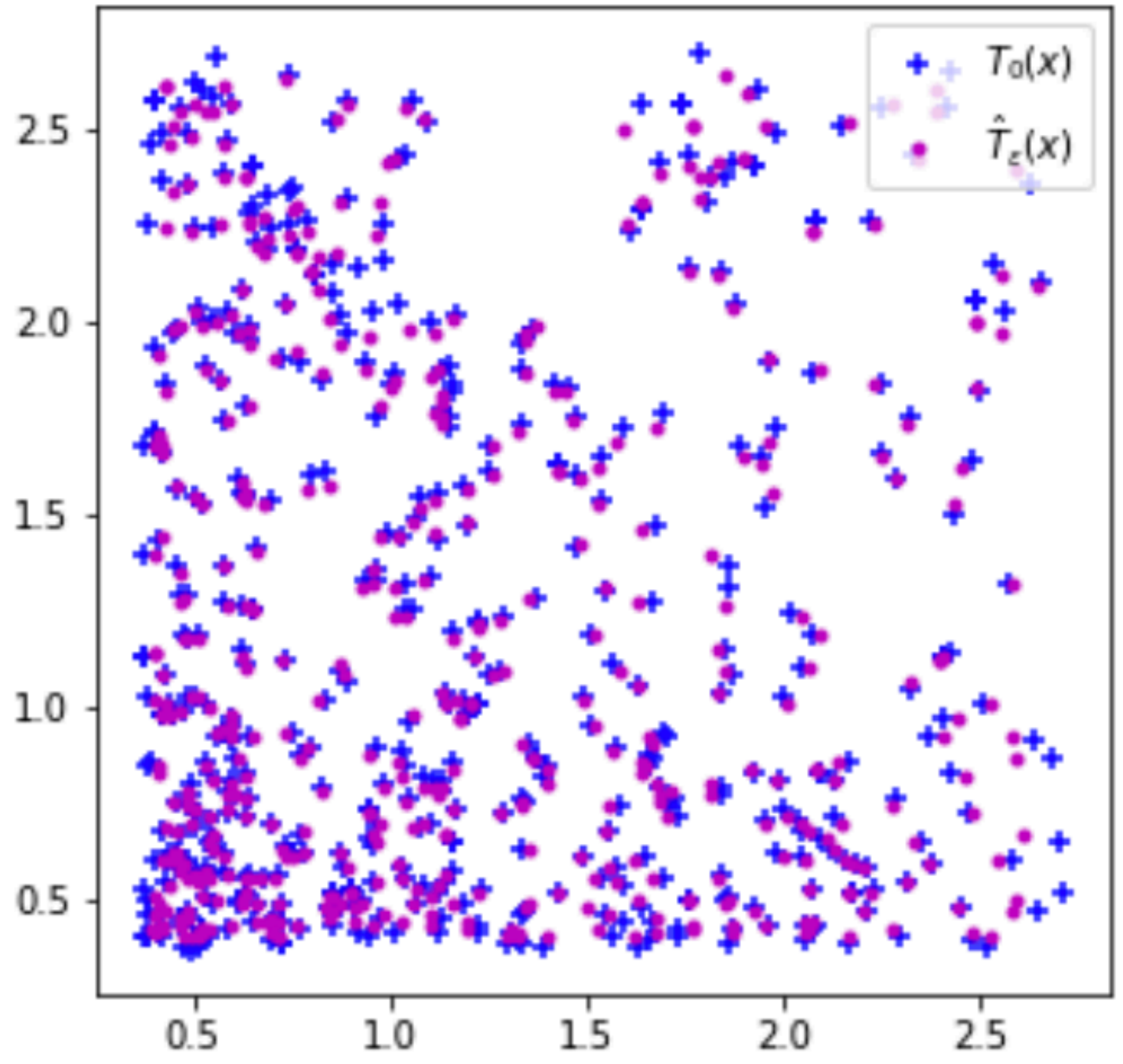
- (Near) minimax optimal
- numerically intractable in $d > 3$; complexity $O(N^d)$

