An entropic generalization of Caffarelli’s contraction theorem

Aram-Alexandre Pooladian
New York University

An entropic generalization of Caffarelli's contraction theorem

Aram-Alexandre Pooladian
New York University

Joint work with Sinho Chewi (PhD student at MIT)

Background

We’ll be working with strongly log-concave distributions

Background

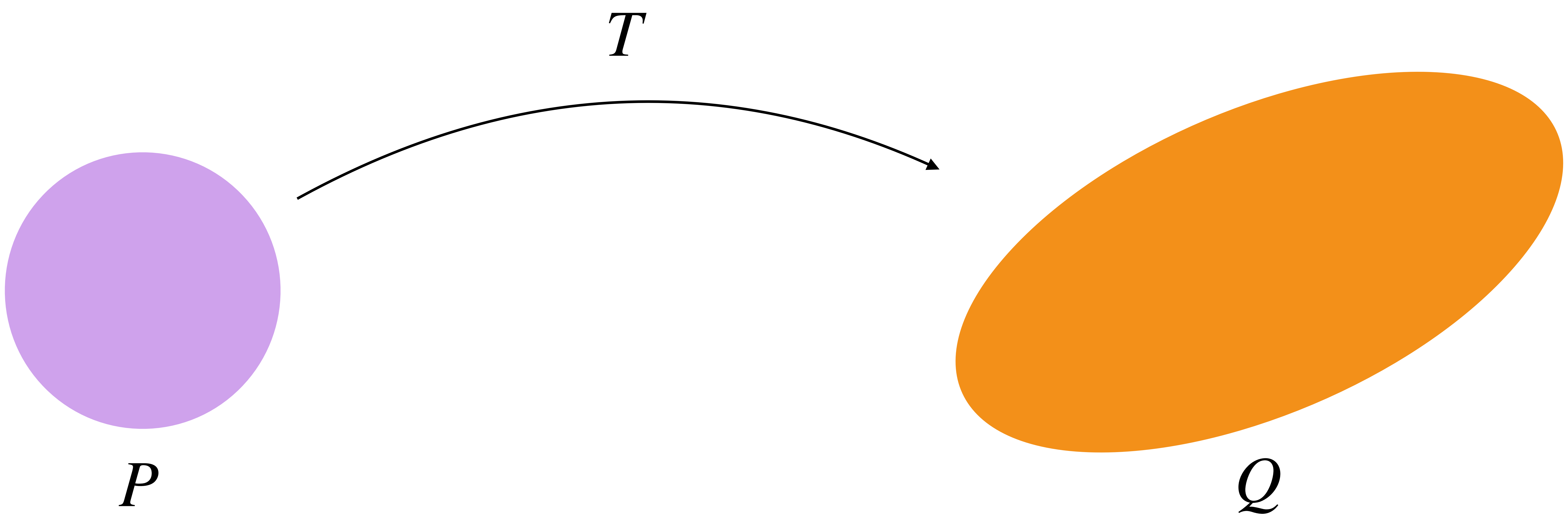
We'll be working with strongly log-concave distributions

Background

Let $P = \exp(-V)$ and $Q = \exp(-W)$ with:

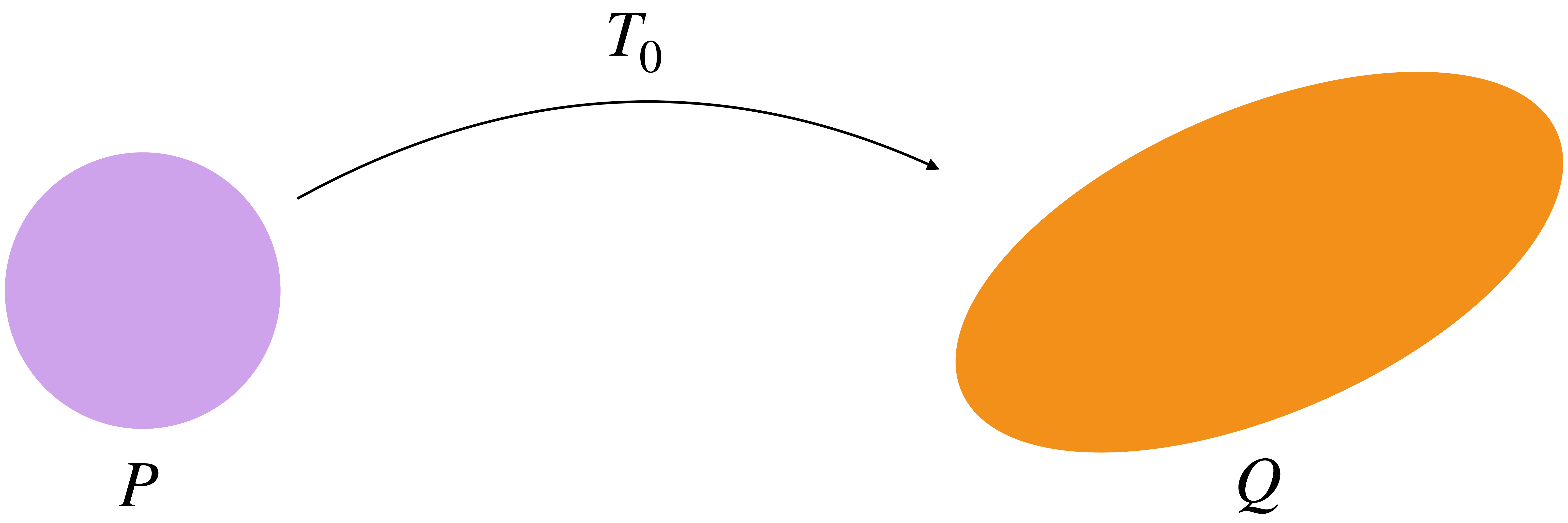
- $\nabla^2 V(x) \preceq \beta I$
- $\nabla^2 W(y) \succeq \alpha I \succ 0$

