Welcome

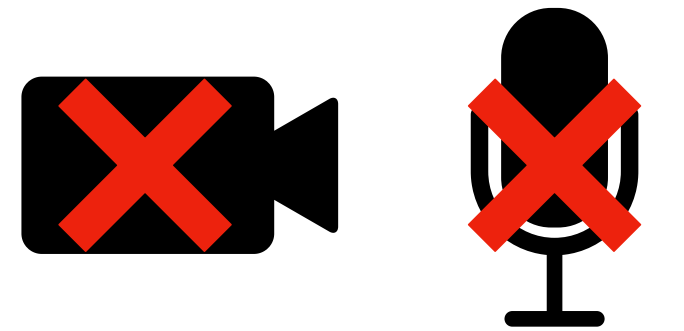
Marcel Klatt

Thesis Defense

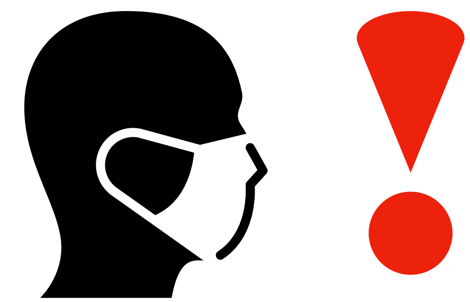
Göttingen, 9th February 2022

Start: 10 am

People attending online: Make sure your camera and microphone are turned off.



People in the Sitzungszimmer: Health and hygiene rules must be followed and face masks are compulsory even in seated areas, regardless of distancing.



GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN

Limit Laws for Empirical Optimal Transport

Marcel Klatt

Thesis Defense

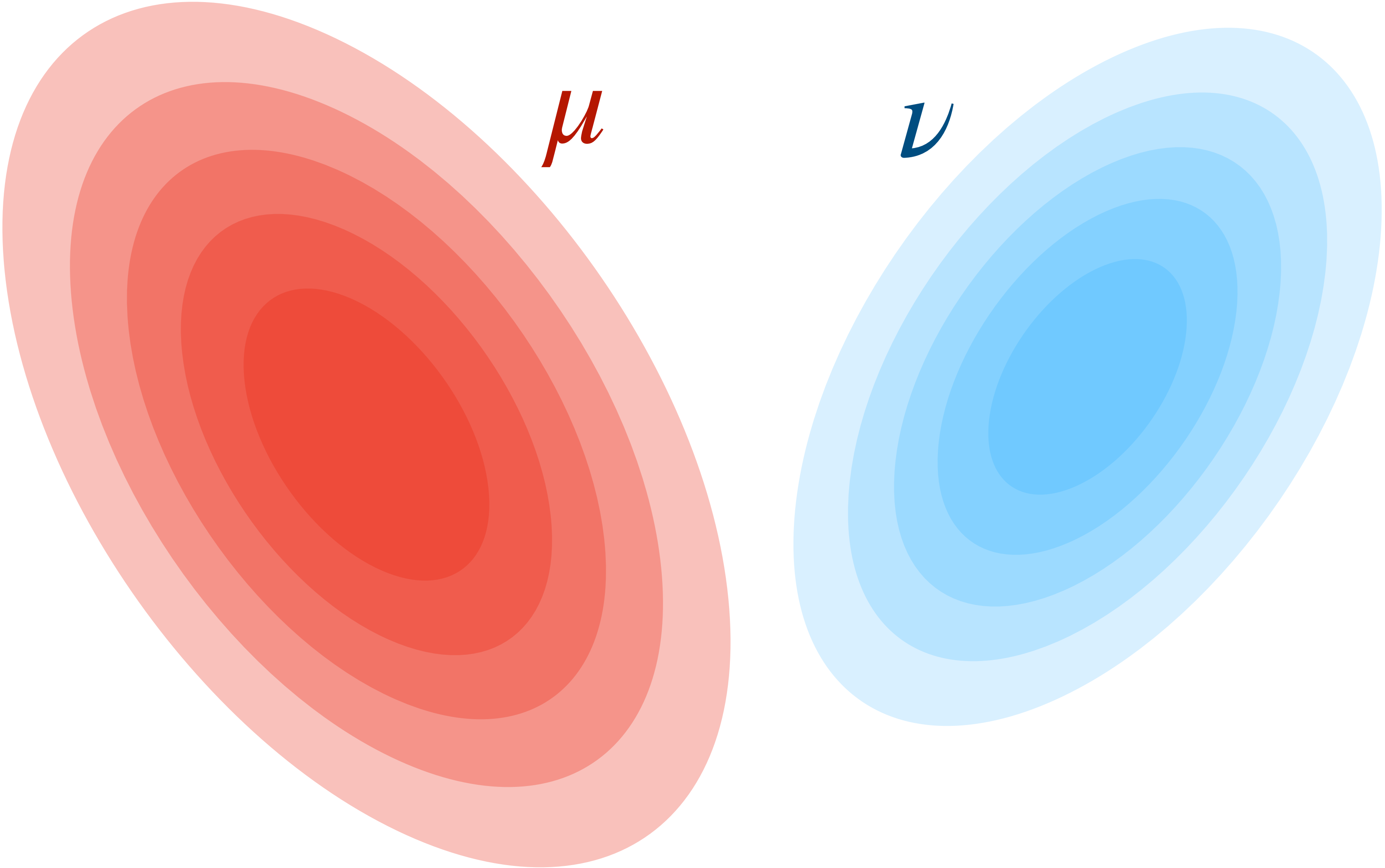
Göttingen, 9th February 2022

Institute for Mathematical Stochastics

RTG 2088

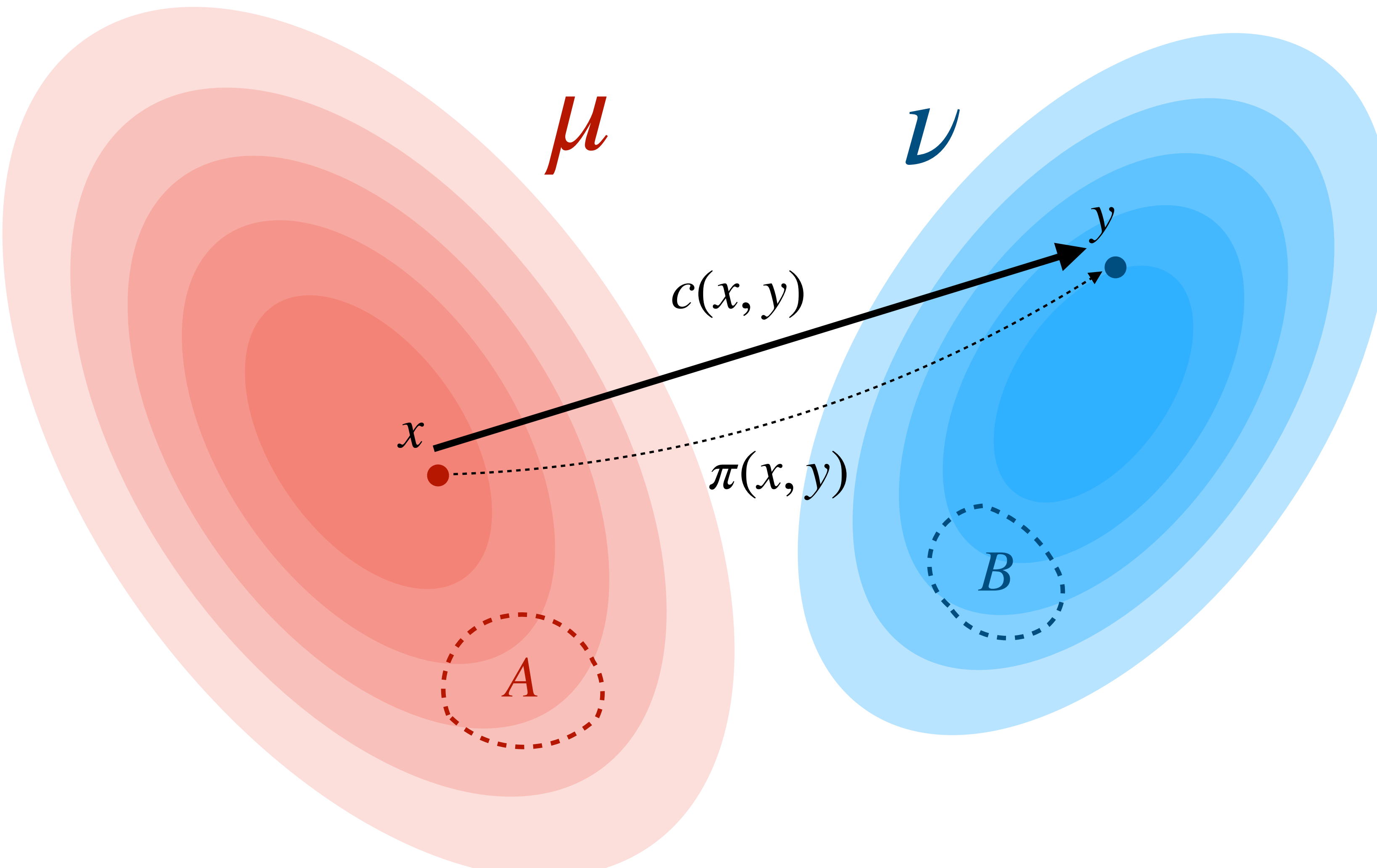
Optimal Transport (OT)

Quantify ***(dis)similarities*** between two probability measures!



Optimal Transport (OT)

Quantify ***(dis)similarities*** between two probability measures!



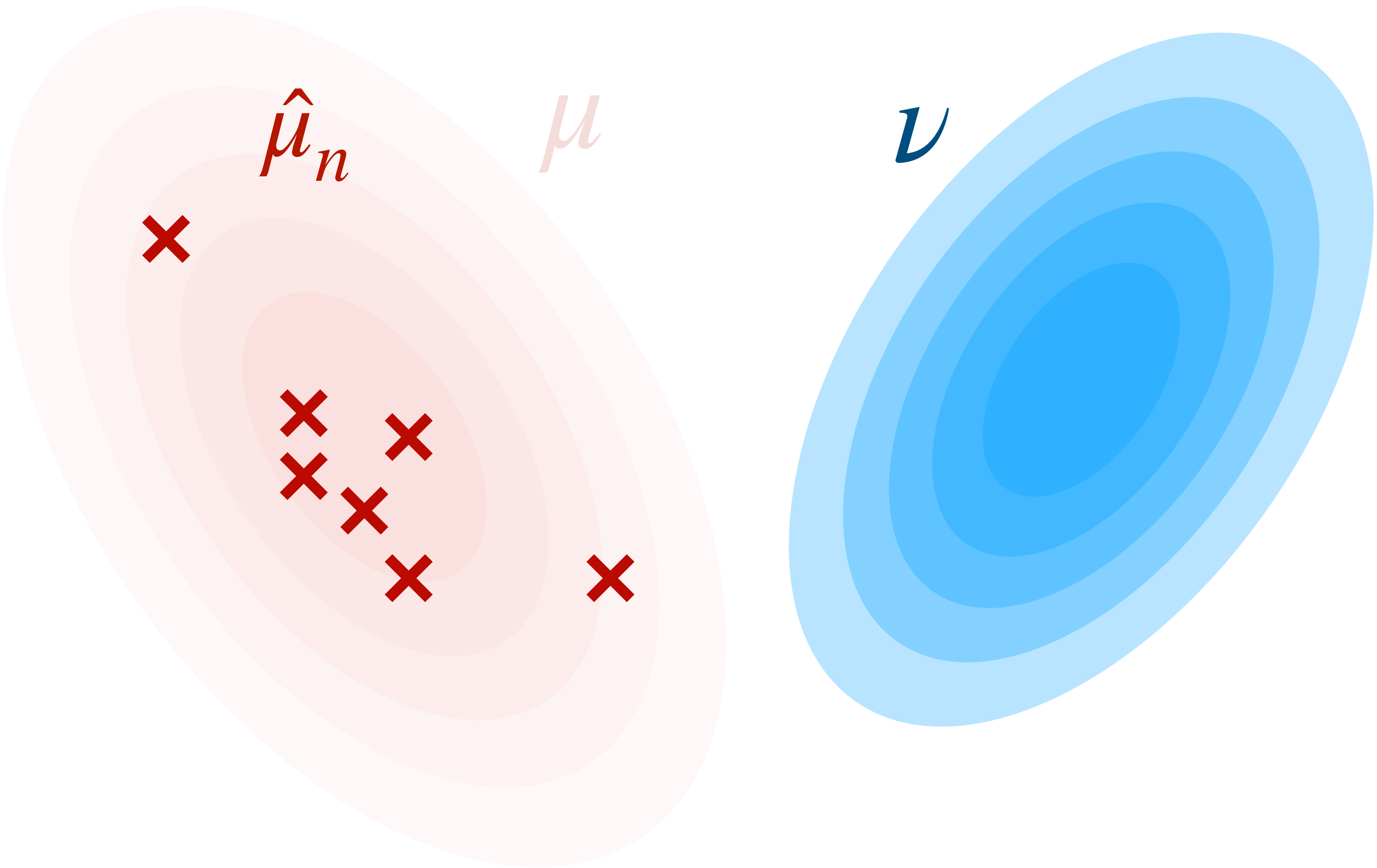
(Dis)similarity equals the ***effort*** to ***transport*** μ to ν.