Estimators from [DGS21] and [MB+21] also achieve minimax rate

- 1-Nearest-Neighbor is computationally tractable in $O(n^3)$

Workaround: entropic map

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

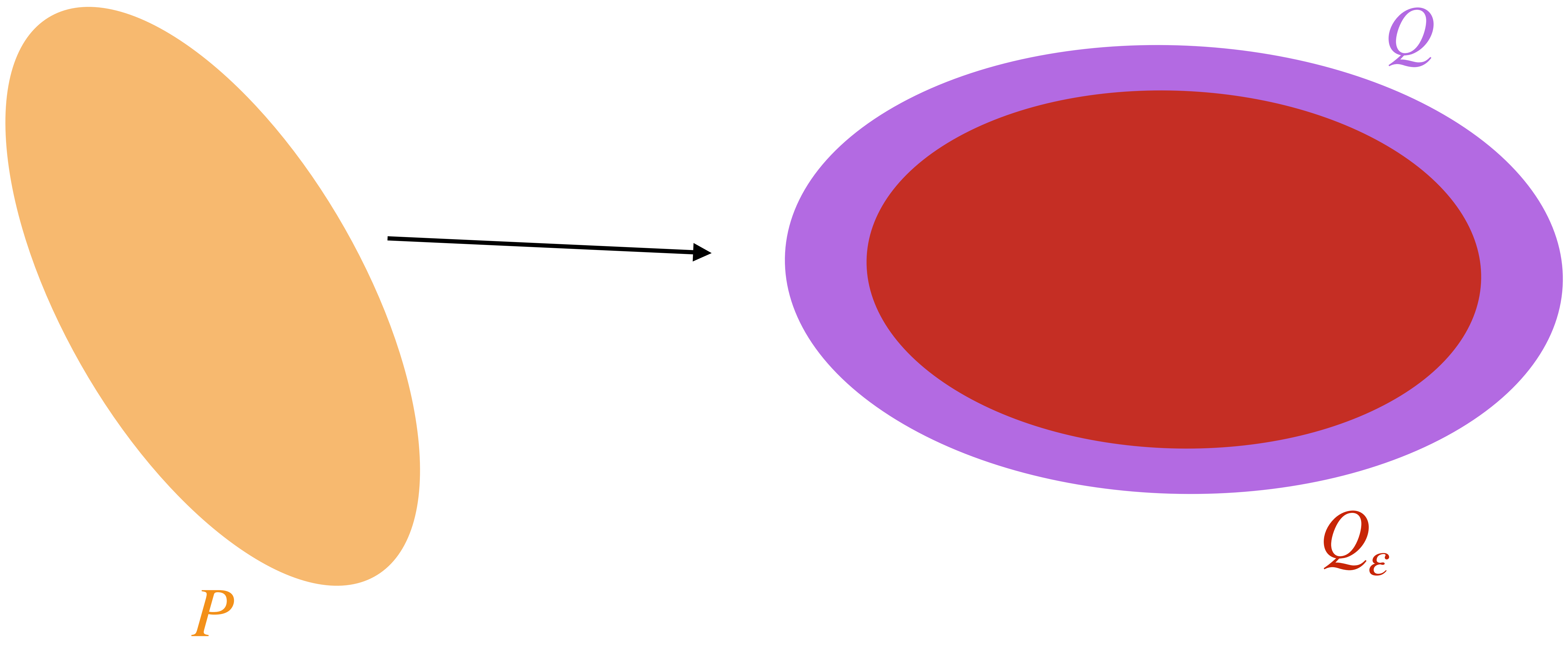


$$T_\varepsilon := \mathbb{E}_{\pi_\varepsilon}[Y \mid X = x]$$

- GPU-friendly implementations
- Complexity: $O(n^2\varepsilon^{-2})$
- $O(n)$ time to evaluate

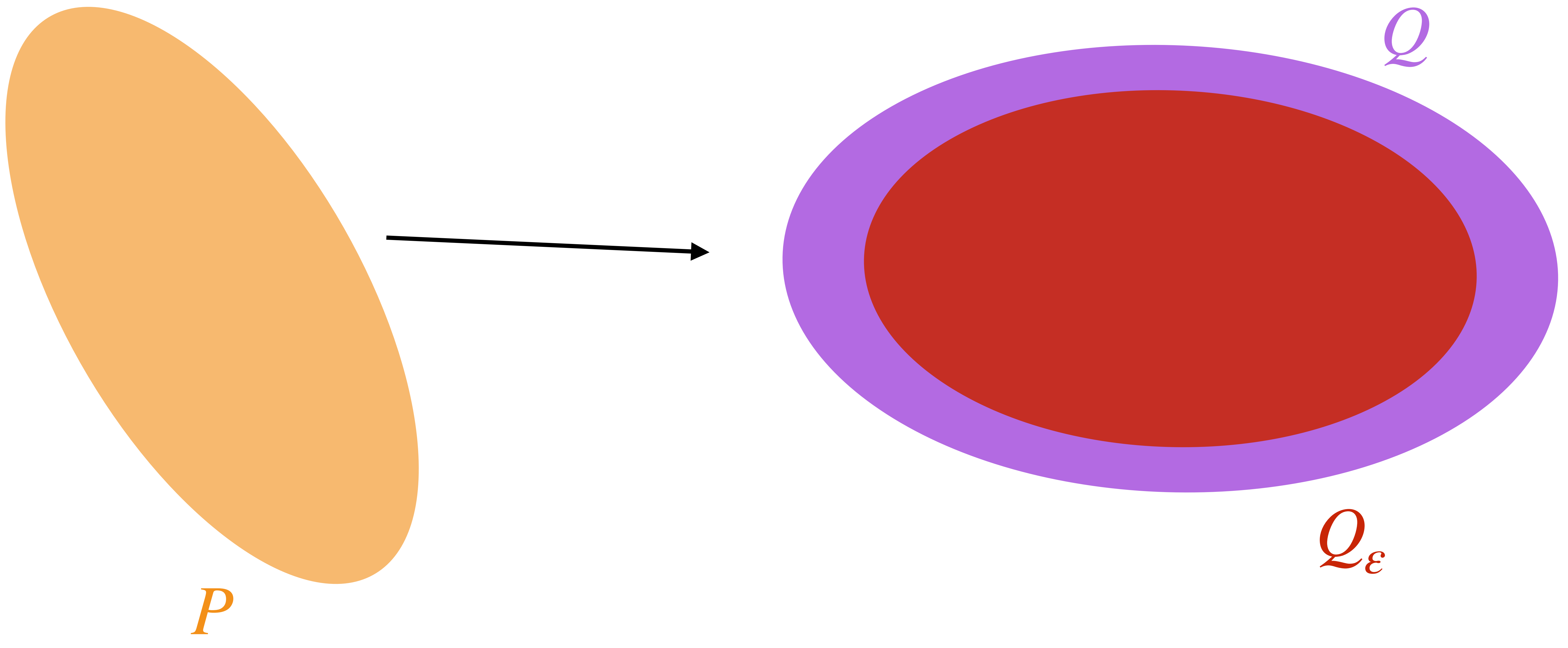
Drawbacks: underdispersed

Drawbacks: underdispersed



Approximation of the target distribution is *underdispersed* for large ε….