Background



Let $\mathscr{T}(P, Q)$ denote the set of valid transport maps from P to Q

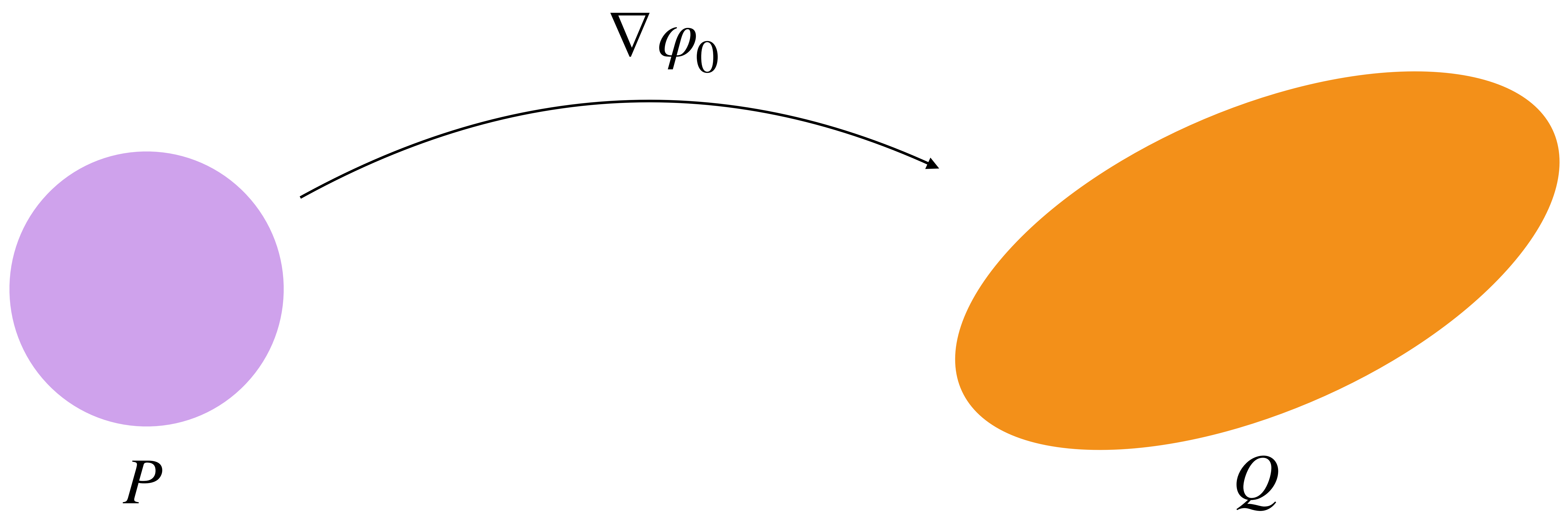
Background



optimal transport map:

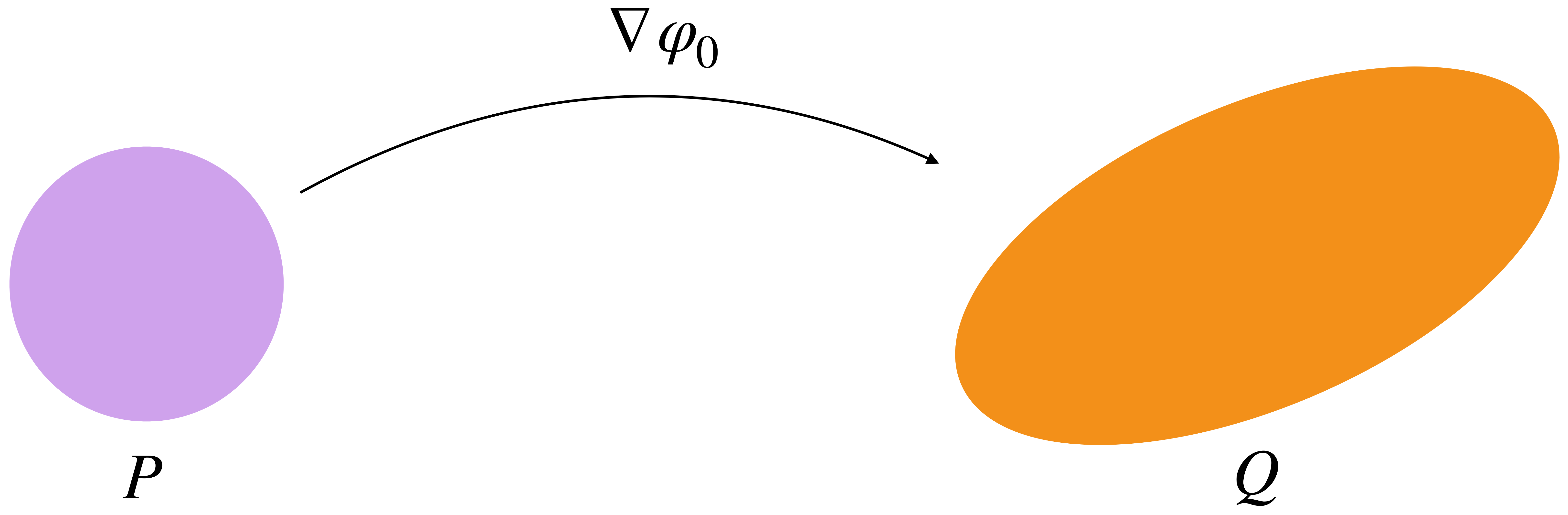
$$T_0 := \operatorname*{argmin}_T \int \frac{1}{2}\|x - T(x)\|_2^2 \,\mathrm{d}P(x) \quad \text{s.t.} \quad T \in \mathcal{T}(P, Q)$$

Background



Brenier's theorem (1991): $T_0 := \nabla \varphi_0$ where φ_0 is a convex function

