Minimax estimation of discontinuous optimal transport maps: The semi-discrete case

Aram-Alexandre Pooladian

New York University

Computational Optimal Transport
Foundations of Computational Mathematics (FoCM)

Minimax estimation of discontinuous optimal transport maps: The semi-discrete case

Aram-Alexandre Pooladian

New York University

Computational Optimal Transport
Foundations of Computational Mathematics (FoCM)

(Supported by NSF grant DMS 2232812)

in collaboration with

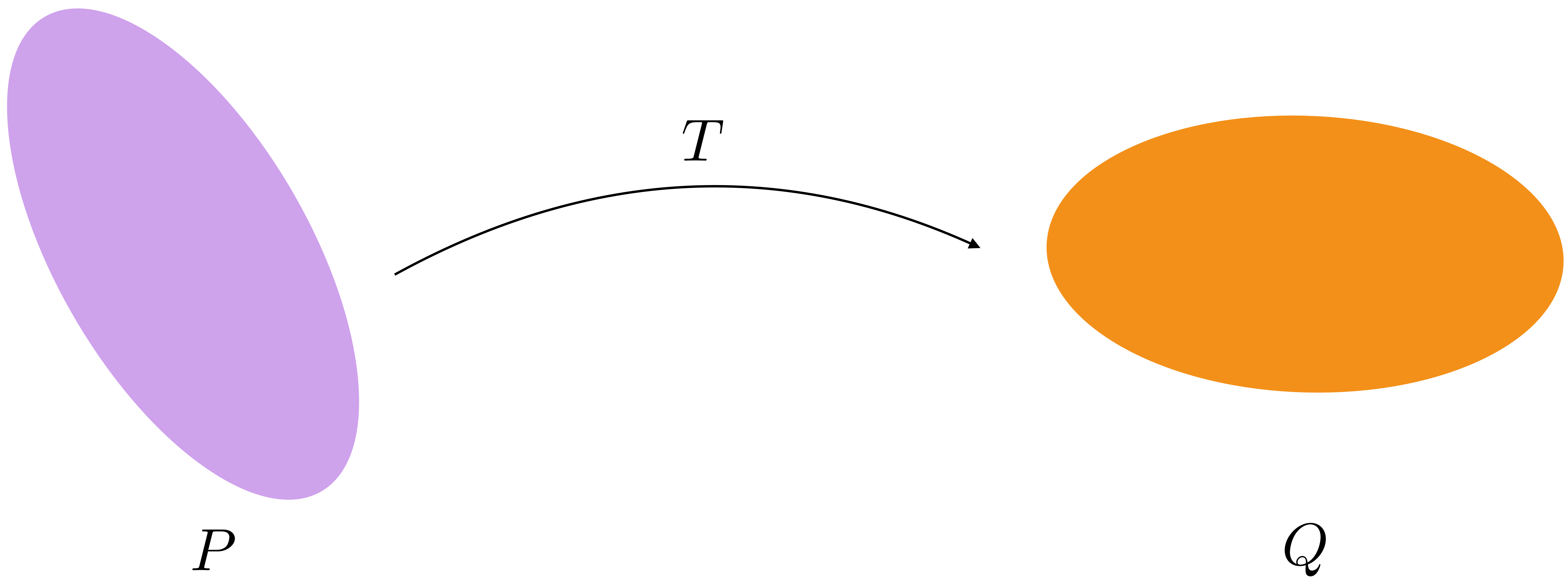
Vincent Divol

Jonathan Niles-Weed

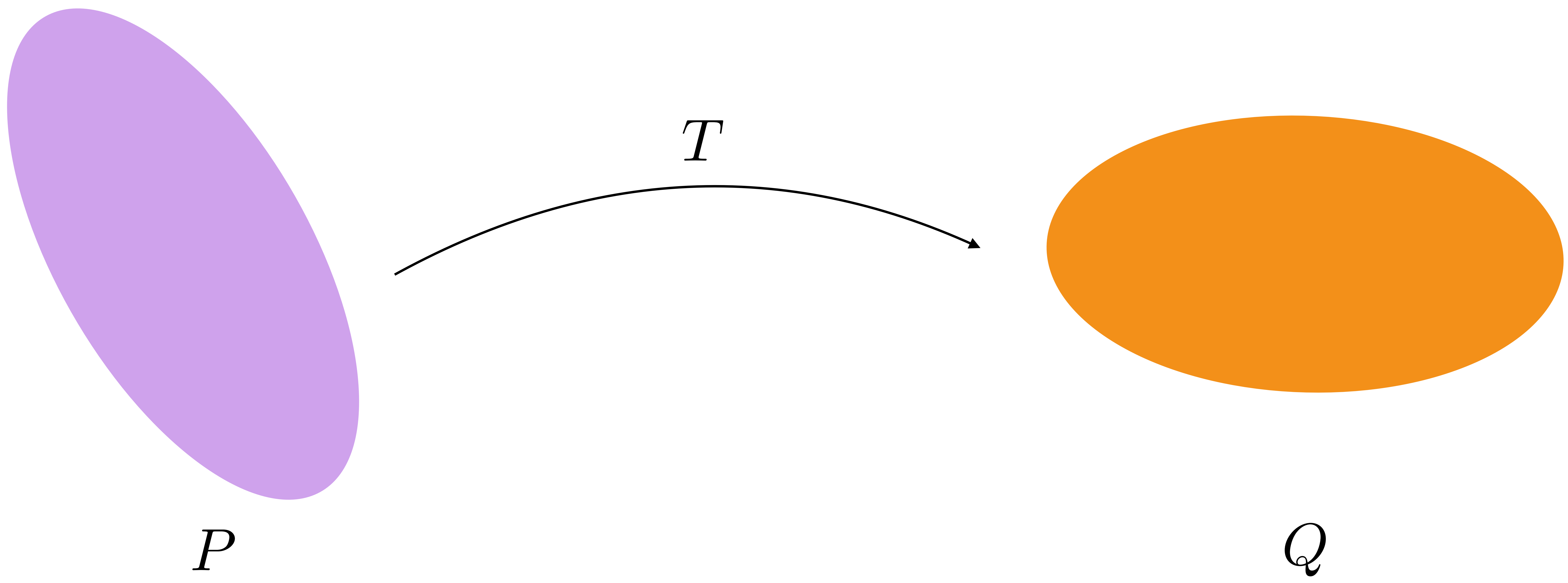
Transport maps

Transport maps

Transport maps

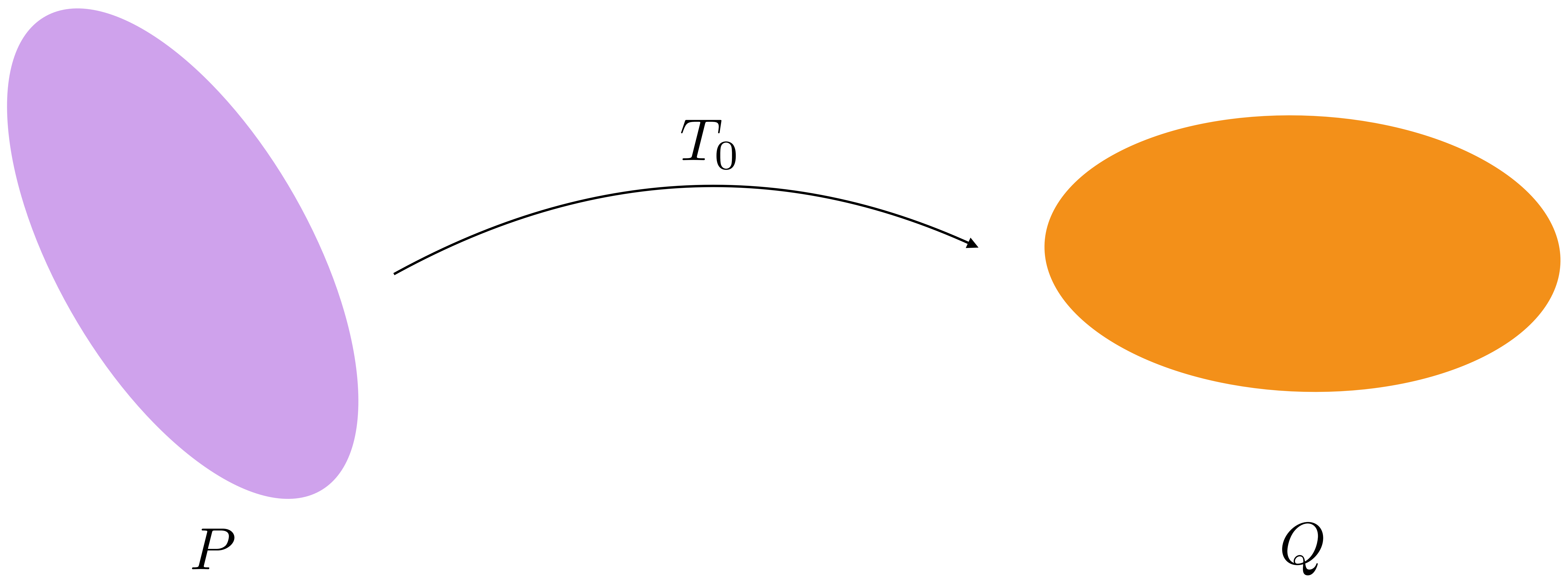


Transport maps



Call T a transport map if $T_\sharp P = Q$ i.e. $X \sim P, T(X) \sim Q$

Optimal transport map

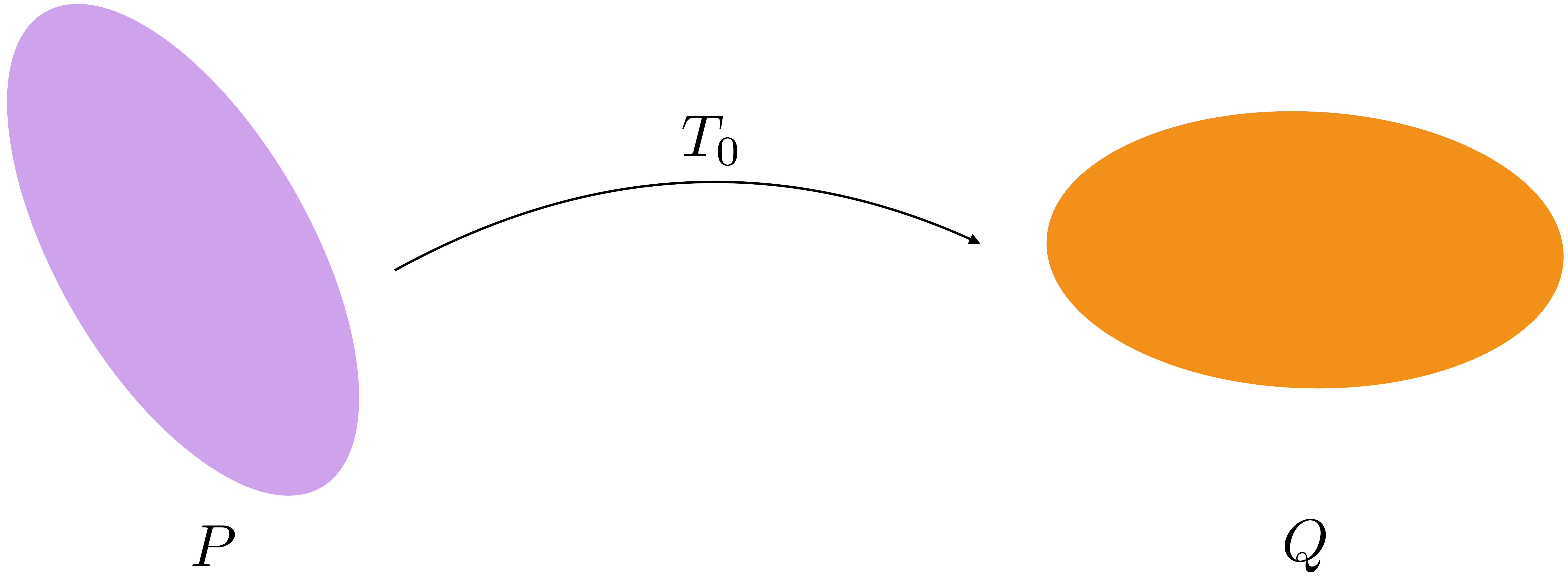


Optimal transport map

P

Monge (1781)
[colorized]

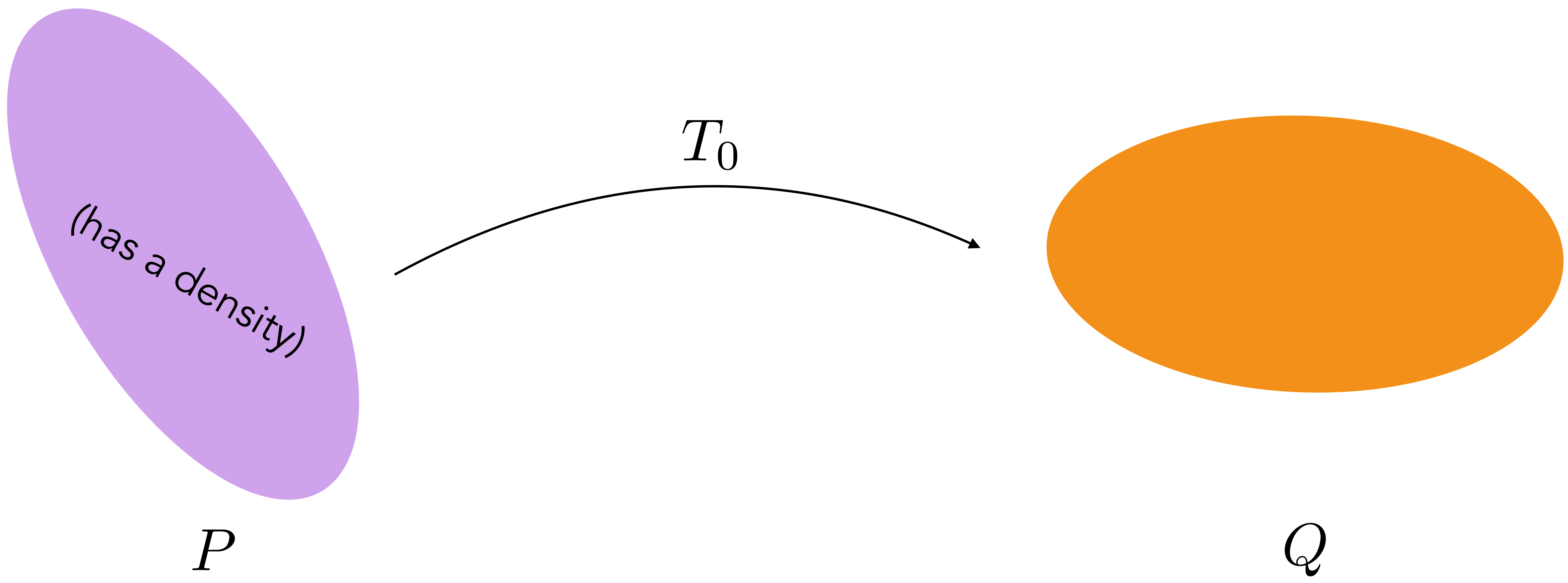
Optimal transport map



Monge (1781)

$$T_0 := \underset{T : T_\sharp P = Q}{\operatorname{argmin}} \mathbb{E}_{X \sim P} \|X - T(X)\|^2$$

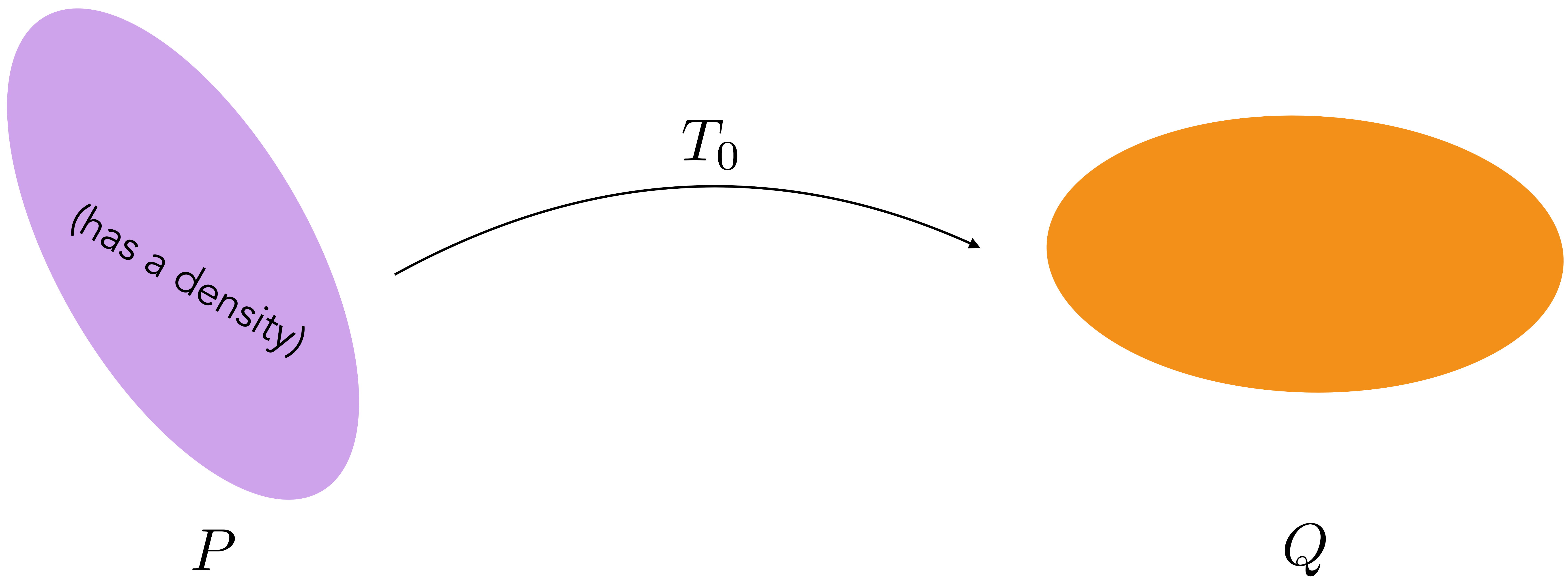
Brenier map



Brenier (1991)
φ_0 is convex

$$\nabla \varphi_0 := \underset{T : T_\sharp P = Q}{\operatorname{argmin}} \mathbb{E}_{X \sim P} \|X - T(X)\|^2$$

Brenier map

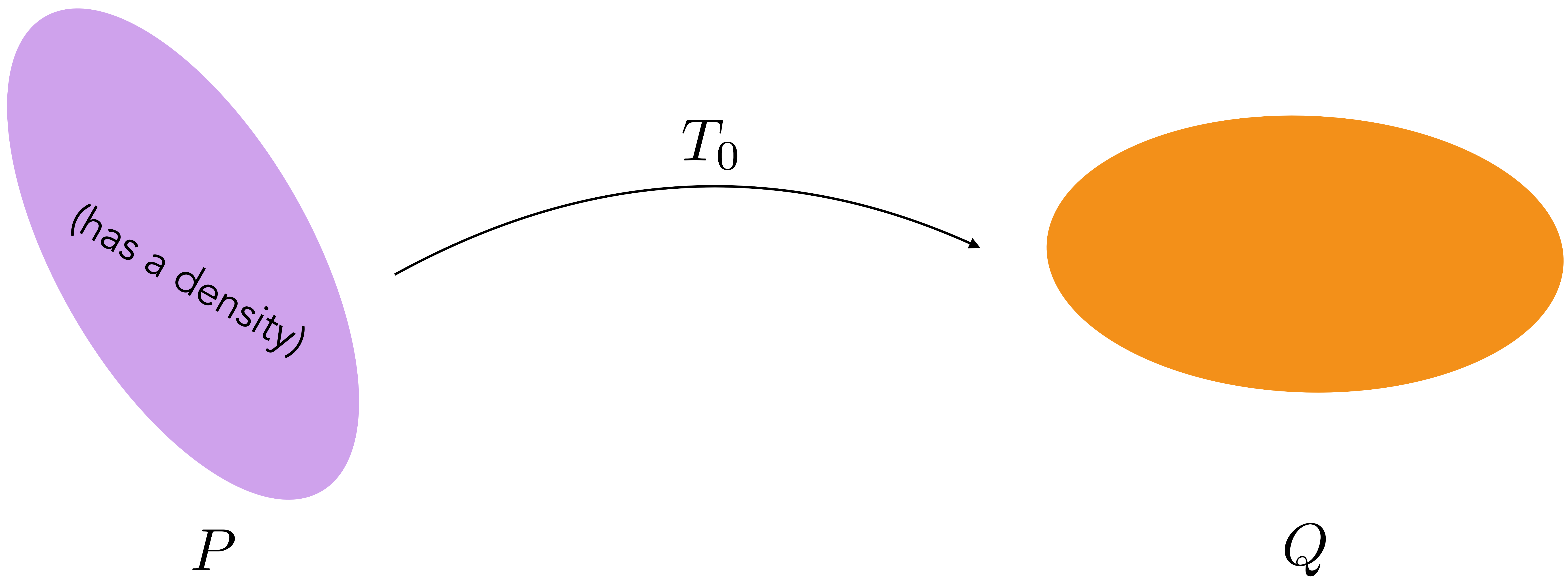


Brenier (1991)
φ_0 is convex

$$\nabla \varphi_0 := \underset{T : T_\sharp P = Q}{\operatorname{argmin}} \mathbb{E}_{X \sim P} \|X - T(X)\|^2$$

$$\varphi_0 = \operatorname{argmin}_\varphi \int \varphi \, \mathrm{d}P + \int \varphi^* \mathrm{d}Q$$

Brenier map



Brenier (1991)
φ_0 is convex

$$\nabla \varphi_0 := \underset{T : T_\sharp P = Q}{\operatorname{argmin}} \mathbb{E}_{X \sim P} \|X - T(X)\|^2$$

$$\varphi_0 = \operatorname{argmin}_\varphi \mathcal{S}(\varphi)$$

Where do transport maps arise?

Where do transport maps arise?

- Generative modelling
  - Huang et al. (2020), Amos (2023), Chen et al. (2023), etc.

Where do transport maps arise?

- Generative modelling
  - Huang et al. (2020), Amos (2023), Chen et al. (2023), etc.
- Multivariate rank and quantile estimation
  - Ghosal & Sen (2022), Hallin et al. (2021), Chernozukov et al. (2017)

Where do transport maps arise?