Drawbacks: underdispersed



Approximation of the target distribution is *underdispersed* for large ε….

but we want ε large!

e.g. [CT+20]

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 $\mu = \nu$, we recover $T_\varepsilon \simeq \mathrm{id}$

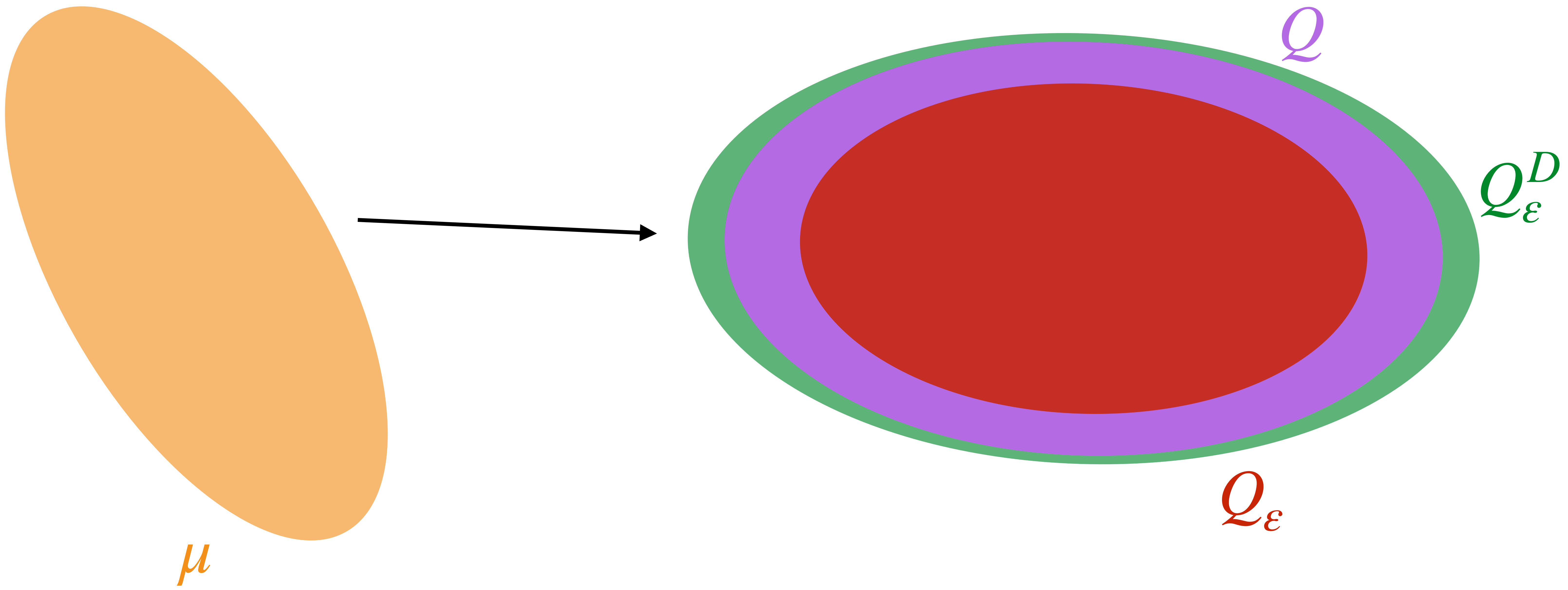
- The correction term is obtained by solving the entropic transport problem from a measure onto itself

