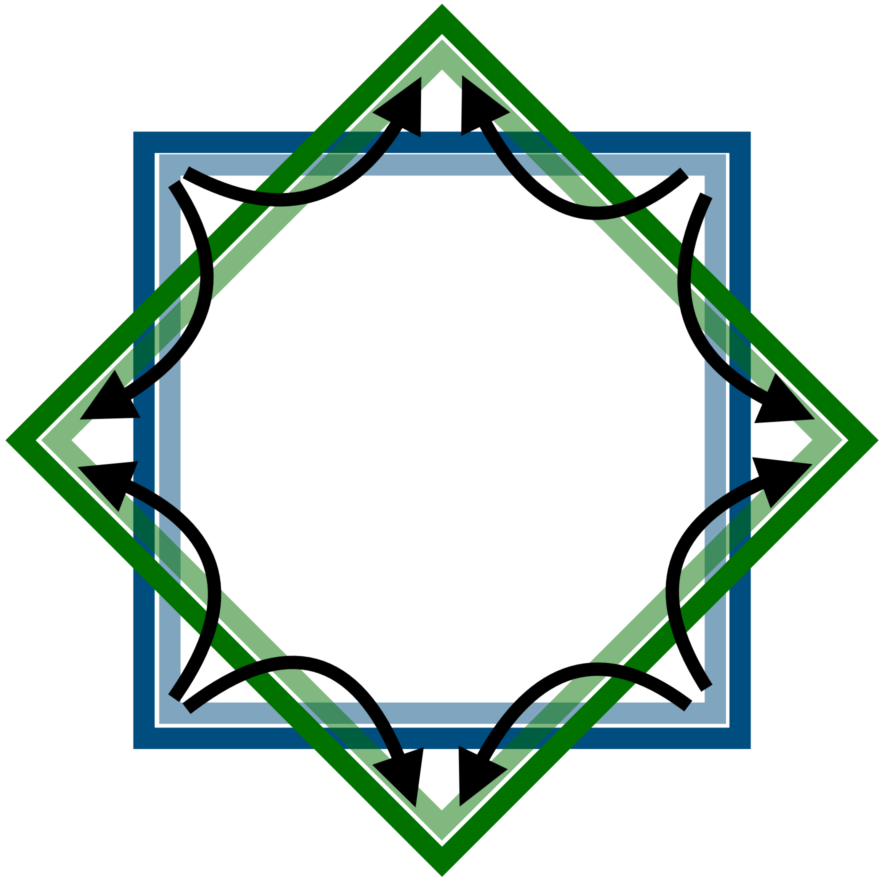


Optimal transport in high-energy physics

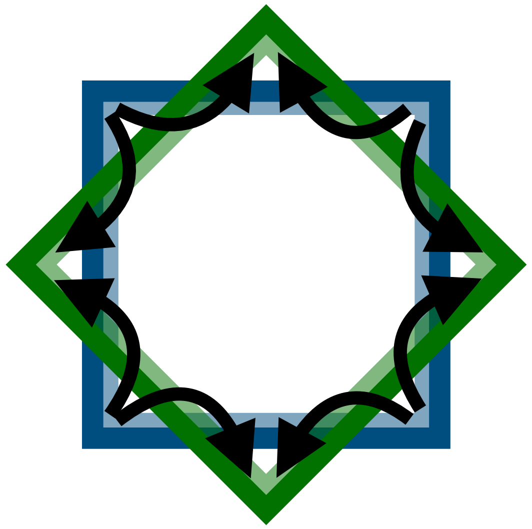
Theory and applications

October 12, 2022

philipp.windischhofer@uchicago.edu

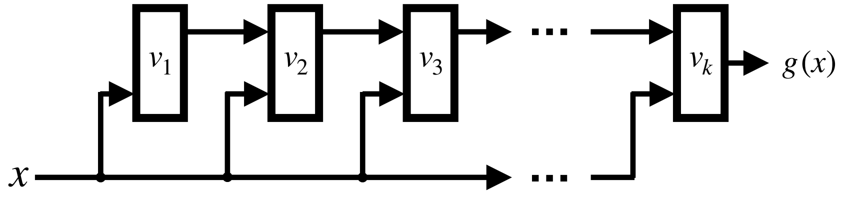
Philipp Windischhofer
University of Chicago

What can you expect?



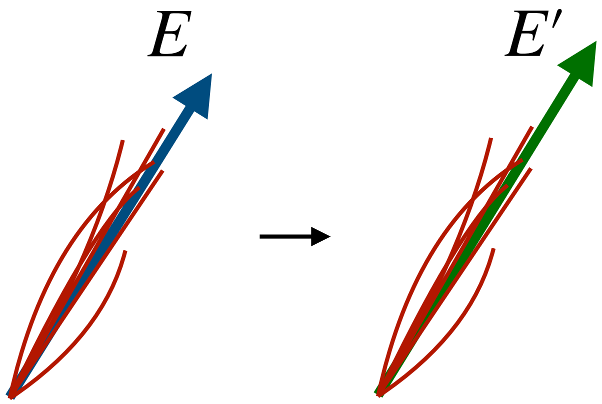
A self-contained introduction to the world of optimal transport

Intuition instead of (complex) mathematics



A glimpse at how to solve optimal transport problems

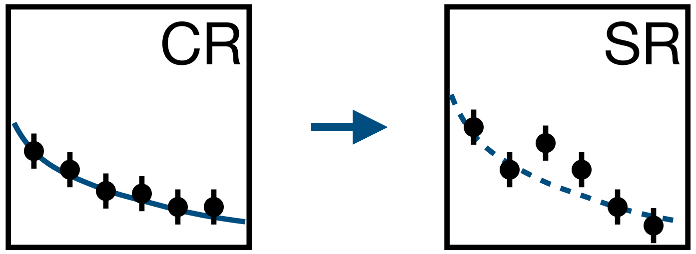
Illustration of one method; interplay with machine learning



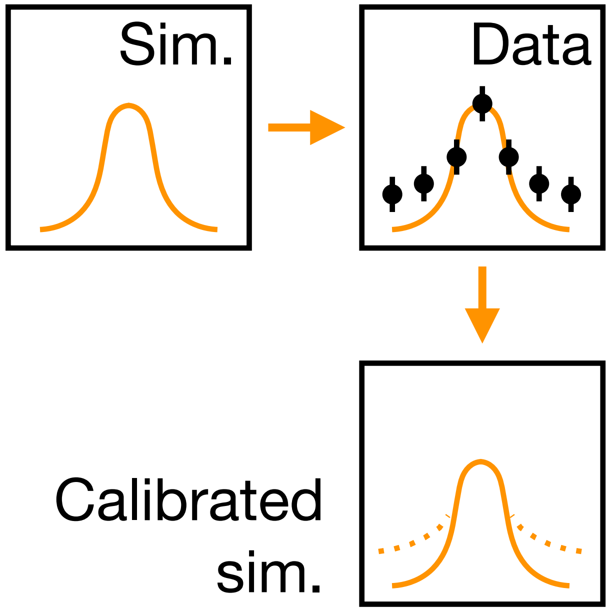
A glimpse at (present and future) applications in particle physics

Why should you care?

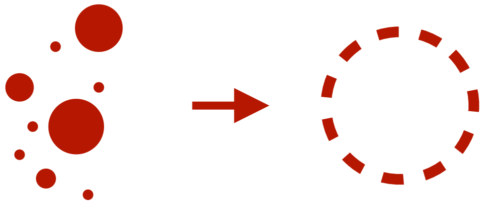
In particle physics, we manipulate (probability) **distributions on a daily basis …**



Extrapolation across phase space
(e.g. control region → signal region)

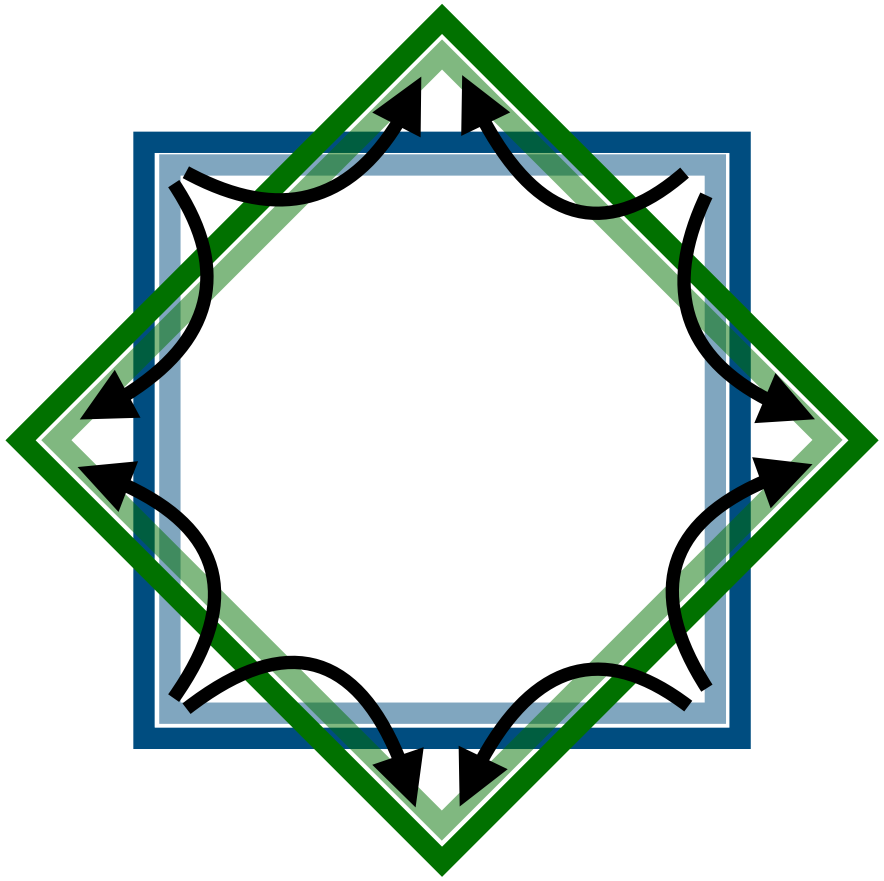


Calibration of simulation
(e.g. Monte Carlo prediction against data side bands)



Interpretation of data
(e.g. jet clustering)

… **optimal transport** provides **useful tools** *(and a unifying perspective)* for many of these!



The theory of optimal transportation

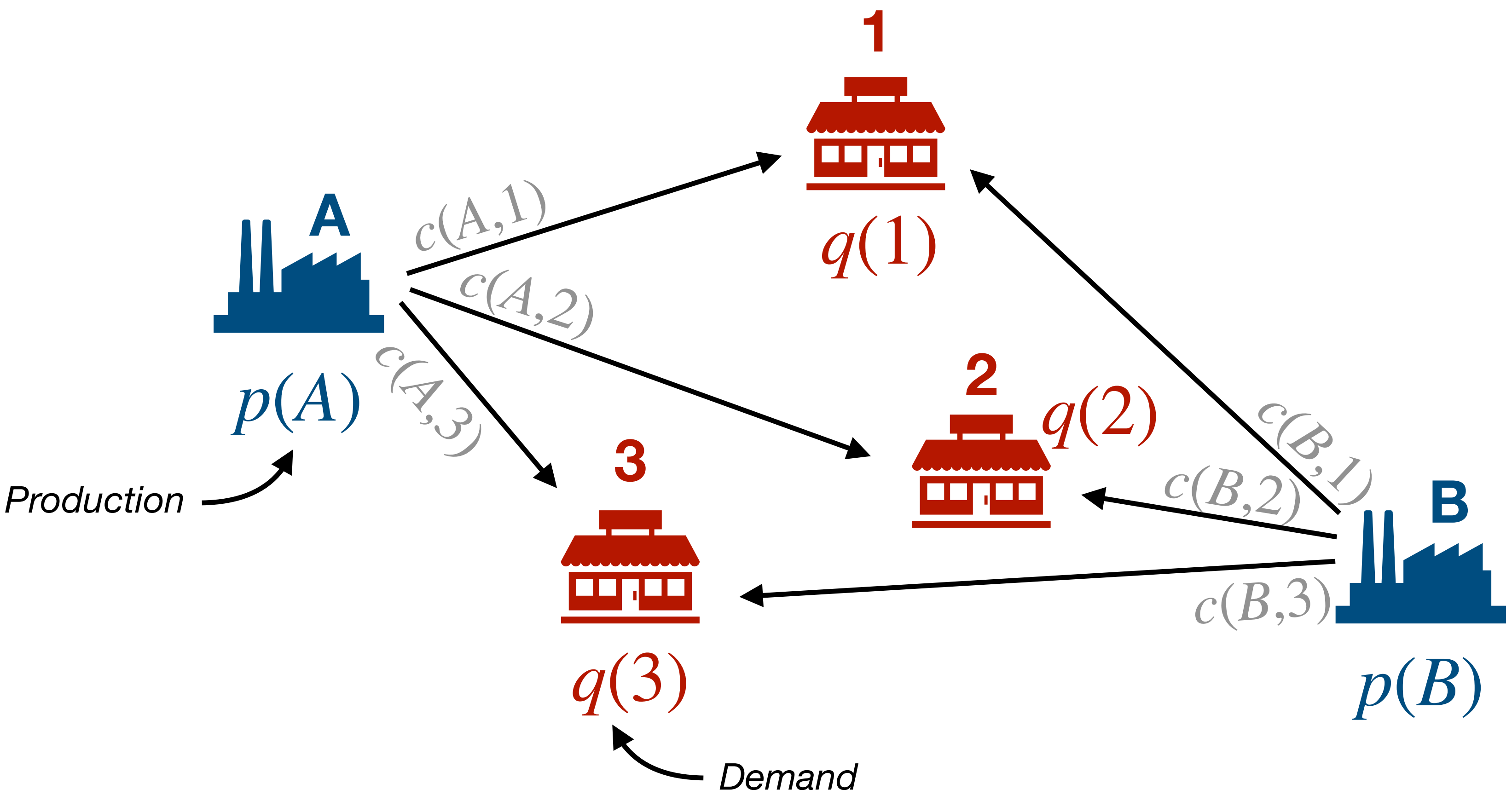
What is optimal transportation?