Monge (1781); Kantorovich (1942)

$$\left.\begin{aligned} \pi(A \times \mathcal{X}) &= \mu(A) \\ \pi(\mathcal{X} \times B) &= \nu(B) \end{aligned}\right\} \pi \in \Pi(\mu, \nu)$$

$$\mathrm{OT}_c(\mu, \nu) = \inf_{\pi \in \Pi(\mu,\nu)} \int_{\mathcal{X}\times\mathcal{X}} c(x, y)\, \mathrm{d}\pi(x, y)$$

Optimal Transport (OT)



Instead of access to μ (and ν), we observe samples:

$$X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mu$$

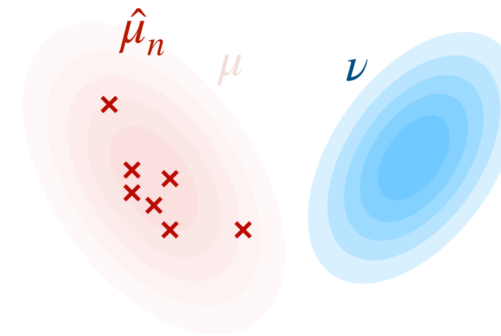
$$\hat{\mu}_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}$$

$$\mathrm{OT}_c(\mu, \nu) = \inf_{\pi \in \Pi(\mu,\nu)} \int_{\mathcal{X}\times\mathcal{X}} c(x,y)\,\mathrm{d}\pi(x,y)$$

↑ ??

$$\mathrm{OT}_c(\hat{\mu}_n, \nu) = \inf_{\pi \in \Pi(\hat{\mu}_n,\nu)} \int_{\mathcal{X}\times\mathcal{X}} c(x,y)\,\mathrm{d}\pi(x,y)$$

Empirical Optimal Transport



- Is $\mathrm{OT}_c(\hat{\mu}_n, \nu)$ a reasonable estimator for $\mathrm{OT}_c(\mu, \nu)$?

Under regularity of c and moment conditions on μ and ν

$$\mathrm{OT}_c(\hat{\mu}_n, \nu) \xrightarrow{n\to\infty} \mathrm{OT}_c(\mu, \nu) \qquad \text{a.s.}$$

Varadarajan (1958); Zolotarev (1975); Bickel & Freedman (1981); Rachev (1982)

- How fast does $\mathrm{OT}_c(\hat{\mu}_n, \nu)$ converge to $\mathrm{OT}_c(\mu, \nu)$?

Depends on regularity of c, μ and ν, e.g.,
bounded support on $\mathcal{X} = \mathbb{R}^d$ and $d \geq 5$,

$$\mathbb{E}\left[\left|\mathrm{OT}_{\|\cdot\|^2}(\hat{\mu}_n, \nu) - \mathrm{OT}_{\|\cdot\|^2}(\mu, \nu)\right|\right] \asymp n^{-2/d}$$

Dudley (1969); Ajtai et al. (1984); Talagrand (1992); Dobrić & Yukich (1995); Fournier & Guillin (2015); Weed & Bach (2019); Manole & Niles-Weed (2021); Hundrieser et al. (2022)

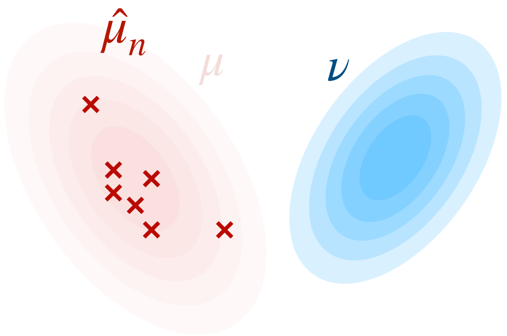
- How does $\mathrm{OT}_c(\hat{\mu}_n, \nu)$ fluctuate asymptotically $(n \to \infty)$ around $\mathrm{OT}_c(\mu, \nu)$?

For $n \to \infty$ and a sequence of real values $r_n \nearrow \infty$,

$$r_n\left(\mathrm{OT}_c(\hat{\mu}_n, \nu) - \mathrm{OT}_c(\mu, \nu)\right) \xrightarrow{\mathcal{D}} Z$$

Munk & Czado (1998); del Barrio et al. (1999, 2005); Freitag et al. (2007); Rippl et al. (2016); Sommerfeld & Munk (2018); Tameling et al. (2019); Berthet et al. (2019, 2020); ; del Barrio & Loubes (2019, 2021); Manole et al. (2021); Sadhu et al. (2021)

Empirical Optimal Transport



$$r_n \left(\mathrm{OT}_c(\hat{\mu}_n, \nu) - \mathrm{OT}_c(\mu, \nu) \right) \xrightarrow{\mathscr{D}} Z$$

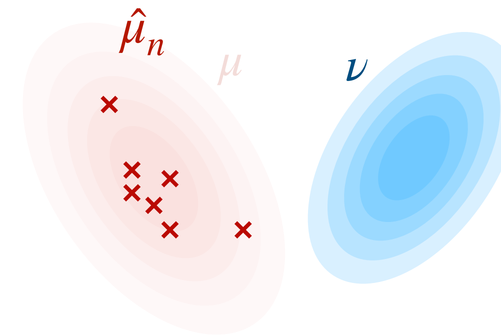
Munk & Czado (1998); del Barrio et al. (1999, 2005); Freitag et al. (2007); Rippl et al. (2016); Sommerfeld & Munk (2018); Tameling et al. (2019); Berthet et al. (2019, 2020); !! ; del Barrio & Loubes (2019, 2021); Manole et al. (2021); Sadhu et al. (2021)

Hundrieser, S., **Klatt, M.**, Staudt, T., and Munk, A. (2021), A unifying approach to central limit theorems for empirical optimal transport. *In preparation*

Klatt, M., Zemel, Y., and Munk, A. (2020), Limit laws for empirical optimal solutions in stochastic linear programs. *Preprint arXiv:2007.13473*

Klatt, M., Tameling, C., and Munk, A. (2020), Empirical regularized optimal transport: Statistical theory and applications. *SIAM Journal on Mathematics of Data Science, 2(2):419-443*

Limit Laws for Empirical OT



Hundrieser, S., **Klatt, M.**, Staudt, T., and Munk, A. (2021), A unifying approach to central limit theorems for empirical optimal transport. *In preparation*

Let $\mathscr{X}$ be a Polish metric space and $\mu, \nu \in \mathscr{P}(\mathscr{X})$.

The cost $c\colon \mathscr{X} \times \mathscr{X} \to \mathbb{R}_+$ is continuous. **(C1)**

The space $\mathscr{X}$ is compact with $\{c(\cdot\,, y) \mid y \in \mathscr{X}\}$ equicontinuous. **(C2)**

The function class $\mathscr{F}_c$ is μ-Donsker. **(E)**

Then, for $n \to \infty$,

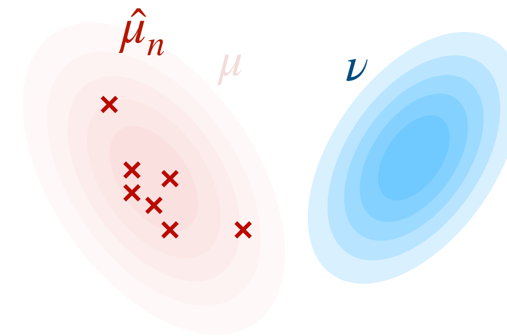
$$\sqrt{n}\left(\mathrm{OT}_c(\hat{\mu}_n, \nu) - \mathrm{OT}_c(\mu, \nu)\right) \xrightarrow{\mathscr{D}} \sup_{f \in S_c(\mu,\nu)} \mathbb{G}_\mu(f)\,.$$

$$\mathscr{F}_c = \left\{ f\colon \mathscr{X} \to \mathbb{R} \mid \exists g\colon \mathscr{X} \to \mathbb{R}, \|g\|_\infty \leq \|c\|_\infty, f(x) = \inf_{y \in \mathscr{X}} c(x,y) - g(y) \right\}$$

$\mathbb{G}_\mu$: A μ-Brownian bridge in the Banach space $l^\infty\left(\mathscr{F}_c\right)$.

$$S_c(\mu,\nu) = \left\{ f \in \mathscr{F}_c \mid \mathrm{OT}_c(\mu,\nu) = \int_{\mathscr{X}} f(x)\, \mathrm{d}\mu(x) + \int_{\mathscr{X}} f^c(y)\, \mathrm{d}\nu(y) \right\}$$

Outline of the Proof



The cost $c\colon \mathcal{X}\times\mathcal{X}\to\mathbb{R}_+$ is continuous. **(C1)**

The space $\mathcal{X}$ is compact with $\{c(\,\cdot\,,y)\mid y\in\mathcal{X}\}$ equicontinuous. **(C2)**

The function class $\mathcal{F}_c$ is μ-Donsker. **(E)**

$$\sqrt{n}\left(\mathrm{OT}_c(\hat{\mu}_n,\nu)-\mathrm{OT}_c(\mu,\nu)\right)\xrightarrow{\mathcal{D}}\sup_{f\in S_c(\mu,\nu)}\mathbb{G}_\mu(f)\,.$$

Kantorovich-Duality yields a functional perspective:

(C1), (C2)

$$\mathrm{OT}_c(\mu,\nu)=\sup_{f\in\mathcal{F}_c}\int_{\mathcal{X}}f(x)\,\mathrm{d}\mu(x)+\int_{\mathcal{X}}f^c(y)\,\mathrm{d}\nu(y)$$

$$f^c(y)=\inf_{x\in\mathcal{X}}c(x,y)-f(x)$$

$\mathrm{OT}_c(\mu,\nu)=\phi(\mu\mid\nu)$ on the subset $\mathcal{P}(\mathcal{X})\subseteq l^\infty(\mathcal{F}_c)$

with ***Hadamard directional derivative:***

(C1), (C2)

$$\phi'_\mu(\Delta\mid\nu)=\sup_{f\in S_c(\mu,\nu)}\Delta(f)\,.$$

Weak convergence of the empirical process in $l^\infty(\mathcal{F}_c)$:

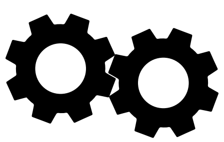
(E)

$$\sqrt{n}\left(\hat{\mu}_n-\mu\right)\xrightarrow{\mathcal{D}}\mathbb{G}_\mu \text{ in } l^\infty(\mathcal{F}_c)$$

with $\mathbb{G}_\mu$ a mean-zero Gaussian process with covariance

$$\mathbb{E}_\mu\left[\mathbb{G}_\mu(f_1)\mathbb{G}_\mu(f_2)\right]=\mu(f_1f_2)-\mu(f_1)\mu(f_2).$$

Delta-Method



$$\sqrt{n}\left(\phi(\hat{\mu}_n\mid\nu)-\phi(\mu\mid\nu)\right)\xrightarrow{\mathcal{D}}\phi'_\mu(\mathbb{G}_\mu\mid\nu)$$

$$S_c(\mu,\nu)=\left\{f\in\mathcal{F}_c\mid \mathrm{OT}_c(\mu,\nu)=\int_{\mathcal{X}}f(x)\,\mathrm{d}\mu(x)+\int_{\mathcal{X}}f^c(y)\,\mathrm{d}\nu(y)\right\}$$

$$\mathcal{F}_c=\left\{f\colon\mathcal{X}\to\mathbb{R}\mid\exists g\colon\mathcal{X}\to\mathbb{R},\ \|g\|_\infty\le\|c\|_\infty,\ f(x)=\inf_{y\in\mathcal{X}}c(x,y)-g(y)\right\}$$

Examples

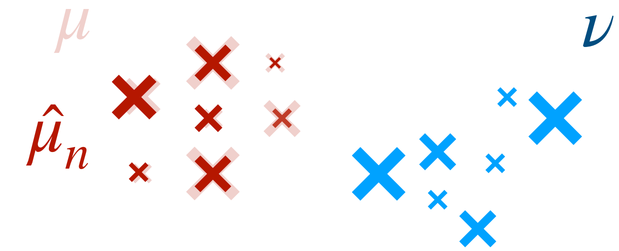
$$\sqrt{n}\left(\mathrm{OT}_c(\hat{\mu}_n, \nu) - \mathrm{OT}_c(\mu, \nu)\right) \xrightarrow{\mathcal{D}} \sup_{f \in S_c(\mu,\nu)} \mathbb{G}_\mu(f).$$

Discrete OT

Bounded cost function c for **(C1)**, **(C2)** and **(E)** to hold.

For weak convergence **(E)**:

$$\sum_{x \in \mathcal{X}} \sqrt{\mu(\{x\})} < \infty$$



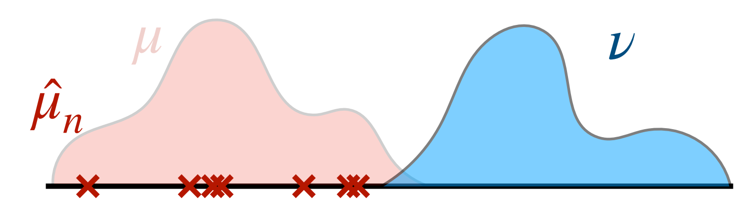
Sommerfeld & Munk (2018); Tameling et al. (2019)

OT on $\mathbb{R}^d$ *for* $d = 1,2,3$

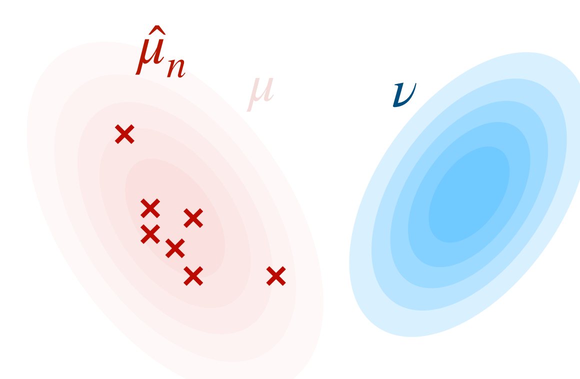
(Reasonable) regularity conditions on c for **(C1)**, **(C2)** and **(E)** to hold.