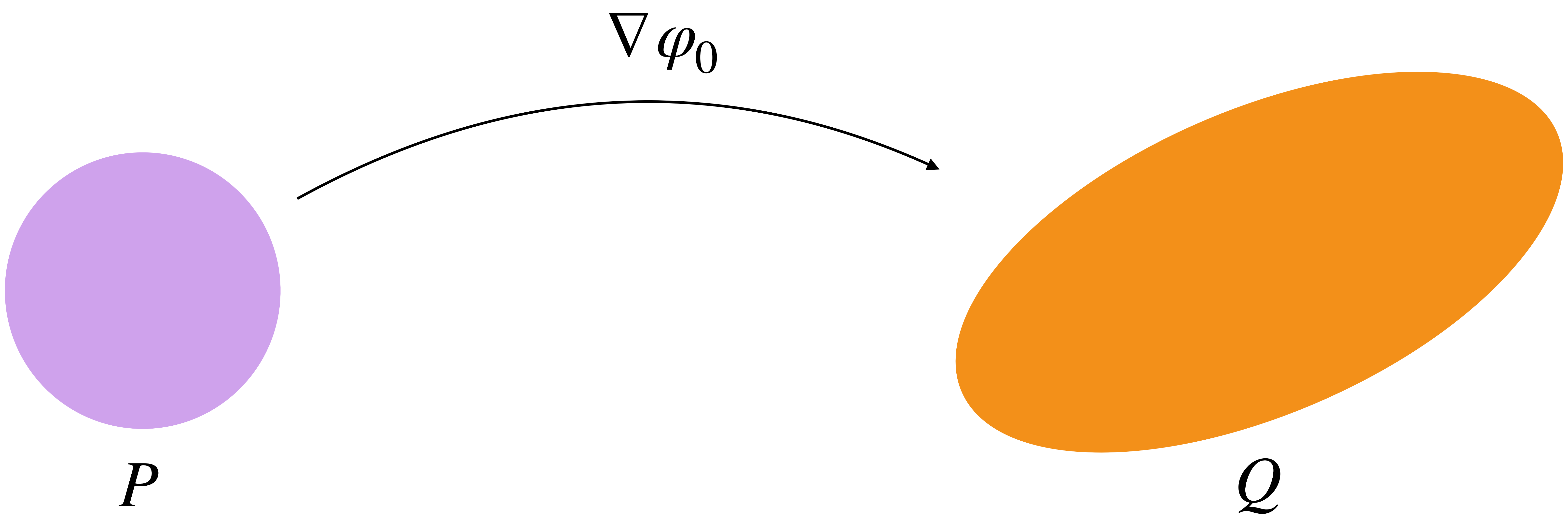
Background



Brenier's theorem (1991): $T_0 := \nabla \varphi_0$ where φ_0 is a convex function

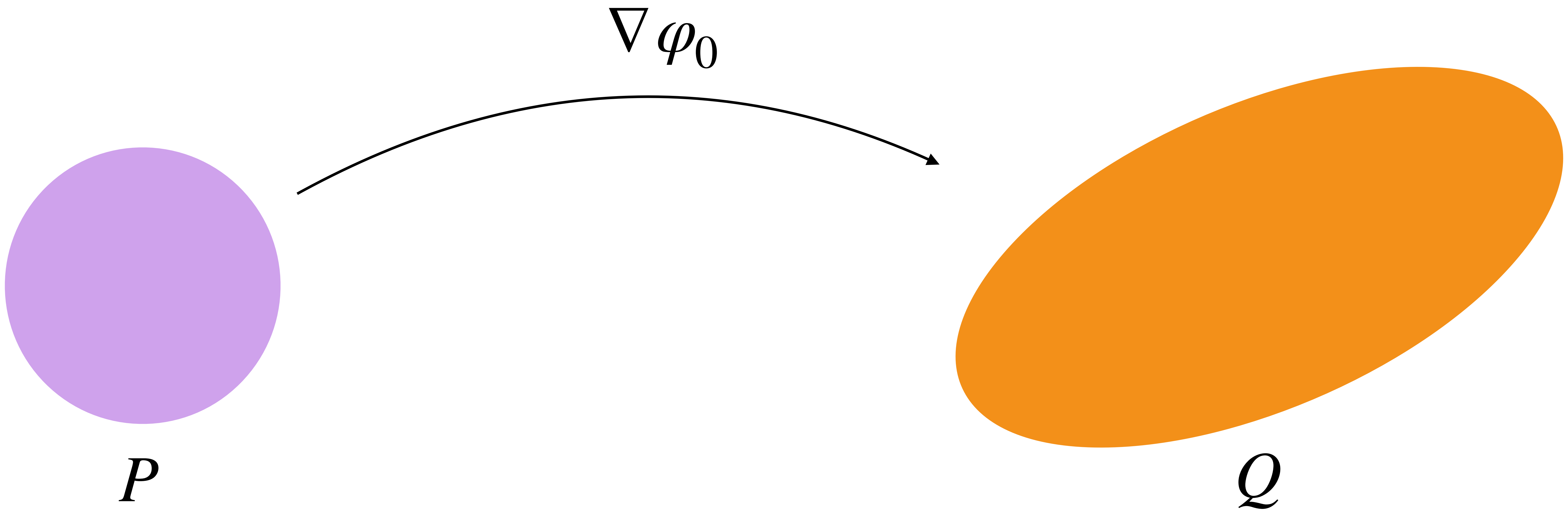
Brenier potential

Background



Caffarelli's contraction theorem (2000): $\|\nabla^2 \varphi_0(x)\|_{\mathrm{op}} \leq \sqrt{\beta/\alpha}$

Background



Caffarelli's contraction theorem (2000): $\|\nabla^2\varphi_0(x)\|_{\text{op}} \leq \sqrt{\beta/\alpha}$

Proof: requires regularity theory, maximum principles from PDEs, etc

Motivation

Caffarelli's contraction theorem (2000): $\|\nabla^2 \varphi_0(x)\|_{\mathrm{op}} \leq \sqrt{\beta/\alpha}$

Why might one be interested in this?

- Establishing functional inequalities (e.g. Poincaré inequality)
- Statistical estimation of OT maps
- Stability of Wasserstein barycenters
- Most applications of OT…

Motivation

Caffarelli's contraction theorem (2000): $\|\nabla^2 \varphi_0(x)\|_{\text{op}} \leq \sqrt{\beta/\alpha}$

<u>Example: Transferring Poincaré inequalities</u>

Let $P = N(0, I)$ and let f be smooth. Then the *Gaussian Poincaré* inequality reads

$$\text{Var}_P(f(X)) \leq 1 \cdot \mathbb{E}_P[\|\nabla f(X)\|^2] .$$

Motivation

Caffarelli's contraction theorem (2000): $\|\nabla^2 \varphi_0(x)\|_{\mathrm{op}} \leq \sqrt{\beta/\alpha}$

<u>Example: Transferring Poincaré inequalities</u>

Let $P = N(0,I)$ and let f be smooth. Then the *Gaussian Poincaré* inequality reads

$$\mathrm{Var}_P(f(X)) \leq 1 \cdot \mathbb{E}_P[\|\nabla f(X)\|^2]\,.$$

($1 = C_{\mathrm{PI}}$)

Motivation

Caffarelli's contraction theorem (2000): $\|\nabla^2 \varphi_0(x)\|_{\text{op}} \leq \sqrt{\beta/\alpha}$

<u>Example: Transferring Poincaré inequalities</u>

Let Q be such that $(\nabla \varphi_0)_\sharp P = Q$. Then via chain-rule

$$\begin{aligned}
\text{Var}_Q(f(Y)) = \text{Var}_P((f \circ \nabla \varphi_0)(X)) &\leq \mathbb{E}_P[\|\nabla (f \circ \nabla \varphi_0)(X)\|^2] \\
&= \mathbb{E}_P[\|(\nabla f \circ \nabla \varphi_0)(X)\, \nabla^2 \varphi_0(X)\|^2] \\
&\leq \|\nabla^2 \varphi_0\|_{\text{op}}^2\, \mathbb{E}_P[\|(\nabla f \circ \nabla \varphi_0)(X)\|^2] \\
&= \|\nabla^2 \varphi_0\|_{\text{op}}^2\, \mathbb{E}_Q[\|\nabla f(Y)\|^2]
\end{aligned}$$

Motivation

Caffarelli's contraction theorem (2000): $\|\nabla^2 \varphi_0(x)\|_{\mathrm{op}} \leq \sqrt{\beta/\alpha}$

Example: Transferring Poincaré inequalities

Let Q be such that $(\nabla \varphi_0)_\sharp P = Q$. Then via chain-rule

$$\mathrm{Var}_Q(f(Y)) \leq \|\nabla^2 \varphi_0\|_{\mathrm{op}}^2 \, \mathbb{E}_Q[\|\nabla f(Y)\|^2]$$

C_{PI}

Motivation

What happens in practice? **Entropic** optimal transport

- OT map estimation
- Computing barycenters
- Pretty much everything

Why?