The answer to a logistics problem!

*"How to transport commodities from N **factories** to M **stores** …*

*… in the presence of a **transportation cost** $c(a, i)$ between factory a and store i …*

… so that the total cost is minimized?



Assume total production $p(A) + p(B)$ *equals total demand* $q(1) + q(2) + q(3)$

What is optimal transportation?

The answer to a logistics problem!

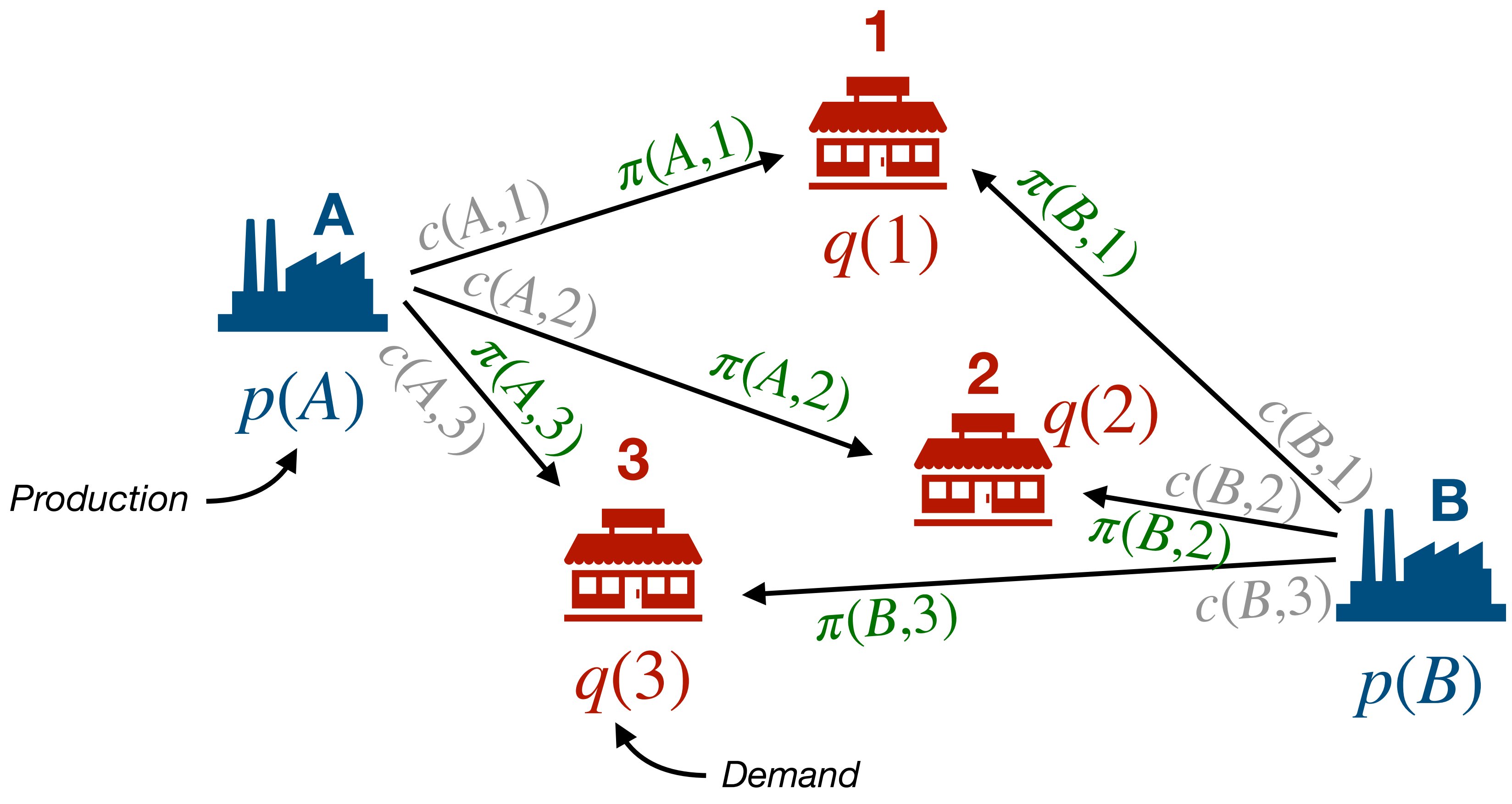
Transportation cost *(per unit mass)*

Optimal transportation plan →

$$\hat{\pi} = \arg\min_{\pi} \sum_a \sum_i \pi(a,i)\, c(a,i)$$

Optimization over all possible transportation plans

Mass transported from factory a to store i *(“transportation plan”)*



Assume total production $p(A) + p(B)$ *equals total demand* $q(1) + q(2) + q(3)$

What is optimal transportation?

The answer to a logistics problem!

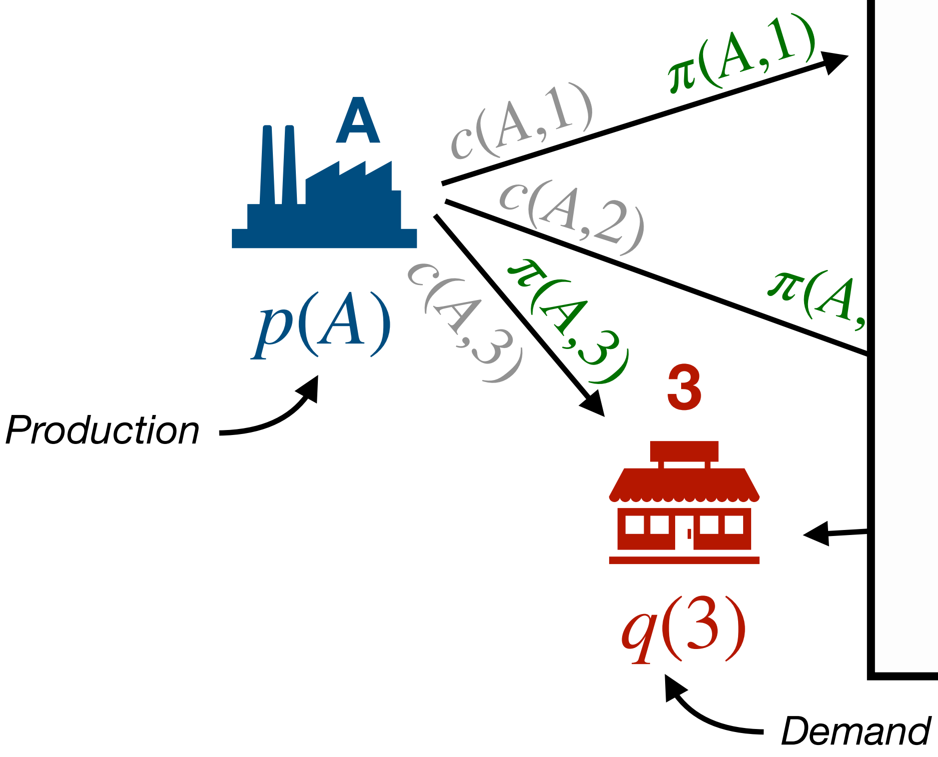
Optimal transportation plan →

$$\hat{\pi} = \arg\min_{\pi} \sum_a \sum_i \pi(a,i)\, c(a,i)$$

Transportation cost
(per unit mass)

Mass transported from factory a to store i
("transportation plan")

Optimization over all possible transportation plans



"All possible" transportation plans?

Must satisfy two conditions:

$$\sum_i \pi(a,i) = p(a) \quad \forall a$$

"Entire production is shipped"

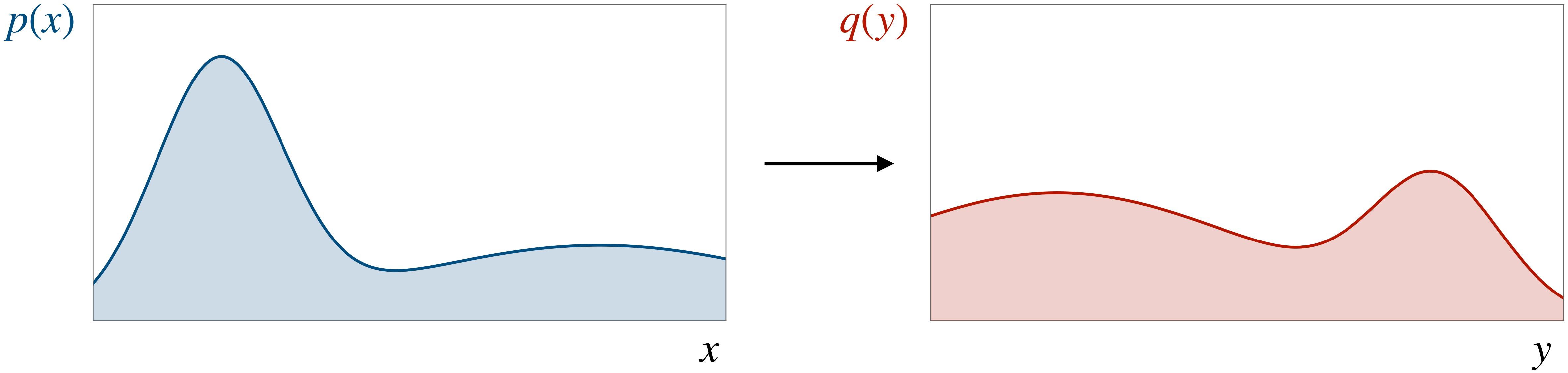
$$\sum_a \pi(a,i) = q(i) \quad \forall i$$

"Entire demand is satisfied"

Assume total production $p(A) + p(B)$ *equals total demand* $q(1) + q(2) + q(3)$

Optimal transport, now continuous

How about a continuous **distribution of production** $p(x)$ and a **continuous distribution of demand** $q(y)$?



Cost to transport one unit of mass from x to y: $c(x, y)$

Transport plan: move an amount $\pi(x, y)$ from x to y

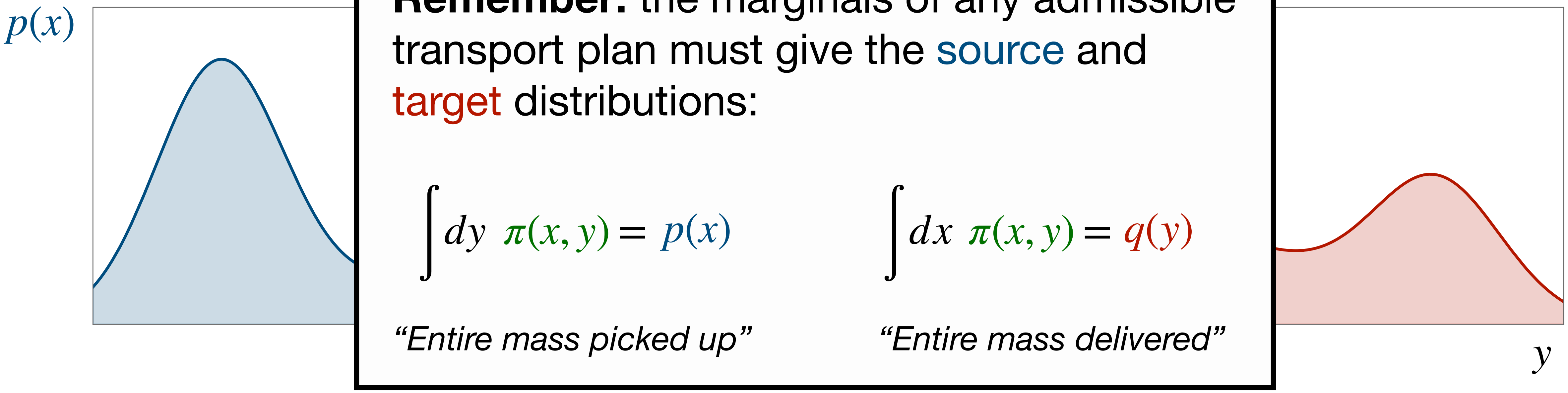
Transport plan with minimal cost:

$$\hat{\pi} = \arg\min_{\pi} \int dx\, dy\ \pi(x, y)\ c(x, y)$$

"Kantorovich optimal transport problem"

Optimal transport, now continuous

How about a continuous **distribution of production** $p(x)$ and a **continuous distribution of demand** $q(y)$?