For weak convergence **(E)**:

$$\sum_{k \in \mathbb{Z}^d} \sqrt{\mu([k, k+1))} < \infty$$



Munk & Czado (1998); Freitag et al. (2007); Berthet & Fort (2019); del Barrio et al. (1999, 2005)



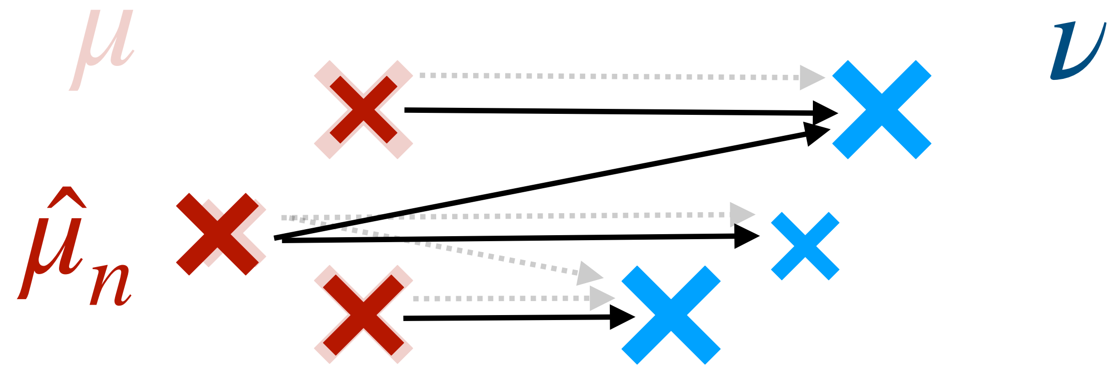
Empirical OT Plan

Klatt, M., Zemel, Y., and Munk, A. (2020), Limit laws for empirical optimal solutions in stochastic linear programs. *Preprint arXiv:2007.13473*

$$\pi \in \arg \min_{\pi \in \Pi(\mu,\nu)} \sum_{i,j}^{N} c_{ij}\pi_{ij}$$

↑ ??

$$\hat{\pi}_n \in \arg \min_{\pi \in \Pi(\hat{\mu}_n,\nu)} \sum_{i,j}^{N} c_{ij}\pi_{ij}$$



Suppose that

Dual solutions for OT are non-degenerate. **(ND)**

Then, for $n \to \infty$,

$$\sqrt{n}\left(\hat{\pi}_n - \pi\right) \xrightarrow{\mathscr{D}} \sum_{k=1}^{K} \mathbf{1}_{\left\{\mathbb{G}_\mu \in H_k\right\}} \pi\left(I_k, [\mathbb{G}_\mu, 0_N]\right).$$

$K =$ | Dual optimal basic solutions |

I_k primal and dual feasible bases

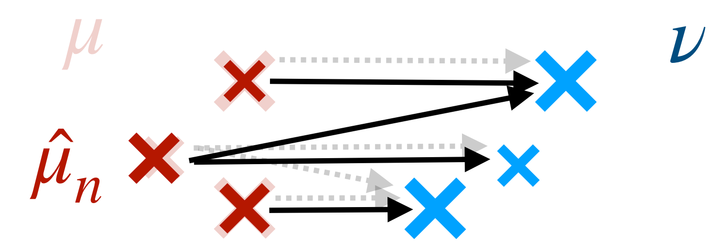
H_k cones of feasible perturbations at μ

$$\sqrt{n}\left(\hat{\mu}_n - \mu\right) \xrightarrow{\mathscr{D}} \mathbb{G}_\mu$$

Outline of the Proof

Dual solutions for OT are non-degenerate. **(ND)**

$$\sqrt{n}\left(\hat{\pi}_n - \pi\right) \xrightarrow{\mathscr{D}} \sum_{k=1}^{K} \mathbf{1}_{\left\{\mathbb{G}_\mu \in H_k\right\}} \pi\left(I_k, [\mathbb{G}_\mu, 0_N]\right).$$

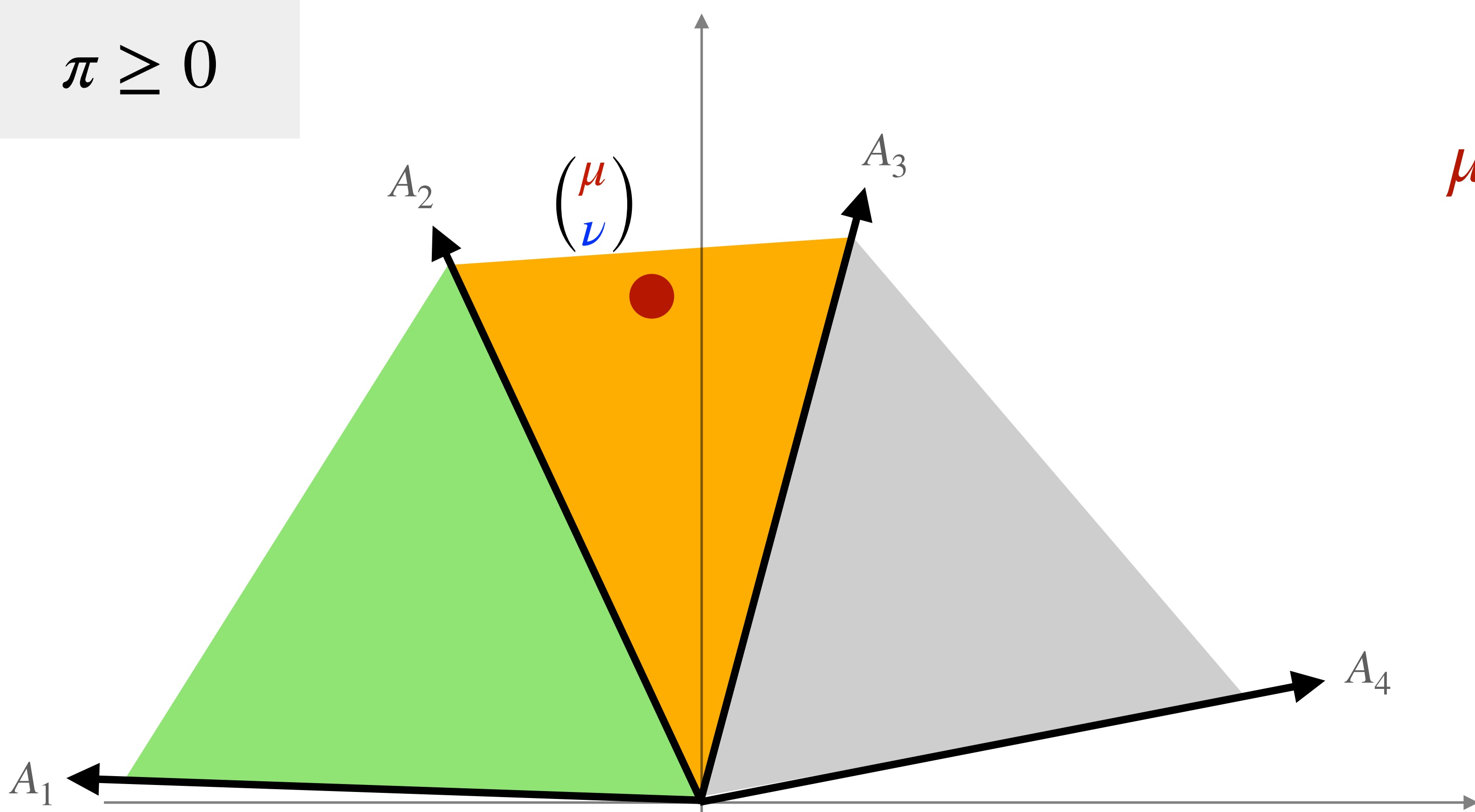


Sensitivity Analysis for linear programs:

$$\min_\pi c^T \pi$$
$$A\pi = \begin{pmatrix}\mu \\ \nu\end{pmatrix}$$
$$\pi \geq 0$$

! !

$$A = [A_1, A_2, A_3, A_4]$$



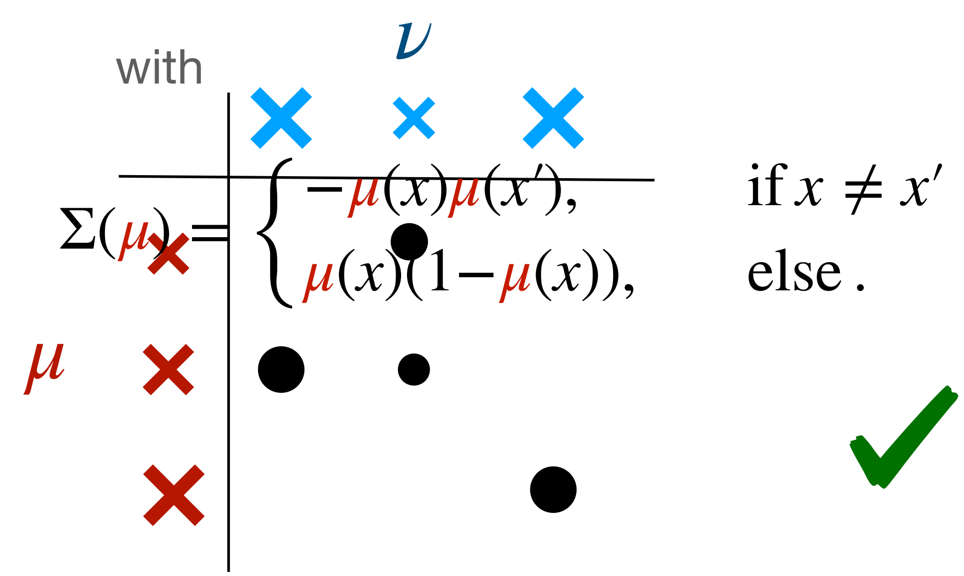
$$\mathrm{pos}(A) = \{x \mid x = A\pi, \pi \geq 0\}$$

Weak convergence of the empirical process:

$$\sqrt{n}\left(\hat{\mu}_n - \mu\right) \xrightarrow{\mathscr{D}} \mathbb{G}_\mu \sim \mathcal{N}\left(0, \Sigma(\mu)\right)$$

with

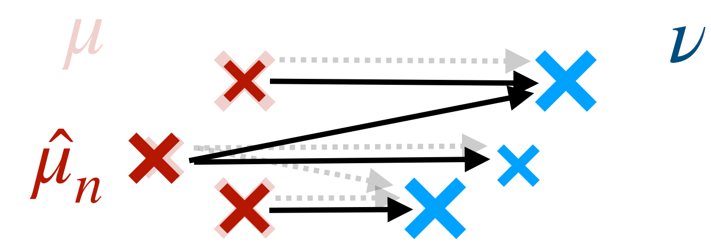
$$\Sigma(\mu) = \begin{cases} -\mu(x)\mu(x'), & \text{if } x \neq x' \\ \mu(x)(1-\mu(x)), & \text{else}. \end{cases}$$



Outline of the Proof

Dual solutions for OT are non-degenerate. **(ND)**

$$\sqrt{n}\left(\hat{\pi}_n - \pi\right) \xrightarrow{\mathcal{D}} \sum_{k=1}^{K} \mathbf{1}_{\left\{\mathbb{G}_{\mu} \in H_k\right\}}\, \pi\left(I_k, [\mathbb{G}_{\mu}, 0_N]\right).$$



Sensitivity Analysis for linear programs:

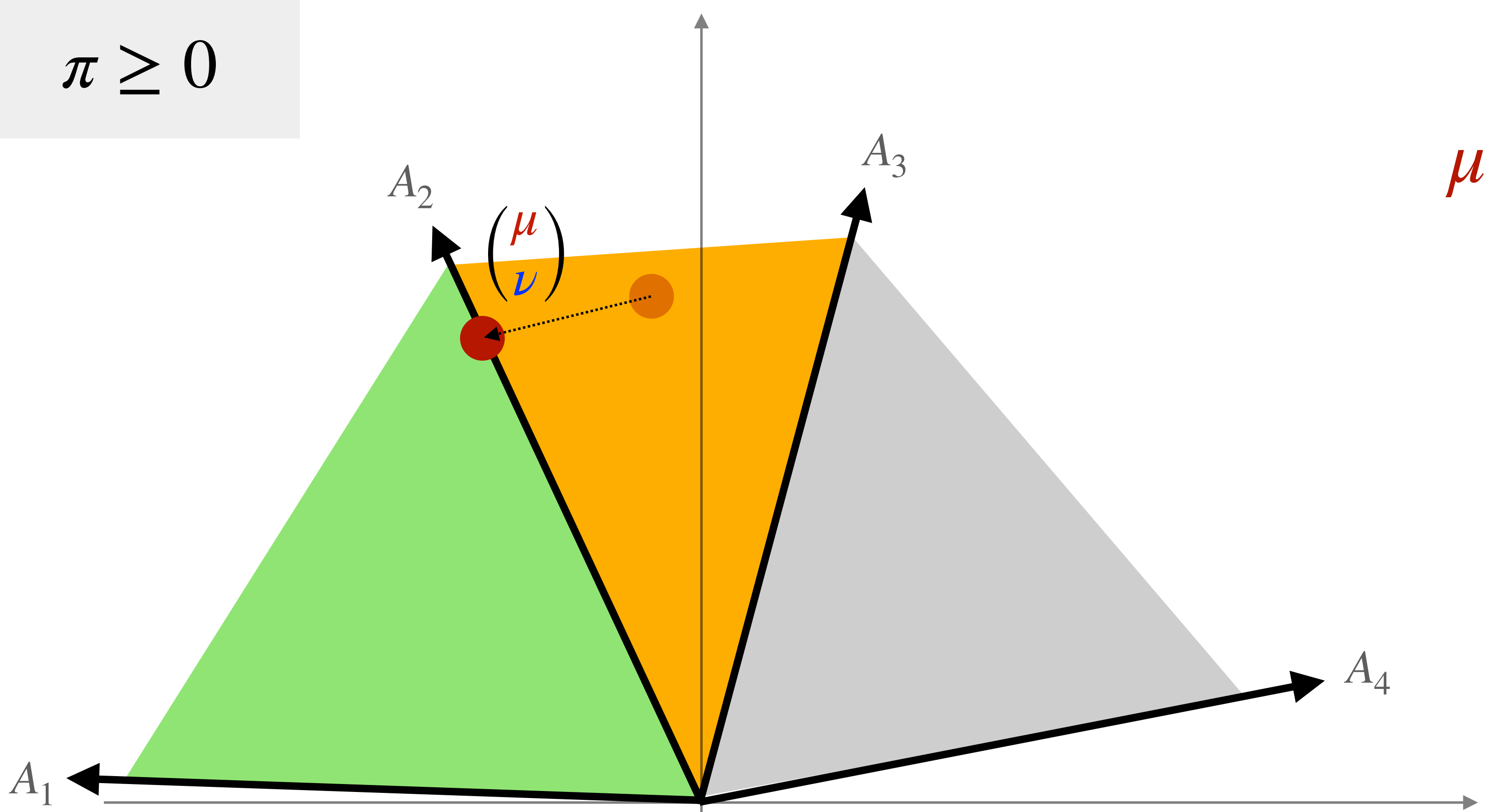
$$\min_{\pi} c^T \pi$$

$$A\pi = \begin{pmatrix} \mu \\ \nu \end{pmatrix}$$

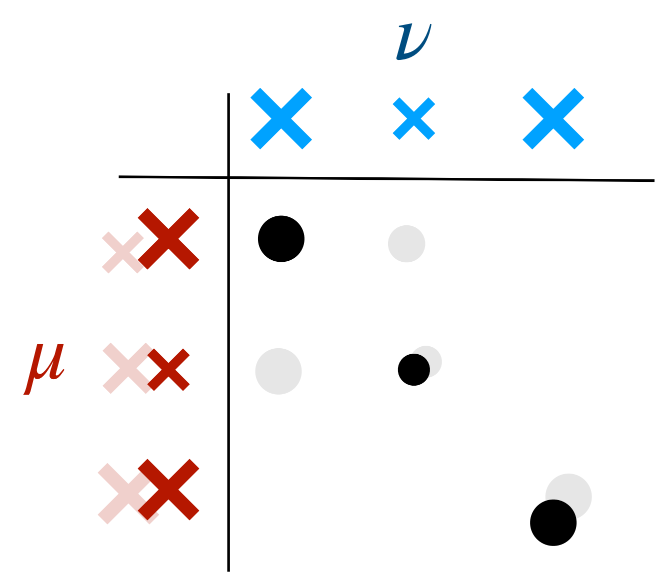
$$\pi \geq 0$$

!