- Sinkhorn's algorithm (Cut13)
- Parallelizable implementation with GPU speedups
- Effective even when $n \simeq 10^4$

Motivation

What happens in practice? **Entropic** optimal transport

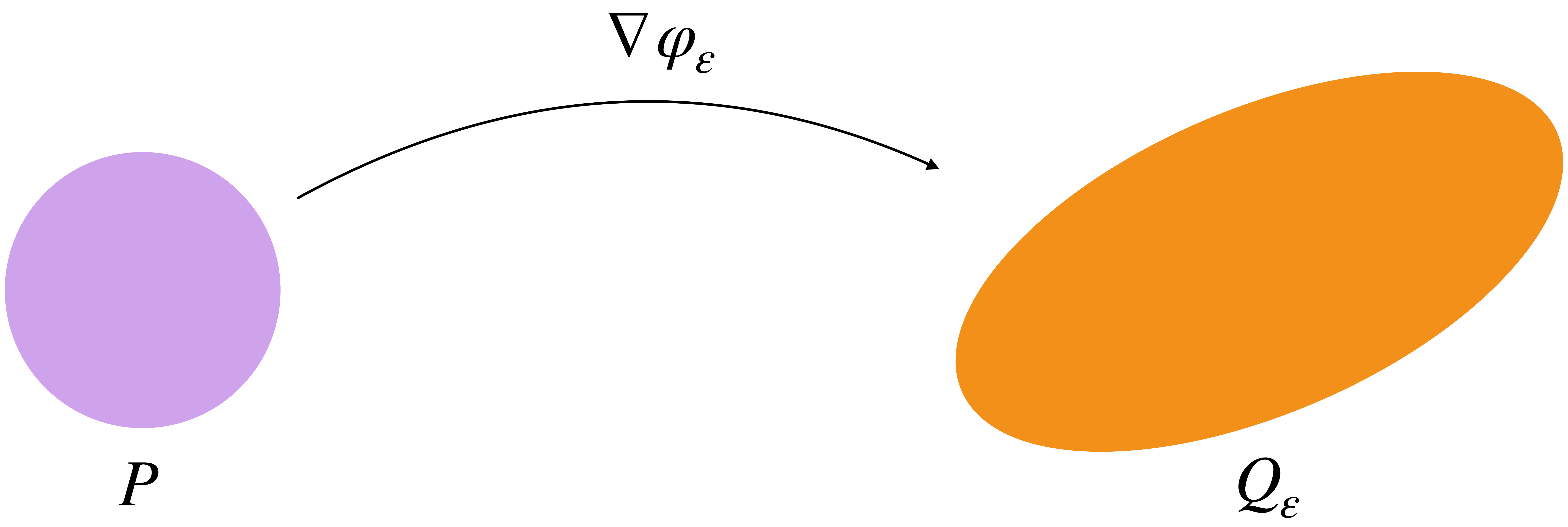
Schrodinger Bridge

- OT map estimation
- Computing barycenters
- Pretty much everything

Why?

- Sinkhorn's algorithm (Cut13)
- Parallelizable implementation with GPU speedups
- Effective even when $n \simeq 10^4$

Our contributions



- Generalize Caffarelli's result to *entropic* Brenier potentials
- Recover Caffarelli's result (shortest proof to date)

Optimal transport

Optimal transport

(Primal) $$\pi_0 = \text{argmin}_{\pi \in \Gamma(P,Q)} \iint \frac{1}{2}\|x - y\|^2 \mathrm{d}\pi(x, y)$$

(Dual) $$(f_0, g_0) = \text{argmax}_{f,g} \int f \,\mathrm{d}P + \int g \,\mathrm{d}Q - \iota_{f(x)+g(y) \leq \frac{1}{2}\|x-y\|^2}$$

(Potential) $$\varphi_0 = \frac{1}{2}\|\cdot\|^2 - f_0 \;,\; \psi_0 = \frac{1}{2}\|\cdot\|^2 - g_0$$

(Map) $$T_0 = \nabla \varphi_0$$

Entropic optimal transport

(Primal) $\pi_\varepsilon = \text{argmin}_{\pi\in\Gamma(P,Q)} \iint \frac{1}{2}\|x-y\|^2 \mathrm{d}\pi(x,y)$ $+\varepsilon \mathrm{D}_{KL}(\pi\|P\otimes Q)$

(Dual) $(f_0, g_0) = \text{argmax}_{f,g} \int f\,\mathrm{d}P + \int g\,\mathrm{d}Q - \iota_{f(x)+g(y)\leq\frac{1}{2}\|x-y\|^2}$?

(Potential) $\varphi_0 = \frac{1}{2}\|\cdot\|^2 - f_0$, $\psi_0 = \frac{1}{2}\|\cdot\|^2 - g_0$?

(Map) $T_0 = \nabla\varphi_0$?

Entropic optimal transport

(Primal) $$\pi_\varepsilon = \text{argmin}_{\pi\in\Gamma(P,Q)} \iint \frac{1}{2}\|x-y\|^2 \mathrm{d}\pi(x,y) \ + \varepsilon \mathrm{D}_{KL}(\pi\|P\otimes Q)$$

(Dual) $$(f_\varepsilon, g_\varepsilon) = \text{argmax}_{f,g} \int f\,\mathrm{d}P + \int g\,\mathrm{d}Q - \varepsilon \iint e^{(f(x)+g(y)-\frac{1}{2}\|x-y\|^2)/\varepsilon}\,\mathrm{d}P(x)\mathrm{d}Q(y) + \varepsilon$$

(Potential) $$\varphi_0 = \frac{1}{2}\|\cdot\|^2 - f_0\ ,\ \psi_0 = \frac{1}{2}\|\cdot\|^2 - g_0$$?

(Map) $$T_0 = \nabla\varphi_0$$?