- Generative modelling
  - Huang et al. (2020), Amos (2023), Chen et al. (2023), etc.
- Multivariate rank and quantile estimation
  - Ghosal & Sen (2022), Hallin et al. (2021), Chernozukov et al. (2017)
- Computational biology (predicting trajectories of single-cell genomics)
  - Schiebinger et al. (2019), Bunne et al (2021, 2022, 2023)

Where do transport maps arise?

- Generative modelling
  - Huang et al. (2020), Amos (2023), Chen et al. (2023), etc.
- Multivariate rank and quantile estimation
  - Ghosal & Sen (2022), Hallin et al. (2021), Chernozukov et al. (2017)
- Computational biology (predicting trajectories of single-cell genomics)
  - Schiebinger et al. (2019), Bunne et al (2021, 2022, 2023)
- Causal inference
  - Gunsilius et al. (2020, 2021)

Where do transport maps arise?

- Generative modelling
  - Huang et al. (2020), Amos (2023), Chen et al. (2023), etc.
- Multivariate rank and quantile estimation
  - Ghosal & Sen (2022), Hallin et al. (2021), Chernozukov et al. (2017)
- Computational biology (predicting trajectories of single-cell genomics)
  - Schiebinger et al. (2019), Bunne et al (2021, 2022, 2023)
- Causal inference
  - Gunsilius et al. (2020, 2021)
- Transfer learning and domain adaptation
  - Alvarez-Melis and Fusi (2020), Flamary et al. (2015)

Where do transport maps arise?

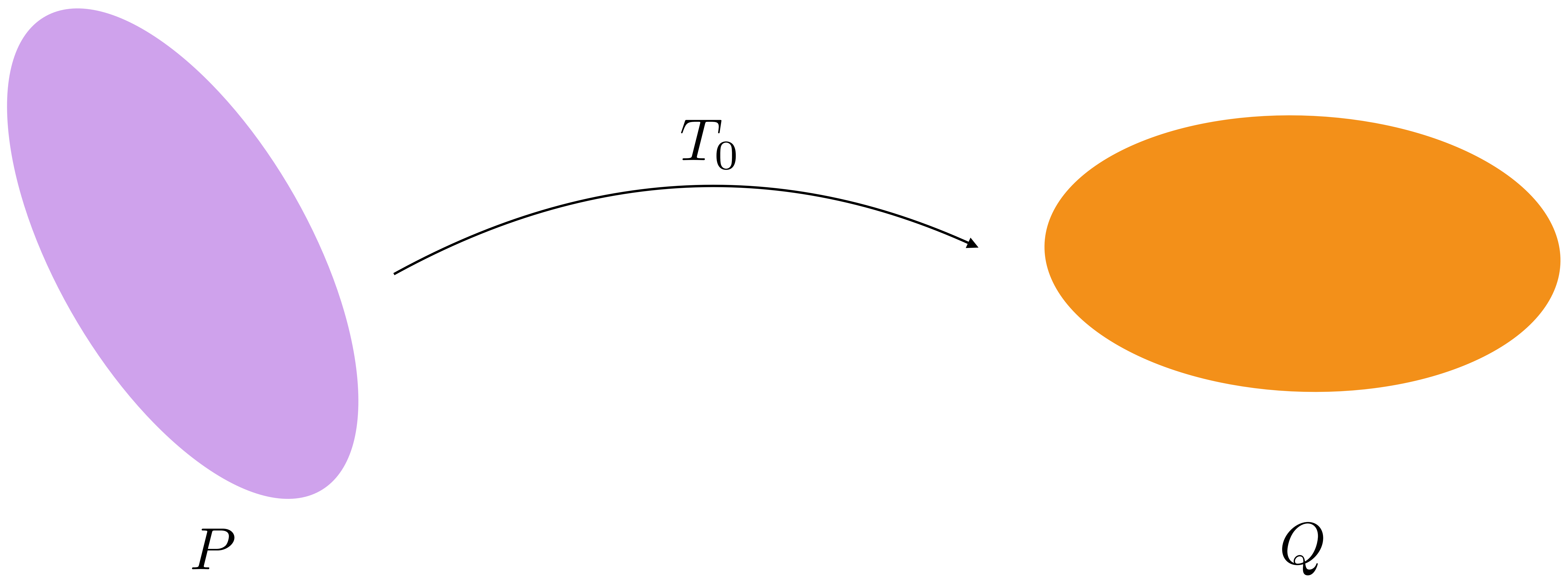
- Generative modelling
  - Huang et al. (2020), Amos (2023), Chen et al. (2023), etc.
- Multivariate rank and quantile estimation
  - Ghosal & Sen (2022), Hallin et al. (2021), Chernozukov et al. (2017)
- Computational biology (predicting trajectories of single-cell genomics)
  - Schiebinger et al. (2019), Bunne et al (2021, 2022, 2023)
- Causal inference
  - Gunsilius et al. (2020, 2021)
- Transfer learning and domain adaptation
  - Alvarez-Melis and Fusi (2020), Flamary et al. (2015)
- among many others!

Statistical estimation of OT maps

Statistical estimation of OT maps

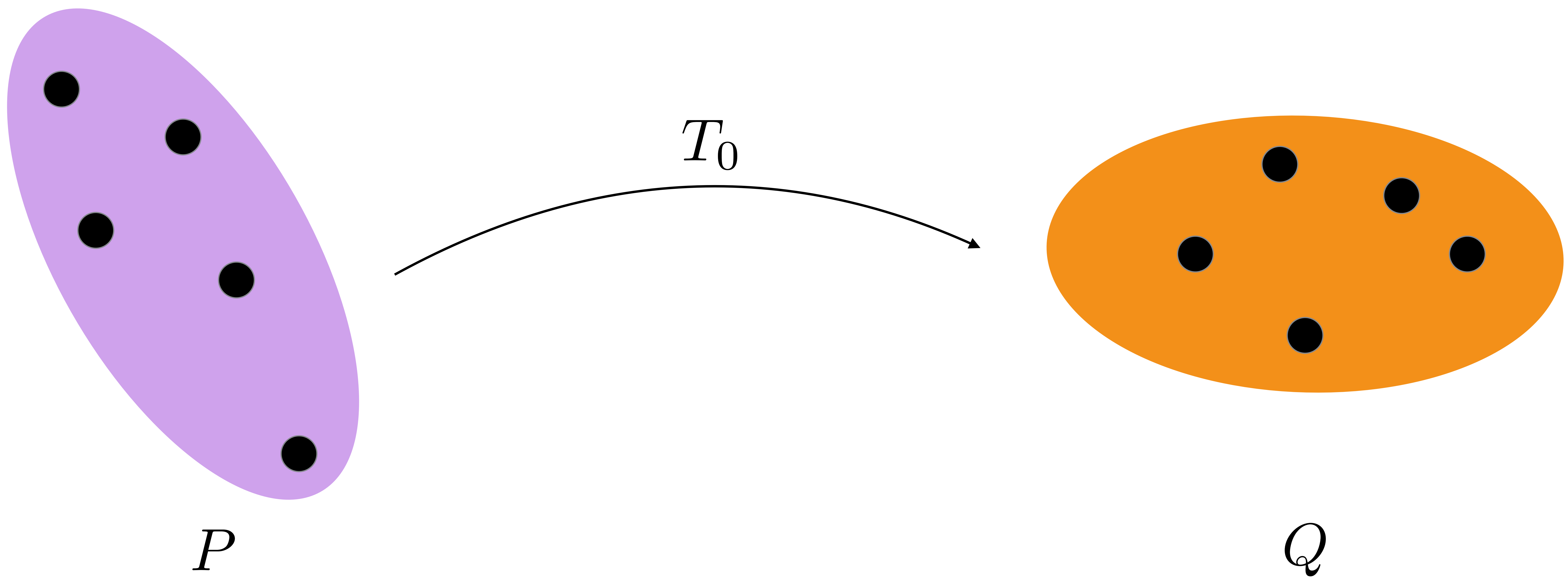
(A1) P has density $0 < p_{\min} \leq p(x) \leq p_{\max}$ with convex support $\mathrm{supp}(P) \subseteq B(0; R)$

Statistical estimation of OT maps



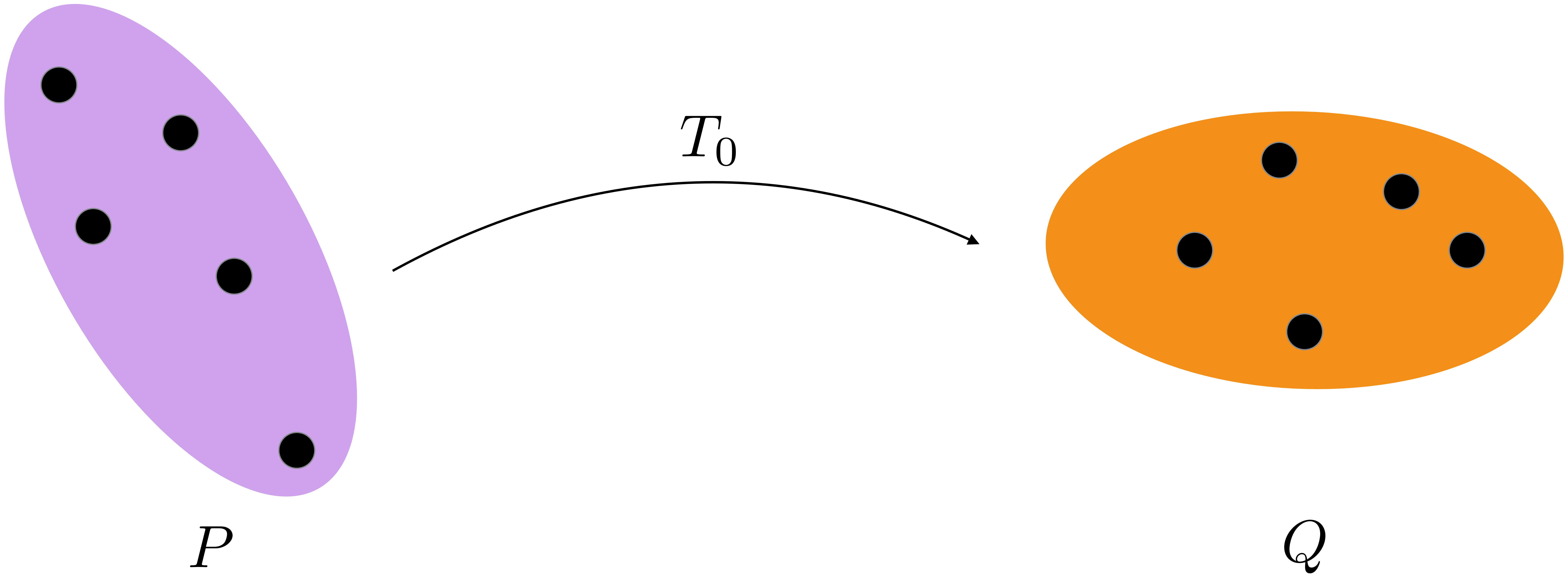
(A1) P has density $0 < p_{\min} \leq p(x) \leq p_{\max}$ with convex support $\text{supp}(P) \subseteq B(0; R)$

Statistical estimation of OT maps



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

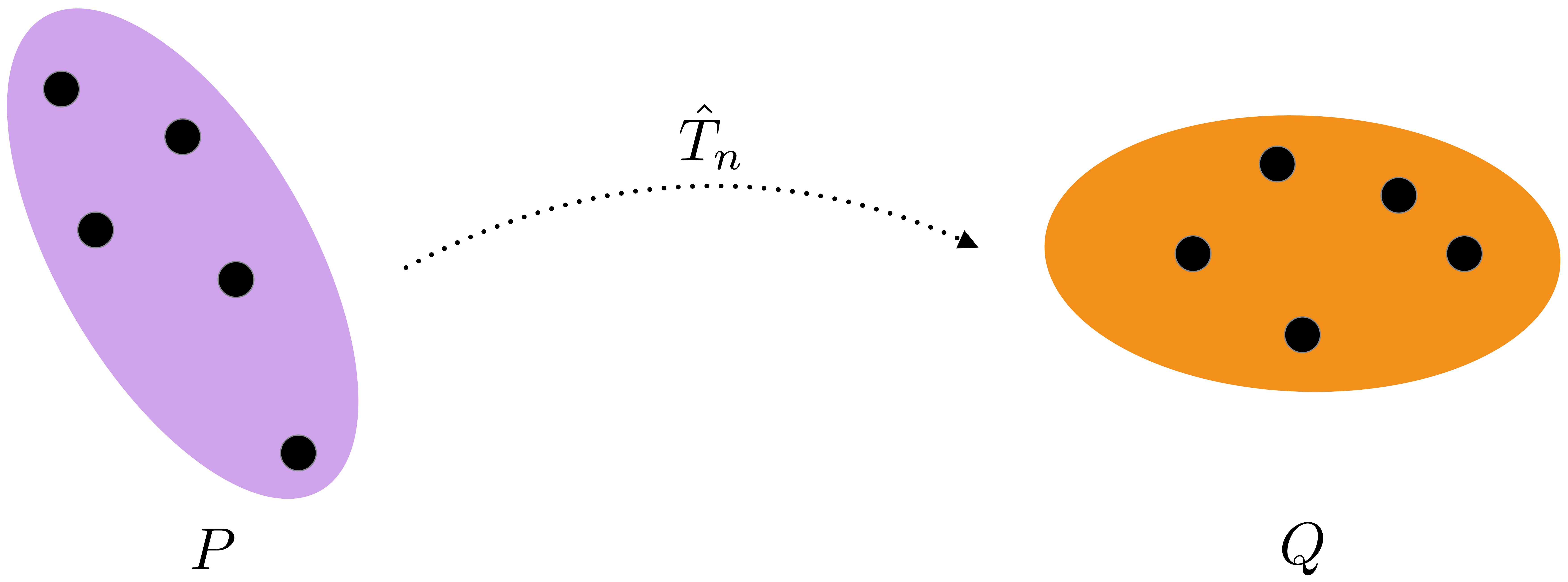
Statistical estimation of OT maps



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

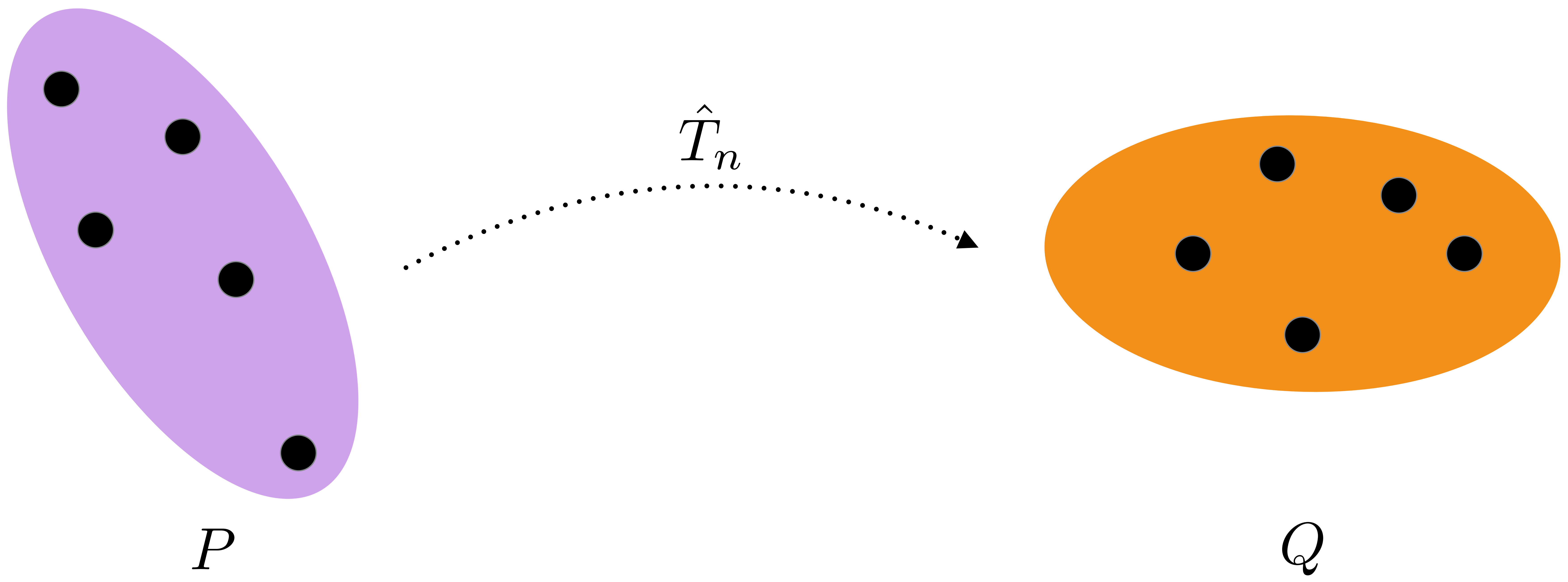
Question: How to estimate T_0 on the basis of samples?

Statistical estimation of OT maps



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

Statistical estimation of OT maps



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

$$\mathbb{E}\|\hat{T}_n - T_0\|^2_{L^2(P)} \lesssim \, ?$$

Prior work

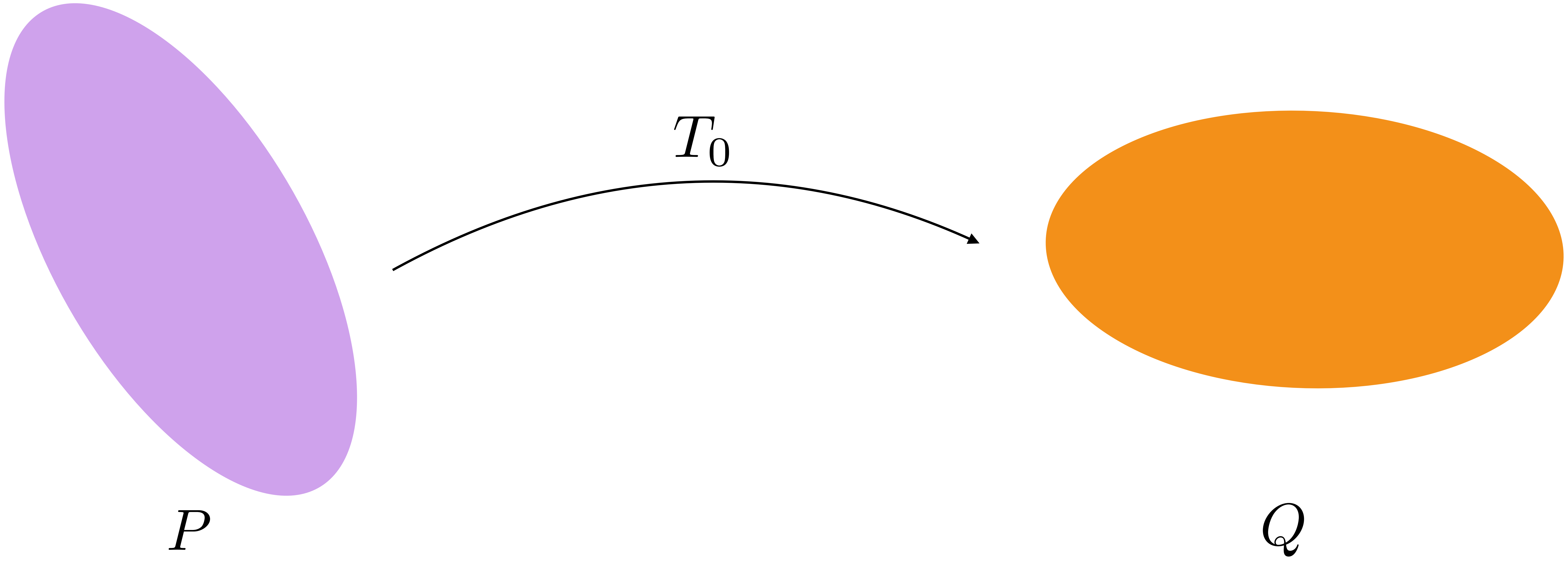
Suppose **(A1)** and T_0 **bi-Lipschitz**,

$$0 \prec \mu I \preceq DT_0 \preceq LI$$

Prior work

Suppose **(A1)** and T_0 **bi-Lipschitz**,

$$0 \prec \mu I \preceq DT_0 \preceq LI$$



Prior work

Suppose **(A1)** and T_0 **bi-Lipschitz**,

$$\mathbb{E}\|\hat{T}_n - T_0\|^2_{L^2(P)} \lesssim n^{-2/d} \qquad 0 \prec \mu I \preceq DT_0 \preceq LI$$

Prior work

Suppose **(A1)** and T_0 **bi-Lipschitz**,

$$\mathbb{E}\|\hat{T}_n - T_0\|^2_{L^2(P)} \lesssim n^{-2/d} \qquad 0 \prec \mu I \preceq DT_0 \preceq LI$$

- Hütter & Rigollet (2019): Estimator based on wavelets
- Deb et al., and Manole et al. (2021): Plug-in estimators
- Muzellec et al. (2021): Kernel SoS (tractable; high smoothness regime)
- Manole et al. (2021): 1-Nearest-Neighbor estimator (tractable, optimal)
- **P.** & Niles-Weed (2021): Entropic optimal transport (tractable, suboptimal)

Prior work

Suppose **(A1)** and T_0 **bi-Lipschitz**,

$$\mathbb{E}\|\hat{T}_n - T_0\|^2_{L^2(P)} \lesssim n^{-2/d} \qquad 0 \prec \mu I \preceq DT_0 \preceq LI$$

- Hütter & Rigollet (2019): Estimator based on wavelets
- Deb et al., and Manole et al. (2021): Plug-in estimators
- Muzellec et al. (2021): Kernel SoS (tractable; high smoothness regime)
- Manole et al. (2021): 1-Nearest-Neighbor estimator (tractable, optimal)
- **P.** & Niles-Weed (2021): Entropic optimal transport (tractable, suboptimal)

How? Crucial lemma: $\|\tilde{T} - T_0\|^2_{L^2(P)} \lesssim L(\mathcal{S}(\tilde{\varphi}) - \mathcal{S}(\varphi_0))$