$$A = [A_1, A_2, A_3, A_4]$$



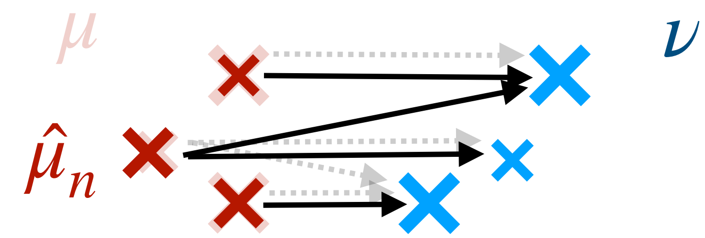
$$\text{pos}(A) = \left\{x \mid x = A\pi, \pi \geq 0\right\}$$



Outline of the Proof

Dual solutions for OT are non-degenerate. **(ND)**

$$\sqrt{n}\left(\hat{\pi}_n - \pi\right) \xrightarrow{\mathcal{D}} \sum_{k=1}^{K} \mathbf{1}_{\left\{\mathbb{G}_\mu \in H_k\right\}}\, \pi\left(I_k, [\mathbb{G}_\mu, 0_N]\right).$$



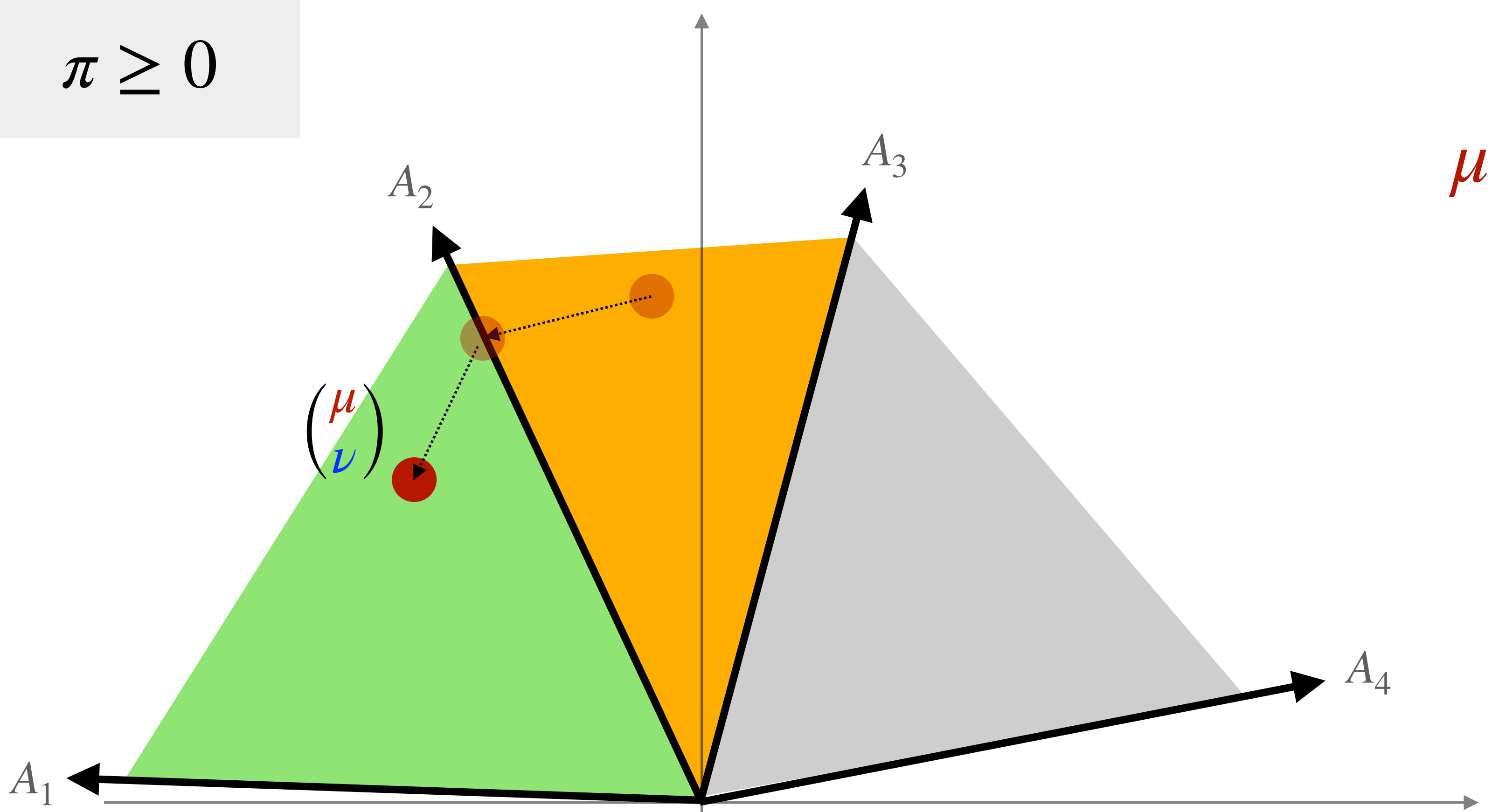
Sensitivity Analysis for linear programs:

$$\min_\pi c^T \pi$$
$$A\pi = \begin{pmatrix} \mu \\ \nu \end{pmatrix}$$
$$\pi \geq 0$$

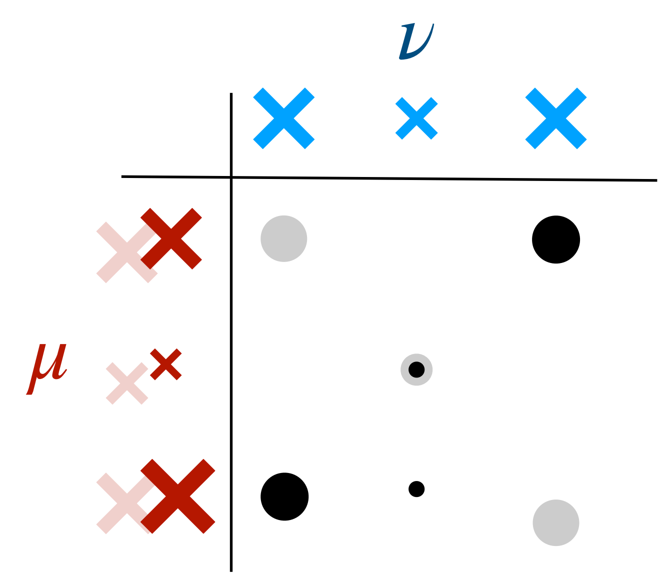
! The OT plan is a *non-local* quantity **!**

! !

$$A = \left[A_1, A_2, A_3, A_4\right]$$



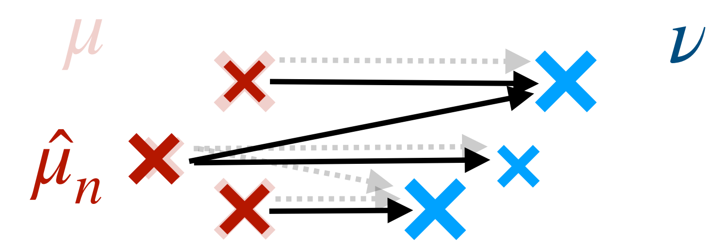
$$\text{pos}(A) = \left\{x \mid x = A\pi, \pi \geq 0\right\}$$



Outline of the Proof

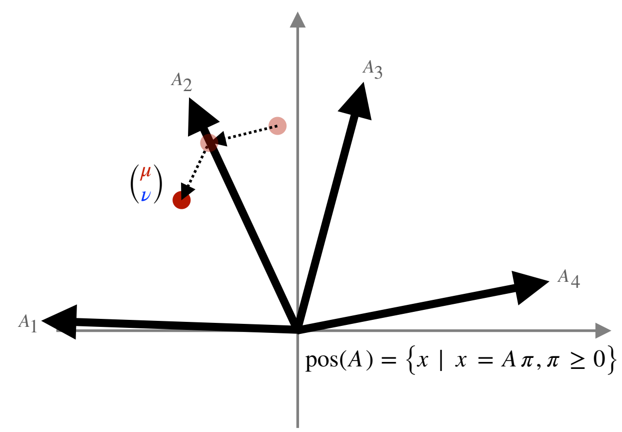
Dual solutions for OT are non-degenerate. **(ND)**

$$\sqrt{n}\left(\hat{\pi}_n - \pi\right) \xrightarrow{\mathscr{D}} \sum_{k=1}^{K} \mathbf{1}_{\left\{\mathbb{G}_\mu \in H_k\right\}}\, \pi\left(I_k, [\mathbb{G}_\mu, 0_N]\right).$$



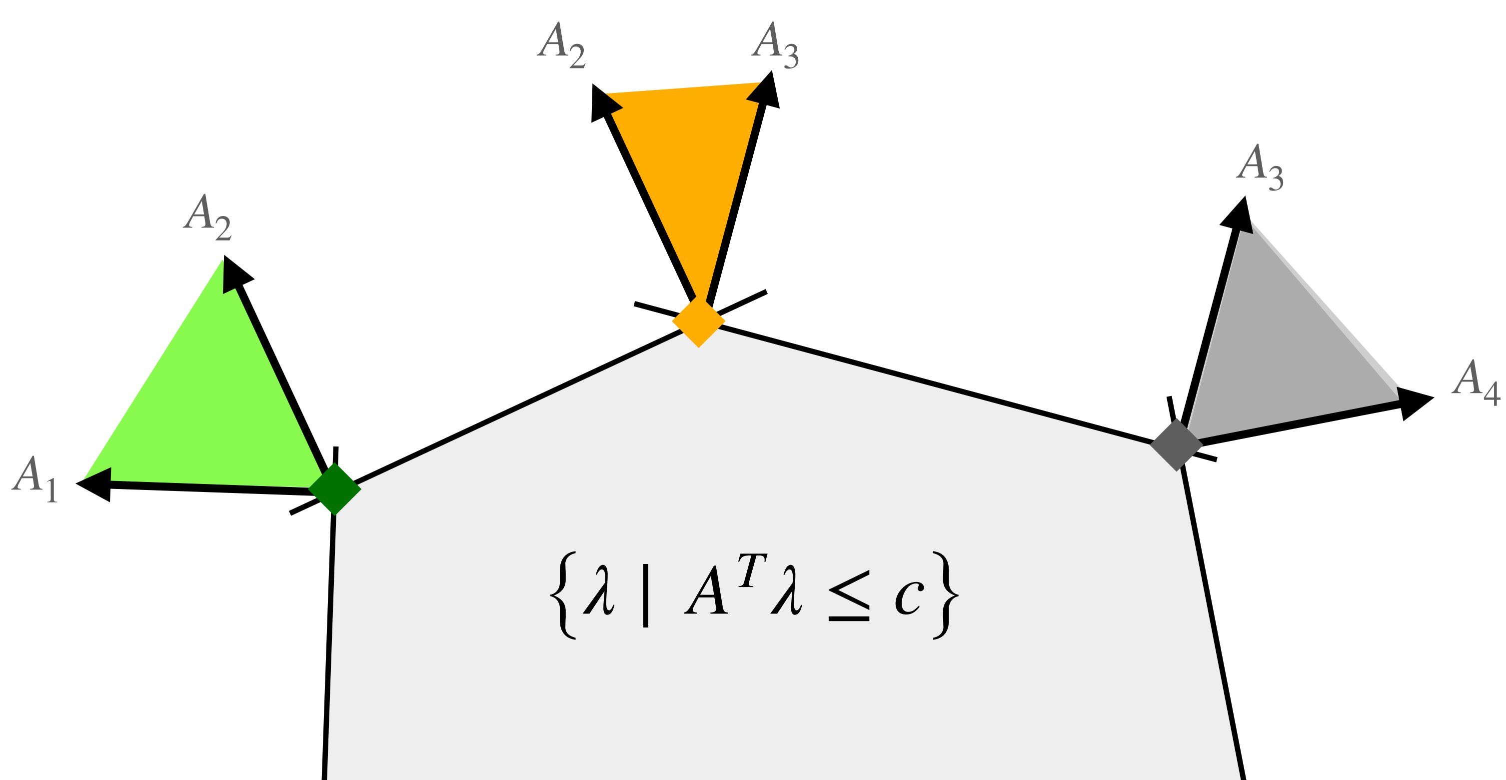
Sensitivity Analysis for linear programs:

$$\min_{\pi} c^T \pi$$
$$A\pi = \begin{pmatrix} \mu \\ \nu \end{pmatrix}$$
$$\pi \geq 0$$