Fix: Debiasing/Centering

Debiased entropic map T_{ε}^{D}

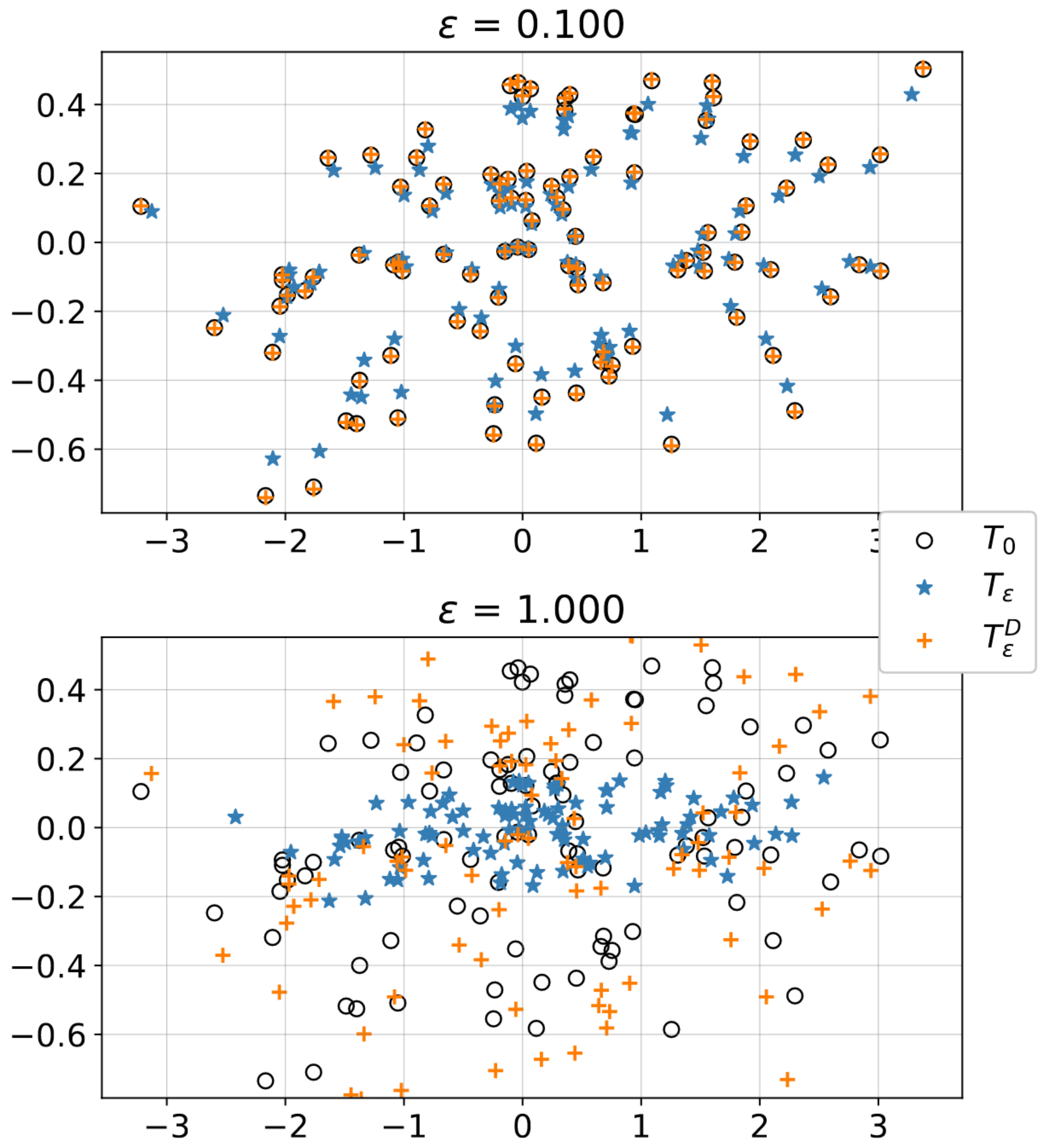
Fix: Debiasing/Centering

Debiased entropic map T_ε^D



<u>*Overdispersed*</u> for large ε

Fix: Debiasing/Centering



Debiased entropic map T_ε^D

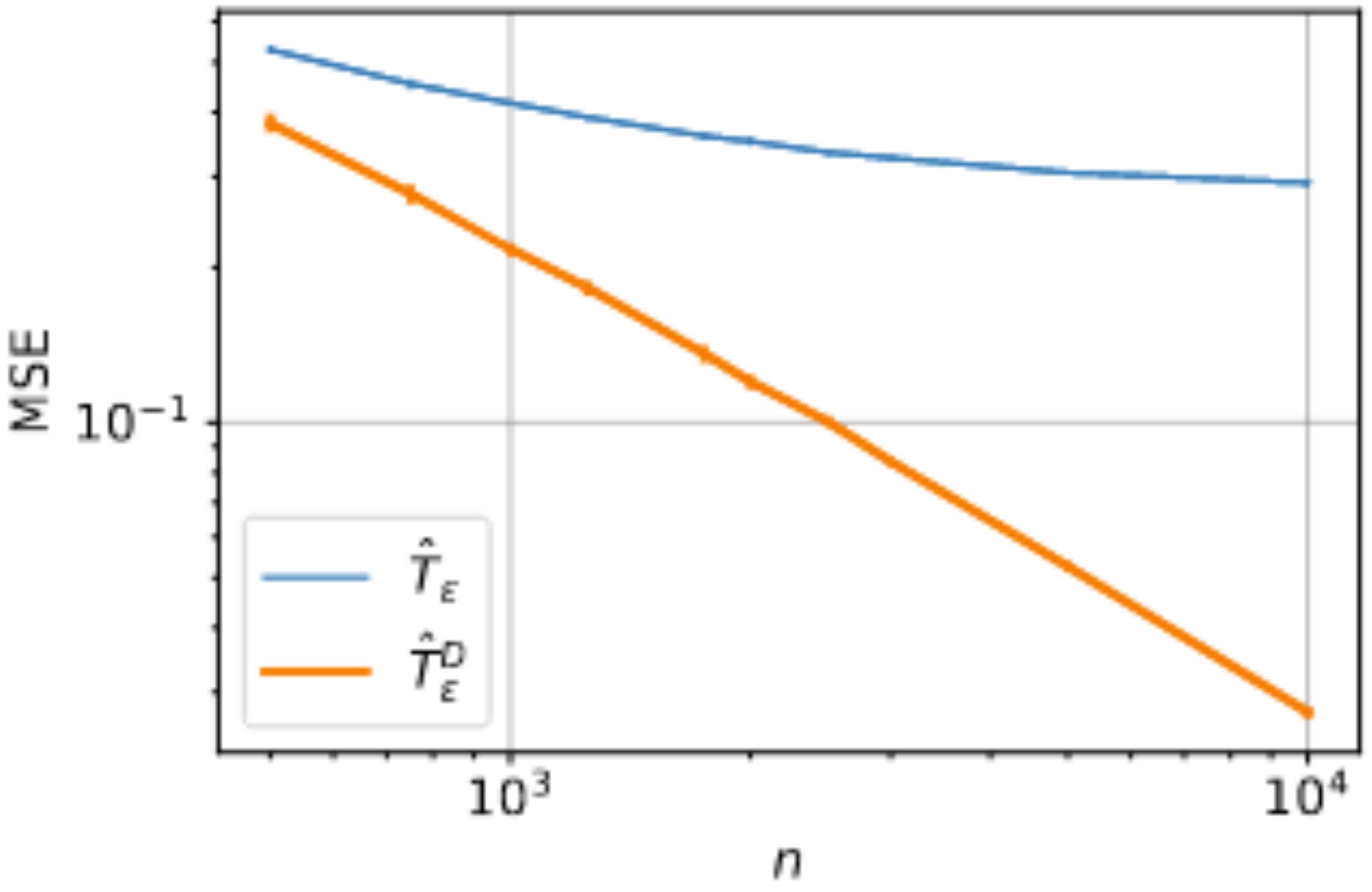