Remember: the marginals of any admissible transport plan must give the source and target distributions:

$$\int dy \;\pi(x,y) = p(x) \qquad \int dx \;\pi(x,y) = q(y)$$

"Entire mass picked up" *"Entire mass delivered"*

Cost to transport one unit of mass from x to y: $c(x,y)$

Transport plan: move an amount $\pi(x,y)$ from x to y

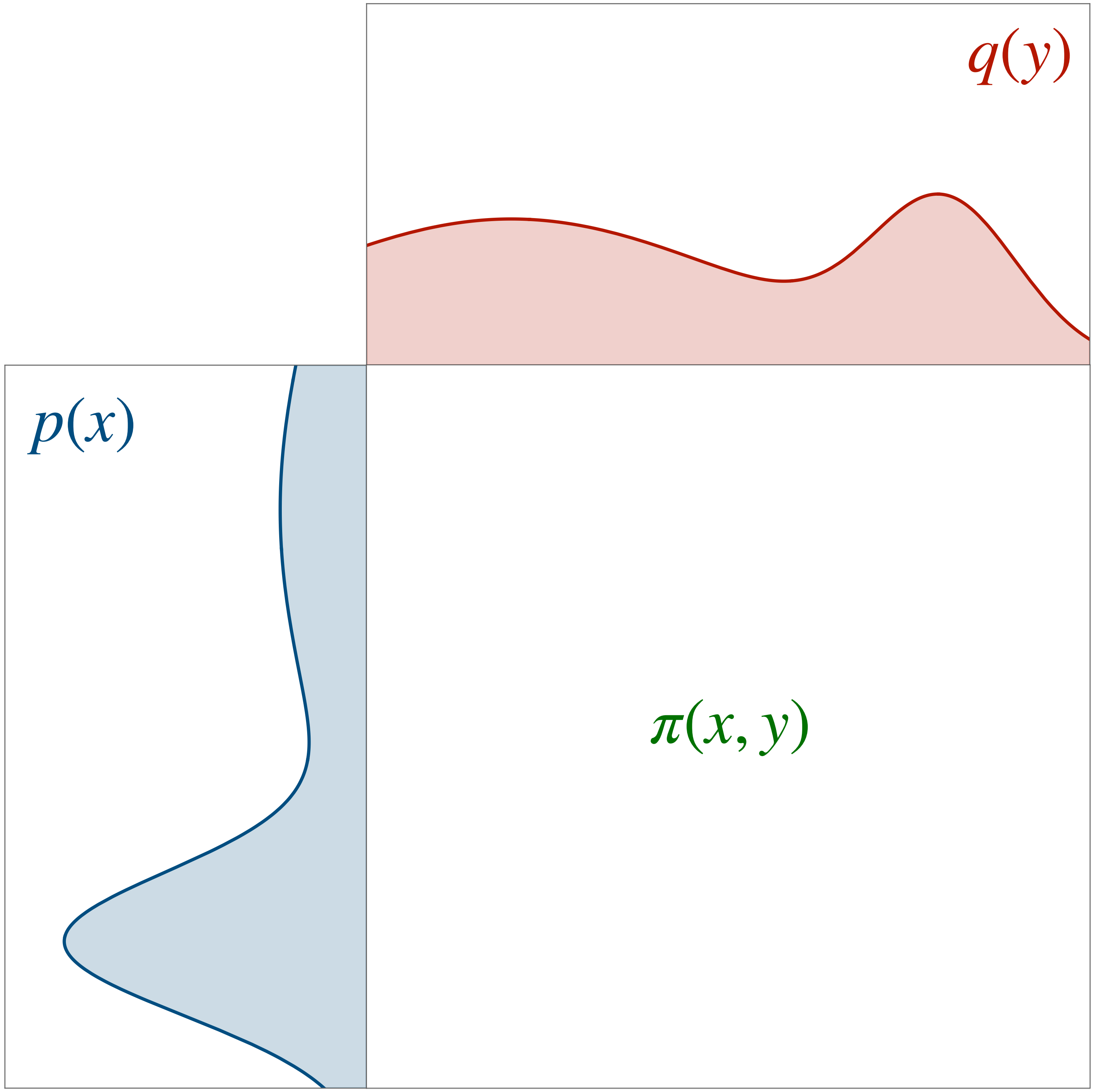
Transport plan with minimal cost:

$$\hat{\pi} = \arg\min_{\pi} \int dx\, dy \;\pi(x,y)\; c(x,y)$$

"Kantorovich optimal transport problem"

Optimal transport, now continuous

How about a continuous **distribution of production** $p(x)$ and a **continuous distribution of demand** $q(y)$?



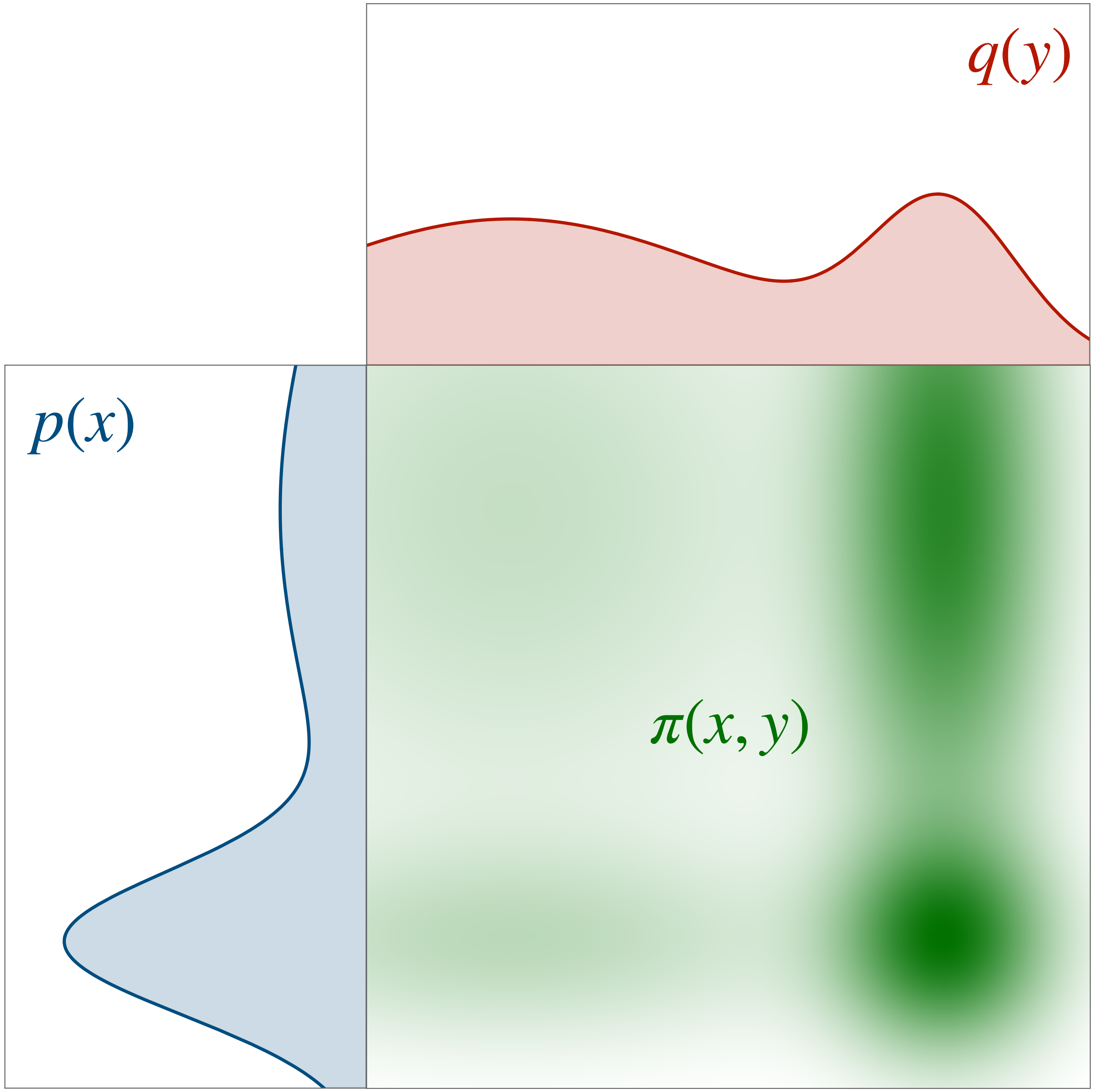
Constraints:

$$\int dy \ \pi(x, y) = p(x)$$

$$\int dx \ \pi(x, y) = q(y)$$

Optimal transport, now continuous

How about a continuous **distribution of production** $p(x)$ and a **continuous distribution of demand** $q(y)$?



It is not difficult to satisfy these constraints!

$$\pi(x, y) = p(x)\, q(y)$$

(Is admissible, but rarely minimal)

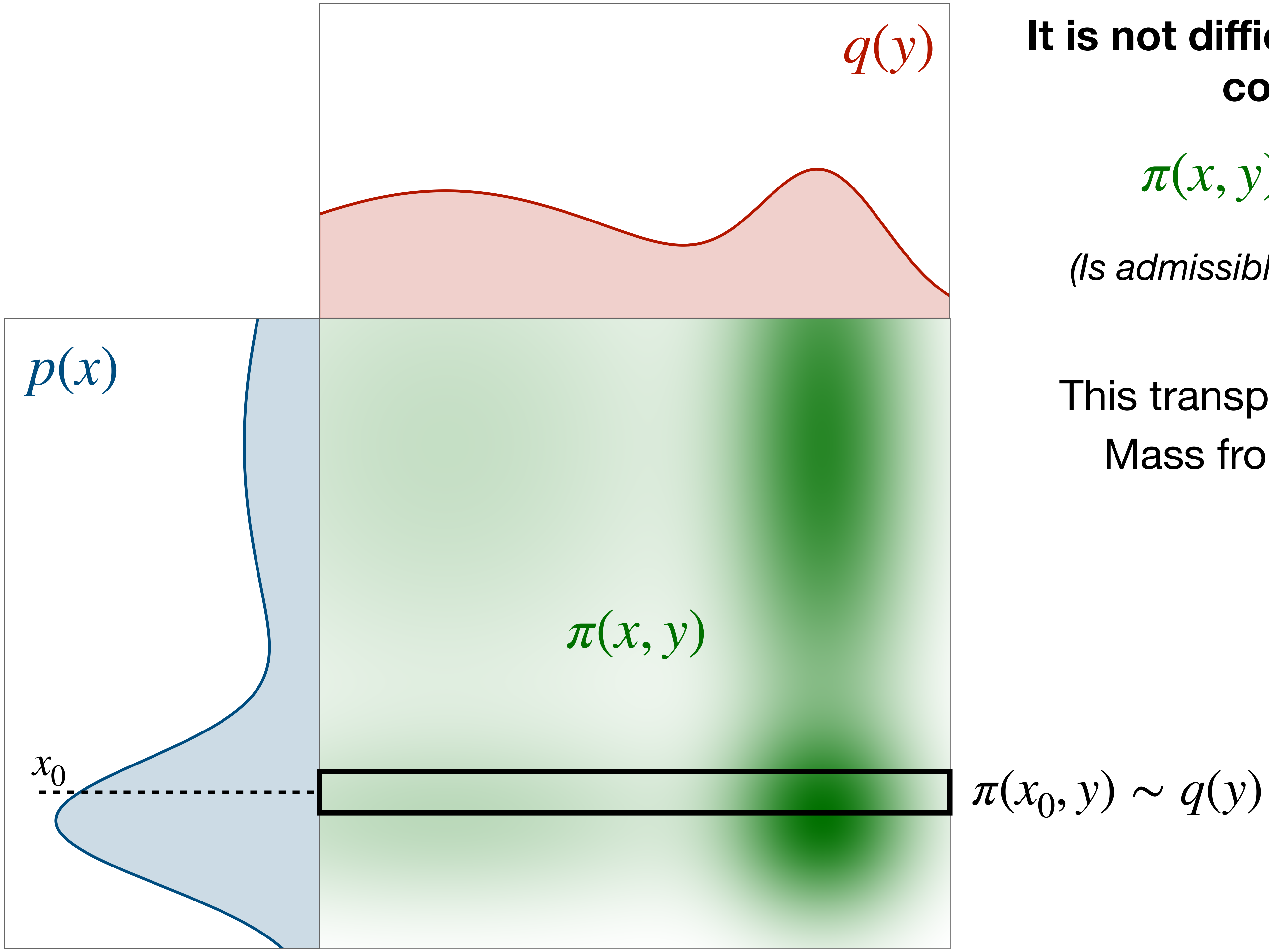
Constraints:

$$\int dy\ \pi(x, y) = p(x)$$

$$\int dx\ \pi(x, y) = q(y)$$

Optimal transport, now continuous

How about a continuous **distribution of production** $p(x)$ and a **continuous distribution of demand** $q(y)$?