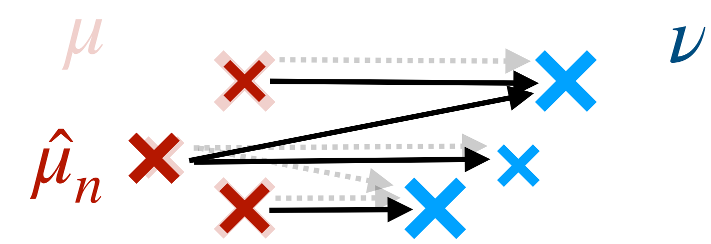
What about assumption **(ND)**?

$$\max_{\lambda \in \mathbb{R}^{2N}} \lambda^T \begin{pmatrix} \mu \\ \nu \end{pmatrix}$$
$$A^T \lambda \leq c$$



Outline of the Proof

Dual solutions for OT are non-degenerate. **(ND)**



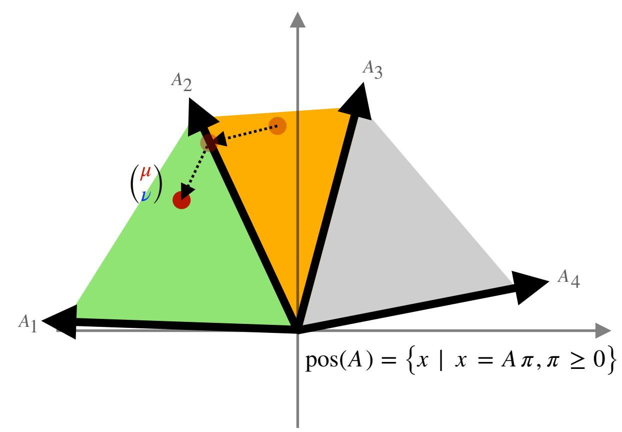
$$\sqrt{n}\left(\hat{\pi}_n - \pi\right) \xrightarrow{\mathscr{D}} \sum_{k=1}^{K} \mathbf{1}_{\left\{\mathbb{G}_\mu \in H_k\right\}}\, \pi\left(I_k, [\mathbb{G}_\mu, 0_N]\right).$$

__Sensitivity Analysis__ for linear programs:

$$\min_{\pi} c^T \pi$$

$$A\pi = \begin{pmatrix} \mu \\ \nu \end{pmatrix}$$

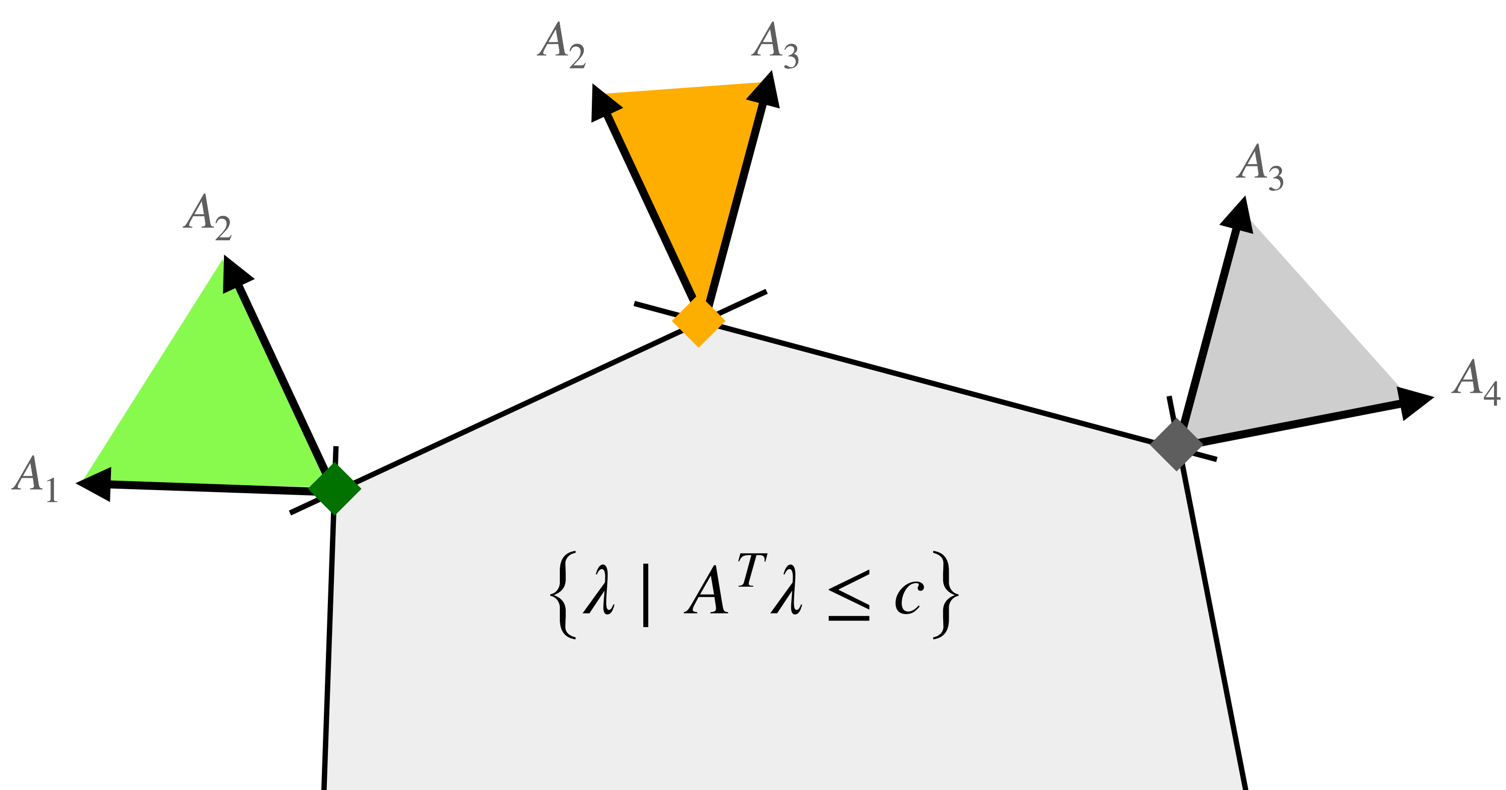
$$\pi \geq 0$$



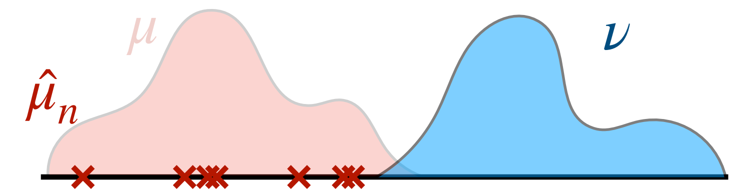
What about assumption **(ND)**?

$$\max_{\lambda \in \mathbb{R}^{2N}} \lambda^T \begin{pmatrix} \mu \\ \nu \end{pmatrix}$$

$$A^T \lambda \leq c$$



Conclusion and Outlook

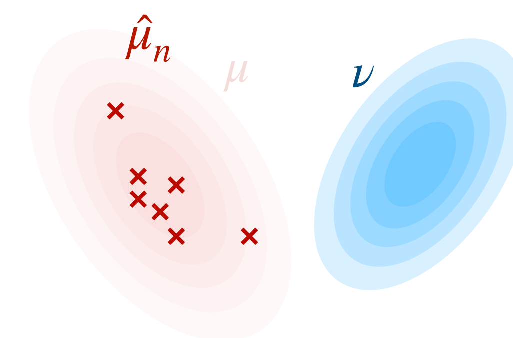


- Asymptotics for the ***empirical OT cost*** are well-understood.

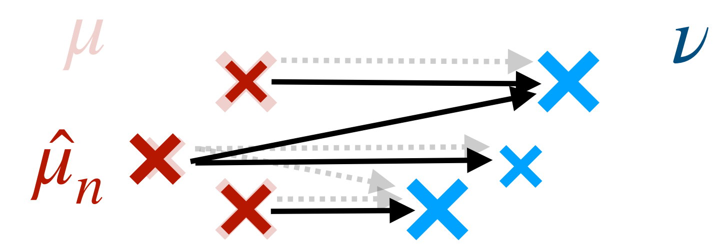
BUT: Limit laws in $\mathbb{R}^d$ for $d \geq 4$ are still challenging.

Curse of dimensionality —> Additional smoothness assumptions?

Goldfeld & Greenwald (2020); Sadhu et. al (2021)
—> Gaussian smoothed OT cost



- Asymptotics for the ***empirical OT plan*** are in their infancy.



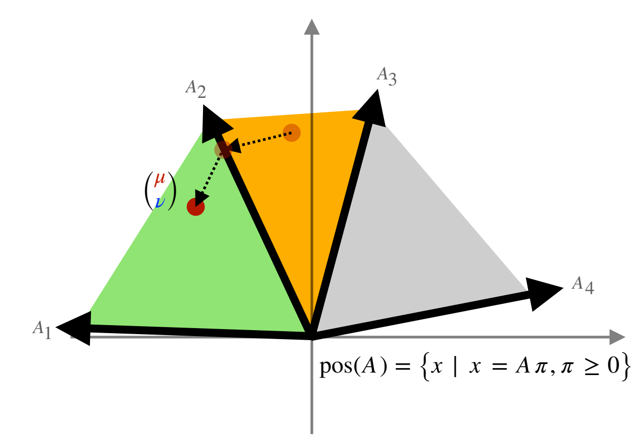
! The OT plan is a ***non-local*** quantity **!**

Beyond discrete settings, recent results focus on the sample complexity inherent in estimating the OT plan.

Manole et al. (2021); Deb et al. (2021)

Statistical different behavior for entropy regularized OT plans.

Klatt et al. (2020)



References

Varadarajan, S. (1958). On the convergence of sample probability distribution. *The Indian Journal of Statistics*, volume 19, pp. 23-26

Dudley, R. M. (1969). The speed of mean Glivenko-Cantelli convergence. *The Annals of Mathematical Statistics*, 40(1), pp. 40-50

Ajtai, M., Komlós, J., and Tusnády, G. (1984). On optimal matchings. *Combinatorica*, 4(4), pp. 259-264

Fournier, N. and Guilin, A. (2015). On the rate of convergence in Wasserstein distance of the empirical measure. *Probability Theory and Related Fields*, 162(3), pp. 707-738

Manole, T. and Niles-Weed, J. (2021). Sharp convergence rates for empirical optimal transport with smooth costs. *preprint arXiv:2106.131181*

Munk, A. and Czado, C. (1998). Nonparametric validation of similar distributions and assessment of goodness of fit. *Journal of the Royal Statistical Society: Series B*, 60(1), pp. 223-241

del Barrio, E., Giné, E., and Matrán, C. (1999). Central limit theorems for the Wasserstein distance between the empirical and the true distributions. *The Annals of Probability*, 27(2), pp. 1009-1071

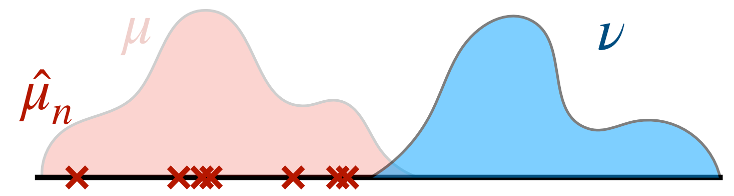
Rippl, T., Munk, A., and Sturm, A. (2016). Limit Laws for the empirical Wasserstein distance: Gaussian distributions. *Journal of Multivariate Analysis*, 151, pp. 90-109

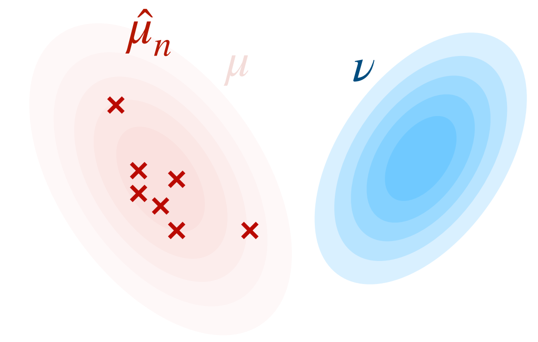
Sommerfeld, M. and Munk, A. (2018). Inference for empirical Wasserstein distances on finite spaces. *Journal of the Royal Statistical Society: Series B*, 80(1), pp. 219-238

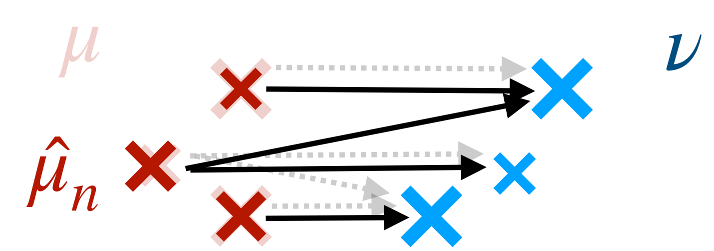
Tameling, C., Sommerfeld, M., and Munk, A. (2019). Empirical optimal transport on countable metric spaces: Distributional limits and statistical applications. *The Annals of Applied Probability*, 29(5), pp. 2744-2781

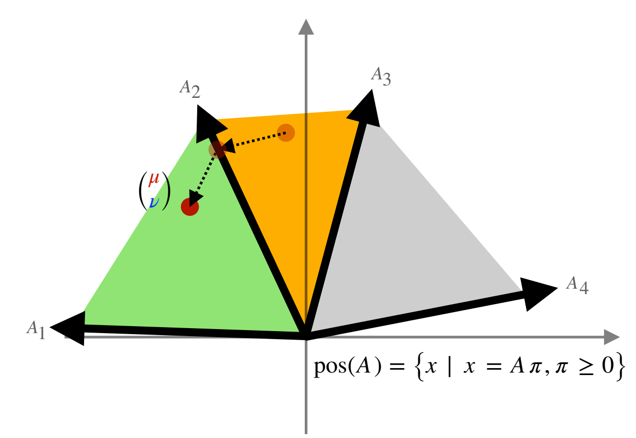
del Barrio, E. and Loubes, J.-M. (2019). Central limit theorems for empirical transportation cost in general dimensions. *The Annals of Probability*, 47(2), pp. 926-951