The case for discontinuity

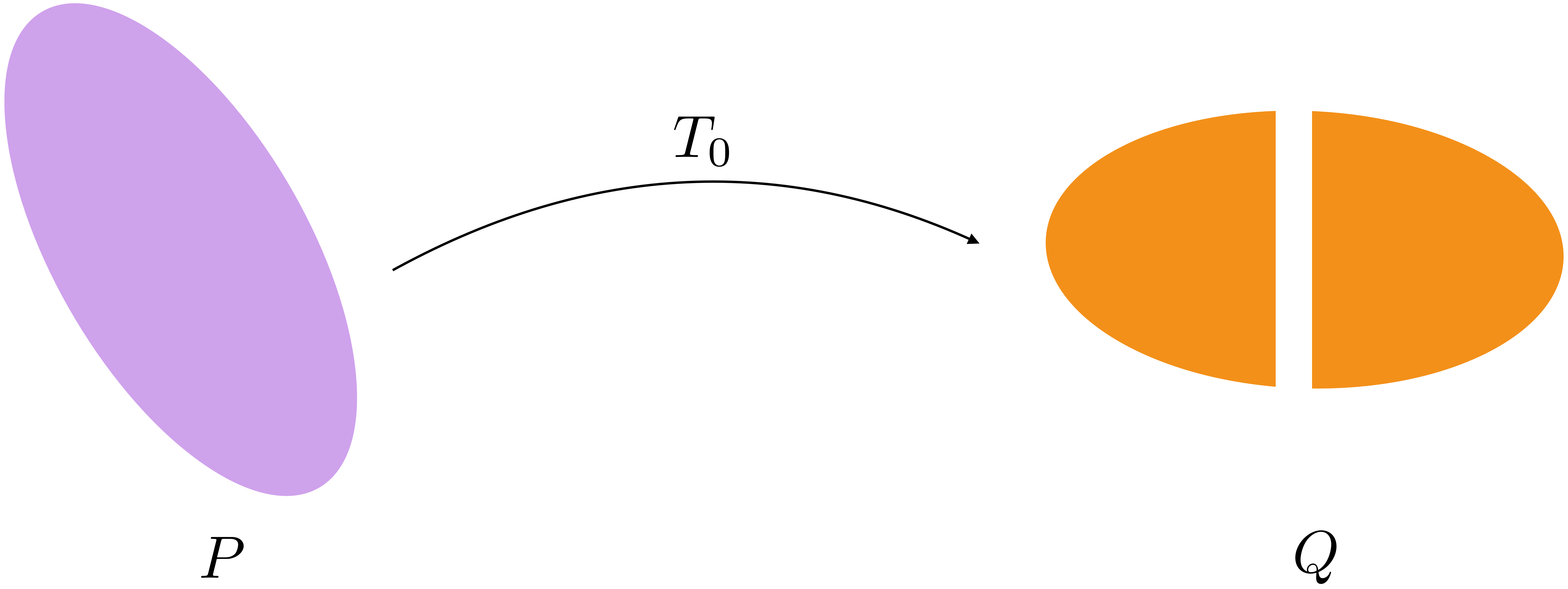
Suppose **(A1)** and T_0 **discontinuous**,

$$\mathbb{E}\|\hat{T}_n - T_0\|^2_{L^2(P)} \lesssim \, ?$$

The case for discontinuity

Suppose **(A1)** and T_0 **discontinuous**,

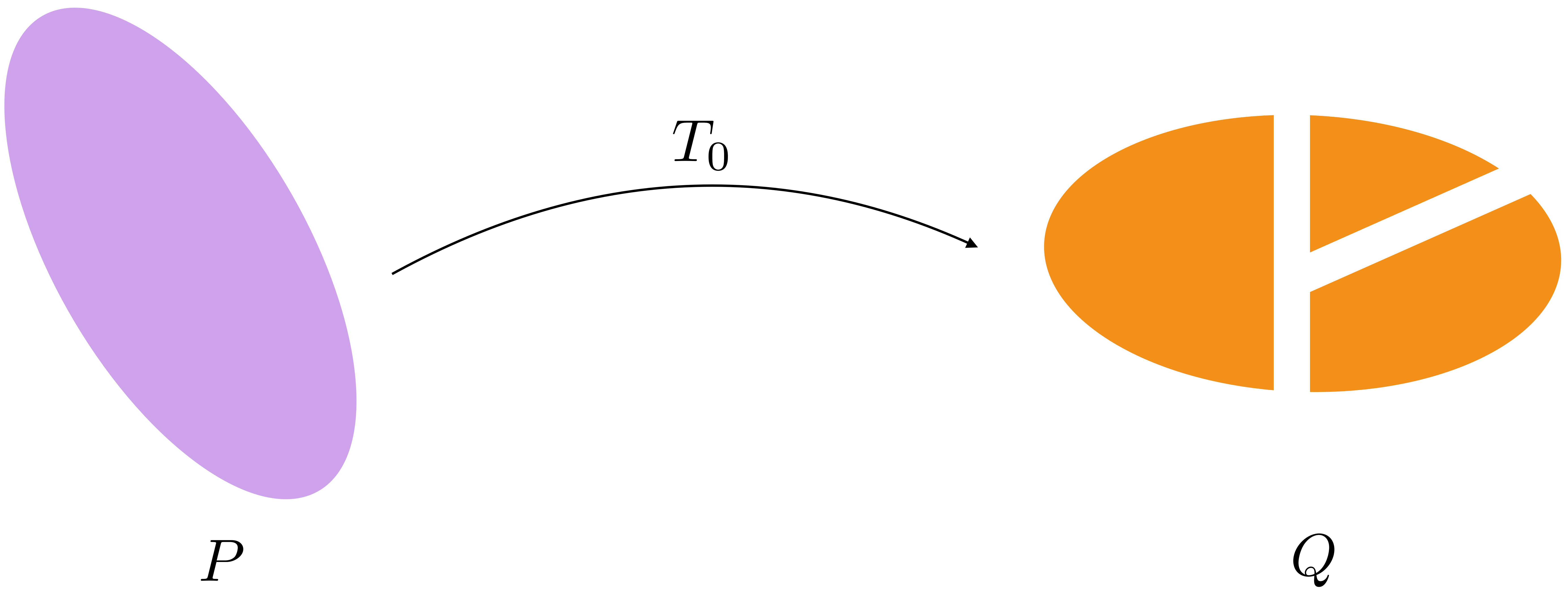
$$\mathbb{E}\|\hat{T}_n - T_0\|^2_{L^2(P)} \lesssim ?$$



The case for discontinuity

Suppose **(A1)** and T_0 **discontinuous**,

$$\mathbb{E}\|\hat{T}_n - T_0\|^2_{L^2(P)} \lesssim ?$$



The case for discontinuity

Suppose **(A1)** and T_0 **discontinuous**,

$$\mathbb{E}\|\hat{T}_n - T_0\|^2_{L^2(P)} \lesssim \, ?$$

- ~~Hütter & Rigollet (2019): Estimator based on wavelets~~
- ~~Deb et al., and Manole et al. (2021): Plug-in estimators~~
- ~~Muzellec et al. (2021): Kernel SoS (tractable; high smoothness regime)~~
- ~~Manole et al. (2021): 1-Nearest-Neighbor estimator (tractable, optimal)~~
- ~~**P.** & Niles-Weed (2021): Entropic optimal transport (tractable, suboptimal)~~

The case for discontinuity

Suppose **(A1)** and T_0 **discontinuous**,

$$\mathbb{E}\|\hat{T}_n - T_0\|^2_{L^2(P)} \lesssim ?$$

- Hütter & Rigollet (2019): Estimator based on wavelets
- Deb et al., and Manole et al. (2021): Plug-in estimators
- Muzellec et al. (2021): Kernel SoS (tractable; high smoothness regime)
- Manole et al. (2021): 1-Nearest-Neighbor estimator (tractable, optimal)
- **P.** & Niles-Weed (2021): Entropic optimal transport (tractable, suboptimal)

Why? Crucial lemma: $\|\tilde{T} - T_0\|^2_{L^2(P)} \lesssim L(\mathcal{S}(\tilde{\varphi}) - \mathcal{S}(\varphi_0))$

The case for discontinuity

Suppose **(A1)** and T_0 **discontinuous**,

$$\mathbb{E}\|\hat{T}_n - T_0\|^2_{L^2(P)} \lesssim \, ?$$

- ~~Hütter & Rigollet (2019): Estimator based on wavelets~~
- ~~Deb et al., and Manole et al. (2021): Plug-in estimators~~
- ~~Muzellec et al. (2021): Kernel SoS (tractable; high smoothness regime)~~
- ~~Manole et al. (2021): 1-Nearest-Neighbor estimator (tractable, optimal)~~
- ~~**P.** & Niles-Weed (2021): Entropic optimal transport (tractable, suboptimal)~~

Why? Crucial lemma: ~~$\|\tilde{T} - T_0\|^2_{L^2(P)} \lesssim L(\mathcal{S}(\tilde{\varphi}) - \mathcal{S}(\varphi_0))$~~

The case for discontinuity

Suppose **(A1)** and T_0 **semi-discrete OT map**,

The case for discontinuity

Suppose **(A1)** and T_0 **semi-discrete OT map**,

$$\mathbb{E}\|\hat{T}_n - T_0\|^2_{L^2(P)} \lesssim n^{-1/2}$$

The case for discontinuity

Suppose **(A1)** and T_0 **semi-discrete OT map**,

$$\mathbb{E}\|\hat{T}_n - T_0\|^2_{L^2(P)} \lesssim n^{-1/2}$$

Surprising! Why? $Q_n \simeq Q$ but P_n is still far from P!

The case for discontinuity

Suppose **(A1)** and T_0 **semi-discrete OT map**,

$$\mathbb{E}\|\hat{T}_n - T_0\|^2_{L^2(P)} \lesssim n^{-1/2}$$

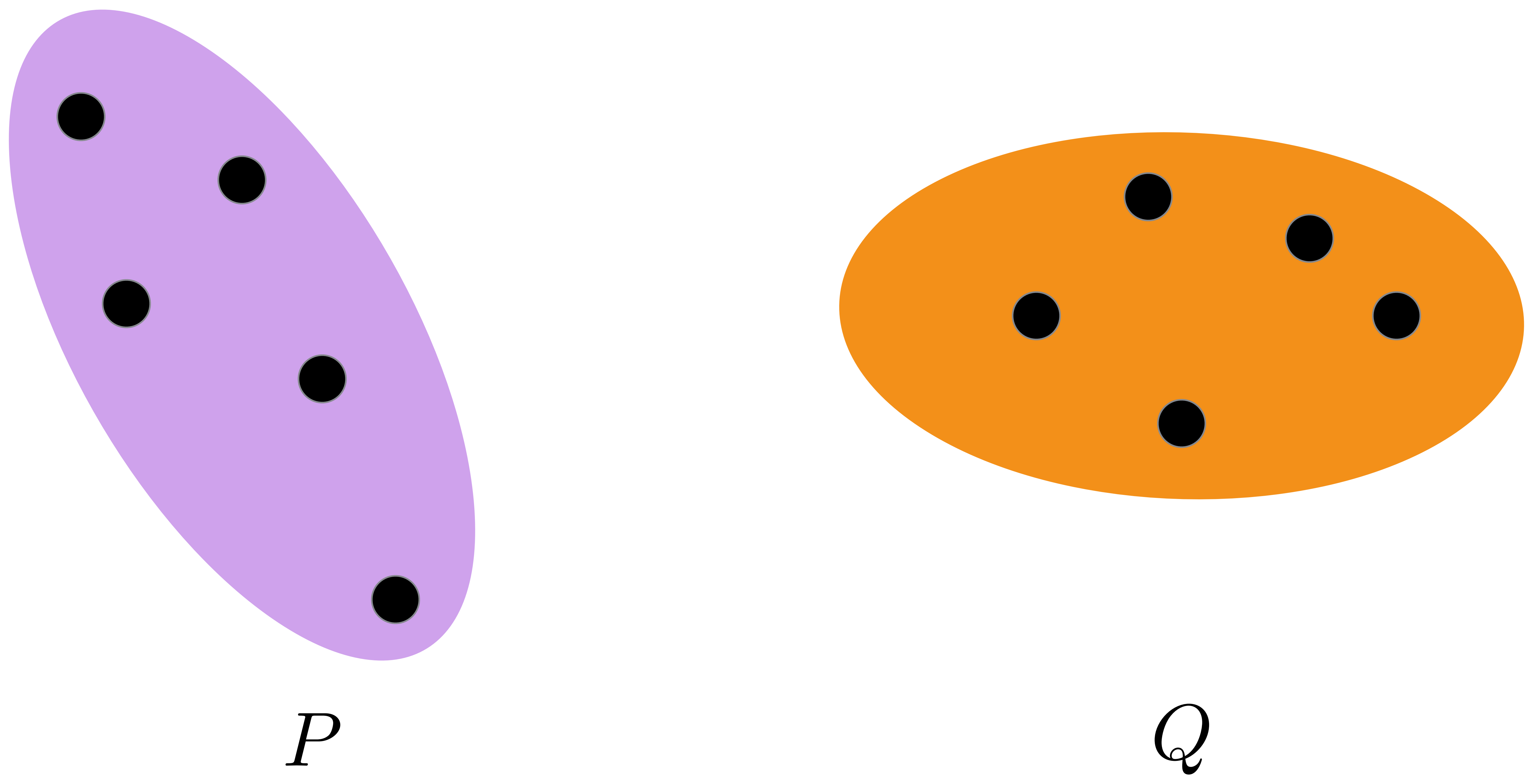
Surprising! Why? $Q_n \simeq Q$ but P_n is still far from P!

Rest of this talk: We will show

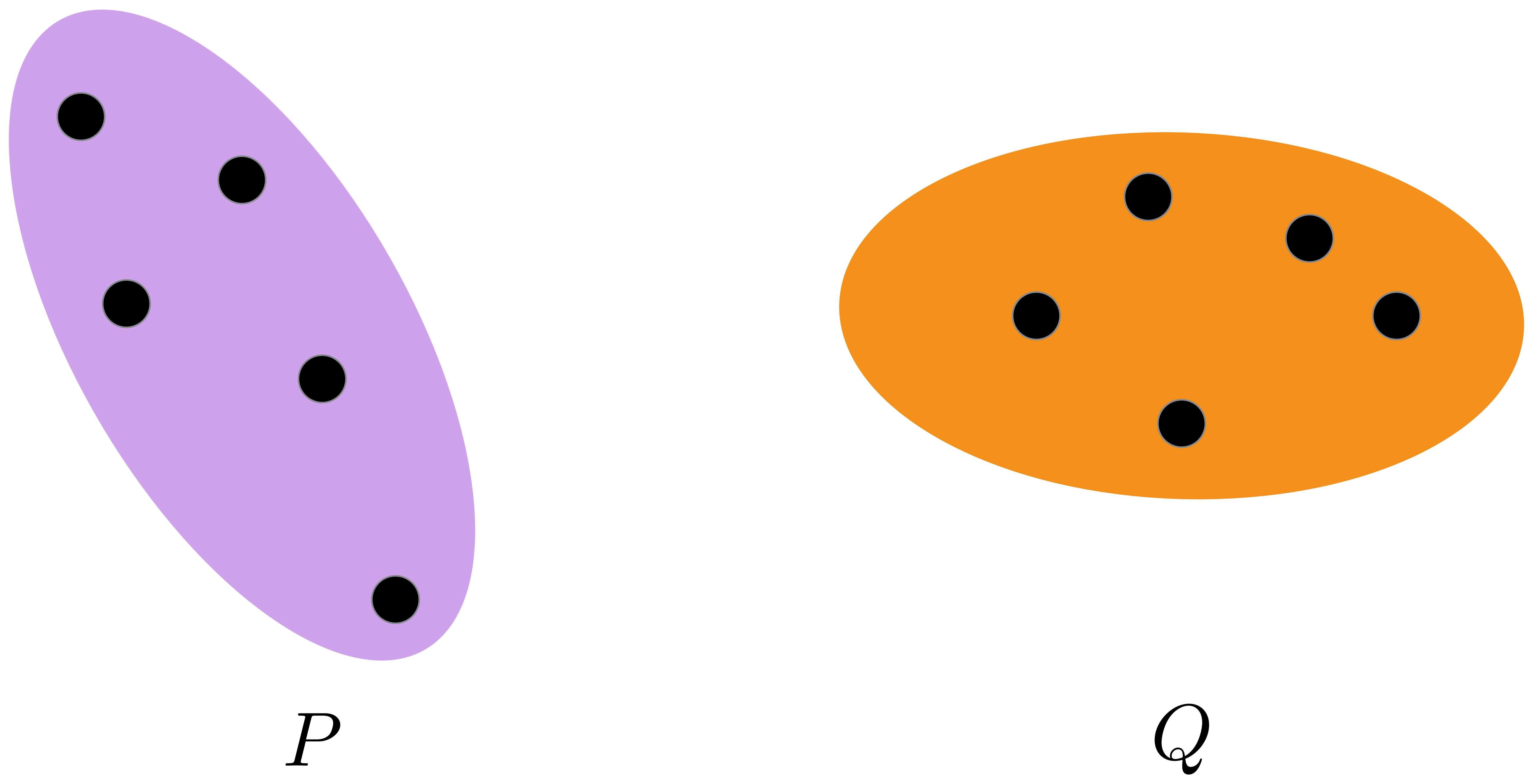
- Manole et al. (2021): 1-Nearest-Neighbor estimator (suffers from c.o.d.!)
- **P.**, Divol, & Niles-Weed (2023): Entropic optimal transport (minimax optimal)

Unpacking the 1NN estimator

Unpacking the 1NN estimator

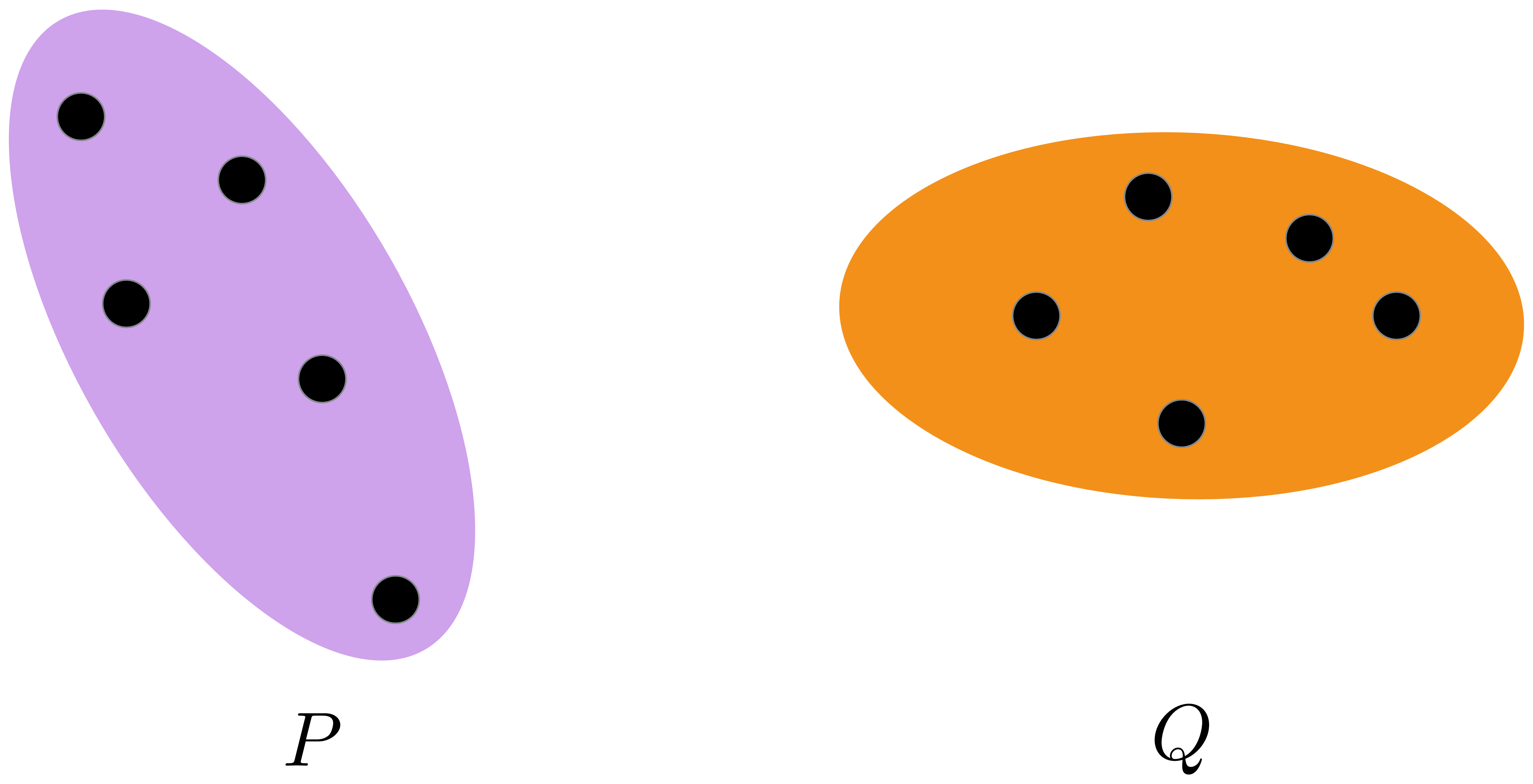


Unpacking the 1NN estimator



Discrete OT: compute $C_{ij} = \|X_i - Y_j\|_2^2$ and solve

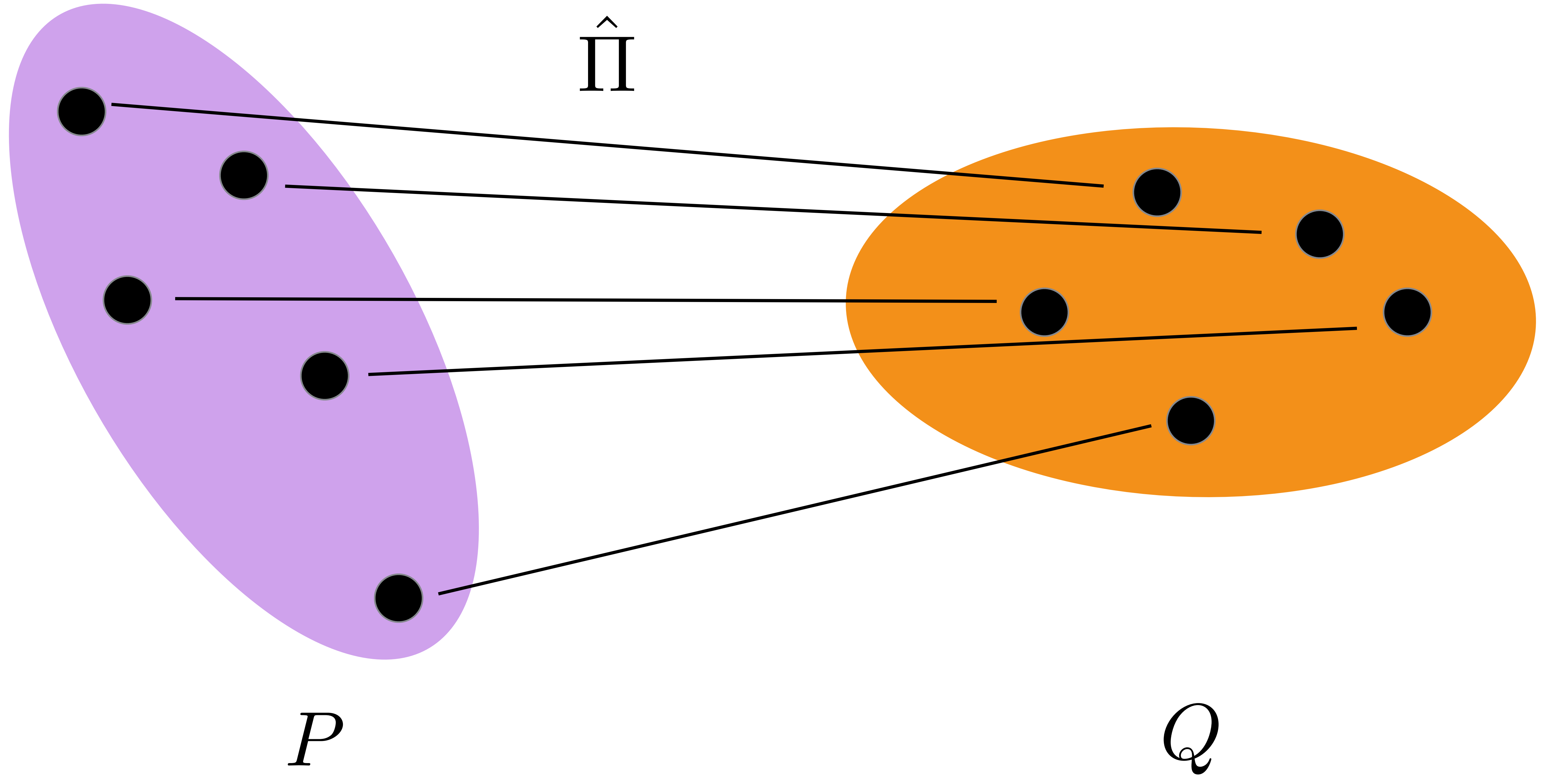
Unpacking the 1NN estimator



Discrete OT: compute $C_{ij} = \|X_i - Y_j\|_2^2$ and solve

$$\min_{\Pi} \langle \Pi, C \rangle \quad \text{s.t.} \quad \Pi \in \mathrm{DS}_n \subseteq \mathbb{R}_+^{n \times n}$$