It is not difficult to satisfy these constraints!

$$\pi(x, y) = p(x)\, q(y)$$

(Is admissible, but rarely minimal)

This transport plan distributes Mass from x_0 across all y

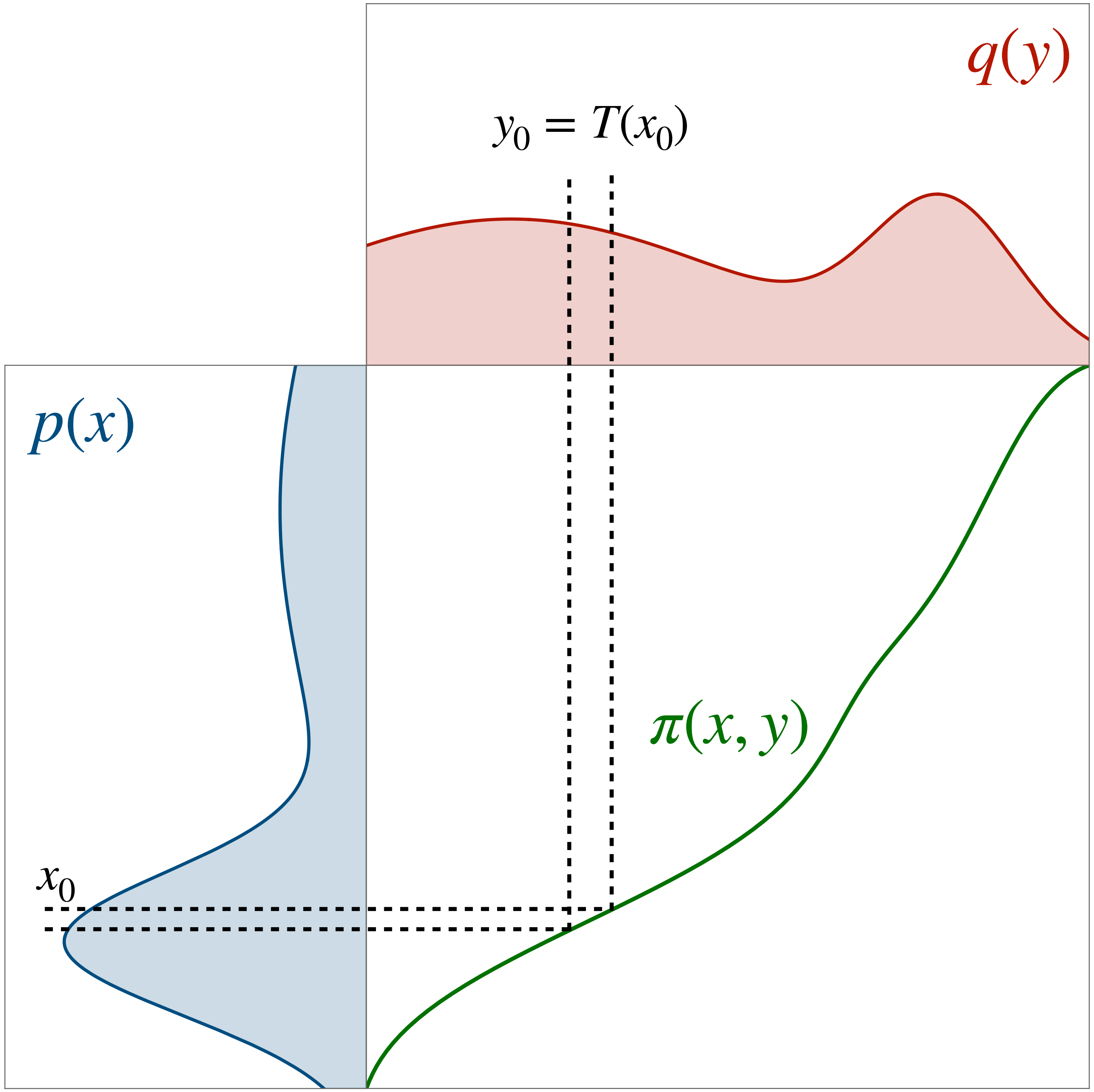
Constraints:

$$\int dy\ \pi(x, y) = p(x)$$

$$\int dx\ \pi(x, y) = q(y)$$

Optimal transport, now continuous

How about a continuous **distribution of production** $p(x)$ and a **continuous distribution of demand** $q(y)$?



Transport plans can also be "deterministic":

$$\pi(x, y) = p(x)\, \delta[y - T(x)]$$

This is a change of coordinates $x \rightarrow y = T(x)$; must satisfy

$$q(y) = p(x) \left(\frac{dT}{dx} \right)^{-1}$$

Constraints:

$$\int dy \;\; \pi(x, y) = p(x)$$

$$\int dx \;\; \pi(x, y) = q(y)$$

Optimal transport à la Monge

Kantorovich problem

$$\hat{\pi} \;= \arg\min_{\pi} \int dx\, dy\; \pi(x,y)\; c(x,y)$$

$$\int dy\;\; \pi(x,y) = p(x) \qquad \int dx\;\; \pi(x,y) = q(y)$$

Optimal transport à la Monge

Kantorovich problem

$$\hat{\pi} = \arg\min_{\pi} \int dx\, dy\ \pi(x,y)\ c(x,y)$$

$$\boxed{\int dy\ \pi(x,y) = p(x) \qquad \int dx\ \pi(x,y) = q(y)}$$

“Monge optimal transport problem”

$$\hat{T} = \arg\min_{T} \int dx\ p(x)\ c(x, T(x))$$

$$\pi(x,y) = p(x)\,\delta[y - T(x)]$$

$$\boxed{q(y) = p(x)\left(\frac{dT}{dx}\right)^{-1}}$$

Optimal transport à la Monge

If both $p(x)$ and $q(y)$ are **sufficiently continuous***, the solution to the Kantorovich problem …

$$\hat{\pi} = \arg\min_{\pi} \int dx\, dy\ \pi(x, y)\ c(x, y)$$

$$\boxed{\int dy\ \pi(x, y) = p(x) \qquad \int dx\ \pi(x, y) = q(y)}$$

… is guaranteed to be of the "deterministic" kind, i.e. it solves the *"Monge optimal transport problem"*

$$\hat{T} = \arg\min_{T} \int dx\ p(x)\ c(x, T(x))$$

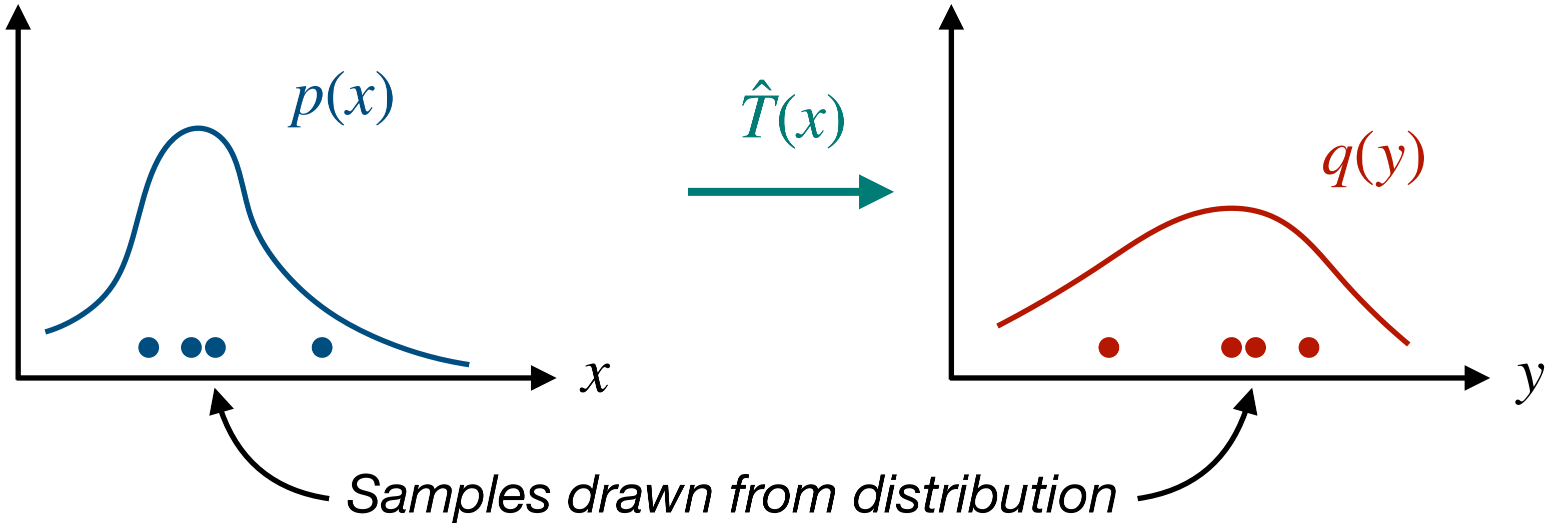
$$\pi(x, y) = p(x)\, \delta[y - T(x)]$$

$$\boxed{q(y) = p(x) \left(\frac{dT}{dx} \right)^{-1}}$$

* and the cost function $c(x, y)$ is a *convex* function

Monge vs. Kantorovich

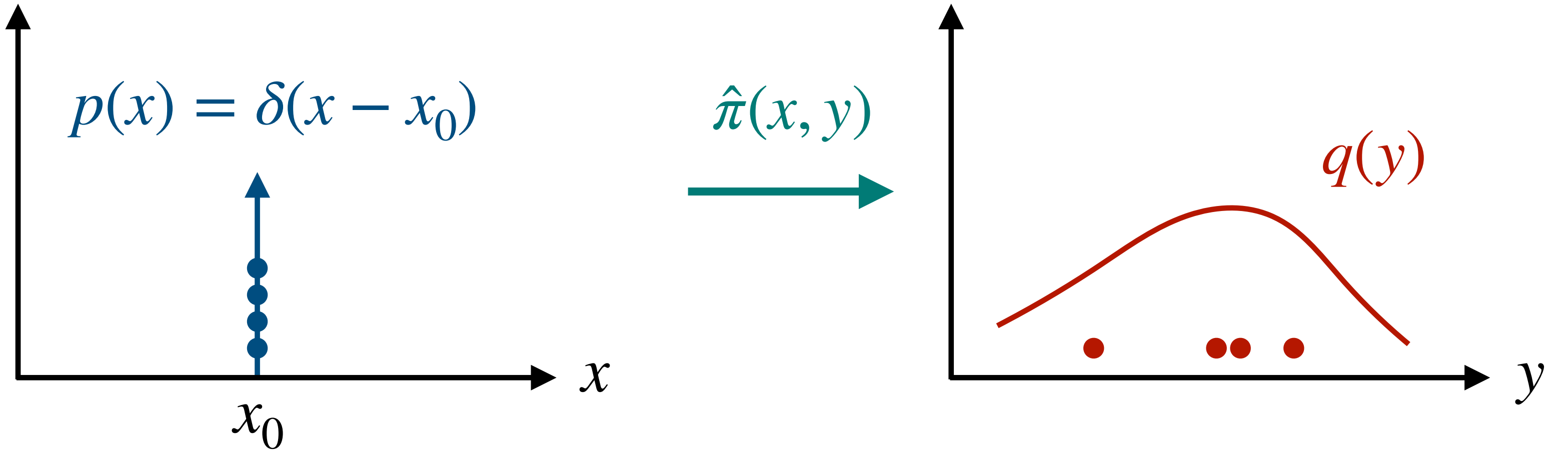
Transport between two smooth distributions:



Samples drawn from distribution

Deterministic transport
(“reordering of samples”) sufficient
→ ***Monge problem***

Transport between non-smooth and smooth distribution:



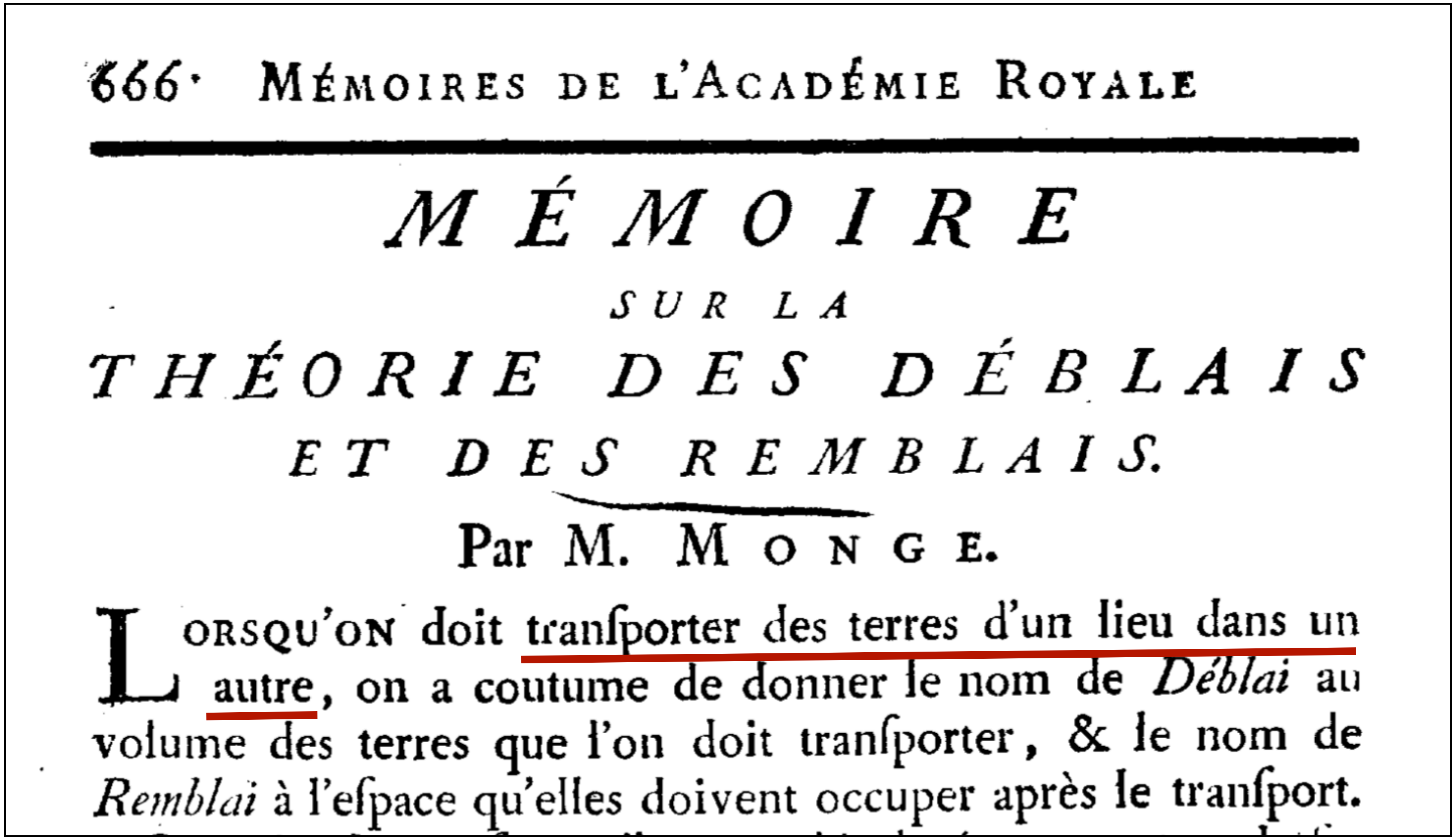
Need stochastic transport
(“random smearing of samples”)
→ ***Kantorovich problem***

Optimal transport à la Monge

If both

Kanto

The beginning of transportation theory > 240 years ago



666. MÉMOIRES DE L'ACADÉMIE ROYALE

MÉMOIRE
SUR LA
THÉORIE DES DÉBLAIS
ET DES REMBLAIS.

Par M. MONGE.

LORSQU'ON doit tranſporter des terres d'un lieu dans un autre, on a coutume de donner le nom de *Déblai* au volume des terres que l'on doit tranſporter, & le nom de *Remblai* à l'eſpace qu'elles doivent occuper après le tranſport.

$\hat{\pi}$

$= q(y)$

… is

$\hat{T}$

* and the cost function $c(x, y)$ is a *convex* function

240 years of optimal transport

Today we know a lot about the structure of optimal transport solutions

(High-profile, Fields-medal winning research!)

The character of the solution depends strongly on the cost function $c(x, y)$

240 years of optimal transport

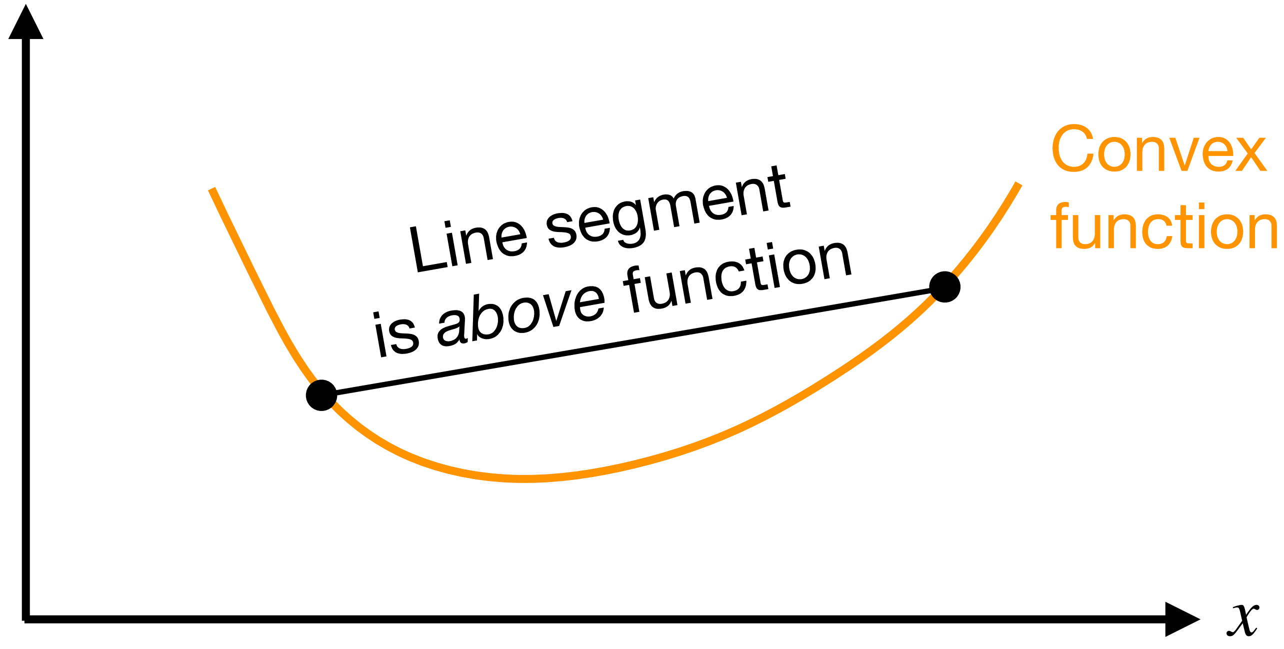
Today we know a lot about the structure of optimal transport solutions

(High-profile, Fields-medal winning research!)

The character of the solution depends strongly on the cost function $c(x, y)$

For “smooth” distributions and convex cost functions:

Solution to Kantorovich problem (“stochastic transport”) = Solution to Monge problem (“deterministic transport”)



The choice of cost function

Many useful cost functions are convex!

E.g. $c(x, y) = |x - y|^p$ for $p > 1$

… let's look at a few examples!

The choice of cost function

Many useful cost functions are convex!

E.g. $c(x, y) = |x - y|^p$ for $p > 1$

… let's look at a few examples!

$p = 2$, i.e. $c(x, y) = |x - y|^2$

The optimal transport function is the gradient of a convex potential!

("Brenier's theorem")

$$\hat{T}(x) = \nabla g(x)$$

"Transport potential"
(Also convex!)

For this case:
Optimal transport ⇔ Electrostatics

The transport vector field $\hat{T}$ has zero curl!

"Don't ship your stuff in circles."

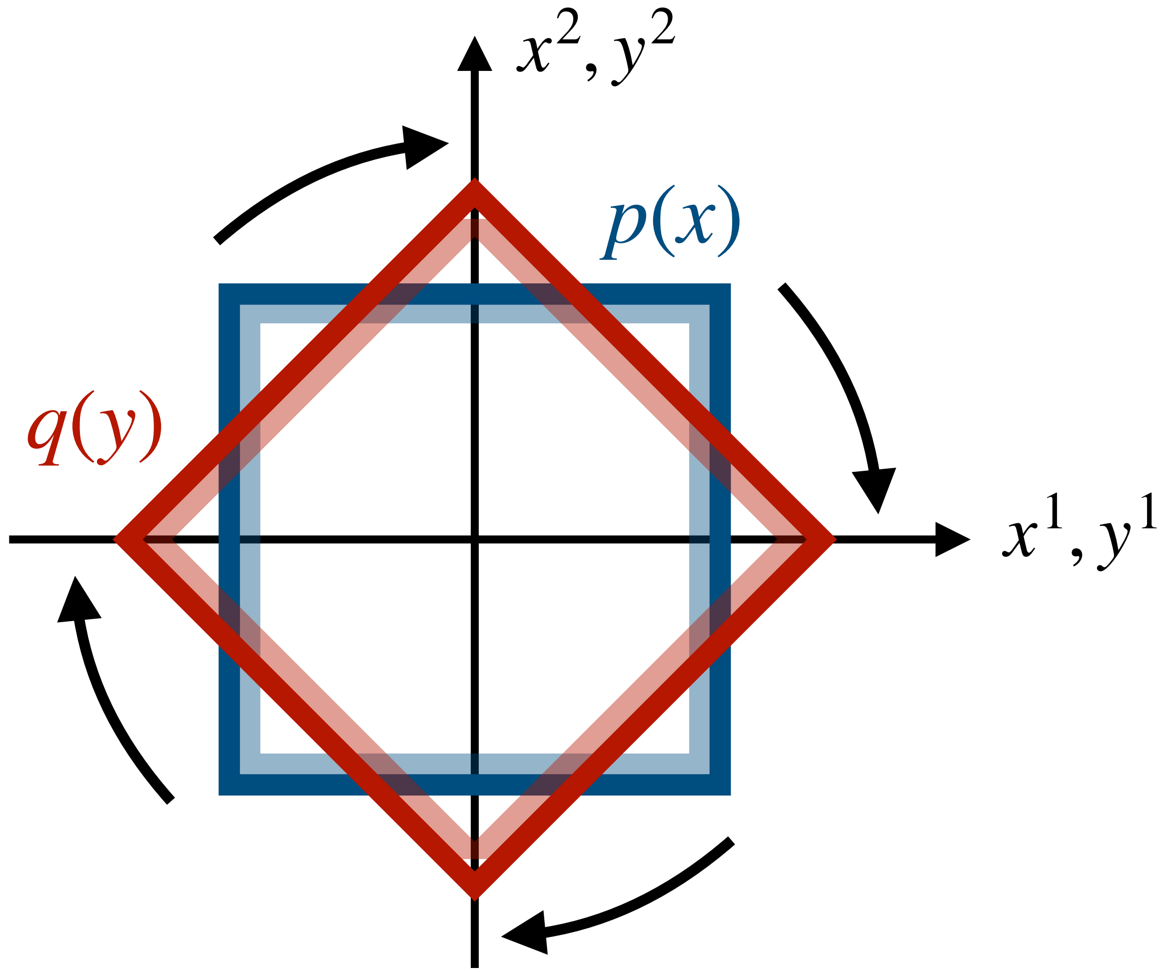
The choice of cost function

Many useful cost functions are convex!

E.g. $c(x, y) = |x - y|^p$ for $p > 1$

… let's look at a few examples!

$p = 2$, i.e. $c(x, y) = |x - y|^2$



Example:

Source distribution $p(x)$ populates inside of axis-aligned square

Target distribution $q(y)$ populates "rotated" square

But: rotation is not a gradient vector field!

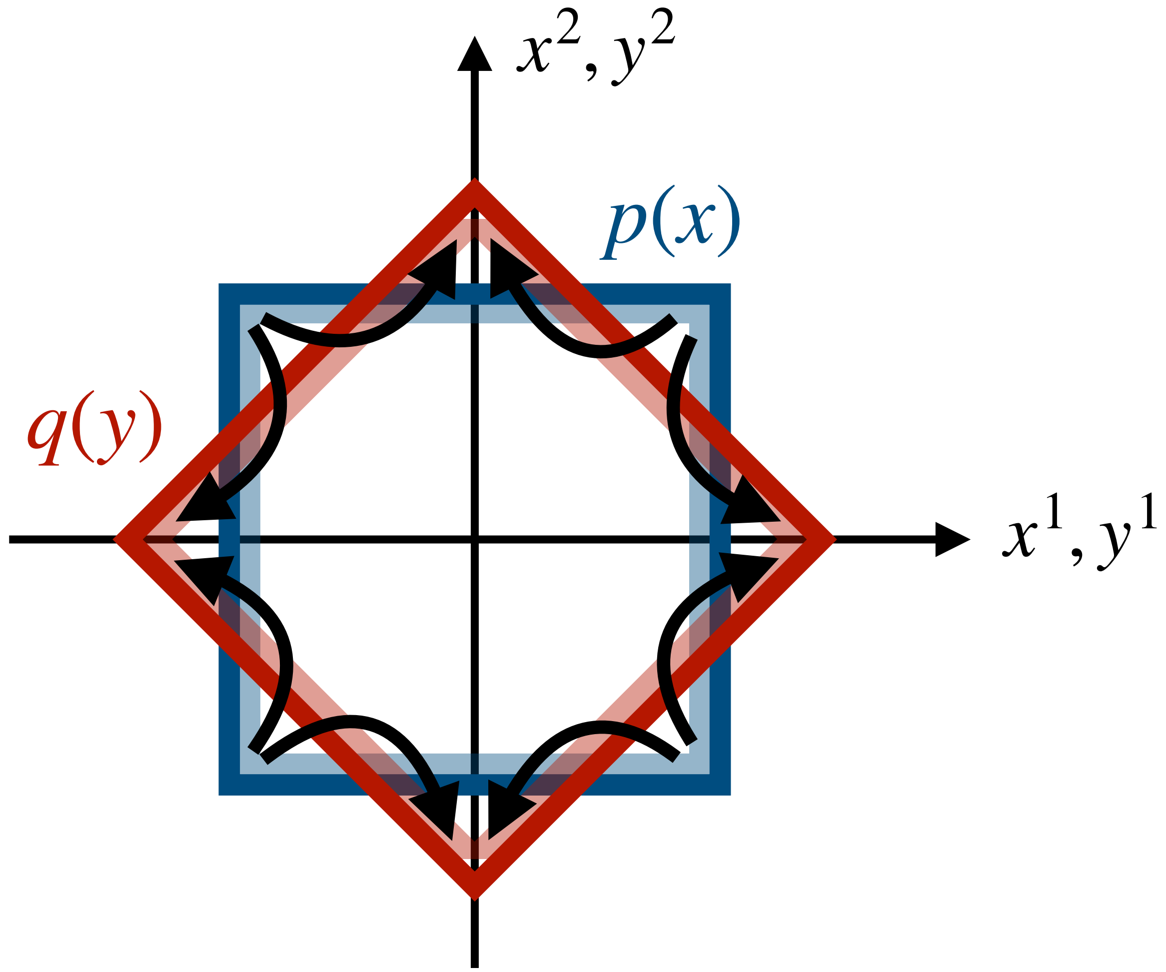
The choice of cost function

Many useful cost functions are convex!