versus (biased) entropic map T_ε

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

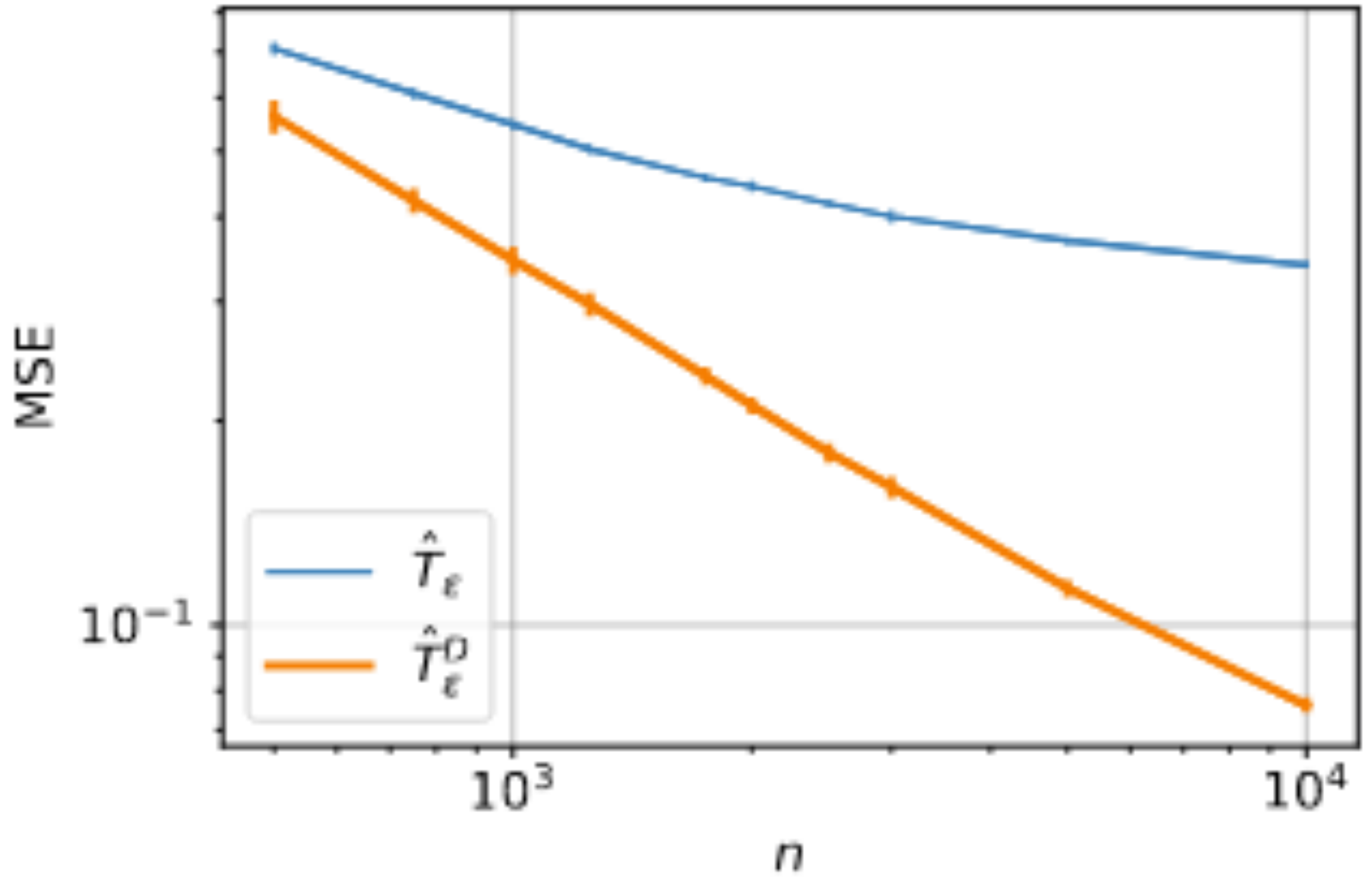
Fix: Debiasing/Centering

For a well-chosen value of ε, we see that debiasing significantly aids in estimating (smooth) optimal transport maps (plots are in $d = 10$)