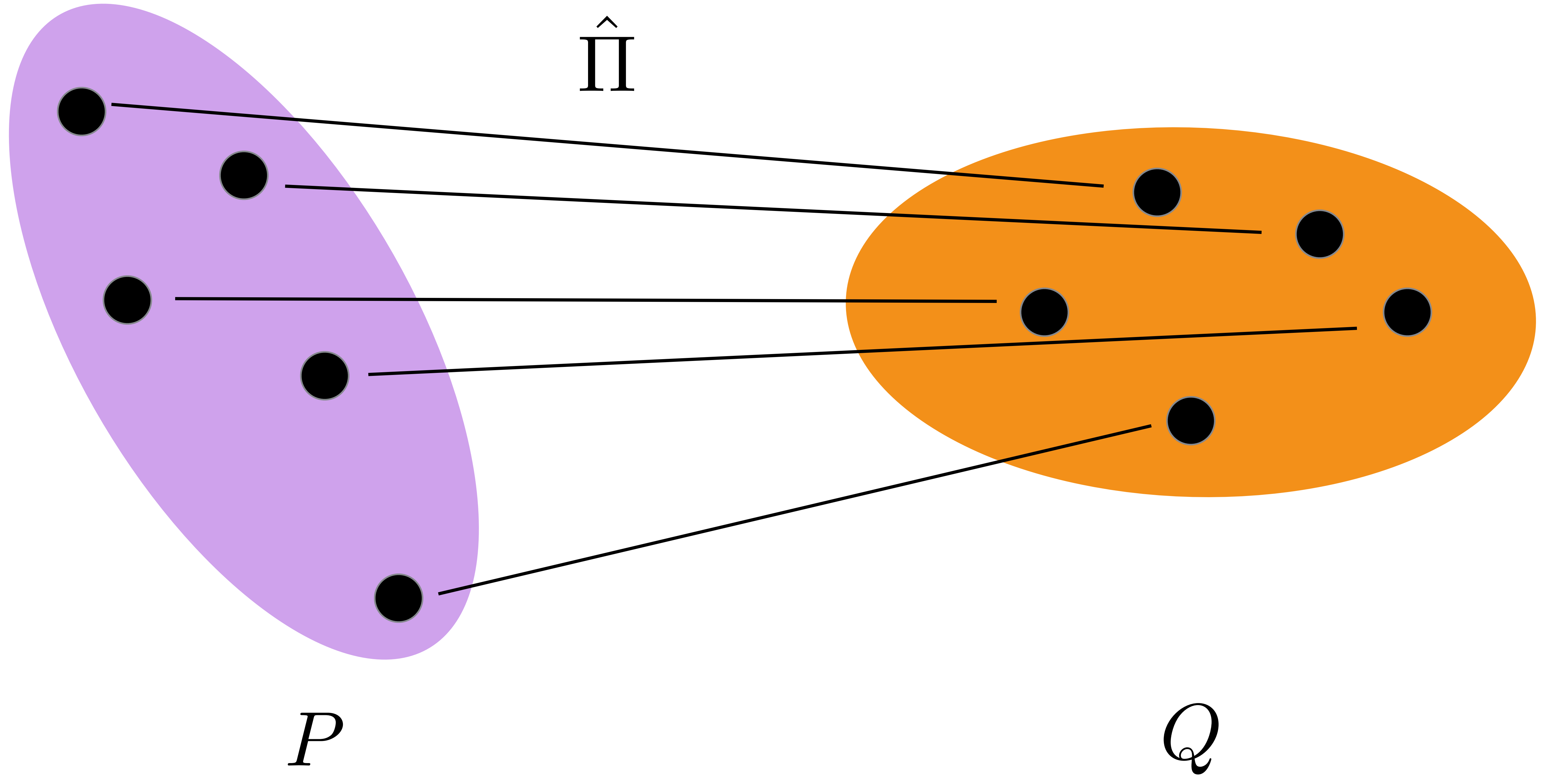
Unpacking the 1NN estimator



Discrete OT: compute $C_{ij} = \|X_i - Y_j\|_2^2$ and solve

$$\hat{\Pi} = \operatorname*{argmin}_{\Pi} \langle \Pi, C \rangle \quad \text{s.t.} \quad \Pi \in \mathrm{DS}_n \subseteq \mathbb{R}_+^{n \times n}$$

Unpacking the 1NN estimator

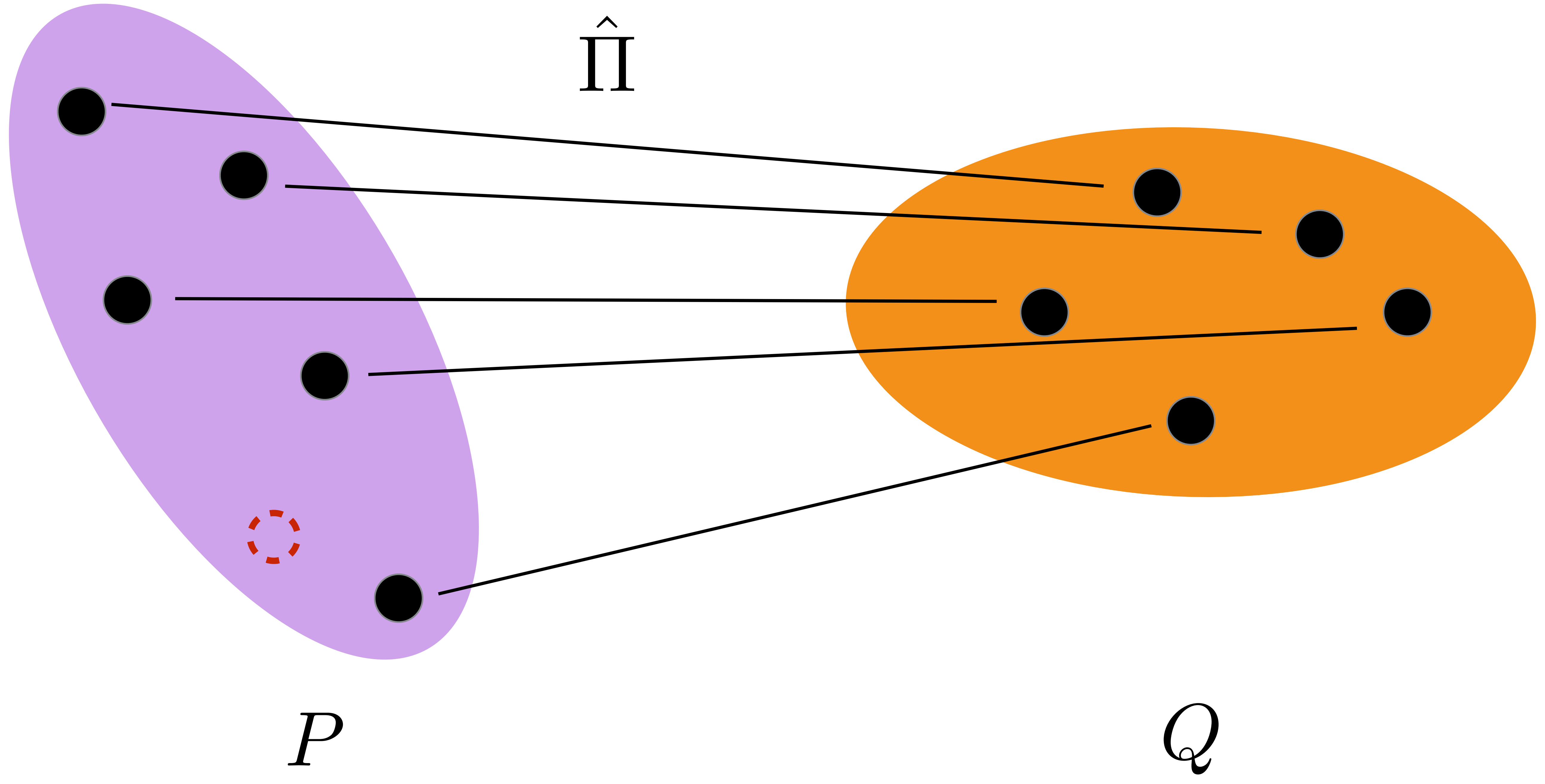


- Need to store $C \in \mathbb{R}_+^{n \times n}$ (i.e. costly)
- Runtime: $O(n^3)$ (i.e. slow)

Discrete OT: compute $C_{ij} = \|X_i - Y_j\|_2^2$ and solve

$$\hat{\Pi} = \underset{\Pi}{\operatorname{argmin}} \langle \Pi, C \rangle \quad \text{s.t.} \quad \Pi \in \mathrm{DS}_n \subseteq \mathbb{R}_+^{n \times n}$$

Unpacking the 1NN estimator

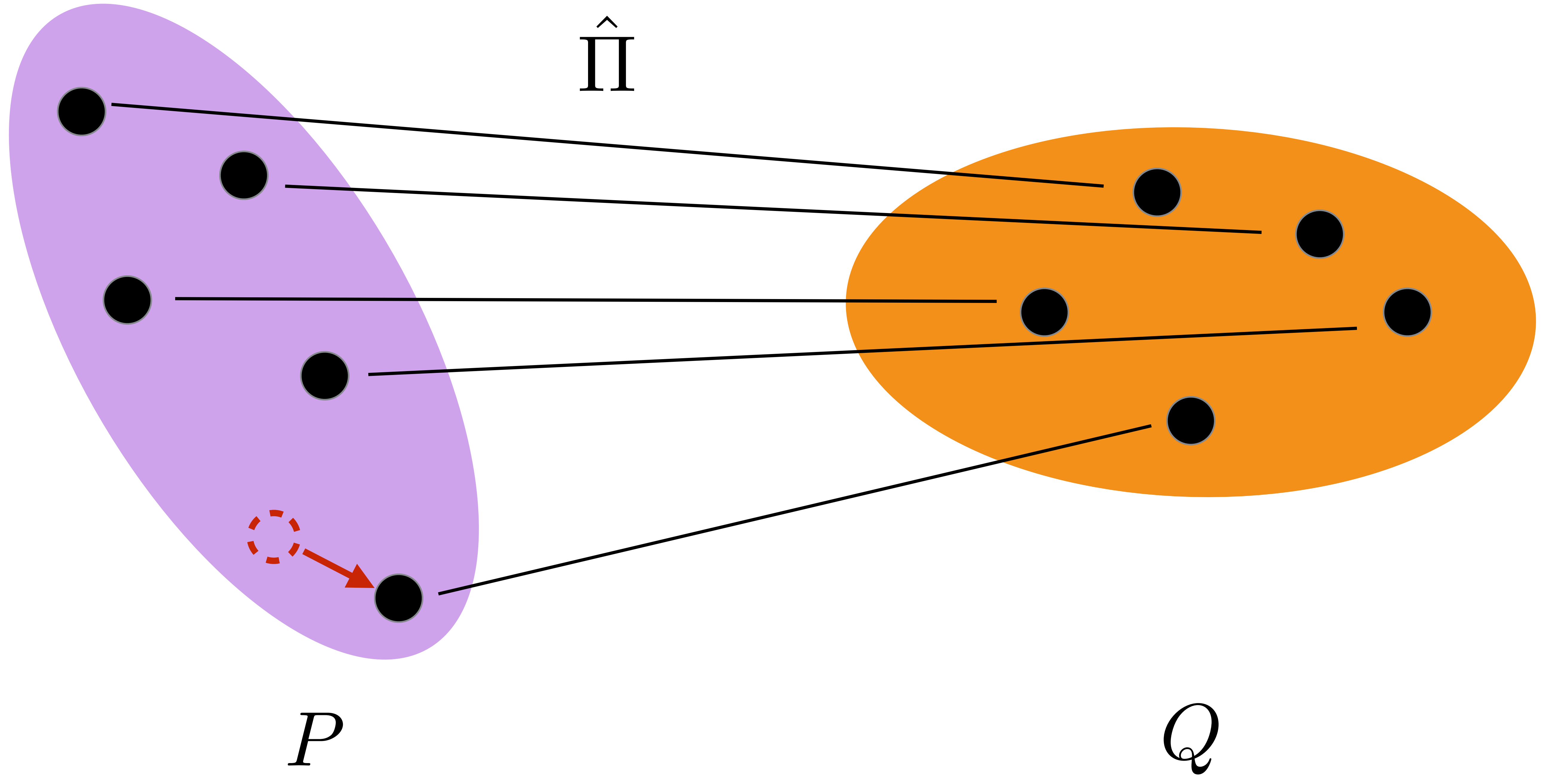


- Need to store $C \in \mathbb{R}_+^{n \times n}$ (i.e. costly)
- Runtime: $O(n^3)$ (i.e. slow)

Discrete OT: compute $C_{ij} = \|X_i - Y_j\|_2^2$ and solve

$$\hat{\Pi} = \operatorname*{argmin}_{\Pi} \langle \Pi, C \rangle \quad \text{s.t.} \quad \Pi \in \mathrm{DS}_n \subseteq \mathbb{R}_+^{n \times n}$$

Unpacking the 1NN estimator

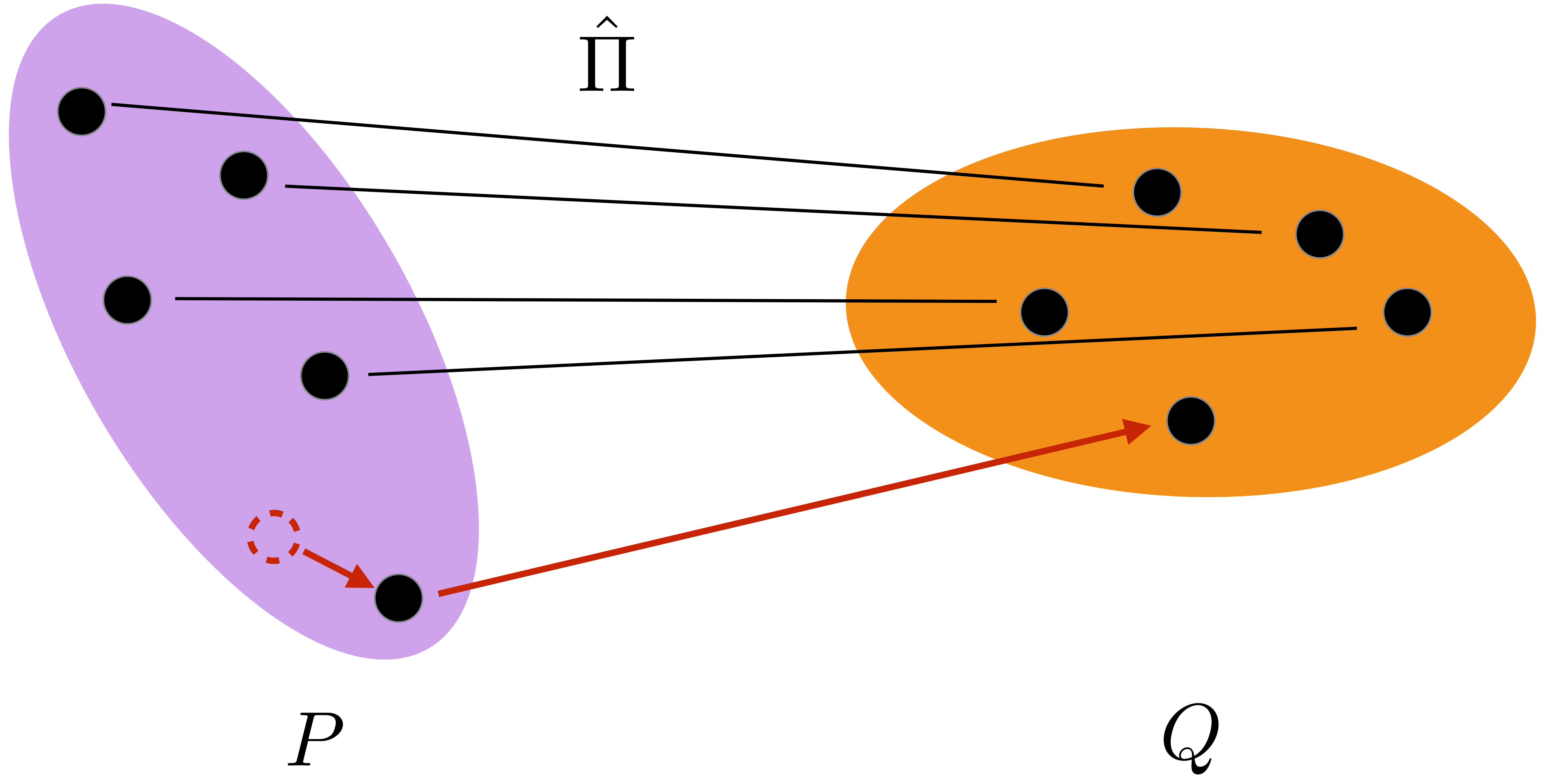


- Need to store $C \in \mathbb{R}_+^{n \times n}$ (i.e. costly)
- Runtime: $O(n^3)$ (i.e. slow)

Discrete OT: compute $C_{ij} = \|X_i - Y_j\|_2^2$ and solve

$$\hat{\Pi} = \operatorname*{argmin}_{\Pi} \langle \Pi, C \rangle \quad \text{s.t.} \quad \Pi \in \mathrm{DS}_n \subseteq \mathbb{R}_+^{n \times n}$$

Unpacking the 1NN estimator

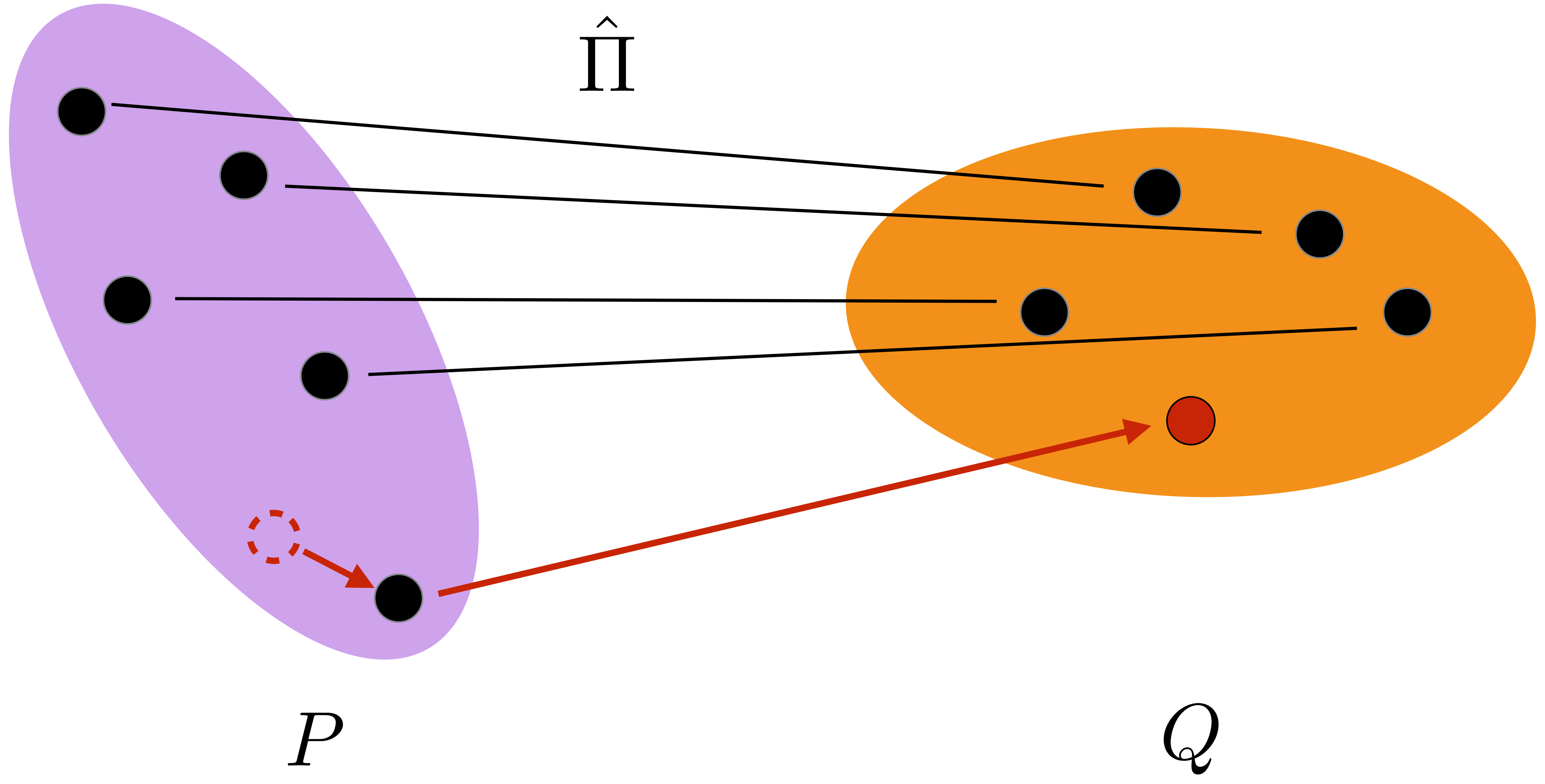


- Need to store $C \in \mathbb{R}_+^{n \times n}$ (i.e. costly)
- Runtime: $O(n^3)$ (i.e. slow)