Entropic optimal transport

(Primal) $\pi_\varepsilon = \text{argmin}_{\pi \in \Gamma(P,Q)} \iint \frac{1}{2}\|x-y\|^2 \mathrm{d}\pi(x,y) + \varepsilon \mathrm{D}_{KL}(\pi \| P \otimes Q)$

(Dual) $(f_\varepsilon, g_\varepsilon) = \text{argmax}_{f,g} \int f \,\mathrm{d}P + \int g \,\mathrm{d}Q - \varepsilon \iint e^{(f(x)+g(y)-\frac{1}{2}\|x-y\|^2)/\varepsilon} \mathrm{d}P(x)\mathrm{d}Q(y) + \varepsilon$

(Potential) $\varphi_\varepsilon = \frac{1}{2}\|\cdot\|^2 - f_\varepsilon$, $\psi_\varepsilon = \frac{1}{2}\|\cdot\|^2 - g_\varepsilon$

(Map) $T_0 = \nabla \varphi_0$ **?**

Entropic optimal transport

(Primal) $$\pi_\varepsilon = \text{argmin}_{\pi\in\Gamma(P,Q)} \iint \frac{1}{2}\|x-y\|^2 \mathrm{d}\pi(x,y) + \varepsilon \mathrm{D}_{KL}(\pi\|P\otimes Q)$$

(Dual) $$(f_\varepsilon, g_\varepsilon) = \text{argmax}_{f,g} \int f\,\mathrm{d}P + \int g\,\mathrm{d}Q - \varepsilon \iint e^{(f(x)+g(y)-\frac{1}{2}\|x-y\|^2)/\varepsilon}\,\mathrm{d}P(x)\mathrm{d}Q(y) + \varepsilon$$

(Potential) $$\varphi_\varepsilon = \frac{1}{2}\|\cdot\|^2 - f_\varepsilon\,, \quad \psi_\varepsilon = \frac{1}{2}\|\cdot\|^2 - g_\varepsilon$$

(Map) $$T_\varepsilon = \nabla\varphi_\varepsilon$$

Entropic optimal transport

Convergence of regularized to unregularized potentials:

- $\varphi_\varepsilon \longrightarrow \varphi_0$ in $L^1(P)$ [Nutz, Weisel '21]
- $\nabla \varphi_\varepsilon \longrightarrow \nabla \varphi_0$ in $L^2(P)$ [**P.**, Niles-Weed '21]

Specifically, $\|\nabla \varphi_\varepsilon - \nabla \varphi_0\|^2_{L^2(P)} \lesssim \varepsilon^2 I_0(P, Q) + \varepsilon^{\min(4,\alpha+1)/2}$ $(\psi_0 \in C^{\alpha+1})$

Entropic optimal transport

(Primal) $$\pi_\varepsilon = \text{argmin}_{\pi \in \Gamma(P,Q)} \iint \frac{1}{2}\|x-y\|^2 \mathrm{d}\pi(x,y) + \varepsilon \mathrm{D}_{KL}(\pi \| P \otimes Q)$$

has the following closed form representation

$$\pi_\varepsilon(x,y) = \exp\left\{ \varepsilon^{-1} \left(f_\varepsilon(x) + g_\varepsilon(y) - \frac{1}{2}\|x-y\|_2^2 \right) \right\} dP(x)dQ(y)$$

Entropic optimal transport

(Primal) $$\pi_\varepsilon = \text{argmin}_{\pi \in \Gamma(P,Q)} \iint \frac{1}{2}\|x-y\|^2 \mathrm{d}\pi(x,y) + \varepsilon \mathrm{D}_{KL}(\pi \| P \otimes Q)$$

has the following closed form representation

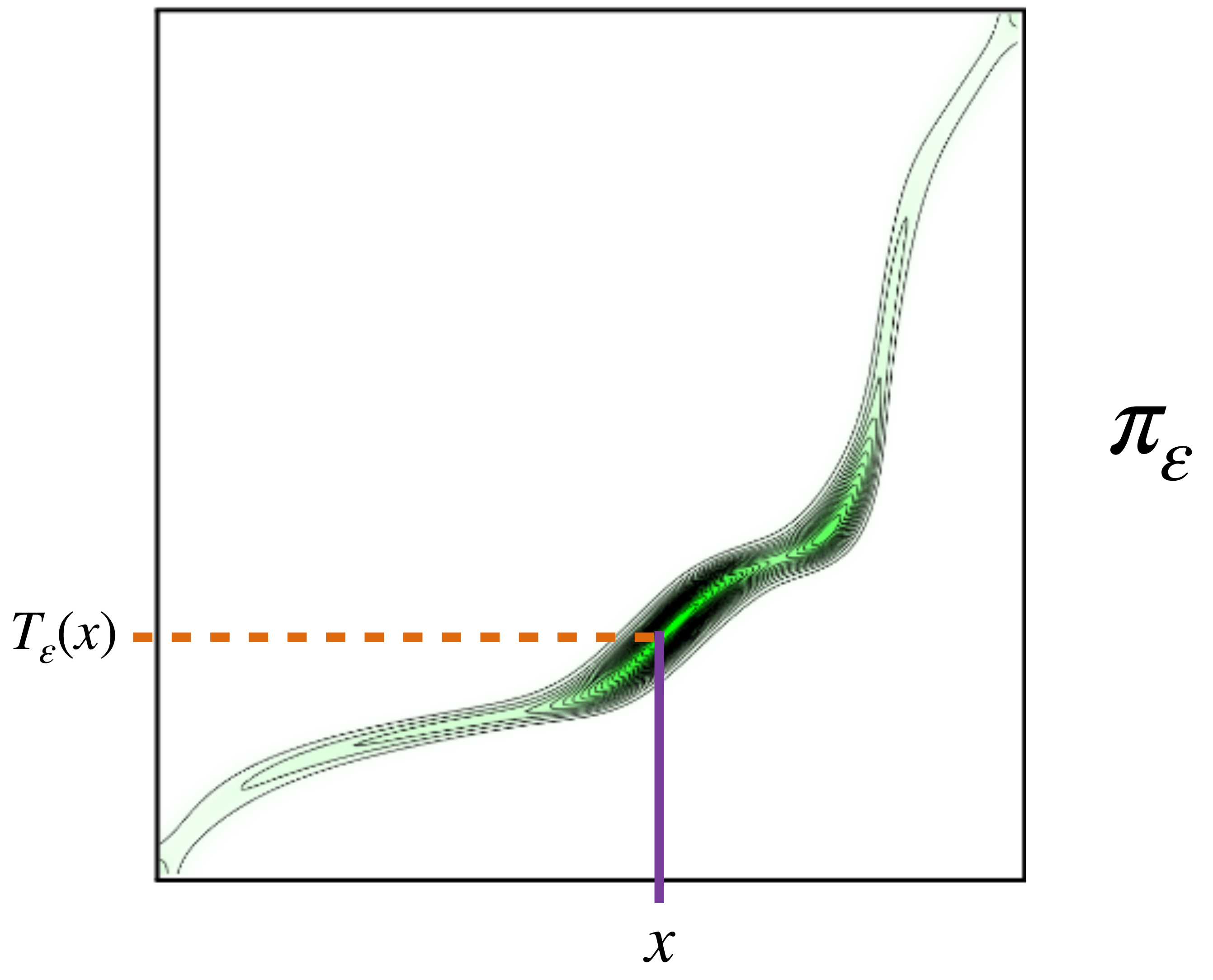
$$\pi_\varepsilon(x,y) = \exp\left\{ -\varepsilon^{-1}\left(\varphi_\varepsilon(x) + \psi_\varepsilon(y) - \langle x, y \rangle\right) - V(x) - W(y) \right\}$$