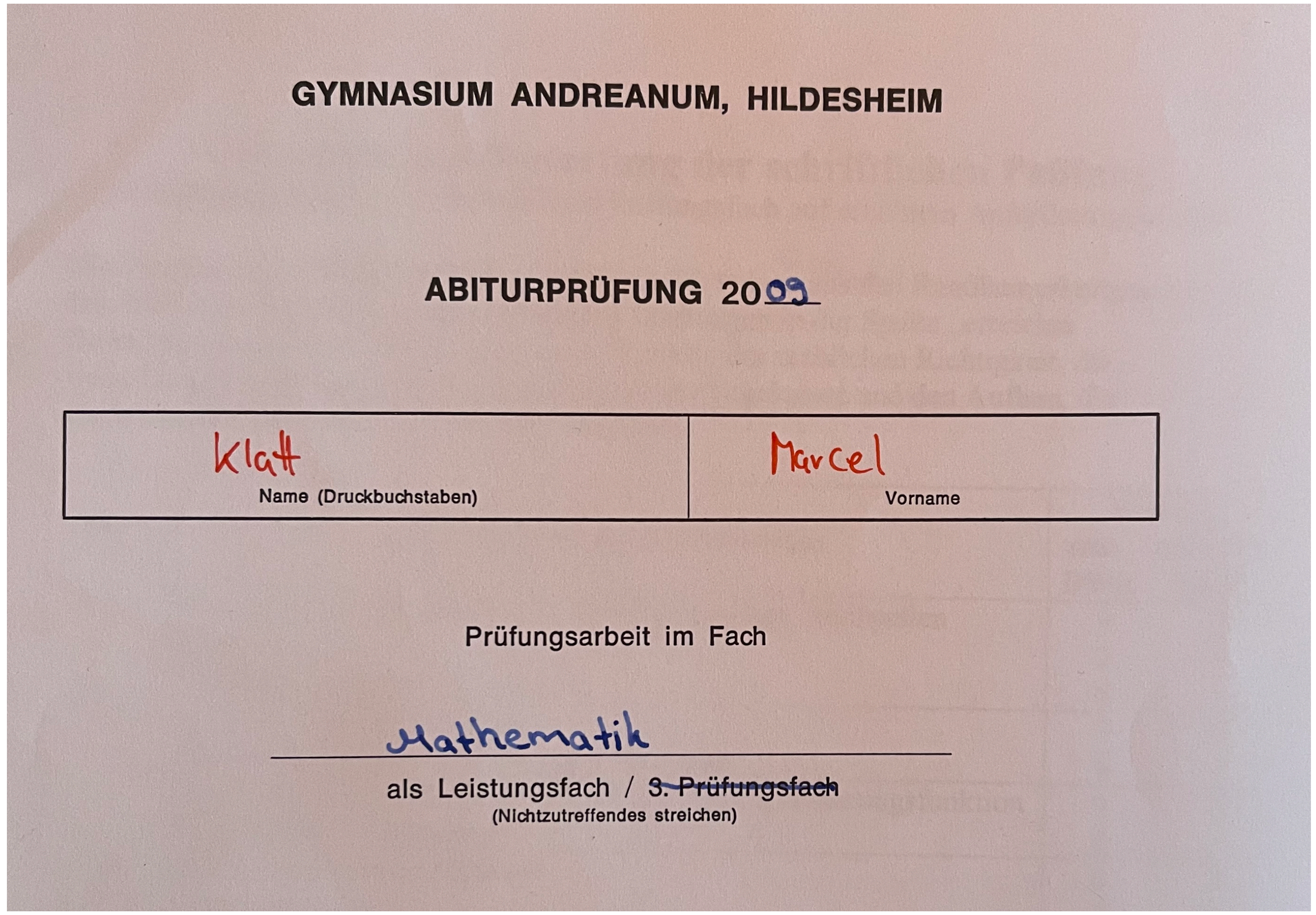
Sadhu, R., Goldfeld, Z., and Kato, K. (2021). Limit distribution theory for the smooth 1-Wasserstein distance with applications. *preprint arXiv:2107.13494*











GYMNASIUM ANDREANUM, HILDESHEIM

ABITURPRÜFUNG 2009

Klatt Name (Druckbuchstaben)	Marcel Vorname

Prüfungsarbeit im Fach

Mathematik

als Leistungsfach / ~~3. Prüfungsfach~~
(Nichtzutreffendes streichen)

[…] Im Analysis und Geometrie-Teil werden Aufgaben auch aus höheren Anforderungsbereichen souverän gelöst. […]

Die Leistungen im Stochastik-Teil weichen von den anderen Bereichen leider ab.

Limit Laws for Empirical OT

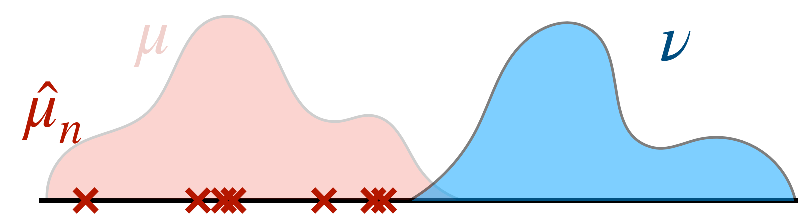
The cost $c : \mathcal{X} \times \mathcal{X} \to \mathbb{R}_+$ is continuous with $\sup_{x,y} |c(x,y)| < \infty$. **(C1')**

The space $\mathcal{X}$ is locally compact with $\{c(\,\cdot\,, y) \mid y \in \mathcal{X}\}$ and $\{c(x, \cdot\,) \mid x \in \mathcal{X}\}$ equicontinuous on $\mathcal{X}$. **(C2')**

OT on $\mathbb{R}^d$

$d = 1$: Bounded cost function, (α, L)-Hölder for $\alpha \in (1/2, 1]$.

$$\left| c(x,y) - c(x',y') \right| \le L \left(|x - x'|^\alpha + |y - y'|^\alpha \right)$$



$\Longrightarrow \mathcal{F}_c$ is a subclass of (α, L)-Hölder functions.

van der Vaart & Wellner (1996)

$\Longrightarrow \mathcal{F}_c$ is μ-Donsker if $\displaystyle\sum_{k \in \mathbb{Z}} \sqrt{\mu([k, k+1))} < \infty$.

---

del Barrio et al. (1999)

If $\displaystyle\int_{-\infty}^{+\infty} \sqrt{F_\mu(t)(1 - F_\mu(t))}\, \mathrm{d}t < \infty$, then

$$\sqrt{n}\,\mathrm{OT}_{|\cdot|}(\hat{\mu}_n, \mu) \xrightarrow{\mathcal{D}} \int_{-\infty}^{+\infty} |\mathbb{B}(F_\mu(t))|\, \mathrm{d}t\,.$$

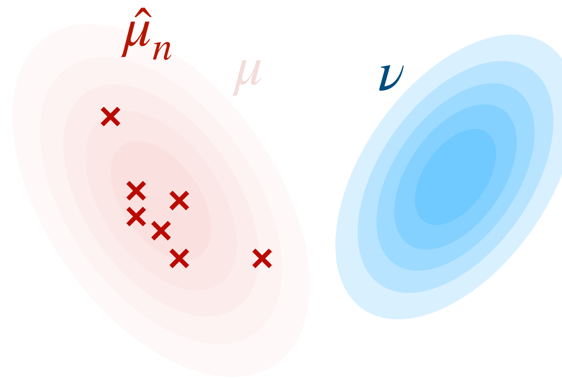
Limit Laws for Empirical OT

The cost $c : \mathscr{X} \times \mathscr{X} \to \mathbb{R}_+$ is continuous with $\sup_{x,y} |c(x,y)| < \infty$. **(C1')**

The space $\mathscr{X}$ is locally compact with $\{c(\cdot, y) \mid y \in \mathscr{X}\}$ and $\{c(x, \cdot) \mid x \in \mathscr{X}\}$ equicontinuous on $\mathscr{X}$. **(C2')**

OT on $\mathbb{R}^d$

$d = 2,3$: Bounded cost function, L-Lipschitz.



$$\left| c(x,y) - c(x',y') \right| \leq L \left(|x - x'| + |y - y'| \right)$$

Suppose that there exists some $\Lambda > 0$ such that for all $k \in \mathbb{Z}^d$ there exists $x_k, y_k \in [k, k+1)$ such that

$$c(\cdot, y) - \Lambda \| \cdot - x_k \|_2^2 \text{ is concave on } [k, k+1) \text{ for all } y \in \mathbb{R}^d,$$

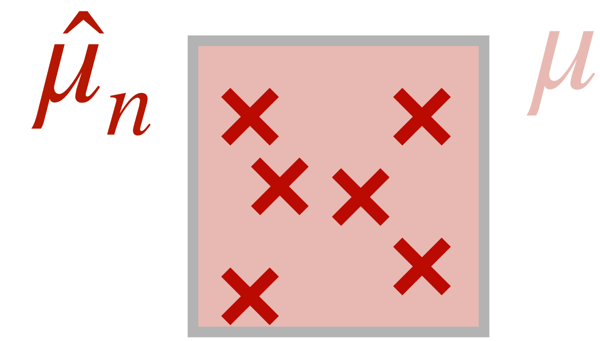
$$c(x, \cdot) - \Lambda \| y_k - \cdot \|_2^2 \text{ is concave on } [k, k+1) \text{ for all } x \in \mathbb{R}^d.$$

$\implies \mathscr{F}_c$ is a subclass of semi-concave and L-Lipschitz functions.

$\implies \mathscr{F}_c$ is μ-Donsker if $\sum_{k \in \mathbb{Z}^d} \sqrt{\mu([k, k+1))} < \infty$.

van der Vaart & Wellner (1996)

High-Dimensional Spaces



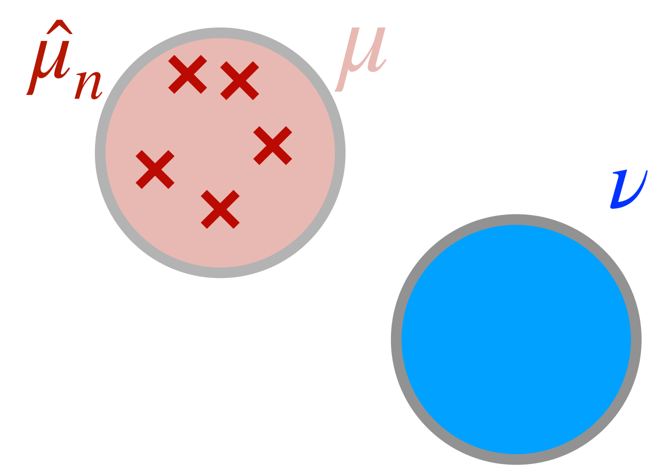
Suppose that μ is a uniform distribution on $[0,1]^d$, then

$$\mathrm{OT}_{\|\cdot\|}(\hat{\mu}_n, \mu) \geq n^{-1/d}\,.$$

Dudley (1969)

Hence, for $d \geq 3$ it holds that

$$\sqrt{n}\,\mathrm{OT}_{\|\cdot\|}(\hat{\mu}_n, \mu) \longrightarrow \infty\,.$$



Suppose that $\mu, \nu \in \mathscr{P}(\mathbb{R}^d)$ have positive density in the interior of their convex support with finite moments of order $4 + \delta$ for some $\delta > 0$, then

$$\sqrt{n}\left(\mathrm{OT}_{\|\cdot\|^2}(\hat{\mu}_n, \nu) - \mathbb{E}\left[\mathrm{OT}_{\|\cdot\|^2}(\hat{\mu}_n, \nu)\right]\right) \xrightarrow{\mathscr{D}} \mathbb{G}_{\mu}(f)\,.$$

del Barrio & Loubes (2019)

Together with

$$\mathbb{E}\left[\mathrm{OT}_{\|\cdot\|^2}(\hat{\mu}_n, \nu)\right] - \mathrm{OT}_{\|\cdot\|^2}(\mu, \nu) \geq n^{-2/d},$$

Manole & Niles-Weed (2021)

it follows for $d \geq 5$ that

$$\sqrt{n}\left(\mathrm{OT}_{\|\cdot\|^2}(\hat{\mu}_n, \nu) - \mathrm{OT}_{\|\cdot\|^2}(\mu, \nu)\right) \longrightarrow \infty\,.$$

Empirical Process Theory (1)

van der Vaart & Wellner (1996)

Let $\mathscr{F}$ be a class of measurable functions from $\mathscr{X}$ to $\mathbb{R}$ such that $\mu(f^2) < \infty$, for every $f \in \mathscr{F}$ and

$$\sup_{f \in \mathscr{F}} |f(x) - \mu(f)| < \infty, \qquad \text{for all } x \in \mathscr{X}.$$

Then,

$$\mathbb{G}_{\mu,n} := \sqrt{n}\left(\mu - \hat{\mu}_n\right) \xrightarrow{\mathscr{D}} \mathbb{G}_\mu \qquad \text{in } l^\infty(\mathscr{F}).$$

$$\Longleftrightarrow$$

There exists a semi-metric $d(\cdot, \cdot)$ on $\mathscr{F}$ such that $(\mathscr{F}, d)$ is totally bounded and

$$\lim_{\delta \to 0} \limsup_{n \to \infty} \mathbb{P}\left(\sup_{d(f,g) \le \delta,\, f,g \in \mathscr{F}} |\mathbb{G}_{\mu,n}(f - g)| > \epsilon \right) = 0, \qquad \text{for every } \epsilon > 0.$$

Empirical Process Theory (2)

van der Vaart & Wellner (1996)

$$\mathbb{P}\left(\sup_{d(f,g)\leq\delta,\, f,g\in\mathcal{F}} |\mathbb{G}_{\mu,n}(f-g)| > \epsilon \right) \leq \frac{1}{\epsilon}\mathbb{E}\left[\sup_{d(f,g)\leq\delta,\, f,g\in\mathcal{F}} |\mathbb{G}_{\mu,n}(f-g)| \right]$$

! Control the expectation via ***maximal inequalities*** (chaining, covering numbers) **!**

$$\int_0^1 \sqrt{\log \mathcal{N}_{[]}(\epsilon, \mathcal{F}\cup\{0\}, L_2(\mu))}\, \mathrm{d}\epsilon < \infty$$

or

$$\int_0^1 \sup_Q \sqrt{\log \mathcal{N}(\epsilon\|F\|_{Q,2}, \mathcal{F}, L_2(Q))}\, \mathrm{d}\epsilon < \infty$$

$\Rightarrow \mathcal{F}$ is μ-Donsker.