Discrete OT: compute $C_{ij} = \|X_i - Y_j\|_2^2$ and solve

$$\hat{\Pi} = \underset{\Pi}{\operatorname{argmin}} \; \langle \Pi, C \rangle \quad \text{s.t.} \quad \Pi \in \mathrm{DS}_n \subseteq \mathbb{R}_+^{n \times n}$$

Unpacking the 1NN estimator

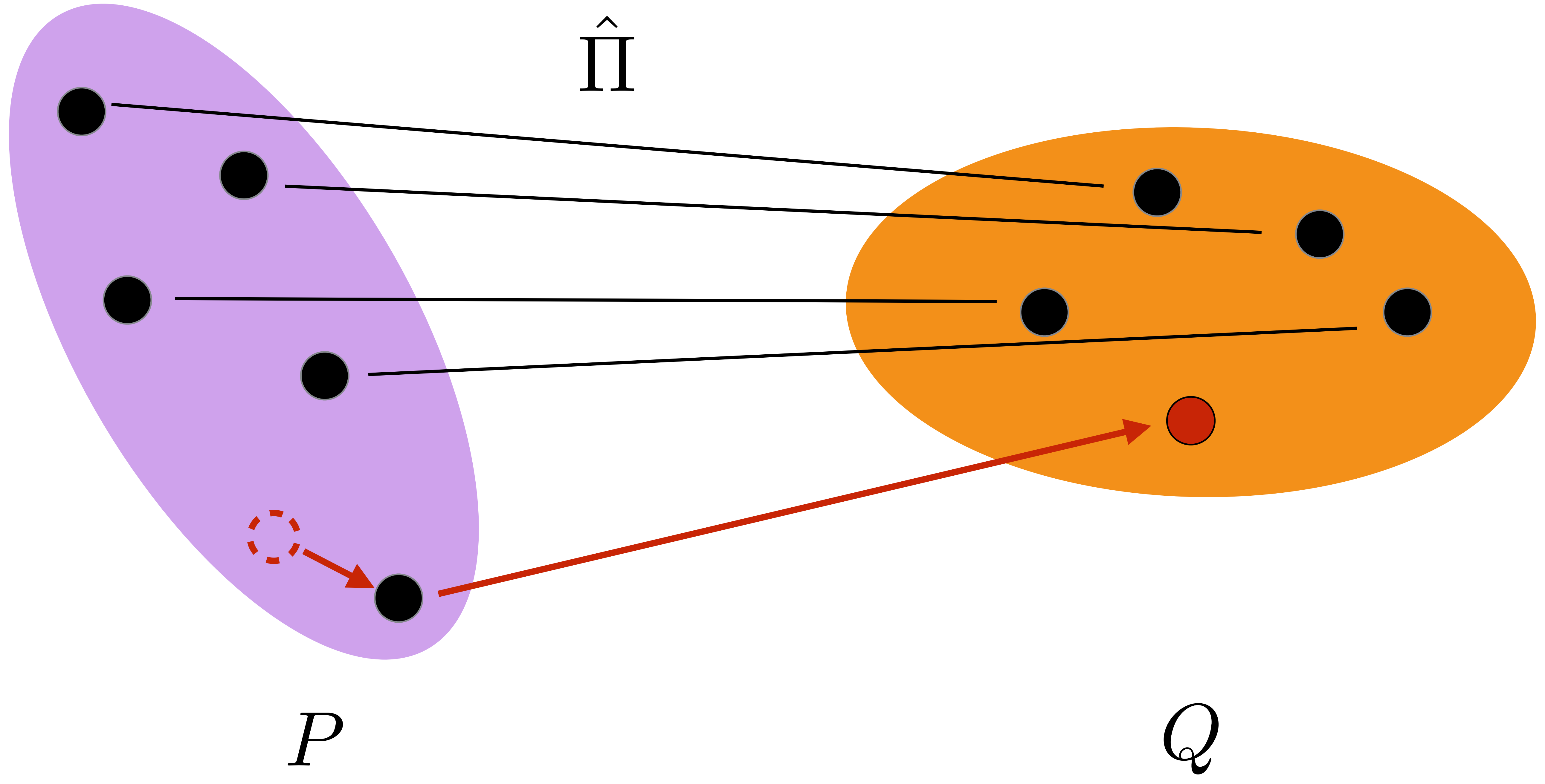


- Need to store $C \in \mathbb{R}_+^{n\times n}$ (i.e. costly)
- Runtime: $O(n^3)$ (i.e. slow)

Discrete OT: compute $C_{ij} = \|X_i - Y_j\|_2^2$ and solve

$$\hat{\Pi} = \operatorname*{argmin}_{\Pi} \langle \Pi, C \rangle \quad \text{s.t.} \quad \Pi \in \mathrm{DS}_n \subseteq \mathbb{R}_+^{n\times n}$$

Unpacking the 1NN estimator



- Need to store $C \in \mathbb{R}_+^{n \times n}$ (i.e. costly)
- Runtime: $O(n^3)$ (i.e. slow)
- Estimator only exists <u>in sample</u>

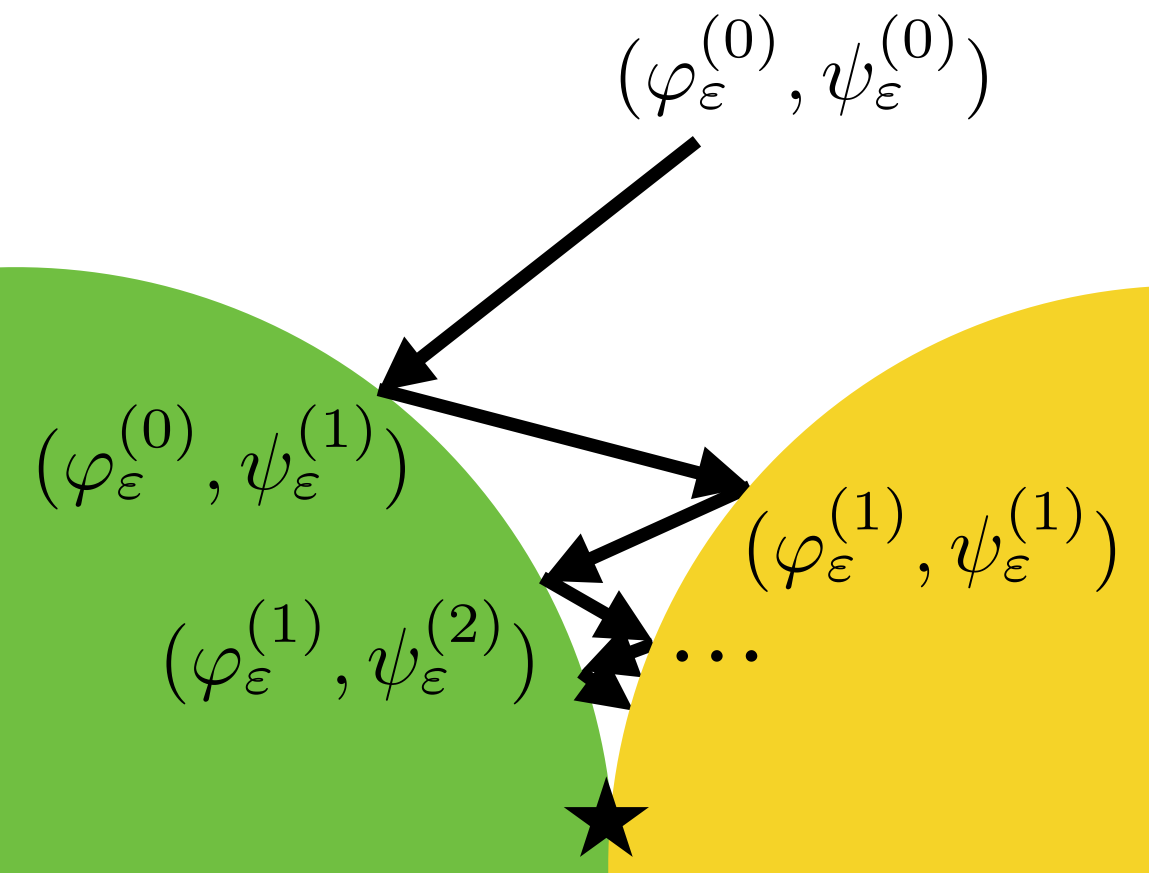
Discrete OT: compute $C_{ij} = \|X_i - Y_j\|_2^2$ and solve

$$\hat{\Pi} = \underset{\Pi}{\operatorname{argmin}} \, \langle \Pi, C \rangle \quad \text{s.t.} \quad \Pi \in \mathrm{DS}_n \subseteq \mathbb{R}_+^{n \times n}$$

Our estimator: entropic Brenier map

Our estimator: entropic Brenier map

Sinkhorn's algorithm: Iteratively fit marginals on the data $(X_i, Y_j)_{i,j=1}^n$

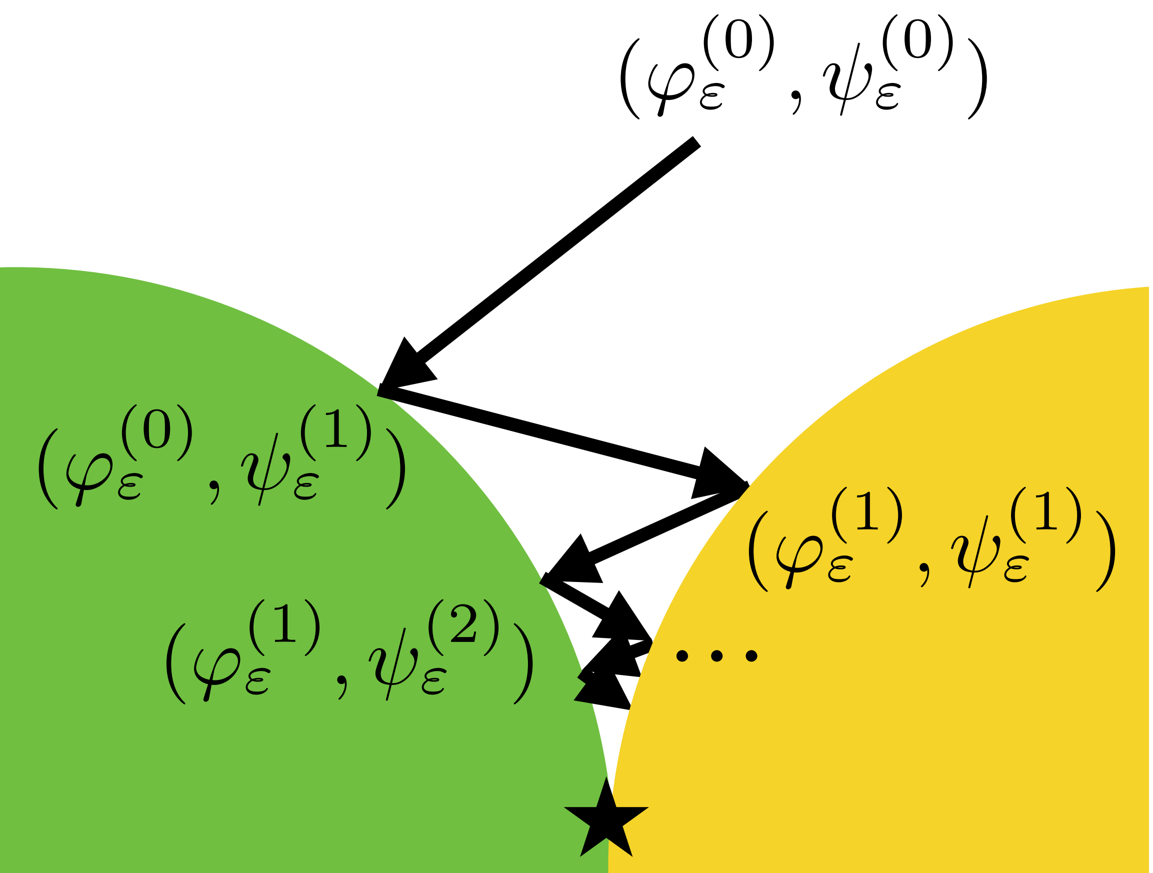


Our estimator: entropic Brenier map

Sinkhorn's algorithm: Iteratively fit marginals on the data $(X_i, Y_j)_{i,j=1}^n$

At optimality: on the data, $(\Pi_\varepsilon)_{ij} = \exp(X_i^\top Y_j - (\hat{\varphi}_\varepsilon)_i - (\hat{\psi}_\varepsilon)_j)$

with the fixed-point relationship



Our estimator: entropic Brenier map

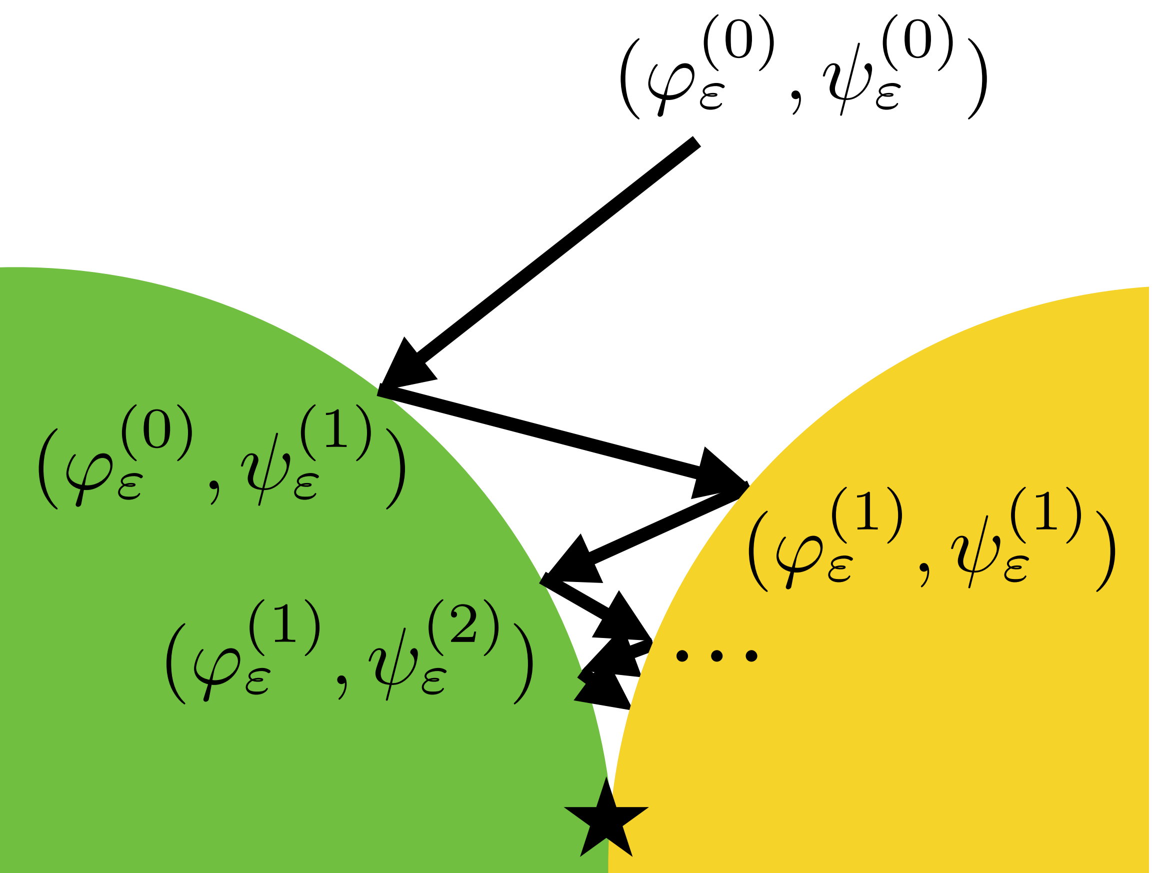
Sinkhorn's algorithm: Iteratively fit marginals on the data $(X_i, Y_j)_{i,j=1}^n$

At optimality: on the data, $(\Pi_\varepsilon)_{ij} = \exp(X_i^\top Y_j - (\hat{\varphi}_\varepsilon)_i - (\hat{\psi}_\varepsilon)_j)$

with the fixed-point relationship

$$(\hat{\varphi}_\varepsilon)_i = \varepsilon \log\left(\frac{1}{n}\sum_{j=1}^n \exp(X_i^\top Y_j - (\hat{\psi}_\varepsilon)_j)/\varepsilon\right)$$

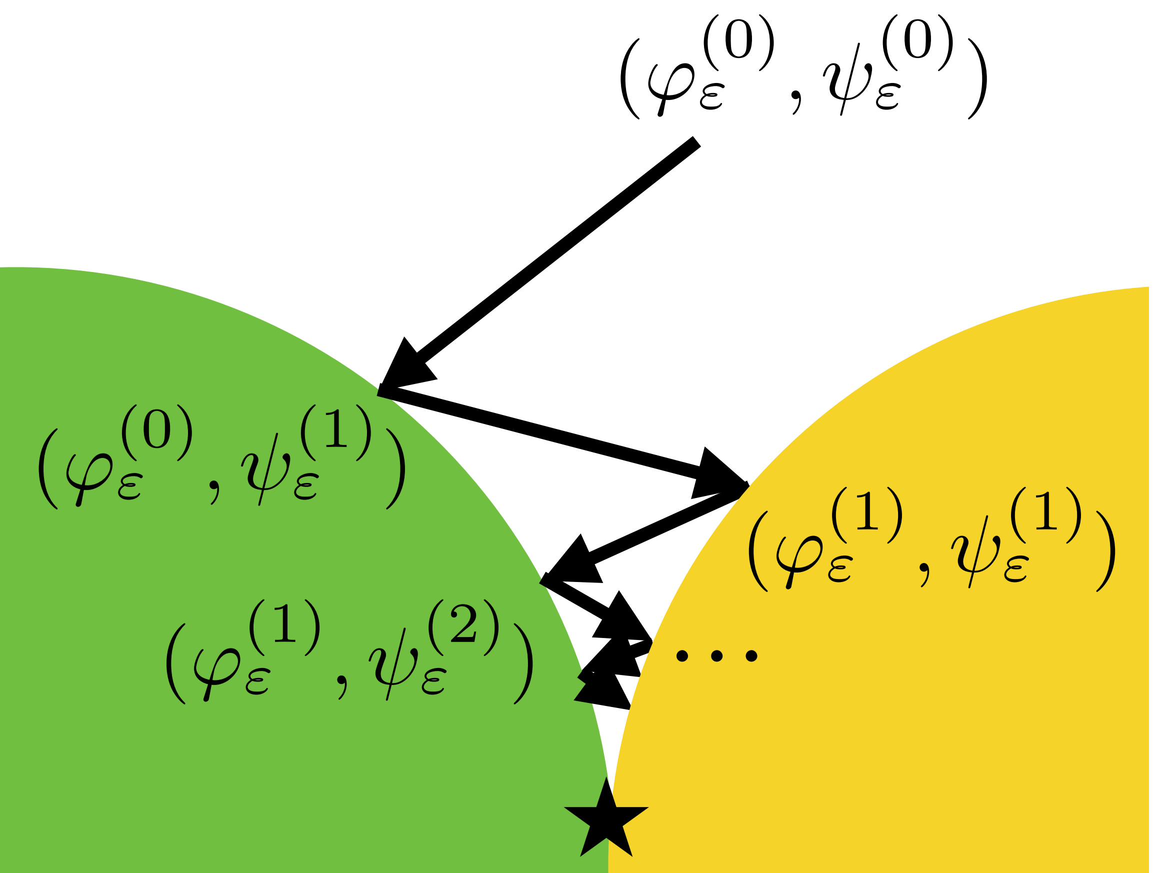
$$(\hat{\psi}_\varepsilon)_j = \varepsilon \log\left(\frac{1}{n}\sum_{i=1}^n \exp(X_i^\top Y_j - (\hat{\varphi}_\varepsilon)_i)/\varepsilon\right)$$



Our estimator: entropic Brenier map

Sinkhorn's algorithm: Iteratively fit marginals on the data $(X_i, Y_j)_{i,j=1}^n$

At optimality: Can extend to functions OFF the data [NW21, MNW19]