Entropic optimal transport

(Map) $T_\varepsilon = \nabla \varphi_\varepsilon$

also expressed as a conditional expectation [Prop 1, **P.**,Niles-Weed '21]

$$\nabla \varphi_\varepsilon(x) = \mathbb{E}_{\pi_\varepsilon}[Y | X = x]$$

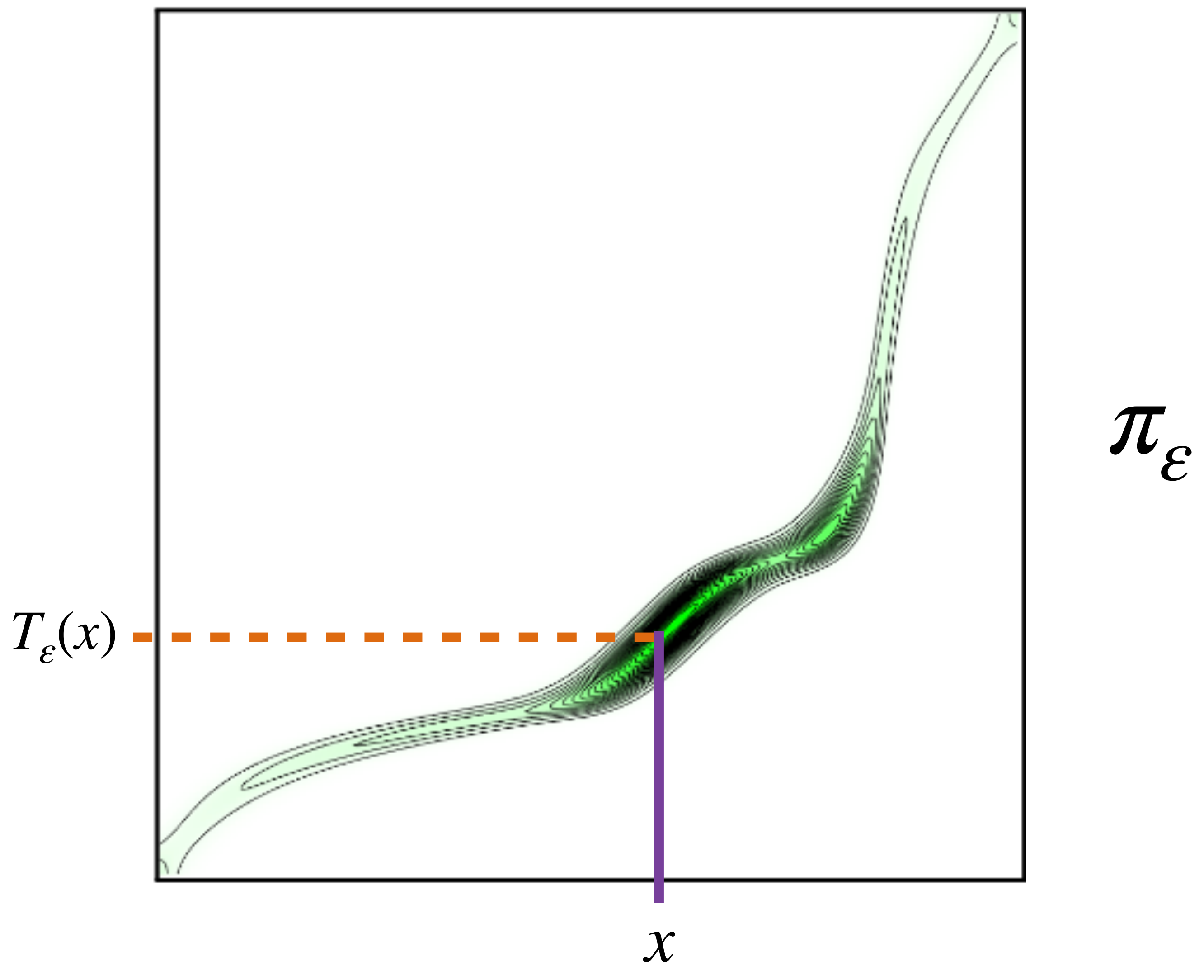


Entropic optimal transport

(Map) $$T_\varepsilon = \nabla \varphi_\varepsilon$$

also expressed as a conditional expectation [Prop 1, **P.**,Niles-Weed '21]

$$\nabla \varphi_\varepsilon(x) = \mathbb{E}_{\pi_\varepsilon}[Y | X = x] =: \mathbb{E}_{\pi_\varepsilon^x}[Y]$$