Empirical Process Theory (3)

van der Vaart & Wellner (1996)

Let $\mathscr{X} = \bigcup_{j=1}^{\infty} \mathscr{X}_j$ be a partition into measurable sets and let $\mathscr{F}_j = \mathscr{F}\mathbf{1}_{\mathscr{X}_j}$. Suppose that for each j the function class $\mathscr{F}_j$ is μ-Donsker such that

$$\mathbb{E}_\mu\left[\|\mathbb{G}_{\mu,n}\|_{\mathscr{F}_j}\right] \leq C c_j$$

for a constant C not depending on j or n. If $\sum_{j=1}^{\infty} c_j < \infty$ and $\mu(F) < \infty$, then the class $\mathscr{F}$ is μ-Donsker.

! Control the expectation via ***maximal inequalities*** (chaining, covering numbers) !

$$\mathbb{E}_\mu\left[\|\mathbb{G}_{\mu,n}\|_{\mathscr{F}_j}\right] \lesssim \int_0^1 \sqrt{1 + \log \mathscr{N}_{[]}(\epsilon\|F_j\|_{\mu,2}, \mathscr{F}, L_2(\mu))}\, \mathrm{d}\epsilon \; \|F_j\|_{\mu,2}$$

$$\sum_{j=1}^{\infty} \|F_j\|_{\mu,2} \lesssim \sum_{j=1}^{\infty} \sqrt{\mu(\mathscr{X}_j)}$$

Hadamard Directional Differentiability

Shapiro (1991); Dümbgen (1993); Römisch (2004)

A map $\Phi: D_\Phi \subset D \to F$ is called ***Hadamard directional differentiable*** at $\theta \in D_\Phi$ if there exists a mapping $\Phi'_\theta: D \to F$ such that

$$\lim_{n\to\infty} \frac{\Phi(\theta + t_n h_n) - \Phi(\theta)}{t_n} = \Phi'_\theta(h)$$

holds for any $h \in D$ and any sequence $t_n \searrow 0$ and h_n with the property that $\theta + t_n h_n \in D_\Phi$ and converging to $h \in D$.

Additionally: …*tangentially* to $\Theta \subset D_\Phi$ if $h_n = \dfrac{\theta_n - \theta}{t_n}$ with $\theta_n \subset \Theta$ and converging to h.

! The derivative $\Phi'_\theta(\,\cdot\,)$ is not required to be linear !

Bootstrap

Dümbgen (1993); Fang & Santos (2019)

! The derivative $\Phi'_\theta(\,\cdot\,)$ is not necessarily required to be linear **!**

⇒***Caution!*** The naive n-out-of-n bootstrap might fail.

$$\sqrt{n}\,\big(\Phi(\mathbb{P}_n^\star) - \Phi(\mathbb{P}_n)\big)$$

$$= \sqrt{n}\,\big(\Phi(\mathbb{P}_n^\star - \mathbb{P}_n + \mathbb{P}_n - \mathbb{P} + \mathbb{P}) - \Phi(\mathbb{P}_n)\big)$$

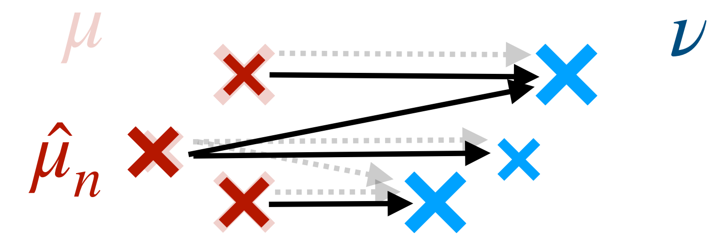
$$\approx \sqrt{n}\,\big(\Phi(\mathbb{P}) - \Phi'_{\mathbb{P}}(\mathbb{P}_n^\star - \mathbb{P}_n + \mathbb{P}_n - \mathbb{P}) - \Phi(\mathbb{P}_n)\big) \quad \xrightarrow{\mathscr{D}} \Phi'_{\mathbb{P}}(\mathbb{G}_1 + \mathbb{G}_2) - \Phi'_{\mathbb{P}}(\mathbb{G}_2)$$

$$\sqrt{n}\Phi'_{\mathbb{P}}(\mathbb{P}_n^\star - \mathbb{P}_n + \mathbb{P}_n - \mathbb{P}) \xrightarrow{\mathscr{D}} \Phi'_{\mathbb{P}}(\mathbb{G}_1 + \mathbb{G}_2)$$

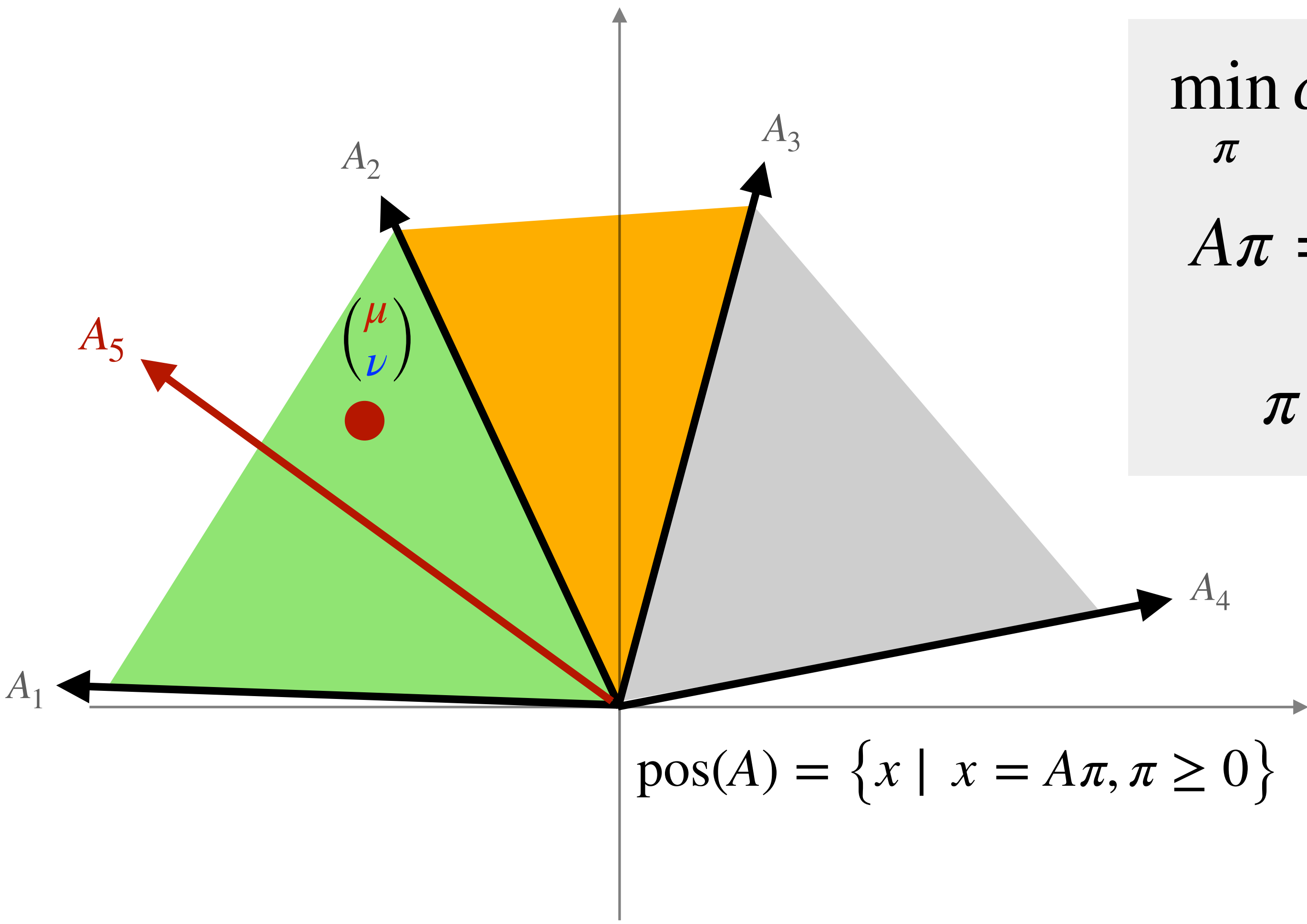
$$\sqrt{n}\,\big(\Phi(\mathbb{P}_n) - \Phi(\mathbb{P})\big) \xrightarrow{\mathscr{D}} \Phi'_{\mathbb{P}}(\mathbb{G}_2)$$

Empirical OT Plan

Dual solutions for OT are non-degenerate. **(ND)**



What if **(ND)** is not satisfied?



$$\text{pos}(A) = \{x \mid x = A\pi, \pi \geq 0\}$$

$$\min_{\pi} c^T \pi$$

$$A\pi = \binom{\mu}{\nu}$$

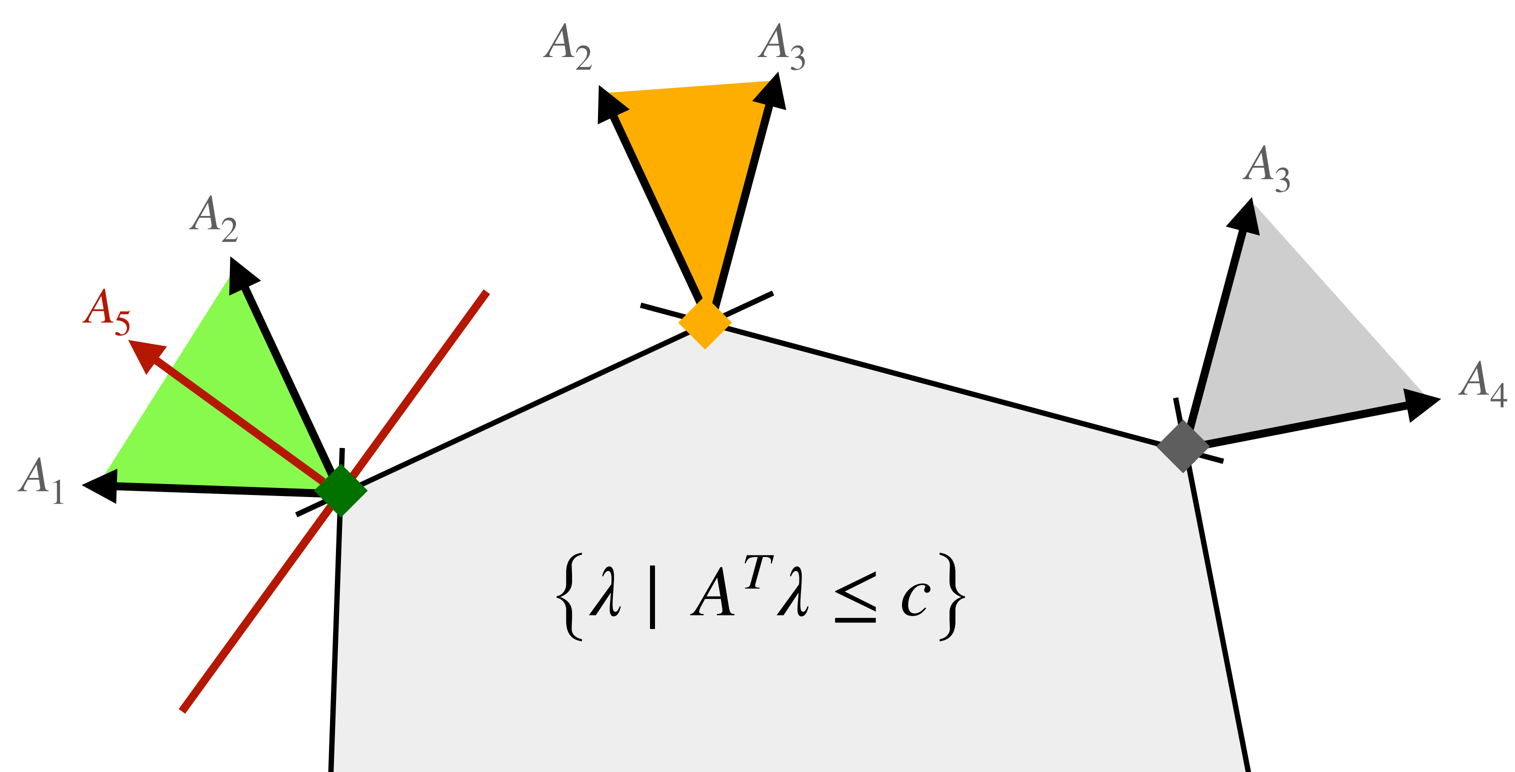
$$\pi \geq 0$$

$$A = [A_1, A_2, A_3, A_4, A_5]$$

! ! ! !