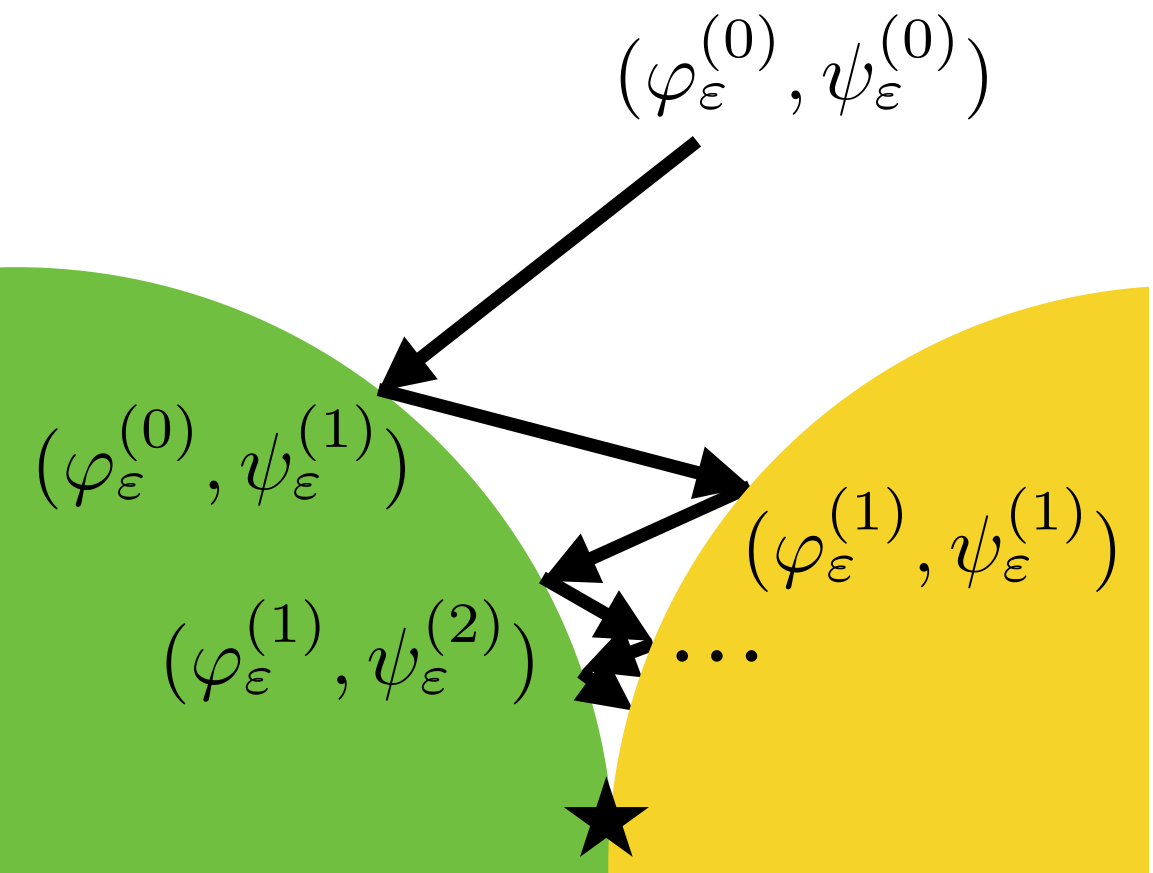
Our estimator: entropic Brenier map

Sinkhorn's algorithm: Iteratively fit marginals on the data $(X_i, Y_j)_{i,j=1}^n$

At optimality: Can extend to functions OFF the data [NW21, MNW19]

$$\hat{\varphi}_\varepsilon(x) = \varepsilon \log \left(\frac{1}{n} \sum_{j=1}^n \exp(x^\top Y_j - (\hat{\psi}_\varepsilon)_j)/\varepsilon \right)$$

$$\hat{\psi}_\varepsilon(y) = \varepsilon \log \left(\frac{1}{n} \sum_{i=1}^n \exp(X_i^\top y - (\hat{\varphi}_\varepsilon)_i)/\varepsilon \right)$$



Our estimator: entropic Brenier map

Our estimator: entropic Brenier map

Recall that we want to estimate a Brenier map $T_0 = \nabla\varphi_0$

Our estimator: entropic Brenier map

Recall that we want to estimate a Brenier map $T_0 = \nabla \varphi_0$

Given $\hat{\varphi}_\varepsilon(x) = \varepsilon \log \left(\frac{1}{n} \sum_{j=1}^{n} \exp(x^\top Y_j - (\hat{\psi}_\varepsilon)_j)/\varepsilon \right)$, our estimator is

Our estimator: entropic Brenier map

Recall that we want to estimate a Brenier map $T_0 = \nabla \varphi_0$

Given $\hat{\varphi}_\varepsilon(x) = \varepsilon \log\left(\frac{1}{n}\sum_{j=1}^n \exp(x^\top Y_j - (\hat{\psi}_\varepsilon)_j)/\varepsilon\right)$, our estimator is

$$\hat{T}_{(\varepsilon,n)}(x) := \nabla \hat{\varphi}_\varepsilon(x) = \sum_{i=1}^n Y_i \frac{e^{(x^\top Y_i - (\hat{\psi}_\varepsilon)_i)/\varepsilon}}{\sum_{k=1}^n e^{(x^\top Y_k - (\hat{\psi}_\varepsilon)_k)/\varepsilon}}$$

Our estimator: entropic Brenier map

Recall that we want to estimate a Brenier map $T_0 = \nabla \varphi_0$

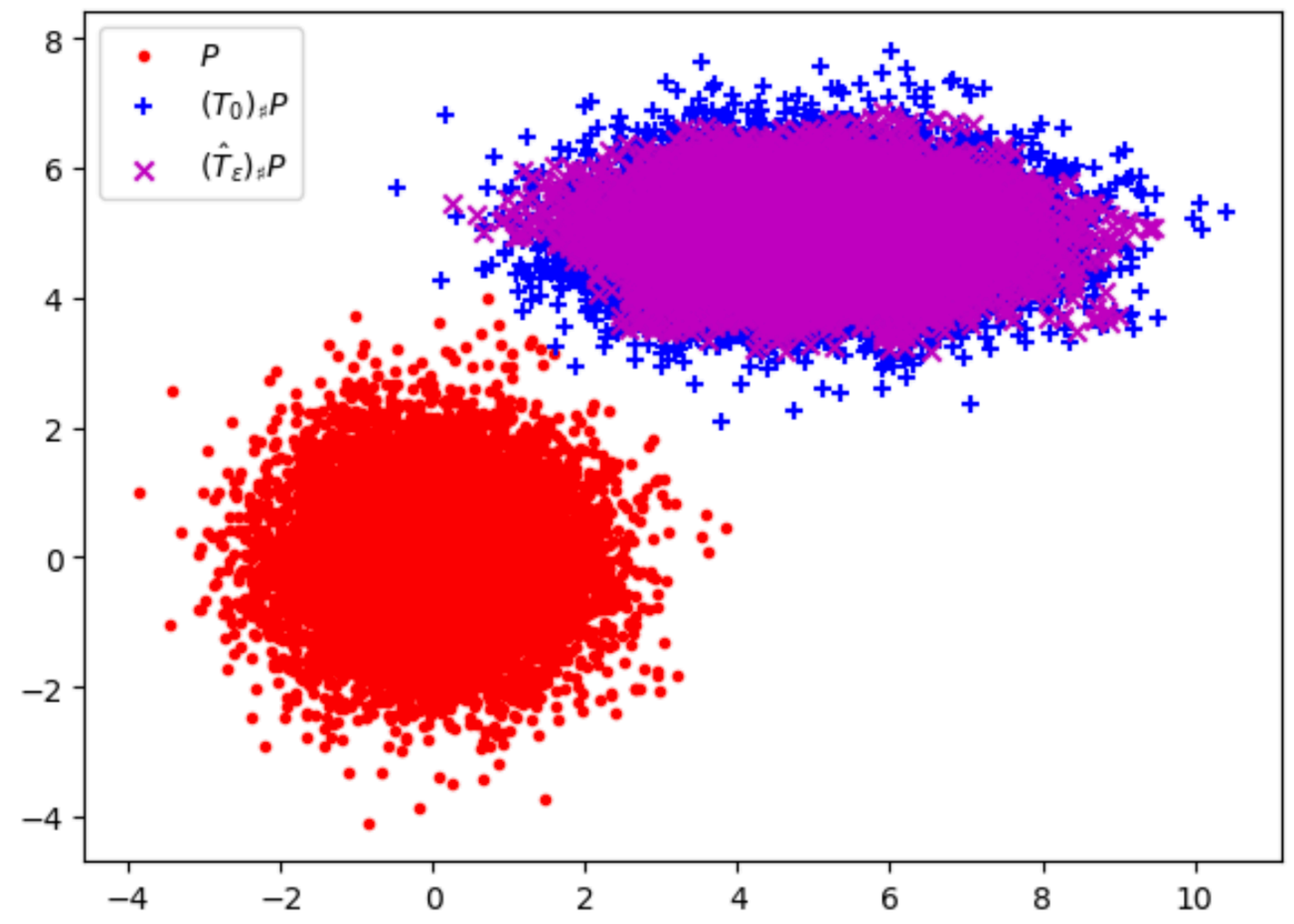
Given $\hat{\varphi}_\varepsilon(x) = \varepsilon \log\left(\frac{1}{n}\sum_{j=1}^n \exp(x^\top Y_j - (\hat{\psi}_\varepsilon)_j)/\varepsilon\right)$, our estimator is

$$\hat{T}_{(\varepsilon,n)}(x) := \nabla\hat{\varphi}_\varepsilon(x) = \sum_{i=1}^n Y_i \frac{e^{(x^\top Y_i - (\hat{\psi}_\varepsilon)_i)/\varepsilon}}{\sum_{k=1}^n e^{(x^\top Y_k - (\hat{\psi}_\varepsilon)_k)/\varepsilon}} \in \mathrm{conv}(\{Y_1, \ldots, Y_n\})$$

2D Visualizations

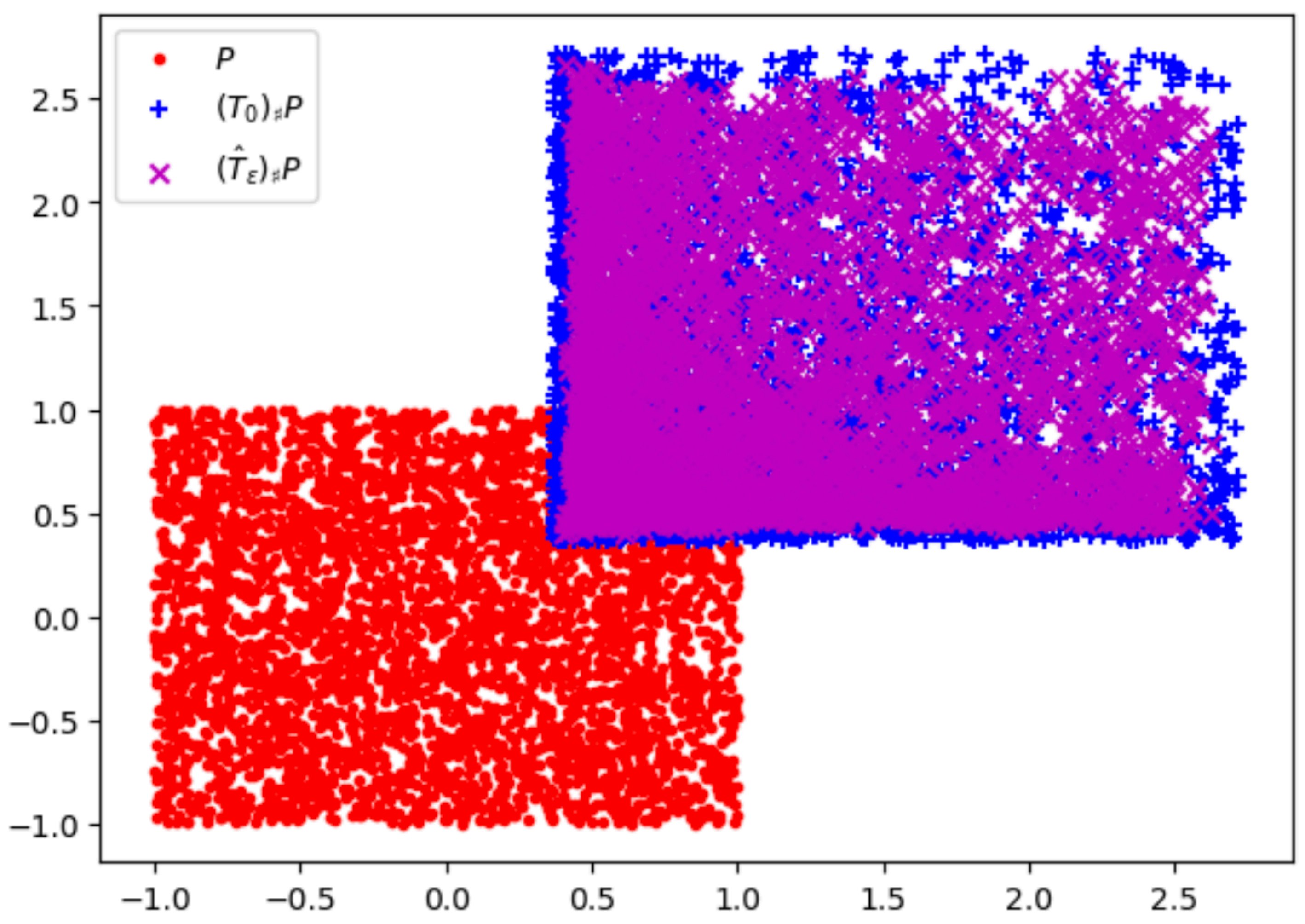
$$P = \mathcal{N}(0, I_2)$$

$$T_0(x) = \Sigma^{1/2}x + a$$



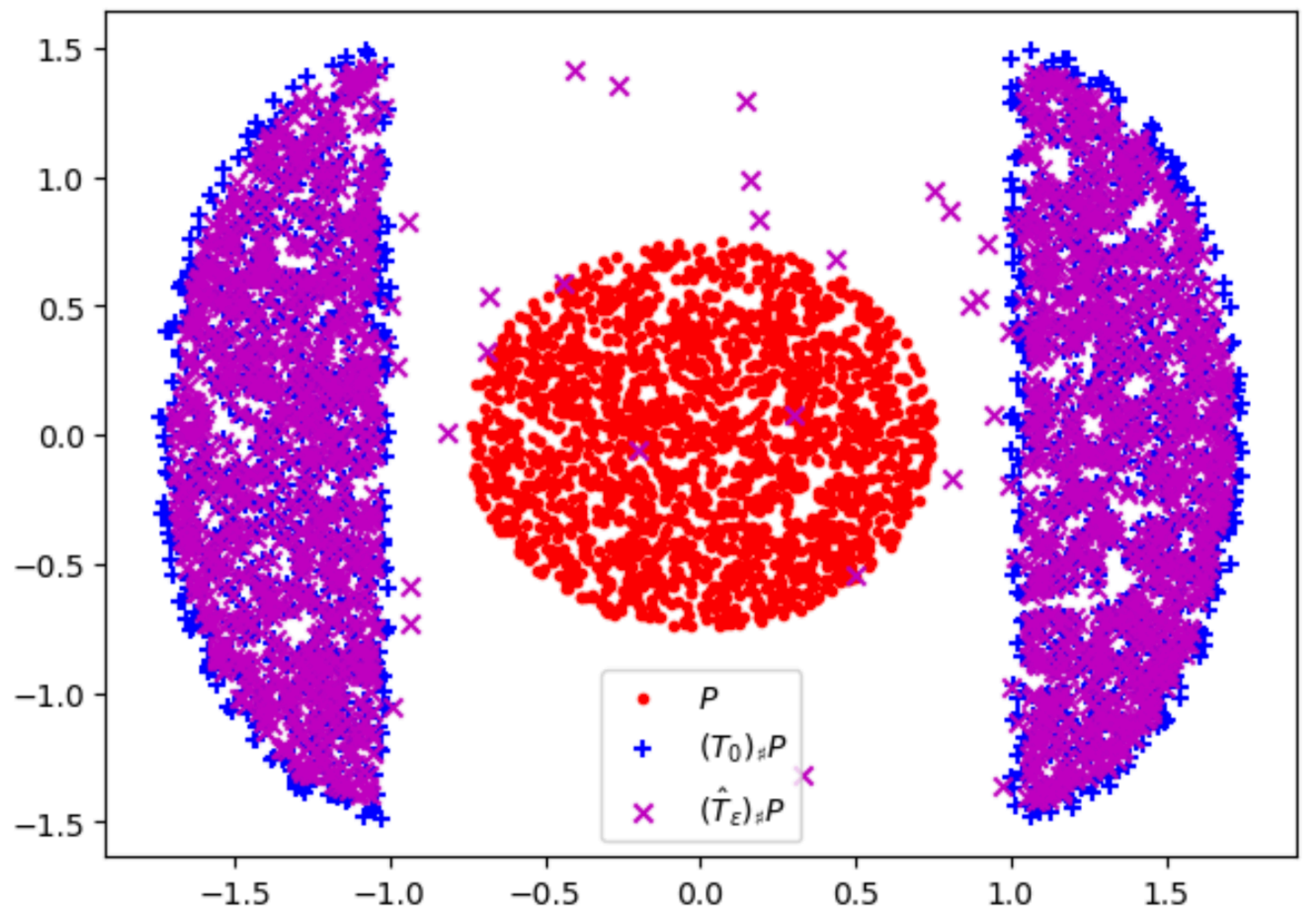
$$P = \mathrm{Unif}([-1, 1]^2)$$

$$T_0(x) = \exp(x)$$



$$P = \mathrm{Unif}(B(0; 1))$$

$$T_0(x) = x + 2\mathrm{sign}(x_1)e_1$$



Our estimator: entropic Brenier map

Recall that we want to estimate a Brenier map $T_0 = \nabla \varphi_0$

Given $\hat{\varphi}_\varepsilon(x) = \varepsilon \log\left(\frac{1}{n}\sum_{j=1}^n \exp(x^\top Y_j - (\hat{\psi}_\varepsilon)_j)/\varepsilon\right)$, our estimator is

$$\hat{T}_{(\varepsilon,n)}(x) := \nabla \hat{\varphi}_\varepsilon(x) = \sum_{i=1}^n Y_i \frac{e^{(x^\top Y_i - (\hat{\psi}_\varepsilon)_i)/\varepsilon}}{\sum_{k=1}^n e^{(x^\top Y_k - (\hat{\psi}_\varepsilon)_k)/\varepsilon}} \in \operatorname{conv}(\{Y_1, \ldots, Y_n\})$$

We will show this is **minimax optimal** in the semi-discrete setting

Semi-discrete optimal transport

P satisfies $\mathbf{(A1)}$ and Q satisfies $\mathbf{(A2)}$

Semi-discrete optimal transport

P satisfies **(A1)** and Q satisfies **(A2)**

(A1) P has density $0 < p_{\min} \leq p(x) \leq p_{\max}$ with convex support $\mathrm{supp}(P) \subseteq B(0; R)$

(A2) For $J \in \mathbb{N}$, $Q = \sum_{j=1}^{J} q_j \delta_{y_j}$, with $\{y_1, \ldots, y_J\} \subseteq B(0; R)$ and $q_j \geq q_{\min} > 0$

Semi-discrete optimal transport

P is "nice" and Q is discrete

Semi-discrete optimal transport

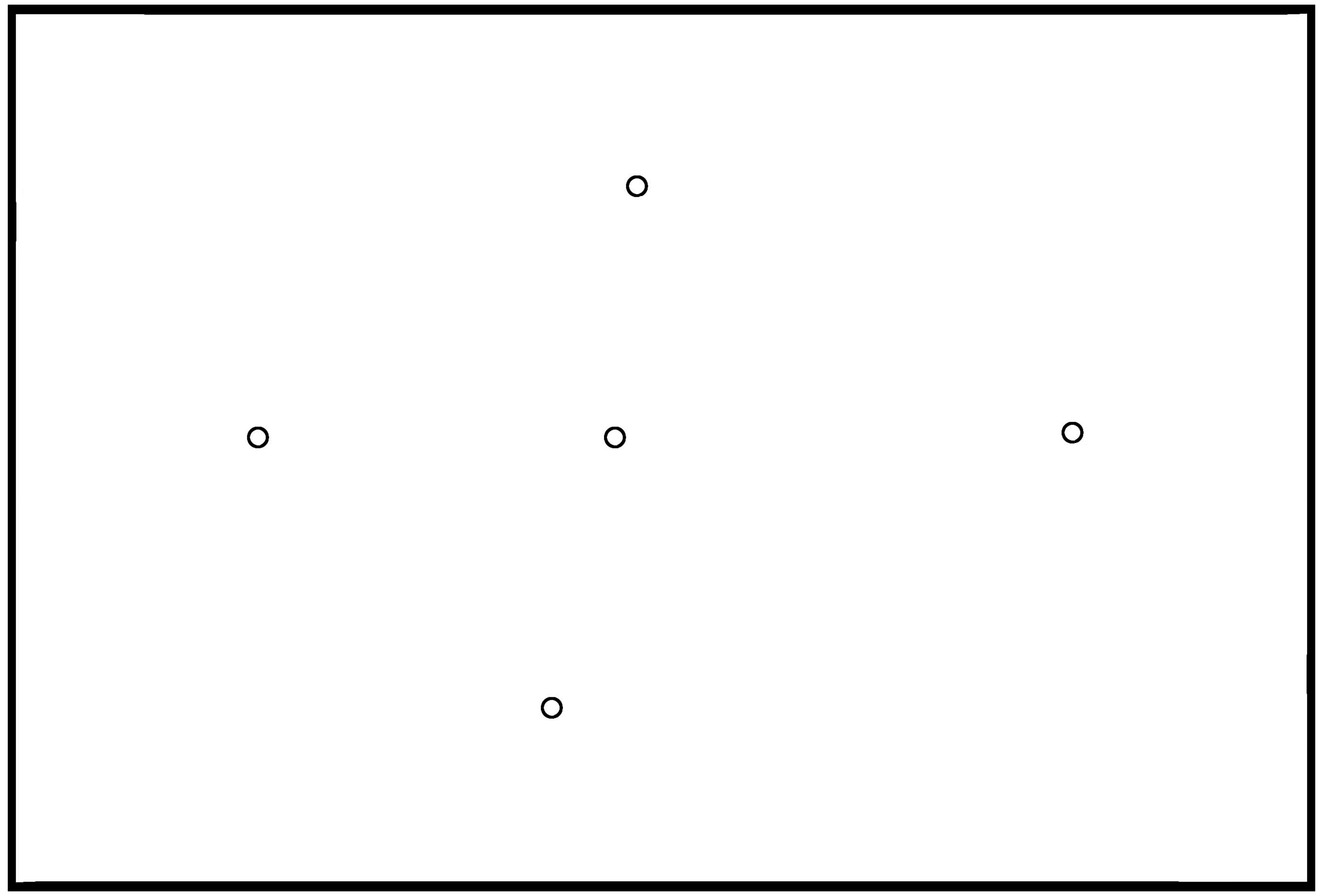
P is "nice" and Q is discrete

Considered in several works

- <u>Computation:</u> Kitagawa et al. (2019), Genevay et al. (2016), etc
- <u>Statistical aspects:</u> del Barrio et al. (2022), Hundreiser et al (2022), etc
- <u>Entropic semi-discrete OT:</u> Altschuler et al. (2021+), Delalande (2021)