E.g. $c(x, y) = |x - y|^p$ for $p > 1$

… let's look at a few examples!

$p = 2$, i.e. $c(x, y) = |x - y|^2$



Example:

Source distribution $p(x)$ populates inside of axis-aligned square

Target distribution $q(y)$ populates "rotated" square

But: rotation is not a gradient vector field!

The optimal transport solution looks like this

The choice of cost function

Many useful cost functions are convex!

E.g. $c(x, y) = |x - y|^p$ for $p > 1$

… let's look at a few examples!

$p = 2$, i.e. $c(x, y) = |x - y|^2$

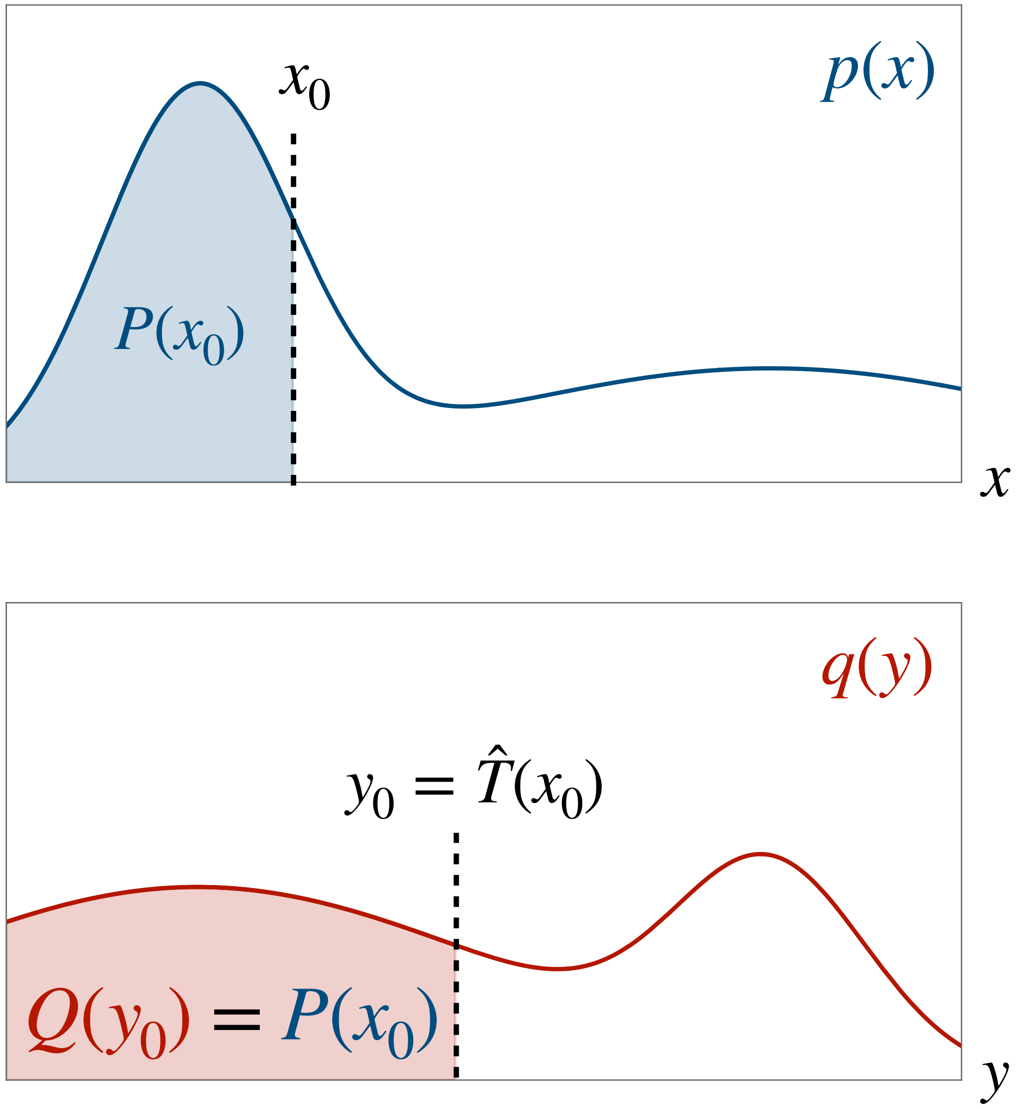
For 1-dimensional distributions:

The optimal transport solution performs quantile-matching *(works for all convex cost functions!)*

$$\hat{T}(x) = Q^{-1}(P(x))$$

Cumulative distributions of $p(x)$, $q(y)$:

Generically: $F(x) = \int_0^x dx' f(x')$



The choice of cost function

Many useful cost functions are convex!

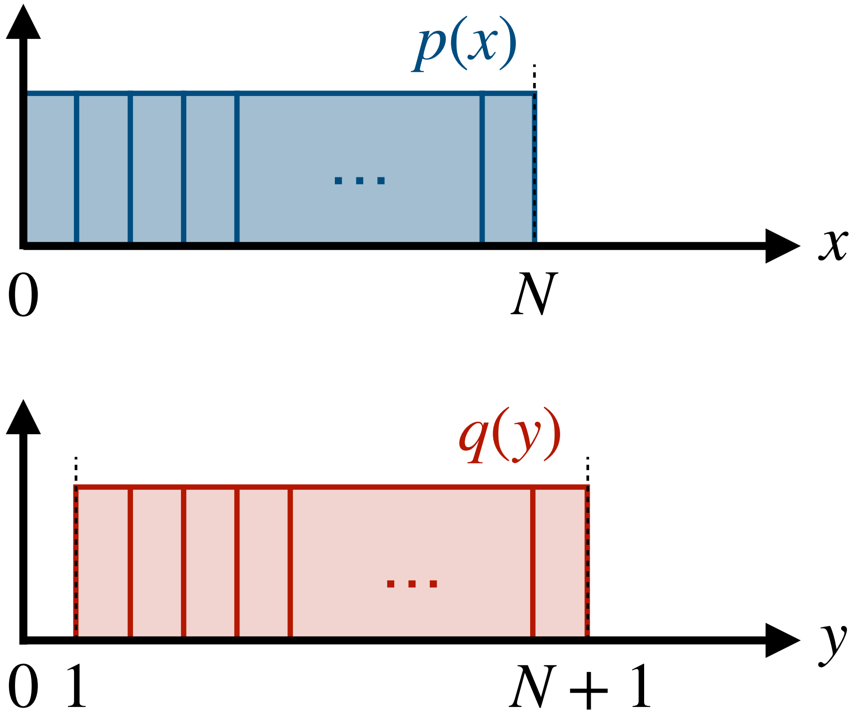
E.g. $c(x, y) = |x - y|^p$ for $p > 1$

… let's look at a few examples!

$p = 1$, i.e. $c(x, y) = |x - y|$ *(Monge's original problem)*

This is a much more complicated case!

Solutions exist for smooth distributions, but no longer unique!



Example:

Uniform source and target distributions
(e.g. rows of N books, shifted by one)

The choice of cost function

Many useful cost functions are convex!

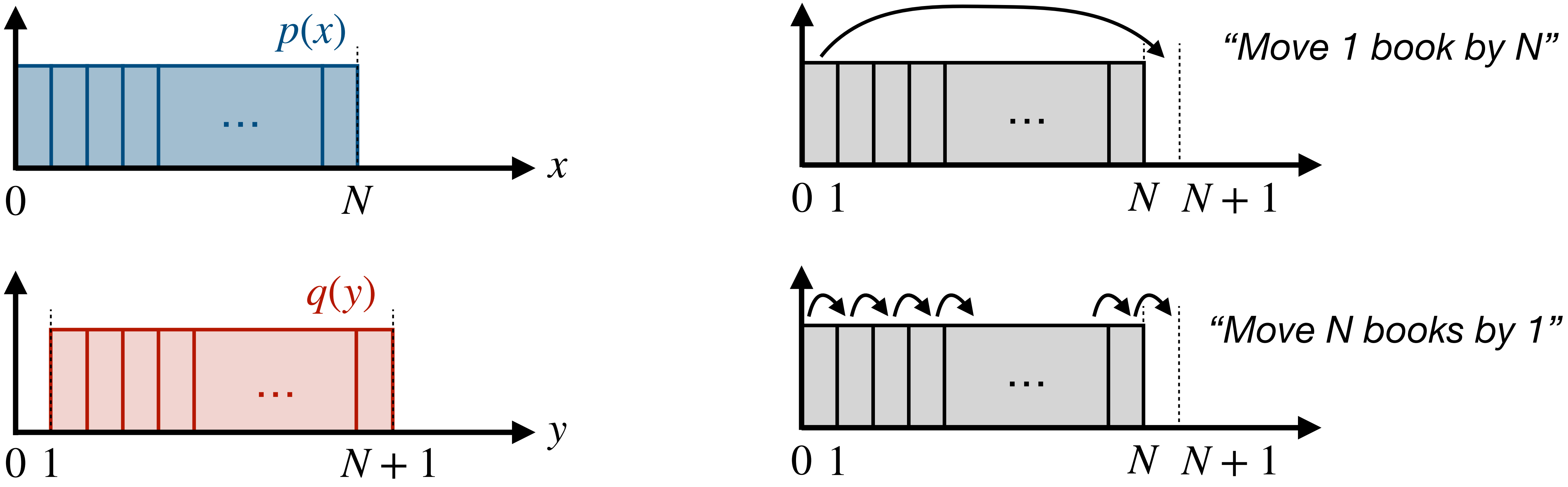
E.g. $c(x, y) = |x - y|^p$ for $p > 1$

… let's look at a few examples!

$p = 1$, i.e. $c(x, y) = |x - y|$ *(Monge's original problem)*

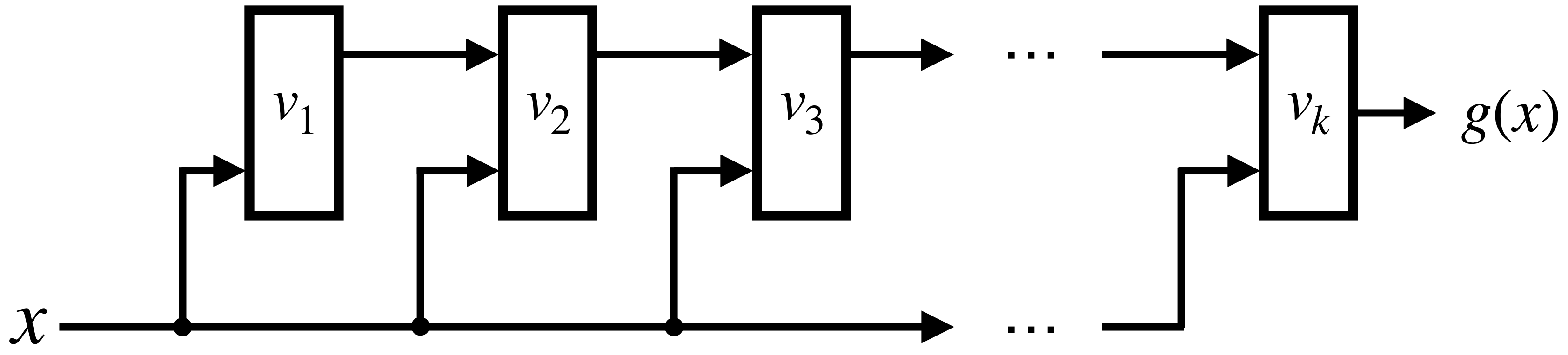
This is a much more complicated case!