$T_0(x) = Ax$



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



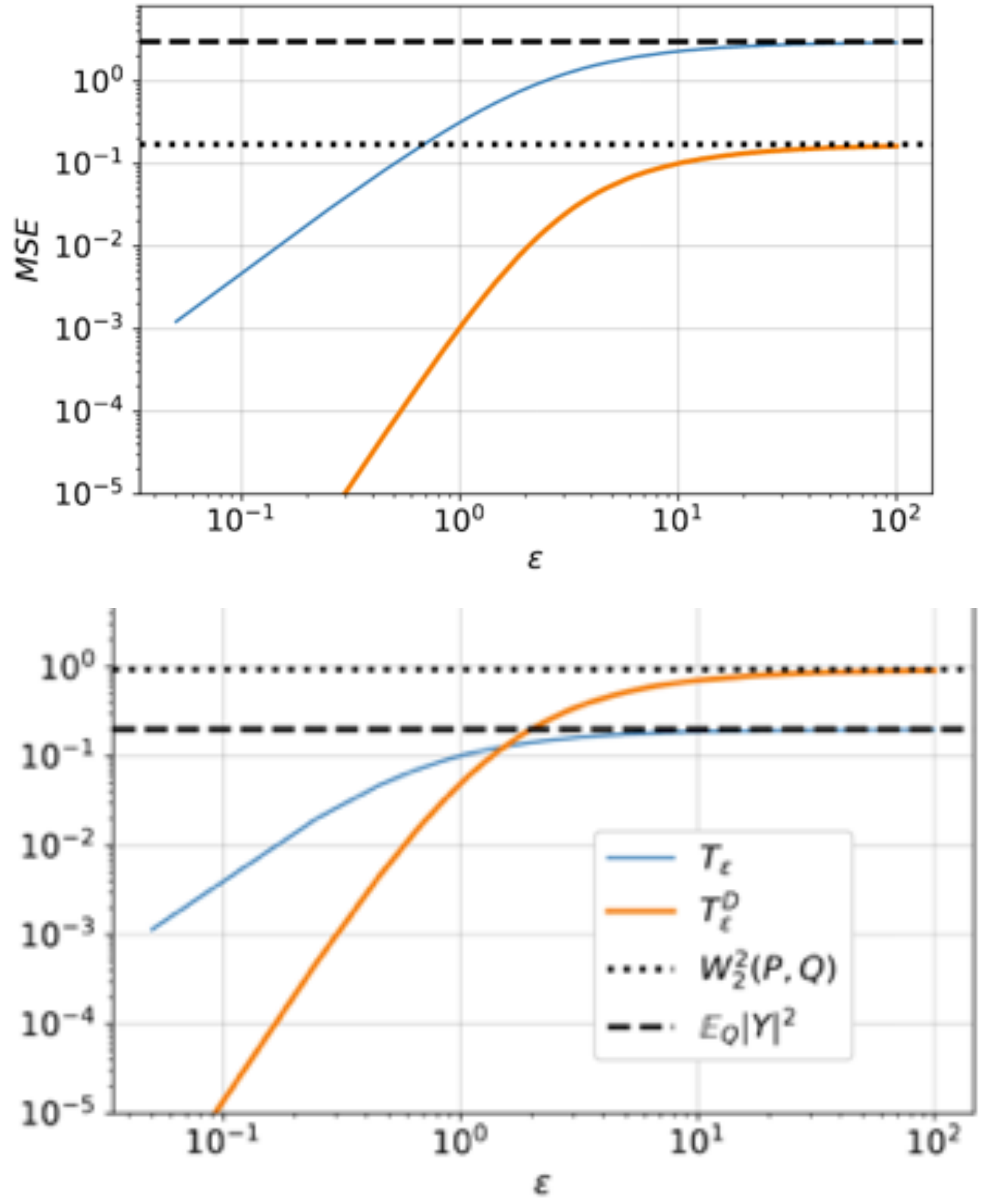
Gaussian-to-Gaussian example

If $P = \mathcal{N}(0,A)$ and $Q = \mathcal{N}(0,B)$, then as $\varepsilon \to 0$