Entropic optimal transport

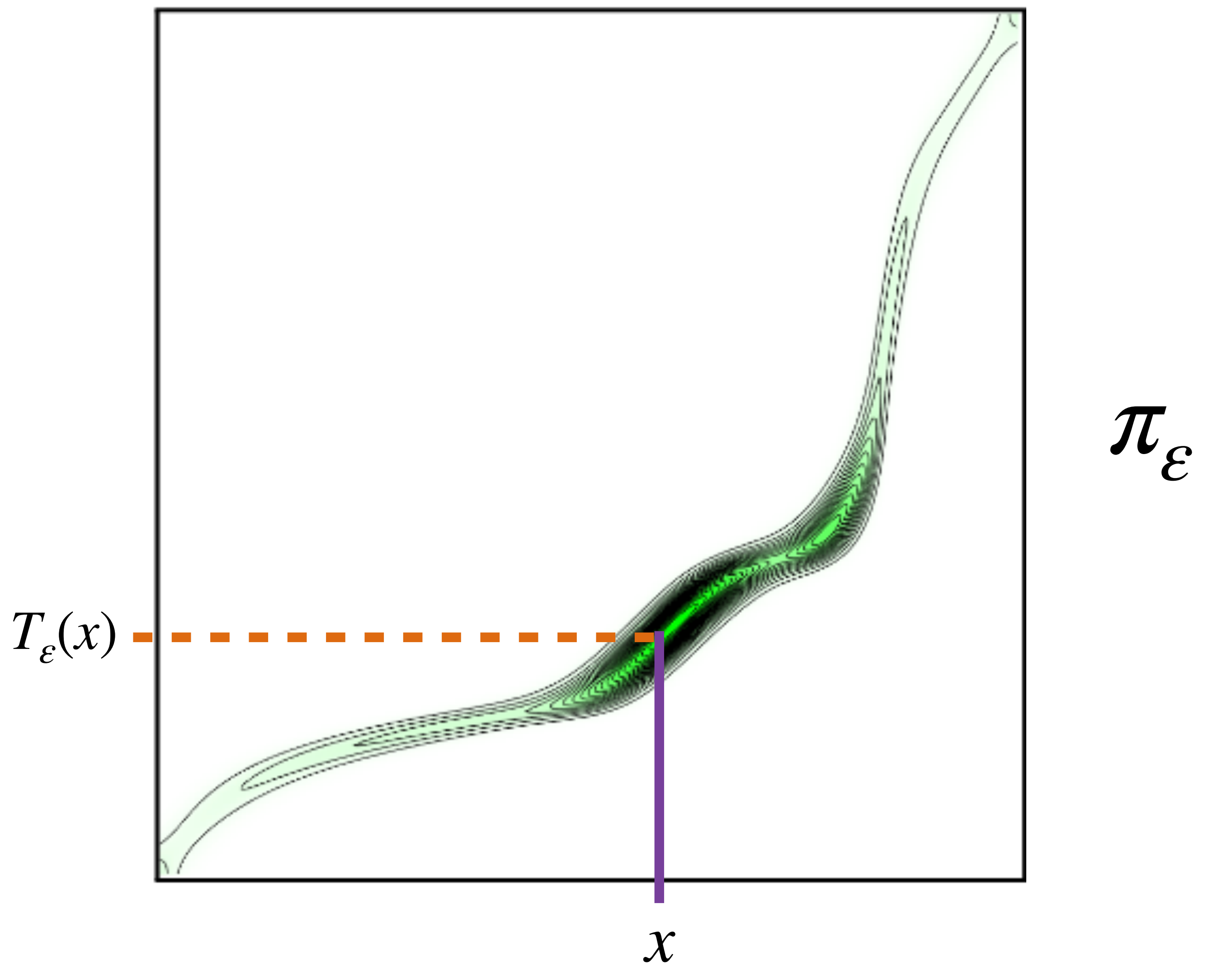
(Map) $$T_\varepsilon = \nabla \varphi_\varepsilon$$

also expressed as a conditional expectation [Prop 1, **P.**,Niles-Weed '21]

$$\nabla \varphi_\varepsilon(x) = \mathbb{E}_{\pi_\varepsilon}[Y | X = x] =: \mathbb{E}_{\pi_\varepsilon^x}[Y]$$

with

$$\pi_\varepsilon^x(y) \propto \exp\left\{ -\varepsilon^{-1}\left(\psi_\varepsilon(y) - \langle x, y \rangle\right) - W(y) \right\}$$



Entropic optimal transport

Lemma 1 (Chewi, **P.**, '22):

- $\nabla^2 \varphi_\varepsilon(x) = \varepsilon^{-1} \mathrm{Cov}_{\pi_\varepsilon}(Y \mid X = x)$, $\nabla^2 \psi_\varepsilon(y) = \varepsilon^{-1} \mathrm{Cov}_{\pi_\varepsilon}(X \mid Y = y)$

- $\nabla^2 \log(1/\pi_\varepsilon^x)(y) = \varepsilon^{-1} \nabla^2 \psi_\varepsilon(y) + \nabla^2 W(y)$

- $\nabla^2 \log(1/\pi_\varepsilon^y)(x) = \varepsilon^{-1} \nabla^2 \varphi_\varepsilon(x) + \nabla^2 V(x)$

Covariance inequalities

Let $P = \exp(-V)$ be a probability measure on $\mathbb{R}^d$ with $V \in C^2$ and convex

(1) *Brascamp-Lieb* inequality: $\mathrm{Cov}_P(X) \preceq \mathbb{E}_P[(\nabla^2 V(X))^{-1}]$ (BL)

(2) *Cramér-Rao* inequality: $(\mathbb{E}_P[\nabla^2 V(X)])^{-1} \preceq \mathrm{Cov}_P(X)$ (CR)

Bounds on Hessians

$$
\begin{aligned}
\nabla^2 \varphi_\varepsilon(x) &= \varepsilon^{-1} \mathrm{Cov}_{\pi_\varepsilon^x}(Y) \\
&\preceq \varepsilon^{-1} \mathbb{E}_{\pi_\varepsilon^x}\left[(\nabla^2 \log(1/\pi_\varepsilon^x)(Y))^{-1} \right] \qquad \text{(BL)} \\
&= \varepsilon^{-1} \mathbb{E}_{\pi_\varepsilon^x}\left[(\varepsilon^{-1} \nabla^2 \psi_\varepsilon(Y) + \nabla^2 W(Y))^{-1} \right]
\end{aligned}
$$

How to progress? *Lower bound* $\nabla^2 \psi_\varepsilon(Y)$ using Cramér-Rao!

Bounds on Hessians

$$
\begin{aligned}
\nabla^2 \varphi_\varepsilon(x) &= \varepsilon^{-1}\mathrm{Cov}_{\pi_\varepsilon^x}(Y) \\
&\preceq \varepsilon^{-1}\mathbb{E}_{\pi_\varepsilon^x}\left[(\nabla^2 \log(1/\pi_\varepsilon^x)(Y))^{-1}\right] && \text{(BL)} \\
&= \varepsilon^{-1}\mathbb{E}_{\pi_\varepsilon^x}\left[(\varepsilon^{-1}\nabla^2 \psi_\varepsilon(Y) + \nabla^2 W(Y))^{-1}\right]
\end{aligned}
$$

$$
\begin{aligned}
\nabla^2 \psi_\varepsilon(Y) &= \varepsilon^{-1}\mathrm{Cov}_{\pi_\varepsilon^Y}(X) \\
&\succeq \varepsilon^{-1}\left(\mathbb{E}_{\pi_\varepsilon^Y}[\nabla^2 \log(1/\pi_\varepsilon^Y)(X)]\right)^{-1} && \text{(CR)} \\
&= \varepsilon^{-1}\left(\mathbb{E}_{\pi_\varepsilon^Y}[\varepsilon^{-1}\nabla^2 \varphi_\varepsilon(X) + \nabla^2 V(X)]\right)^{-1}
\end{aligned}
$$