Solutions exist for smooth distributions, but no longer unique!



"Move 1 book by N"

"Move N books by 1"



Solving
optimal transport problems

Solving the Monge problem

Want to find $\hat{T}$ to solve

$$x \mapsto y = T(x)$$

$$\hat{T} = \arg\min_{T} \int dx \; p(x) \; c(x, T(x))$$ **, subject to the constraint**

$$q(y) = p(x) \left(\frac{dT}{dx} \right)^{-1}$$, starting from samples drawn from $p(x)$ and $q(y)$.

(Highly nonlinear!)

(Continuous, as usual in particle physics)

In general, this is a very difficult problem!

Many different algorithms exist! Two main classes:

Solving the Monge problem

Want to find $\hat{T}$ to solve

$x \mapsto y = T(x)$

$$\hat{T} = \arg\min_{T} \int dx \; p(x) \; c(x, T(x))$$ **, subject to the constraint**

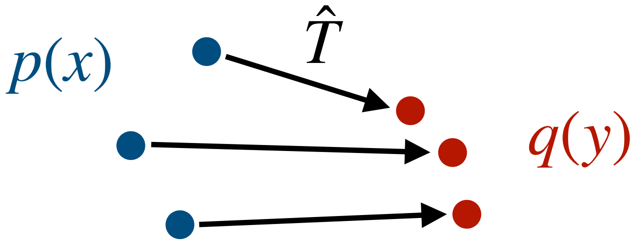
$$q(y) = p(x)\left(\frac{dT}{dx}\right)^{-1}$$, starting from samples drawn from $p(x)$ and $q(y)$.

(Highly nonlinear!)

(Continuous, as usual in particle physics)

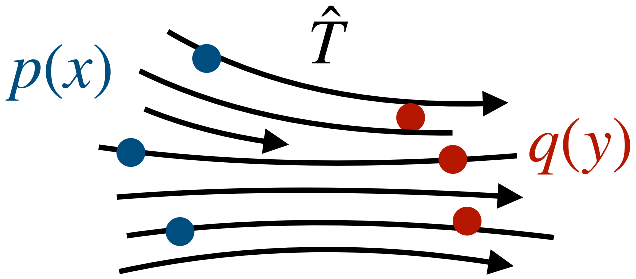
In general, this is a very difficult problem!

Many different algorithms exist! Two main classes:



"Discrete"
optimal transport

Transport empirical distributions
by pairing up samples ~ $\mathcal{O}(N^2)$



"Continuous"
optimal transport

Construct (continuous) transport function,
implicit regularization

Continuous optimal transport: an illustration

In the following: look at **continuous optimal transport** with **quadratic cost function**

(Theoretically well-understood, synergies with modern machine learning)

$$\hat{T} = \arg\min_{T} \int dx \; p(x) \; c(x, T(x))$$

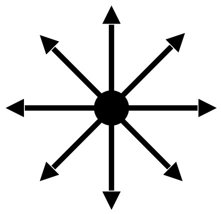
Cost function: $c(x, y) = |x - y|^2$

Constraint: $q(y) = p(x) \left(\frac{dT}{dx} \right)^{-1}$

Reminder: solution is a gradient field

$$\hat{T}(x) = \nabla g(x)$$

Convex potential



"Electrostatics"

Still not trivial to solve: highly problem-dependent and nonlinear constraint!

→ try to find an alternative formulation with simpler constraints

A solution sketch

Monge problem

Nonlinear constraint

Equivalence for smooth distributions

Kantorovich problem

Linear constraints!

$$\hat{T} = \arg\min_{T} \int dx \ p(x) \ c(x, T(x))$$

$$\pi(x, y) = p(x)\, \delta[y - T(x)] \qquad q(y) = p(x) \left(\frac{dT}{dx} \right)^{-1}$$

$$\hat{\pi} = \arg\min_{\pi} \int dx\, dy \ \pi(x, y) \ c(x, y)$$

$$\int dy \ \pi(x, y) = p(x) \qquad \int dx \ \pi(x, y) = q(y)$$

A solution sketch

Monge problem

Nonlinear constraint

$$\hat{T} = \arg\min_{T} \int dx\;\; p(x)\;\; c(x,\, T(x)\,)$$

$$\pi(x, y) = p(x)\,\delta[y - T(x)] \qquad q(y) = p(x)\left(\frac{dT}{dx}\right)^{-1}$$

↕ *Equivalence for smooth distributions*

Kantorovich problem

Linear constraints!

$$\hat{\pi} \;\; = \arg\min_{\pi} \int dx\, dy\;\; \pi(x, y)\;\; c(x, y)$$

$$\int dy\;\; \pi(x, y) = \; p(x) \qquad \int dx\;\; \pi(x, y) = q(y)$$

↕ *Kantorovich-Rubinstein duality*

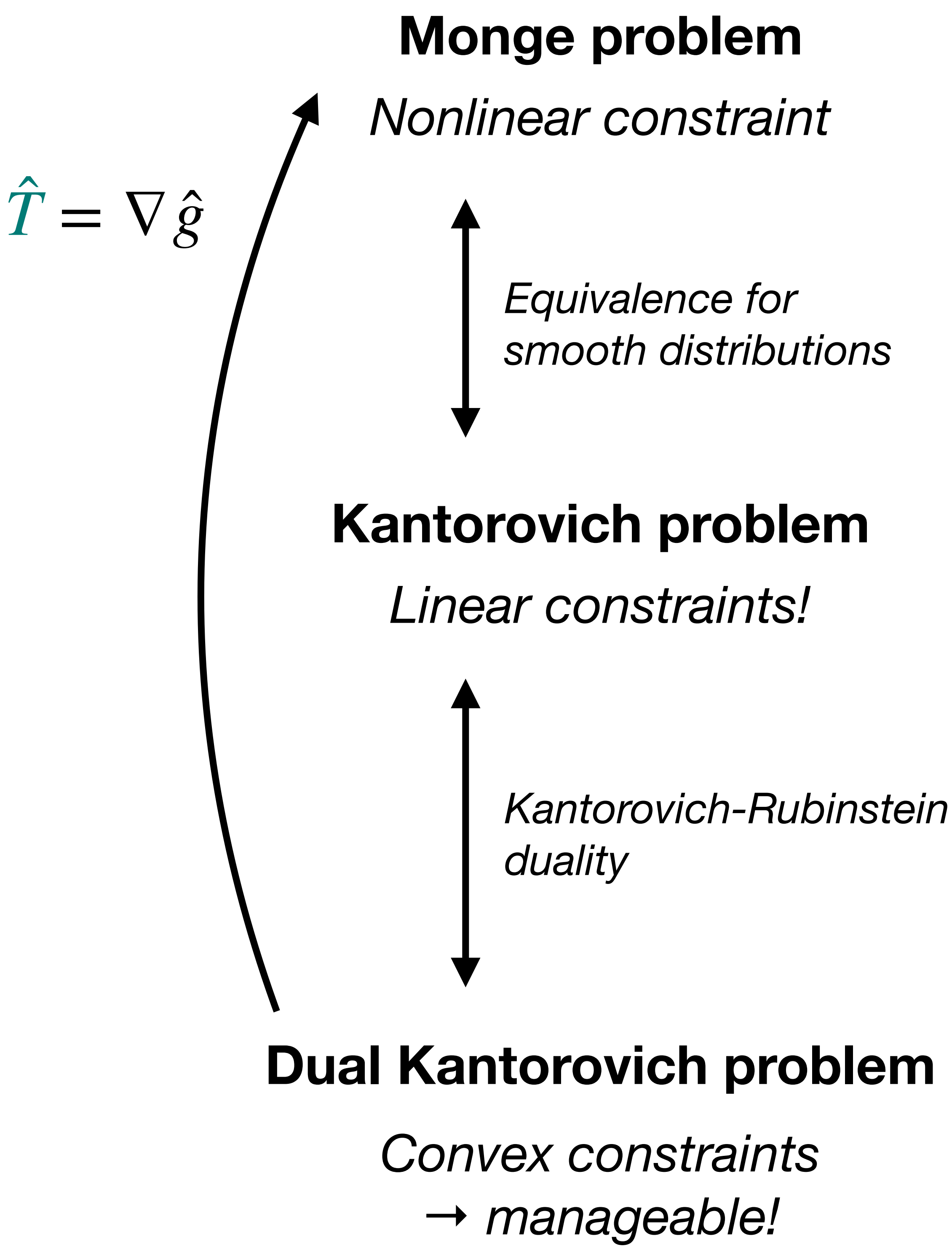
Dual Kantorovich problem

Convex constraints
→ *manageable!*

$$\hat{f}, \hat{g} = \arg\max_{f,g} \int dy\, q(y)\, f(y) + \\ + \int dx\, p(x) g(x)$$

$$g(x) + f(y) \leq c(x, y)$$

A solution sketch



$\hat{T} = \nabla \hat{g}$

Monge problem

Nonlinear constraint

Equivalence for smooth distributions

Kantorovich problem

Linear constraints!

Kantorovich-Rubinstein duality

Dual Kantorovich problem

Convex constraints
→ *manageable!*

$$\hat{T} = \arg\min_{T} \int dx \; p(x) \; c(x, T(x))$$

$$\pi(x,y) = p(x)\,\delta[y - T(x)] \qquad q(y) = p(x)\left(\frac{dT}{dx}\right)^{-1}$$

$$\hat{\pi} \;\; = \arg\min_{\pi} \int dx\, dy \; \pi(x,y) \; c(x,y)$$

$$\int dy \;\; \pi(x,y) = p(x) \qquad \int dx \;\; \pi(x,y) = q(y)$$

$$\hat{f}, \hat{g} = \arg\max_{f,g} \int dy\, q(y)\, f(y) + \int dx\, p(x) g(x)$$

$$g(x) + f(y) \leq c(x,y)$$

The Kantorovich-Rubinstein duality

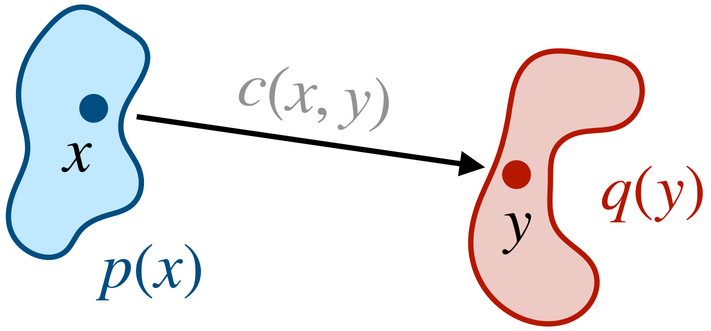
Primal problem:

$$\hat{\pi} = \arg\min_{\pi} \int dx\, dy\ \pi(x, y)\ c(x, y)$$

$$\int dy\ \pi(x, y) = p(x) \qquad \int dx\ \pi(x, y) = q(y)$$

"Operative perspective":

Optimise transportation plan based on point-to-point cost $c(x, y)$



The Kantorovich-Rubinstein duality

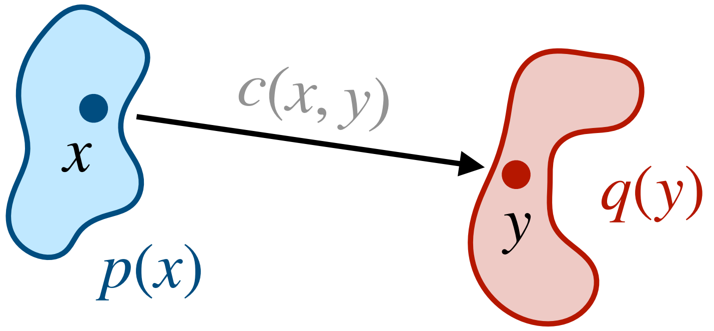
Primal problem:

$$\hat{\pi} = \arg\min_{\pi} \int dx\, dy\ \pi(x,y)\, c(x,y)$$

$$\int dy\ \pi(x,y) = p(x) \qquad \int dx\ \pi(x,y) = q(y)$$

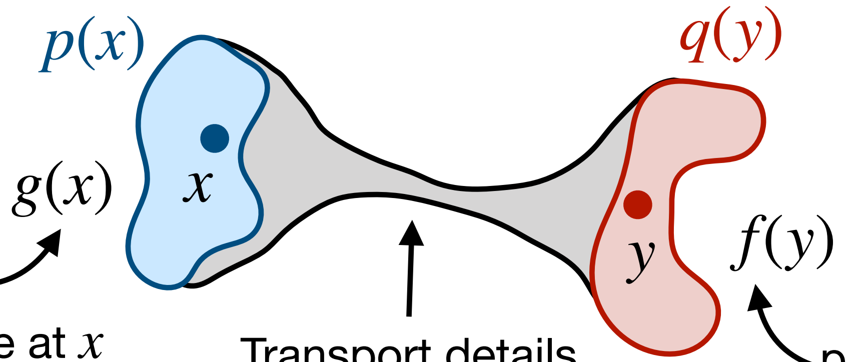
"Operative perspective":

Optimise transportation plan based on point-to-point cost $c(x,y)$



"Black-box perspective":

Optimize prices $g(x)$ and $f(y)$: maximize revenue while underbidding point-to-point transport



Price to depopulate at x
("pick up")

Transport details hidden!