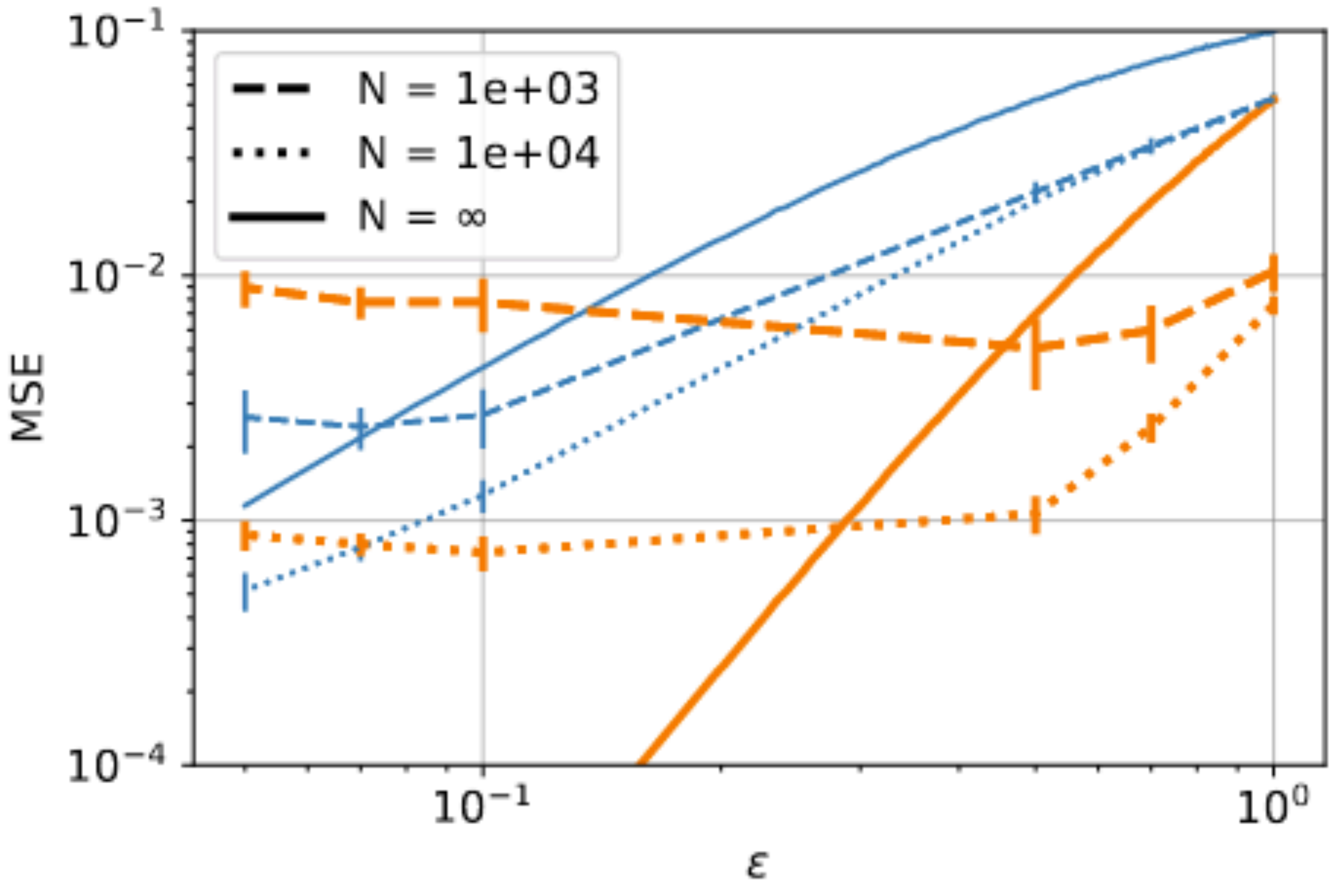
$$\|T_\varepsilon - T_0\|^2_{L^2(P)} \lesssim \varepsilon^2 + O(\varepsilon^4)$$

vs.

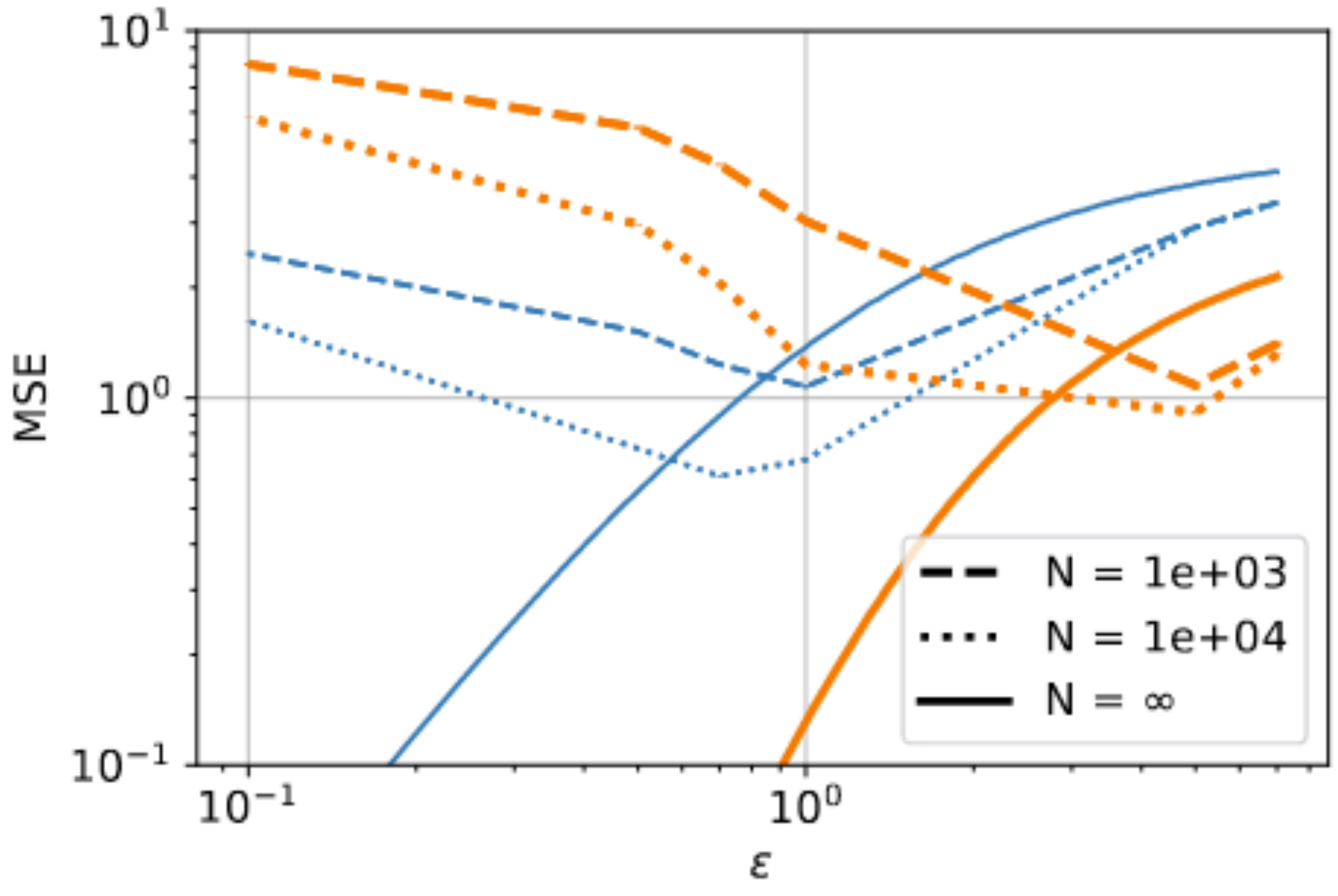
$$\|T^D_\varepsilon - T_0\|^2_{L^2(P)} \lesssim \varepsilon^4 + O(\varepsilon^6)$$