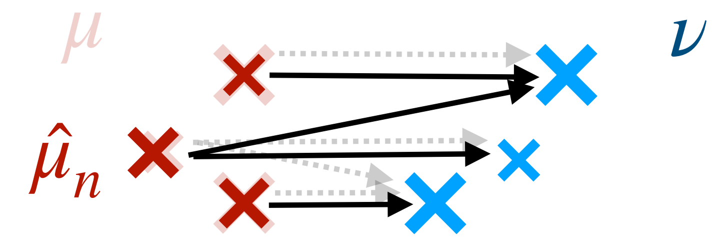
$$\max_{\lambda \in \mathbb{R}^{2N}} \lambda^T \binom{\mu}{\nu}$$

$$A^T \lambda \leq c$$



Entropy OT

Klatt et al. (2020)

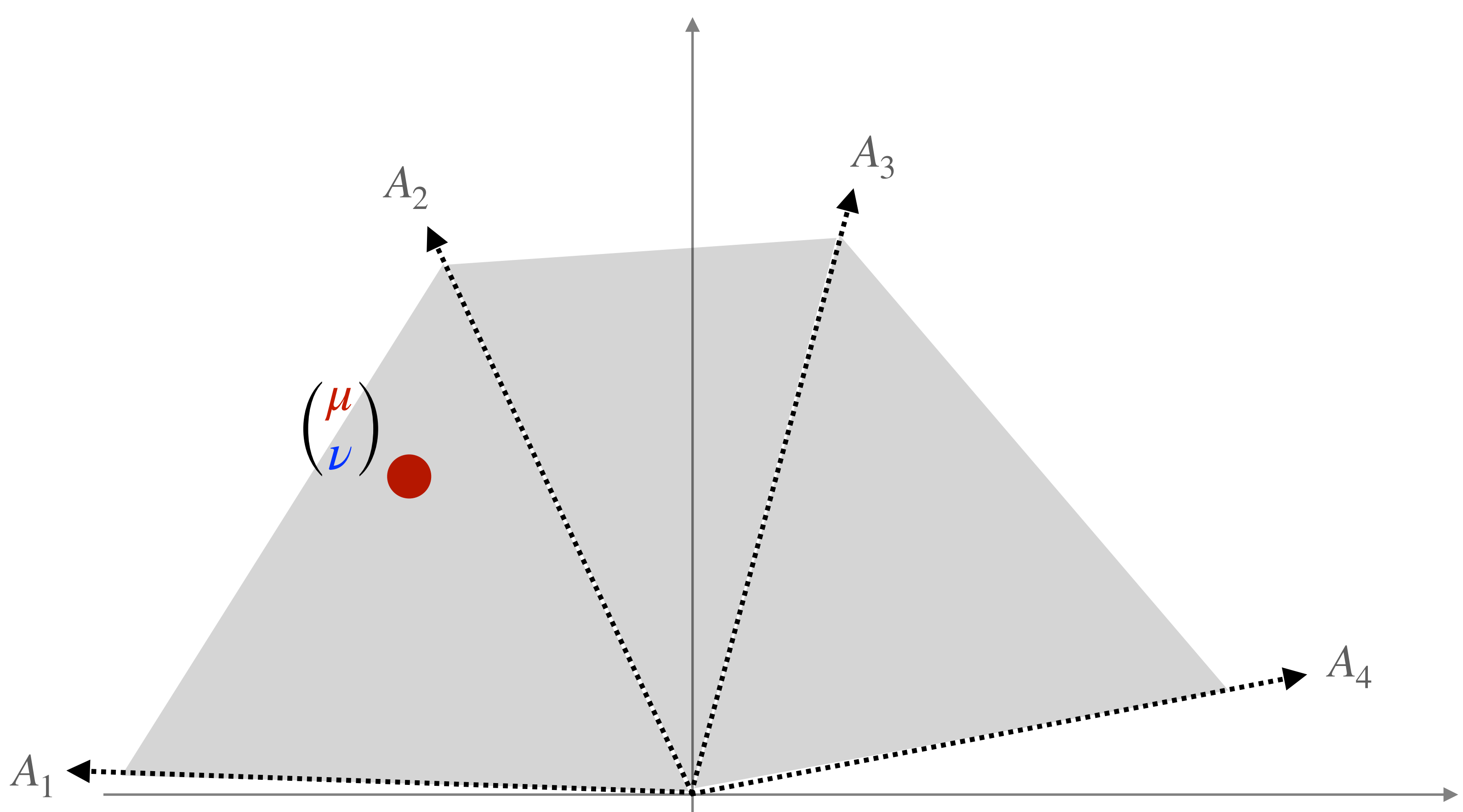


$$\pi_\lambda = \arg \min_{\pi \in \Pi(\mu,\nu)} \sum_{i,j}^{N} c_{ij}\pi_{ij} - \lambda E(\pi), \quad \lambda > 0$$

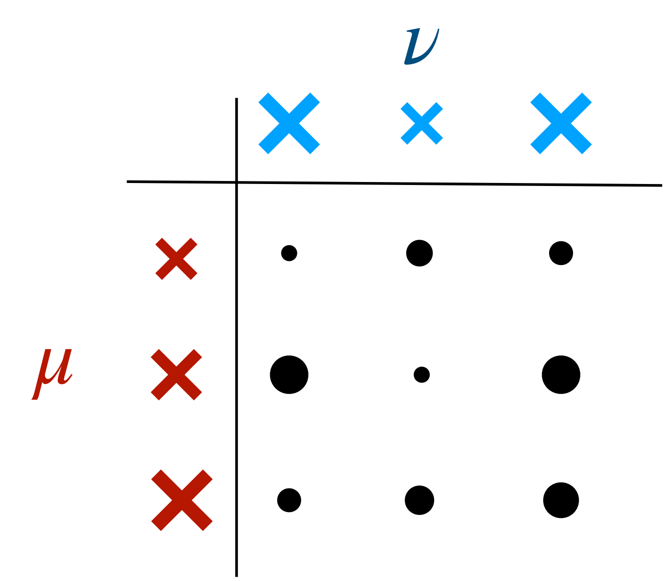
Entropy: $E(\pi) = - \sum_{i,j} \pi_{ij} \log(\pi_{ij})$

Then, for $n \to \infty$,

$$\sqrt{n}\left(\hat{\pi}_\lambda - \pi_\lambda\right) \xrightarrow{\mathscr{D}} Z \sim \mathcal{N}_{N^2}\left(0, \Sigma_\lambda(\mu \mid \nu)\right).$$



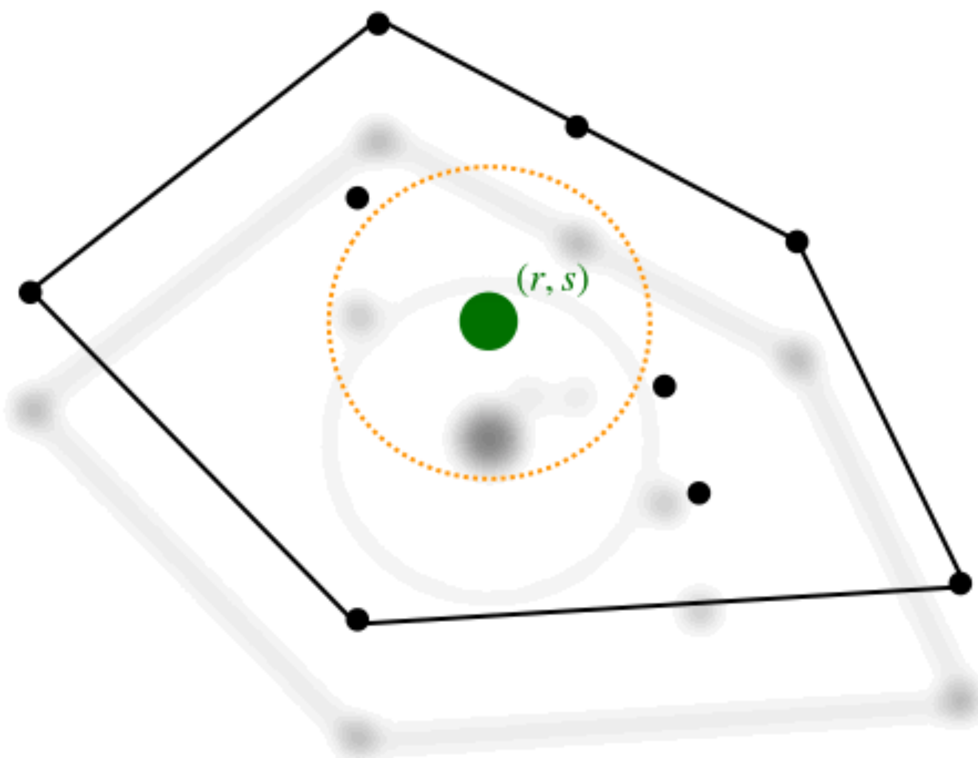
$$\text{pos}(A) = \left\{x \mid x = A\pi, \pi \geq 0\right\}$$

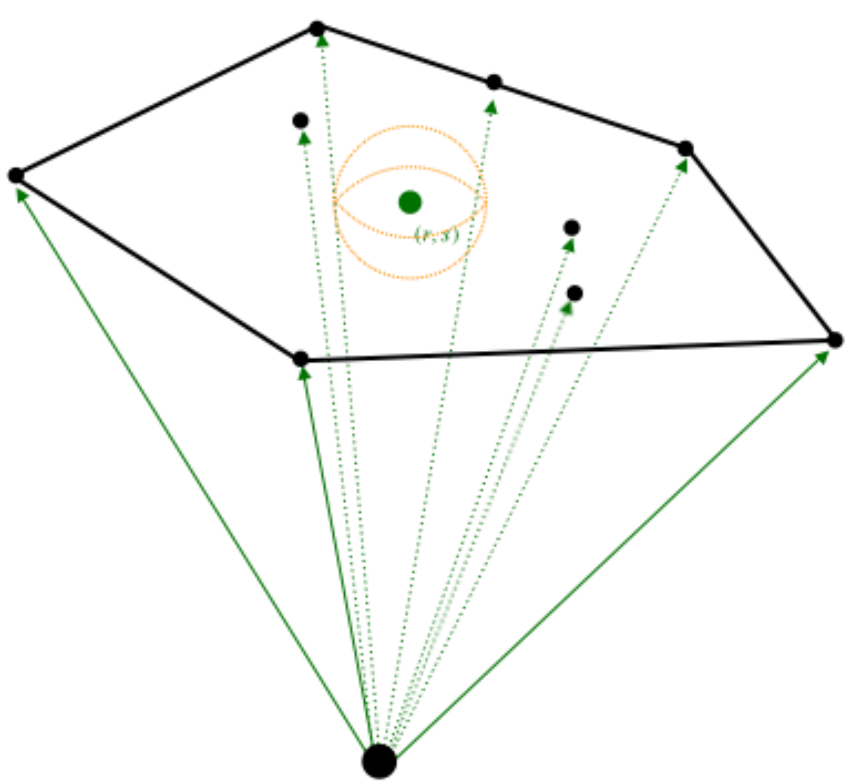


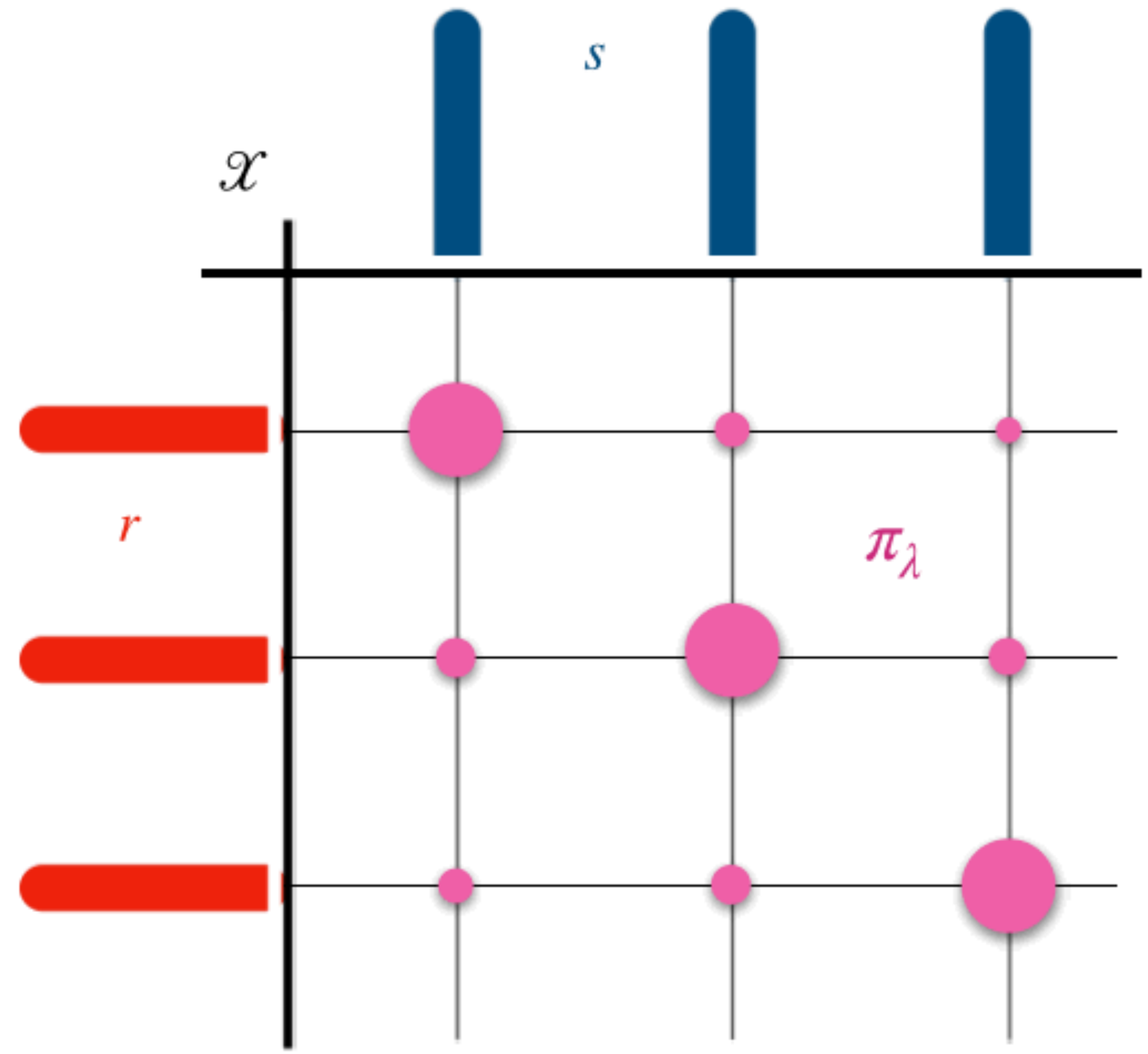
Entropy OT

Why Gaussian fluctuation?

$$\sqrt{n}\left(\hat{\pi}_\lambda - \pi_\lambda\right) \xrightarrow{\mathscr{D}} Z \sim \mathcal{N}_{N^2}\left(0, \Sigma_\lambda(\mu \mid \nu)\right)$$







$$\min_{\pi} c^T\pi - \lambda E(\pi)$$

$$A\,\pi = \begin{pmatrix} r \\ s \end{pmatrix}$$

$$\pi \geq 0$$

Entropy Regularized OT

$\lambda > 0$