Semi-discrete optimal transport

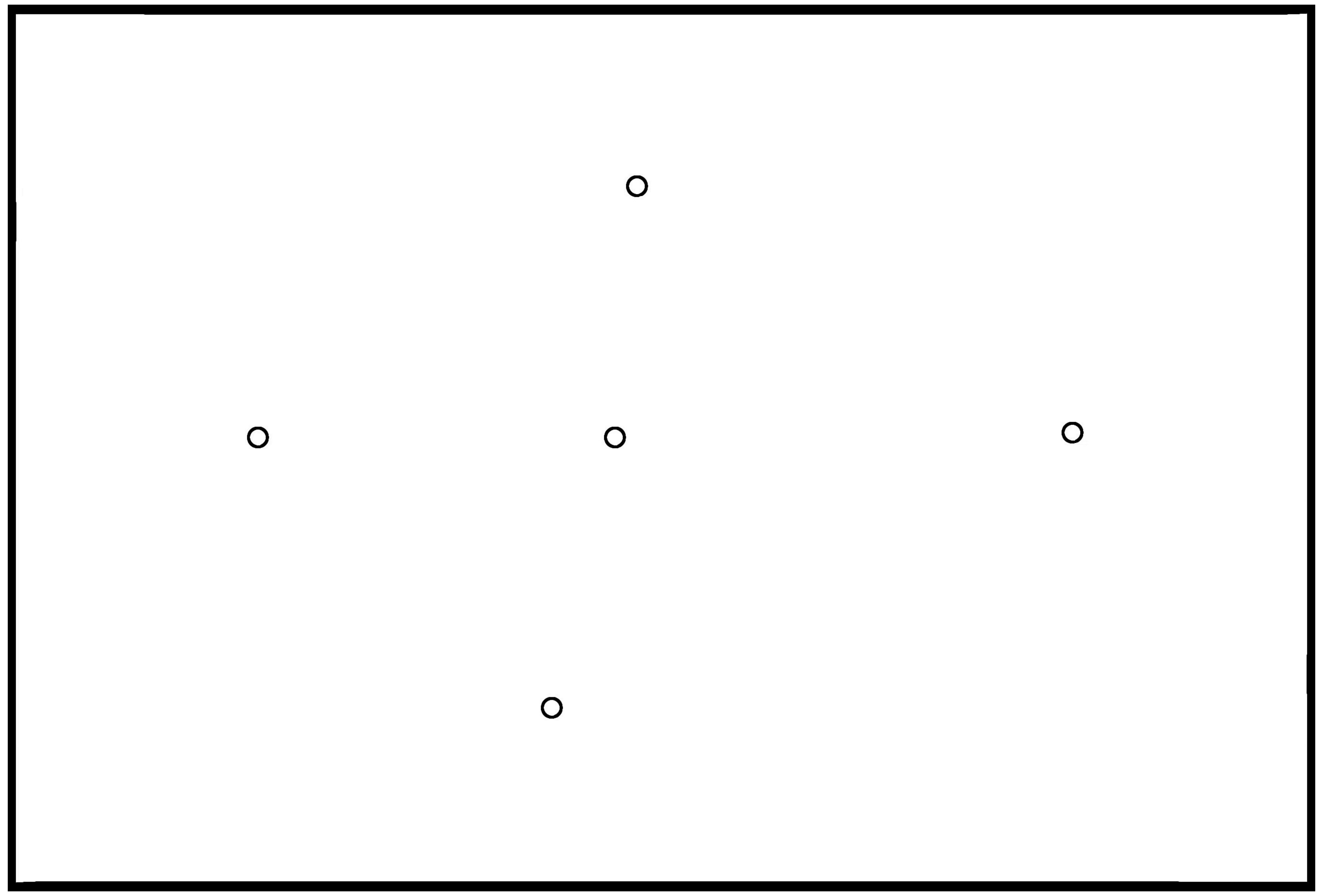
P is "nice" and Q is discrete



Semi-discrete optimal transport

P is "nice" and Q is discrete

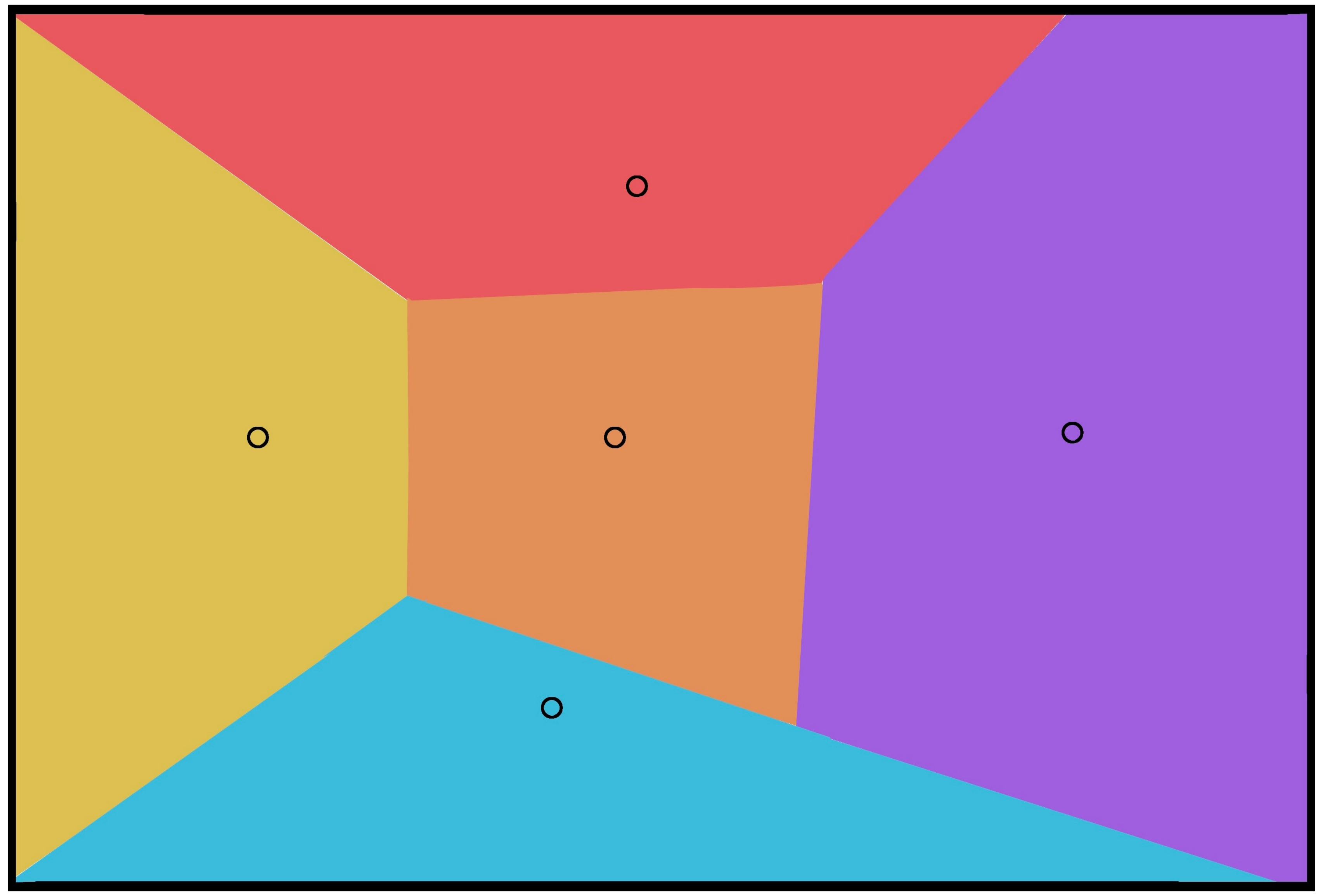
What is the optimal transport map?



Semi-discrete optimal transport

P is "nice" and Q is discrete

What is the optimal transport map?



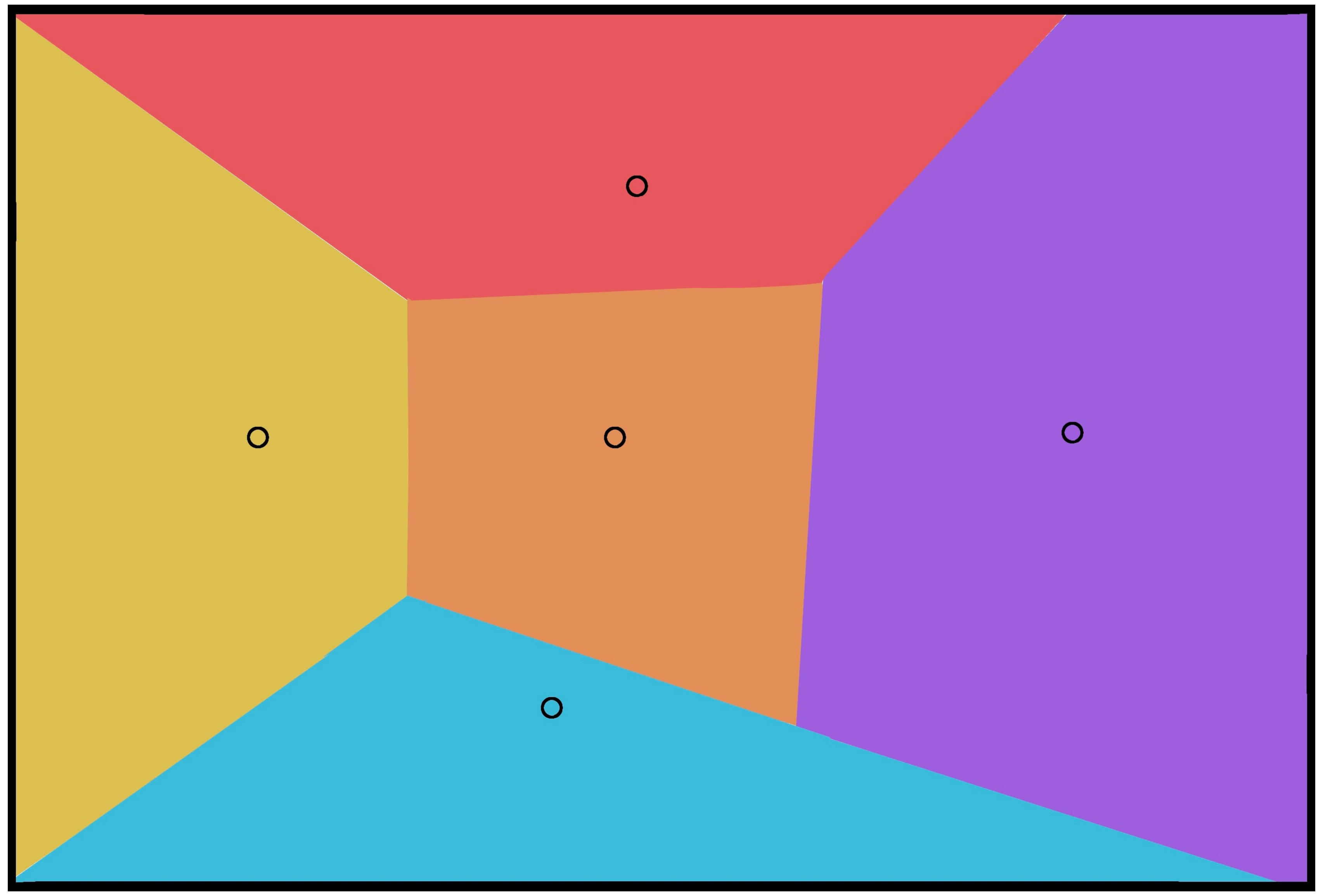
Semi-discrete optimal transport

P is "nice" and Q is discrete

What is the optimal transport map?

$$T_0(x) = \nabla \varphi_0(x) = \underset{j \in [J]}{\operatorname{argmax}} \{x^\top y_j - (\psi_0)_j\}$$

where $\psi_0 \in \mathbb{R}^J$ represents the boundaries of the *Laguerre cells*



Semi-discrete optimal transport

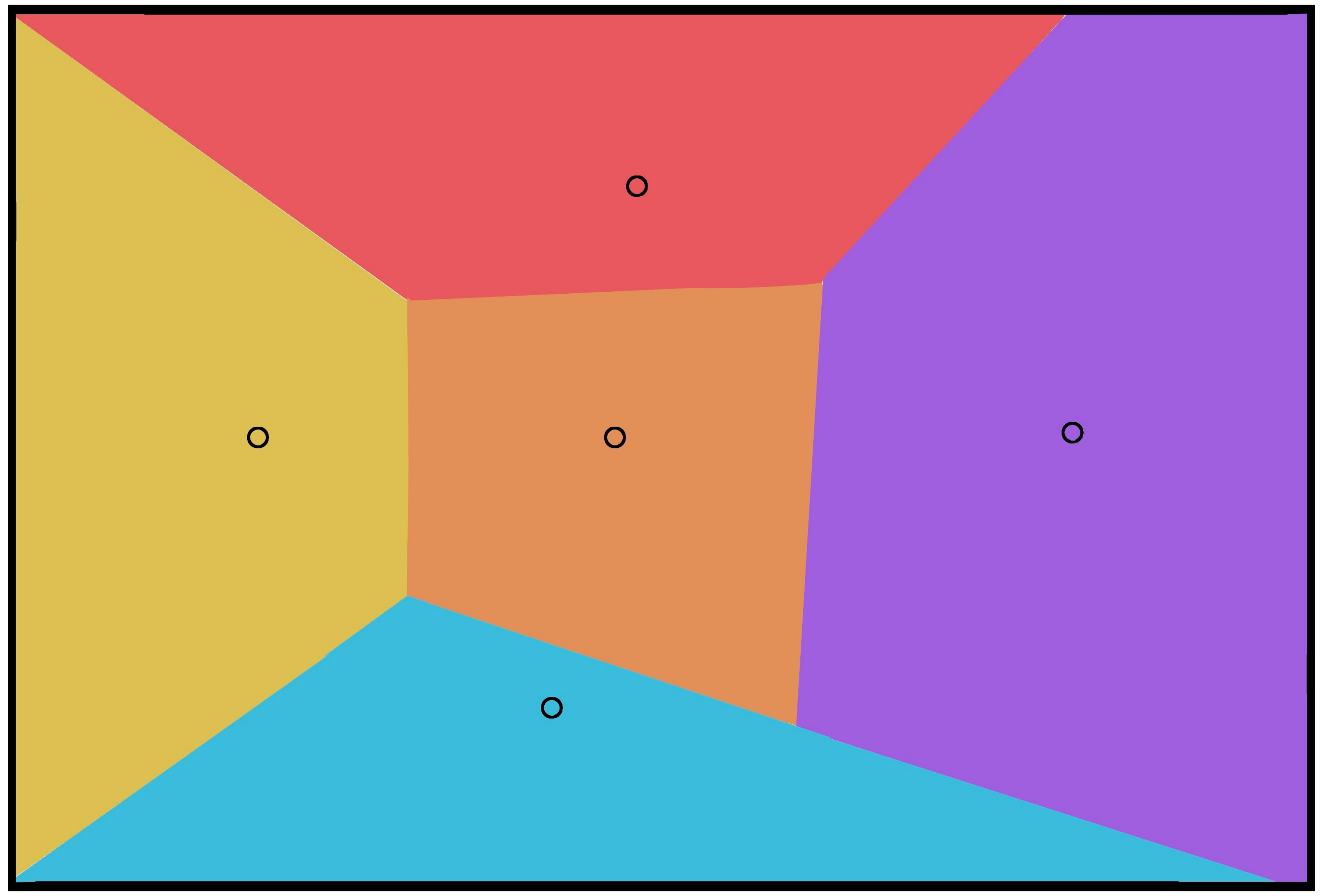
P is "nice" and Q is discrete

What is the optimal transport map?

$$T_0(x) = \nabla \varphi_0(x) = \underset{j \in [J]}{\operatorname{argmax}} \{x^\top y_j - (\psi_0)_j\}$$

where $\psi_0 \in \mathbb{R}^J$ represents the boundaries of the *Laguerre cells*

Hard to get on the basis of samples!



EOT gives minimax estimation rates

Theorem 1 (Informal) *Suppose* ***(A1)*** *and* ***(A2)****. Given i.i.d samples* $X_1, \ldots, X_n \sim P$ *and* $Y_1, \ldots, Y_n \sim Q$*, the entropic Brenier map is minimax optimal. Moreover, the 1NN estimator suffers from the curse of dimensionality.*

EOT gives minimax estimation rates

Theorem 1 (Informal) *Suppose* ***(A1)*** *and* ***(A2)****. Given i.i.d samples* $X_1, \ldots, X_n \sim P$ *and* $Y_1, \ldots, Y_n \sim Q$*, the entropic Brenier map is minimax optimal. Moreover, the 1NN estimator suffers from the curse of dimensionality.*

- Lower bound (Le Cam's method): $\inf_{\hat{T}} \sup_{Q \in \mathcal{Q}_J} \mathbb{E}\|\hat{T} - T_0\|^2_{L^2(P)} \gtrsim n^{-1/2}$,

EOT gives minimax estimation rates

Theorem 1 (Informal) *Suppose* ***(A1)*** *and* ***(A2)****. Given i.i.d samples* $X_1, \ldots, X_n \sim P$ *and* $Y_1, \ldots, Y_n \sim Q$*, the entropic Brenier map is minimax optimal. Moreover, the 1NN estimator suffers from the curse of dimensionality.*

- Lower bound (Le Cam's method): $\inf_{\hat{T}} \sup_{Q \in \mathcal{Q}_J} \mathbb{E}\|\hat{T} - T_0\|^2_{L^2(P)} \gtrsim n^{-1/2}$,
- Upper bound: $\mathbb{E}\|\hat{T}_{\varepsilon,n} - T_0\|^2_{L^2(P)} \lesssim n^{-1/2}$ with $\varepsilon \asymp n^{-1/2}$,

EOT gives minimax estimation rates

Theorem 1 (Informal) *Suppose **(A1)** and **(A2)**. Given i.i.d samples* $X_1, \ldots, X_n \sim P$ *and* $Y_1, \ldots, Y_n \sim Q$*, the entropic Brenier map is minimax optimal. Moreover, the 1NN estimator suffers from the curse of dimensionality.*

- Lower bound (Le Cam's method): $\inf_{\hat{T}} \sup_{Q \in \mathcal{Q}_J} \mathbb{E}\|\hat{T} - T_0\|^2_{L^2(P)} \gtrsim n^{-1/2}$,
- Upper bound: $\mathbb{E}\|\hat{T}_{\varepsilon,n} - T_0\|^2_{L^2(P)} \lesssim n^{-1/2}$ with $\varepsilon \asymp n^{-1/2}$,
- 1NN suffers: there exists P and Q such that $\mathbb{E}\|\hat{T}_{\text{1nn}} - T_0\|^2_{L^2(P)} \gtrsim n^{-1/d}$.

EOT gives minimax estimation rates

Theorem 1 (Informal) *Suppose* ***(A1)*** *and* ***(A2)****. Given i.i.d samples* $X_1, \ldots, X_n \sim P$ *and* $Y_1, \ldots, Y_n \sim Q$*, the entropic Brenier map is minimax optimal. Moreover, the 1NN estimator suffers from the curse of dimensionality.*

- Lower bound (Le Cam's method): $\inf_{\hat{T}} \sup_{Q \in \mathcal{Q}_J} \mathbb{E}\|\hat{T} - T_0\|^2_{L^2(P)} \gtrsim n^{-1/2}$,
- Upper bound: $\mathbb{E}\|\hat{T}_{\varepsilon,n} - T_0\|^2_{L^2(P)} \lesssim n^{-1/2}$ with $\varepsilon \asymp n^{-1/2}$,
- 1NN suffers: there exists P and Q such that $\mathbb{E}\|\hat{T}_{1\text{nn}} - T_0\|^2_{L^2(P)} \gtrsim n^{-1/d}$.