Beware of pitfalls



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

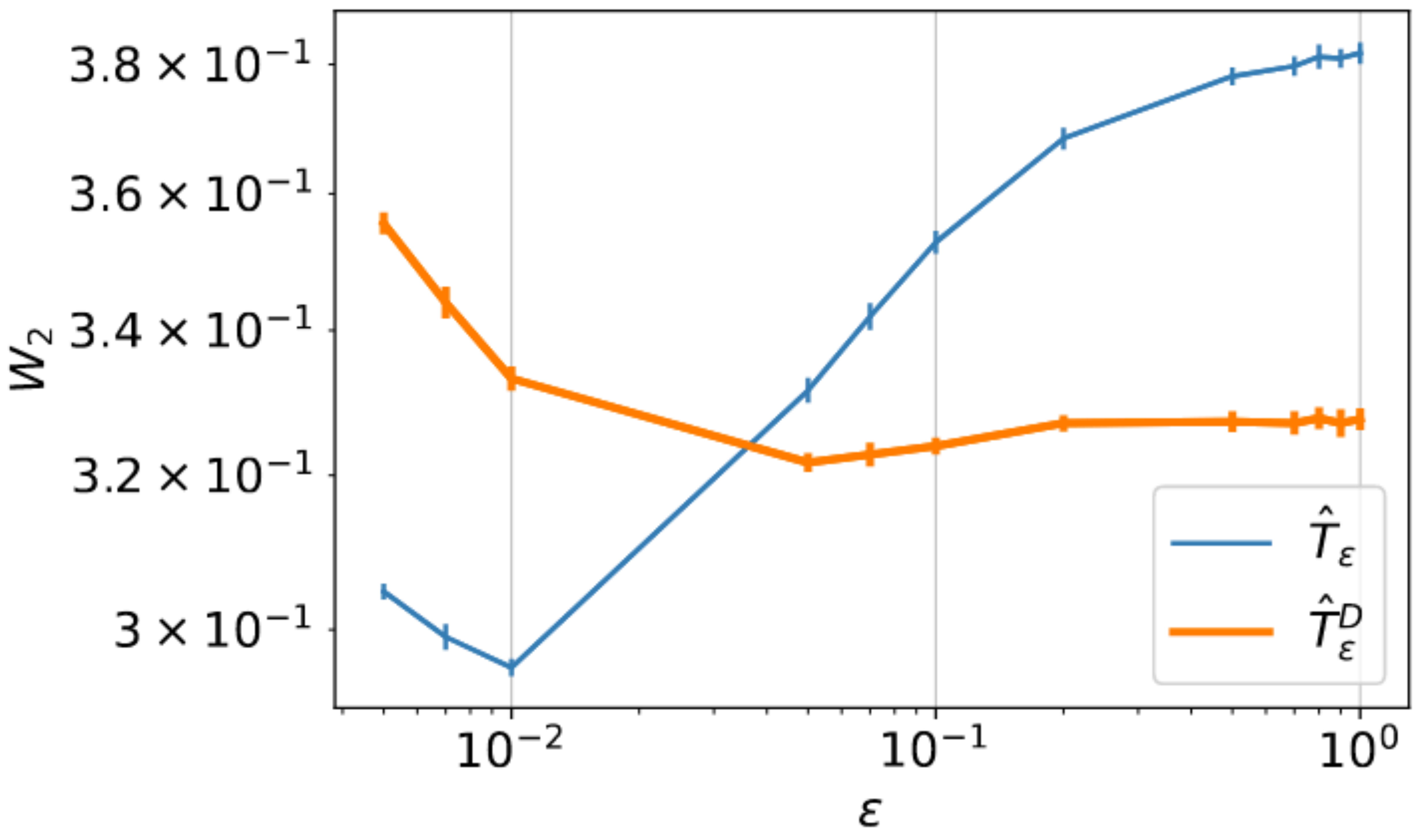


(b) $\hat{T}_\varepsilon$ vs. $\hat{T}_\varepsilon^D$ with Σ concentrated in $d = 15$

Beware of pitfalls

- Predicting genomic trajectories in stem cells [SS+19]

- Tradeoff in performance as $\varepsilon \to 0$



Key takeaways

- Question the convention wisdom in optimal transport, that suggests debiasing is always better

- Empirically and theoretically complicate this conventional wisdom

- Important for practitioners to be wary of these phenomena in downstream tasks

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