Bounds on Hessians

$$
\begin{aligned}
\nabla^2 \varphi_\varepsilon(x) &= \varepsilon^{-1} \mathrm{Cov}_{\pi_\varepsilon^x}(Y) \\
&\preceq \varepsilon^{-1} \mathbb{E}_{\pi_\varepsilon^x}\left[(\nabla^2 \log(1/\pi_\varepsilon^x)(Y))^{-1} \right] && \text{(BL)} \\
&= \varepsilon^{-1} \mathbb{E}_{\pi_\varepsilon^x}\left[(\varepsilon^{-1} \nabla^2 \psi_\varepsilon(Y) + \nabla^2 W(Y))^{-1} \right] \\
&= \mathbb{E}_{\pi_\varepsilon^x}\left[(\nabla^2 \psi_\varepsilon(Y) + \varepsilon \nabla^2 W(Y))^{-1} \right] \\
&\preceq \mathbb{E}_{\pi_\varepsilon^x}\left[\left(\left(\mathbb{E}_{\pi_\varepsilon^Y}[\nabla^2 \varphi_\varepsilon(X) + \varepsilon \nabla^2 V(X)] \right)^{-1} + \varepsilon \nabla^2 W(Y) \right)^{-1} \right] && \text{(CR)}
\end{aligned}
$$

Bounds on Hessians

$$\nabla^2 \varphi_\varepsilon(x) \preceq \mathbb{E}_{\pi_\varepsilon^x}\left[\left(\left(\mathbb{E}_{\pi_\varepsilon^Y}[\nabla^2 \varphi_\varepsilon(X) + \varepsilon \nabla^2 V(X)]\right)^{-1} + \varepsilon \nabla^2 W(Y)\right)^{-1}\right]$$

Define $L_\varepsilon := \sup_x \lambda_{\max}(\nabla^2 \varphi_\varepsilon(x))$ and use $\nabla^2 W(y) \succeq \alpha I, \nabla^2 V(x) \preceq \beta I$, then

$$\lambda_{\max}(\nabla^2 \varphi_\varepsilon(x)) \leq ((L_\varepsilon + \varepsilon\beta)^{-1} + \varepsilon\alpha)^{-1}$$

Bounds on Hessians

$$\nabla^2 \varphi_\varepsilon(x) \preceq \mathbb{E}_{\pi_\varepsilon^x}\left[\left(\left(\mathbb{E}_{\pi_\varepsilon^Y}[\nabla^2 \varphi_\varepsilon(X) + \varepsilon \nabla^2 V(X)]\right)^{-1} + \varepsilon \nabla^2 W(Y)\right)^{-1}\right]$$

Define $L_\varepsilon := \sup_x \lambda_{\max}(\nabla^2 \varphi_\varepsilon(x))$ and use $\nabla^2 W(y) \succeq \alpha I, \nabla^2 V(x) \preceq \beta I$, then

$$L_\varepsilon \leq ((L_\varepsilon + \varepsilon\beta)^{-1} + \varepsilon\alpha)^{-1}$$

Bounds on Hessians

$$\nabla^2 \varphi_\varepsilon(x) \preceq \mathbb{E}_{\pi_\varepsilon^x}\left[\left(\left(\mathbb{E}_{\pi_\varepsilon^Y}[\nabla^2 \varphi_\varepsilon(X) + \varepsilon \nabla^2 V(X)]\right)^{-1} + \varepsilon \nabla^2 W(Y)\right)^{-1}\right]$$

Define $L_\varepsilon := \sup_x \lambda_{\max}(\nabla^2 \varphi_\varepsilon(x))$ and use $\nabla^2 W(y) \succeq \alpha I, \nabla^2 V(x) \preceq \beta I$, then

$$L_\varepsilon \leq \frac{1}{2}\left(\sqrt{4\beta/\alpha + \beta^2\varepsilon^2} - \varepsilon\beta\right)$$

Bounds on Hessians

We showed: $\nabla^2 \varphi_\varepsilon(x) \preceq \frac{1}{2}\left(\sqrt{4\beta/\alpha + \beta^2\varepsilon^2} - \beta\varepsilon\right) I$

Caffarelli's contraction theorem (2000): $\|\nabla^2 \varphi_0(x)\|_{\mathrm{op}} \leq \sqrt{\beta/\alpha}$

Proof: $\lim_{\varepsilon\to 0} \|\nabla^2 \varphi_\varepsilon(x)\|_{\mathrm{op}} \leq \lim_{\varepsilon\to 0} \frac{1}{2}\left(\sqrt{4\beta/\alpha + \beta^2\varepsilon^2} - \beta\varepsilon\right) = \sqrt{\beta/\alpha}$

(Can be made formal using results from [Nutz, Weisel '21])

Bounds on Hessians

We showed: $\nabla^2 \varphi_\varepsilon(x) \preceq \frac{1}{2}\left(\sqrt{4\beta/\alpha + \beta^2\varepsilon^2} - \beta\varepsilon\right) I$

<u>Extensions</u>

- Symmetric proof for lower bounds for all $\varepsilon > 0$
- Generalization of Caffarelli conditions: $\nabla^2 V \preceq A^{-1}, \nabla^2 W \succeq B^{-1}$ with A, B commuting PD matrices

$$\nabla^2 \varphi_0(x) \preceq A^{-1/2} B^{1/2}$$

Thanks!

Bibliography

- [HR21] J-C. Hütter, and P. Rigollet. *Minimax rates of estimation for smooth optimal transport maps.* Annals of Statistics
- [DGS21] N. Deb, P. Ghosal, and B. Sen. *Rates of Estimation of Optimal Transport Maps using Plug-in Estimators via Barycentric Projections.* NeurIPS 2021
- [MB+21] T. Manole, S. Balakrishnan, J. Niles-Weed, and L. Wasserman. *Plugin Estimation of Smooth Optimal Transport Maps.* ArXiv 2021
- [Gen19] A. Genevay. *Entropy-regularized optimal transport for machine learning*. PhD Thesis, 2019
- [SDF+18] V. Seguy, B. Damodaran, R. Flamary, N. Courty, A. Rolet, and M. Blondel. *Large-scale optimal transport and mapping estimation.* ICLR 2018
- [Pal19] S. Pal. *On the difference between entropic cost and the optimal transport cost.* ArXiv 2019
- [C13] M. Cuturi. *Sinkhorn distances: Lightspeed computation of optimal transport*. NIPS, 2013
- [CRL+2020] L. Chizat, P. Roussillon, F. Léger, F-X. Vialard, G. Peyré. *Faster Wasserstein Distance Estimation with the Sinkhorn Divergence.* NeurIPS, 2020