Upper bound proof sketch:

Want to control

statistical error

$$\mathbb{E}\|\hat{T}_{(\varepsilon,n)} - T_0\|^2_{L^2(P)} \lesssim \|T_\varepsilon - T_0\|^2_{L^2(P)} + \mathbb{E}\|\hat{T}_{(\varepsilon,n)} - T_\varepsilon\|^2_{L^2(P)}$$

approximation error

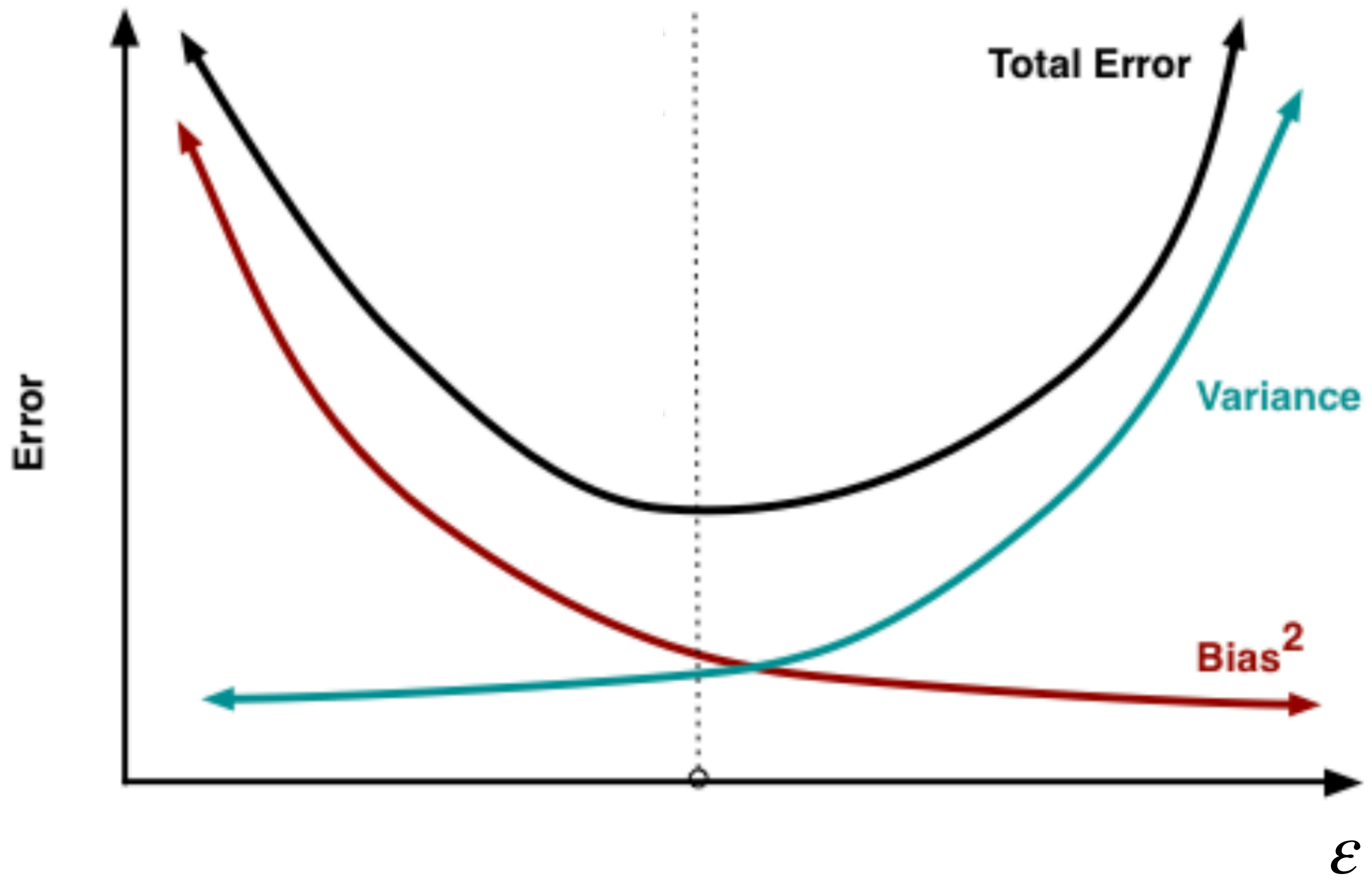
Upper bound proof sketch:

Want to control

statistical error

$$\mathbb{E}\|\hat{T}_{(\varepsilon,n)} - T_0\|^2_{L^2(P)} \lesssim \|T_\varepsilon - T_0\|^2_{L^2(P)} + \mathbb{E}\|\hat{T}_{(\varepsilon,n)} - T_\varepsilon\|^2_{L^2(P)}$$

approximation error



Upper bound proof sketch:

Want to control

statistical error

$$\mathbb{E}\|\hat{T}_{(\varepsilon,n)} - T_0\|^2_{L^2(P)} \lesssim \|T_\varepsilon - T_0\|^2_{L^2(P)} + \mathbb{E}\|\hat{T}_{(\varepsilon,n)} - T_\varepsilon\|^2_{L^2(P)}$$

approximation error

Step 1: **[Thm 3.2]** We show $\|T_\varepsilon - T_0\|^2_{L^2(P)} \lesssim_J \varepsilon$, and is tight

Upper bound proof sketch:

Want to control

statistical error

$$\mathbb{E}\|\hat{T}_{(\varepsilon,n)} - T_0\|^2_{L^2(P)} \lesssim \|T_\varepsilon - T_0\|^2_{L^2(P)} + \mathbb{E}\|\hat{T}_{(\varepsilon,n)} - T_\varepsilon\|^2_{L^2(P)}$$

$$\lesssim_J \quad \varepsilon \quad + \quad ?$$

Step 1: **[Thm 3.2]** We show $\|T_\varepsilon - T_0\|^2_{L^2(P)} \lesssim_J \varepsilon$, and is tight

Upper bound proof sketch:

Want to control

statistical error

$$\mathbb{E}\|\hat{T}_{(\varepsilon,n)} - T_0\|^2_{L^2(P)} \lesssim \|T_\varepsilon - T_0\|^2_{L^2(P)} + \mathbb{E}\|\hat{T}_{(\varepsilon,n)} - T_\varepsilon\|^2_{L^2(P)}$$

$$\lesssim_J \quad \varepsilon \quad + \quad ?$$

Step 1: **[Thm 3.2]** We show $\|T_\varepsilon - T_0\|^2_{L^2(P)} \lesssim_J \varepsilon$, and is tight

Step 2: **[Prop 3.7]** Prove new stability results for statistical error

Upper bound proof sketch:

Want to control

$$\mathbb{E}\|\hat{T}_{(\varepsilon,n)} - T_0\|^2_{L^2(P)} \lesssim \|T_\varepsilon - T_0\|^2_{L^2(P)} + \mathbb{E}\|\hat{T}_{(\varepsilon,n)} - T_\varepsilon\|^2_{L^2(P)}$$
$$\lesssim_J \quad \varepsilon \quad + \quad \varepsilon^{-1} n^{-1}$$

Step 1: **[Thm 3.2]** We show $\|T_\varepsilon - T_0\|^2_{L^2(P)} \lesssim_J \varepsilon$, and is tight

Step 2: **[Prop 3.7]** Prove new stability results for statistical error

Upper bound proof sketch:

Want to control

$$\mathbb{E}\|\hat{T}_{(\varepsilon,n)} - T_0\|^2_{L^2(P)} \lesssim \|T_\varepsilon - T_0\|^2_{L^2(P)} + \mathbb{E}\|\hat{T}_{(\varepsilon,n)} - T_\varepsilon\|^2_{L^2(P)}$$
$$\lesssim_J \quad \varepsilon \quad + \quad \varepsilon^{-1}n^{-1} \qquad \text{(then optimize in } \varepsilon\text{!)}$$

Step 1: [**Thm 3.2**] We show $\|T_\varepsilon - T_0\|^2_{L^2(P)} \lesssim_J \varepsilon$, and is tight

Step 2: [**Prop 3.7**] Prove new stability results for statistical error

Upper bound proof sketch:

Want to control

$$\begin{aligned}\mathbb{E}\|\hat{T}_{(\varepsilon,n)} - T_0\|^2_{L^2(P)} &\lesssim \|T_\varepsilon - T_0\|^2_{L^2(P)} + \mathbb{E}\|\hat{T}_{(\varepsilon,n)} - T_\varepsilon\|^2_{L^2(P)} \\ &\lesssim_J \quad \varepsilon \quad + \quad \varepsilon^{-1}n^{-1} \qquad \text{(then optimize in } \varepsilon\text{!)} \\ &\lesssim_J n^{-1/2}\,.\end{aligned}$$

Step 1: **[Thm 3.2]** We show $\|T_\varepsilon - T_0\|^2_{L^2(P)} \lesssim_J \varepsilon$, and is tight

Step 2: **[Prop 3.7]** Prove new stability results for statistical error

EOT gives minimax estimation rates

Theorem 1 (Informal) *Suppose* ***(A1)*** *and* ***(A2)****. Given i.i.d samples* $X_1, \ldots, X_n \sim P$ *and* $Y_1, \ldots, Y_n \sim Q$*, the entropic Brenier map is minimax optimal. Moreover, the 1NN estimator suffers from the curse of dimensionality.*

- Lower bound (Le Cam's method): $\inf_{\hat{T}} \sup_{Q \in \mathcal{Q}_J} \mathbb{E}\|\hat{T} - T_0\|^2_{L^2(P)} \gtrsim n^{-1/2}$,
- Upper bound: $\mathbb{E}\|\hat{T}_{\varepsilon,n} - T_0\|^2_{L^2(P)} \lesssim n^{-1/2}$ with $\varepsilon \asymp n^{-1/2}$,
- 1NN suffers: there exists P and Q such that $\mathbb{E}\|\hat{T}_{1\mathrm{nn}} - T_0\|^2_{L^2(P)} \gtrsim n^{-1/d}$.

Takeaways and future directions

Takeaways and future directions

- Big picture: Our understanding of discontinuous maps is very limited
  - Rates in general are entirely unknown

Takeaways and future directions

- Big picture: Our understanding of discontinuous maps is very limited
  - Rates in general are entirely unknown
- This work: stepping-stone to understand general discontinuities
  - Bridge computational tractability and statistical efficiency

Takeaways and future directions

- Big picture: Our understanding of discontinuous maps is very limited
  - Rates in general are entirely unknown
- This work: stepping-stone to understand general discontinuities
  - Bridge computational tractability and statistical efficiency
- Can the entropic Brenier map adapt to other forms of structure?
  - Empirical results are promising
  - Should be possible

Takeaways and future directions

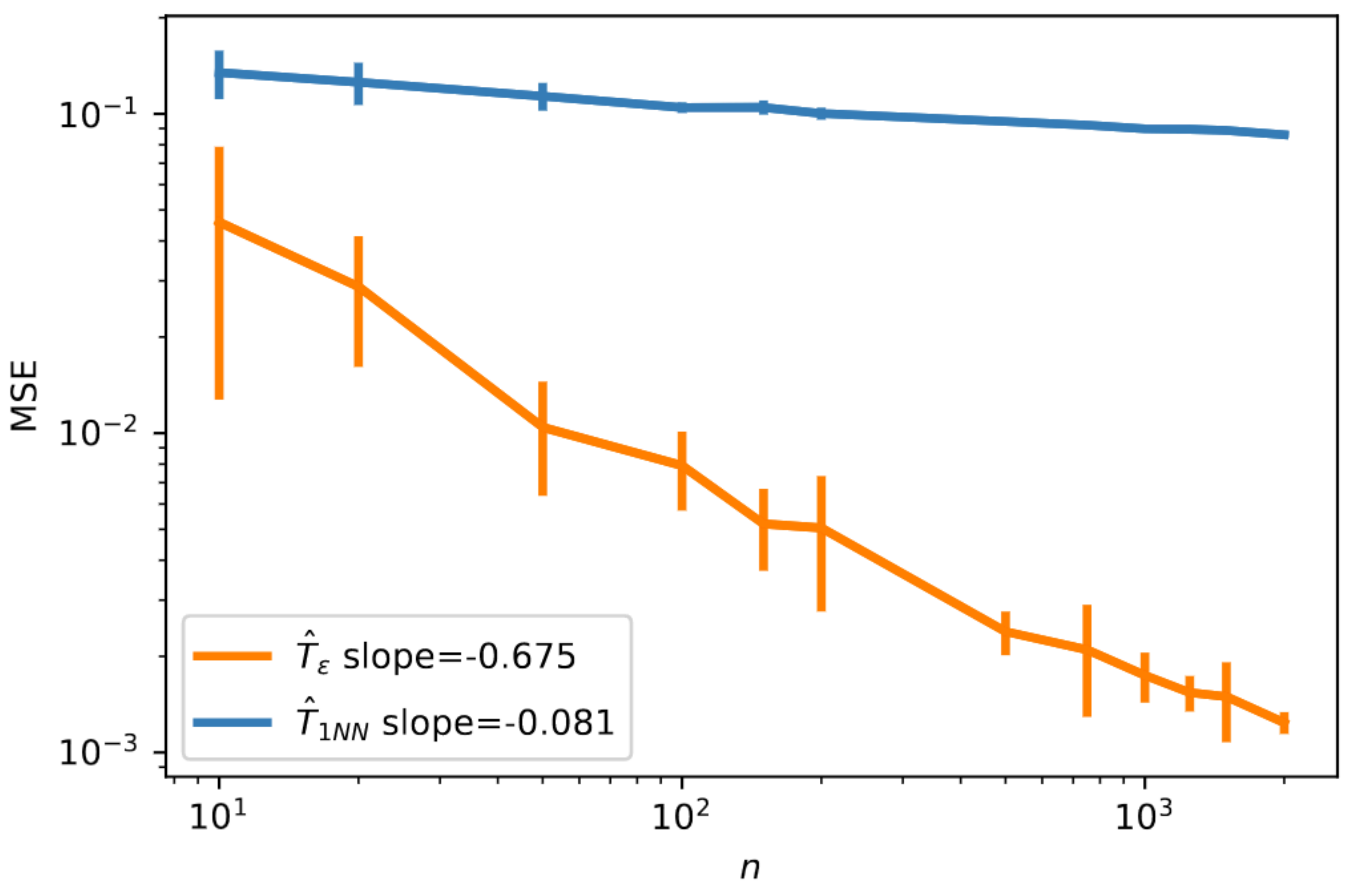
- Big picture: Our understanding of discontinuous maps is very limited
  - Rates in general are entirely unknown
- This work: stepping-stone to understand general discontinuities
  - Bridge computational tractability and statistical efficiency
- Can the entropic Brenier map adapt to other forms of structure?
  - Empirical results are promising
  - Should be possible :)

Thanks!

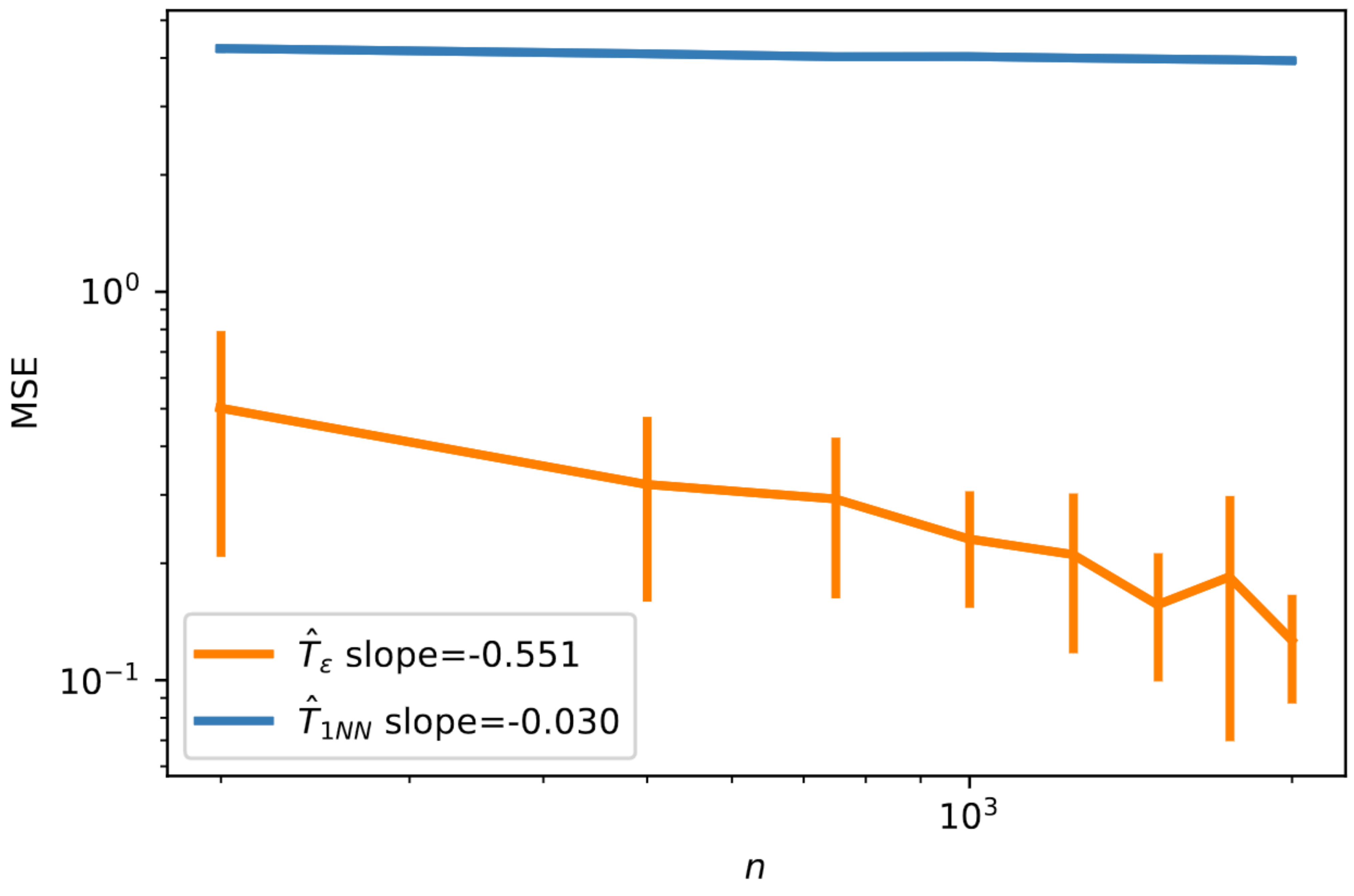
Numerics on synthetic data

Recall $T_0(x) = \text{argmax}_{j\in[J]}\{x^\top y_j - (\psi_0)_j\}$; fix $J = 10$ and $d = 50$

Case 1: data is from a regular grid



Case 2: data is randomly generated



Stability bound

Proposition 3.7. *Let μ, ν, μ', ν' be four probability measures supported in $B(0; R)$. Then the entropic maps $T_\varepsilon^{\mu\to\nu}$ and $T_\varepsilon^{\mu'\to\nu'}$ satisfy*

$$\frac{\varepsilon}{8R^2}\|T_\varepsilon^{\mu\to\nu} - T_\varepsilon^{\mu'\to\nu'}\|_{L^2(\mu)}^2 \leq \int (\varphi_\varepsilon^{\mu'\to\nu'} - \varphi_\varepsilon^{\mu\to\nu})\,\mathrm{d}\mu + \int (\psi_\varepsilon^{\mu'\to\nu'} - \psi_\varepsilon^{\mu\to\nu})\,\mathrm{d}\nu + \varepsilon KL(\nu\|\nu').$$

Sketch: One-sample case

$$\mathbb{E}\|T_\varepsilon^{P\to Q_n} - T_\varepsilon^{P\to Q}\|_{L^2(P)}^2 \lesssim \varepsilon^{-1}\mathbb{E}\left(\int (\psi_\varepsilon^{P\to Q} - \psi_\varepsilon^{P\to Q_n})\,\mathrm{d}(Q_n - Q)\right) + \mathbb{E}[\mathrm{KL}(Q_n\|Q)]$$

1. Application of Prop 3.7 to the empirical measure Q_n and Q

Sketch: One-sample case

$$\begin{aligned}\mathbb{E}\|T_\varepsilon^{P\to Q_n} - T_\varepsilon^{P\to Q}\|_{L^2(P)}^2 &\lesssim \varepsilon^{-1}\mathbb{E}\left(\int (\psi_\varepsilon^{P\to Q} - \psi_\varepsilon^{P\to Q_n})\,\mathrm{d}(Q_n - Q)\right) + \mathbb{E}[\mathrm{KL}(Q_n\|Q)] \\ &\lesssim \varepsilon^{-1}\left(\mathbb{E}[\mathrm{Var}_Q(\psi_\varepsilon^{P\to Q} - \psi_\varepsilon^{P\to Q_n})] + \mathbb{E}[\chi^2(Q_n\|Q)]\right) + \mathbb{E}[\chi^2(Q_n\|Q)]\end{aligned}$$

1. Application of Prop 3.7 to the empirical measure Q_n and Q
2. Young's inequality with the fact $\mathrm{KL} \leq \chi^2$

Sketch: One-sample case

$$\begin{aligned}\mathbb{E}\|T_\varepsilon^{P\to Q_n} - T_\varepsilon^{P\to Q}\|_{L^2(P)}^2 &\lesssim \varepsilon^{-1}\mathbb{E}\left(\int (\psi_\varepsilon^{P\to Q} - \psi_\varepsilon^{P\to Q_n})\,\mathrm{d}(Q_n - Q)\right) + \mathbb{E}[\mathrm{KL}(Q_n\|Q)] \\ &\lesssim \varepsilon^{-1}\left(\mathbb{E}[\mathrm{Var}_Q(\psi_\varepsilon^{P\to Q} - \psi_\varepsilon^{P\to Q_n})] + \mathbb{E}[\chi^2(Q_n\|Q)]\right) + \mathbb{E}[\chi^2(Q_n\|Q)] \\ &\lesssim (1+\varepsilon^{-1})\mathbb{E}[\chi^2(Q_n\|Q)]\end{aligned}$$

1. Application of Prop 3.7 to the empirical measure Q_n and Q
2. Young's inequality with the fact $\mathrm{KL} \le \chi^2$
3. Strong convexity of empirical semidual: $\mathrm{Var}_Q(\psi_\varepsilon - \psi_\varepsilon') \lesssim \chi^2(Q'\|Q)$

Sketch: One-sample case

$$\begin{aligned}
\mathbb{E}\|T_\varepsilon^{P\to Q_n} - T_\varepsilon^{P\to Q}\|_{L^2(P)}^2 &\lesssim \varepsilon^{-1}\mathbb{E}\left(\int (\psi_\varepsilon^{P\to Q} - \psi_\varepsilon^{P\to Q_n})\,\mathrm{d}(Q_n - Q)\right) + \mathbb{E}[\mathrm{KL}(Q_n\|Q)] \\
&\lesssim \varepsilon^{-1}\left(\mathbb{E}[\mathrm{Var}_Q(\psi_\varepsilon^{P\to Q} - \psi_\varepsilon^{P\to Q_n})] + \mathbb{E}[\chi^2(Q_n\|Q)]\right) + \mathbb{E}[\chi^2(Q_n\|Q)] \\
&\lesssim (1+\varepsilon^{-1})\mathbb{E}[\chi^2(Q_n\|Q)] \\
&\lesssim \varepsilon^{-1}n^{-1}\,.
\end{aligned}$$

1. Application of Prop 3.7 to the empirical measure Q_n and Q
2. Young's inequality with the fact $\mathrm{KL} \leq \chi^2$
3. Strong convexity of empirical semidual: $\mathrm{Var}_Q(\psi_\varepsilon - \psi'_\varepsilon) \lesssim \chi^2(Q'\|Q)$
4. Finally, a calculation gives $\mathbb{E}[\chi^2(Q_n\|Q) \lesssim n^{-1}$.

Thanks (part 2)!