Price to populate at y
("deliver")

Dual problem:

$$\hat{f}, \hat{g} = \arg\max_{f,g} \int dy\, q(y)\, f(y) + \int dx\, p(x) g(x)$$

$$g(x) + f(y) \leq c(x,y)$$

The dual problem

The dual problem is (much) easier to solve numerically:

$$\hat{f}, \hat{g} = \arg\max_{f,g} \int dy\, q(y)\, f(y) + \int dx\, p(x) g(x)$$

$$g(x) + f(y) \leq c(x, y)$$

Legendre transform in classical mechanics:

$$\underset{\text{Hamiltonian}}{H(p)} + \underset{\text{Lagrangian}}{L(\dot{q})} = p\dot{q}$$

The dual problem

The dual problem is (much) easier to solve numerically:

$$\hat{f}, \hat{g} = \underset{f,g}{\arg\max} \int dy\, q(y)\, f(y) + \int dx\, p(x) g(x)$$

$$g(x) + f(y) \leq c(x, y)$$

For $c(x,y) = |x - y|^2$, $\hat{f}$ and $\hat{g}$ are Legendre-conjugates!

Legendre transform in classical mechanics:

$$H(p) + L(\dot{q}) = p\dot{q}$$

Hamiltonian *Lagrangian*

$$\hat{g} = \underset{g \in \mathrm{cvx}}{\arg\max} \int dy\, q(y)\, g^*(y) + \int dx\, p(x) g(x)$$

Legendre transform: $g^*(y) = \max_x \left[x \cdot y - g(x)\right]$

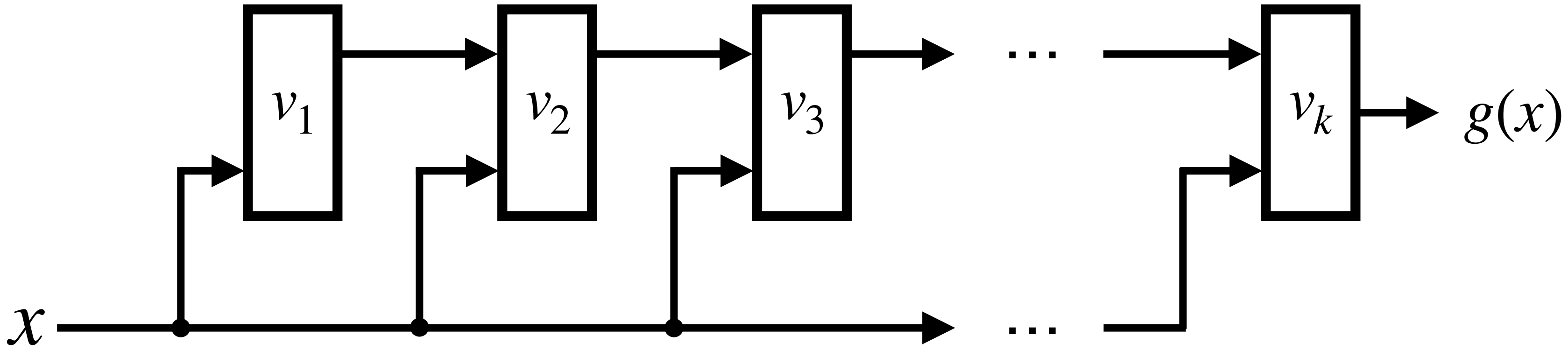
Maximise this "loss function" over all convex functions $g(x)$

Recover optimal transport function $\hat{T} = \nabla \hat{g}$

A numerical solution

Idea: parameterize set of convex functions, find maximum numerically

Input-convex neural networks [1609.07152]



"Compositions of convex functions with convex nondecreasing functions remain convex"

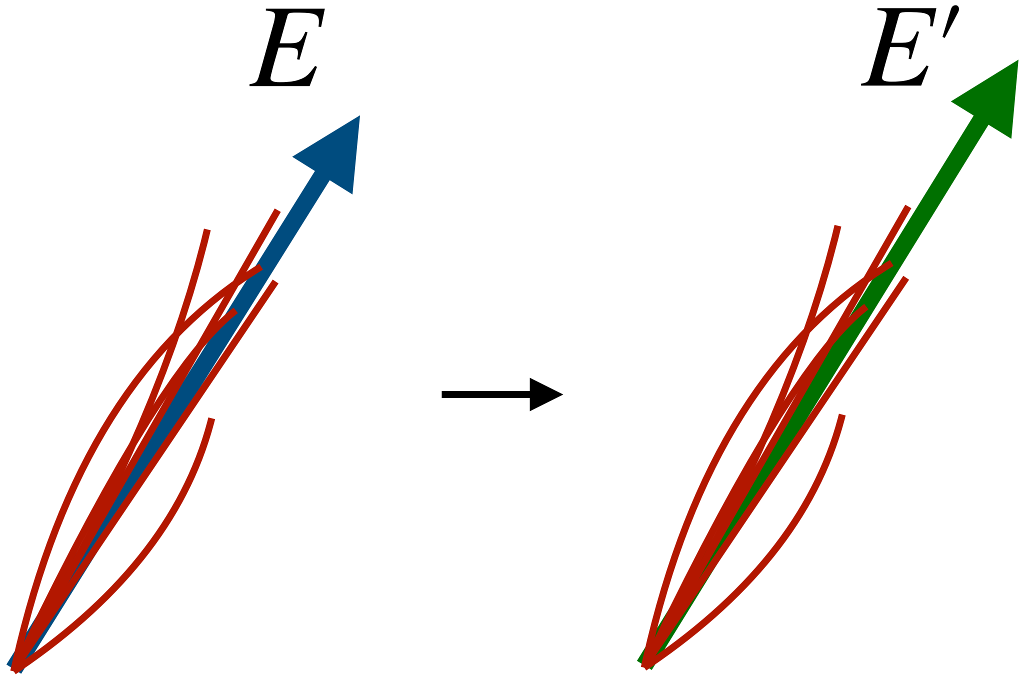
Very similar to standard feedforward networks,
but require convex nondecreasing activation functions and nonnegative weights

Optimal transport becomes tractable with modern Machine Learning infrastructure

"Just another loss function"

Very recent!

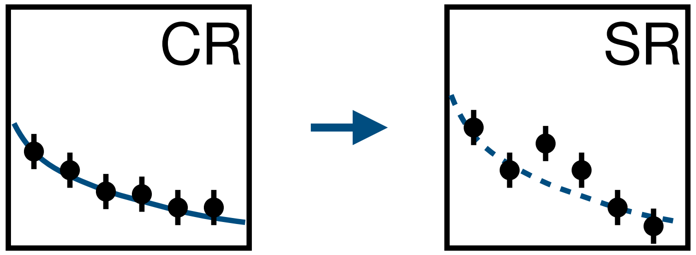
Takes both mathematical groundwork and modern neural network architectures to make large-scale optimal transport feasible in practice!



Applications in high-energy physics

Synergies with high-energy physics

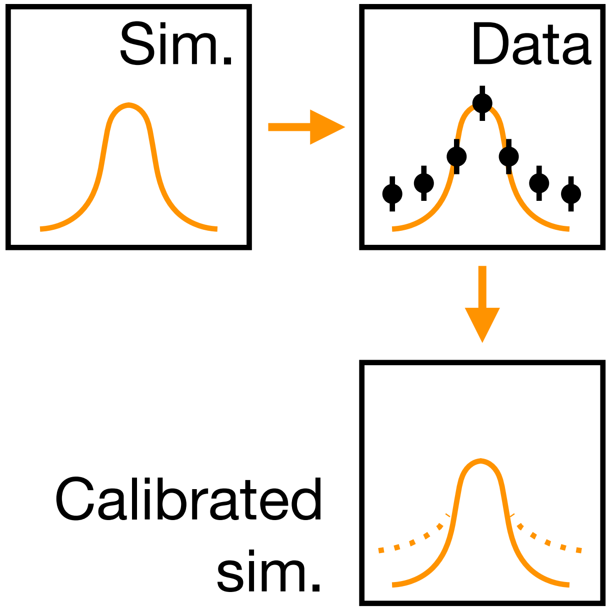
In particle physics, we manipulate (probability) **distributions on a daily basis …**



Extrapolation across phase space

(e.g. control region → signal region)

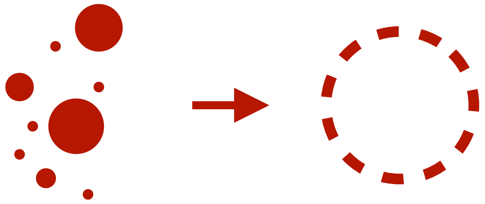
[2208.02807]



Calibration of simulation

(e.g. Monte Carlo prediction against data side bands)

[2107.08648]



Interpretation of data

(e.g. jet clustering)

[2004.04159]

… **optimal transport** provides **useful tools** *(and a unifying perspective)* for many of these!

Calibrating stochastic simulation

Collider-based particle physics is in a simulation-driven era!

Detailed **simulation models** encompass collective **domain knowledge**, from matrix elements (*TeV*) to detector signals (*eV*) … *tuned over decades!*

→ *Maximizes physics potential of our instruments*

But: simulations still need to be fine-tuned *("calibrated")* **to faithfully represent reality**

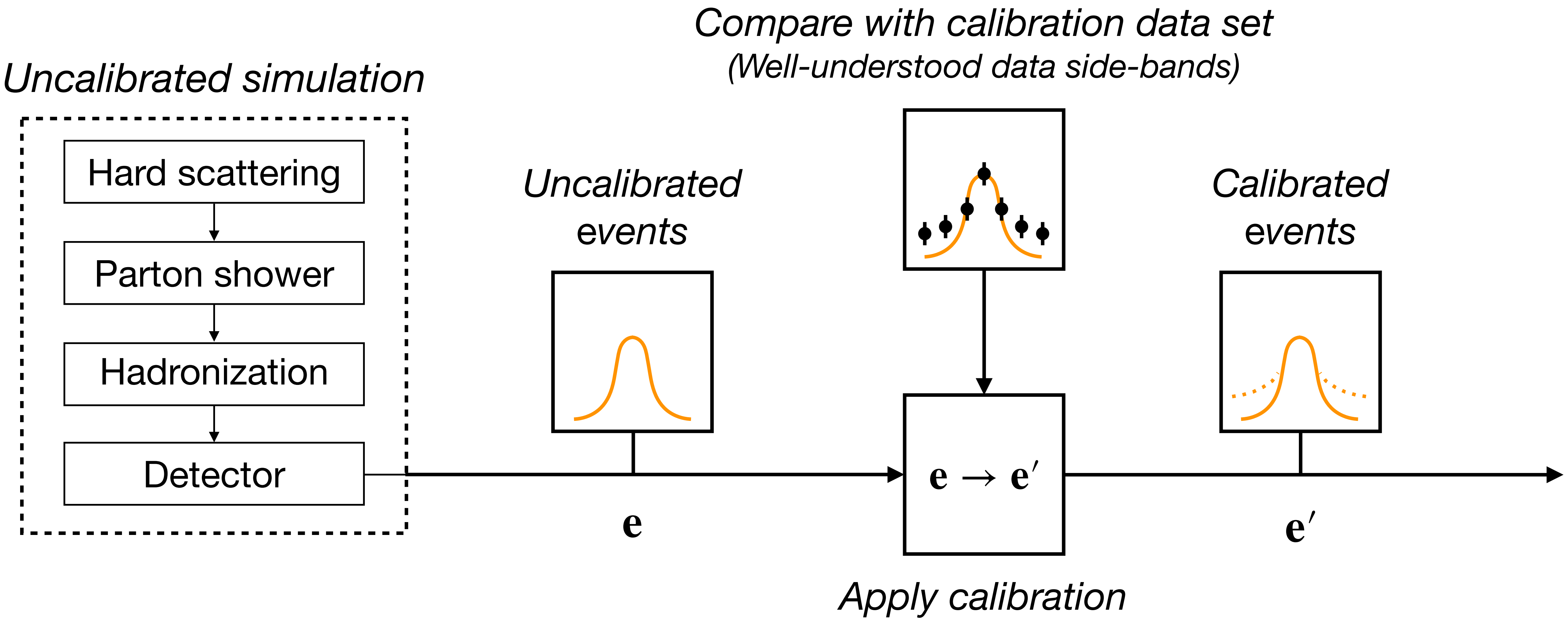
Calibrating stochastic simulation

Collider-based particle physics is in a simulation-driven era!

Detailed **simulation models** encompass collective **domain knowledge**, from matrix elements (*TeV*) to detector signals (*eV*) … *tuned over decades!*

→ *Maximizes physics potential of our instruments*

But: simulations still need to be fine-tuned (*"calibrated"*) **to faithfully represent reality**



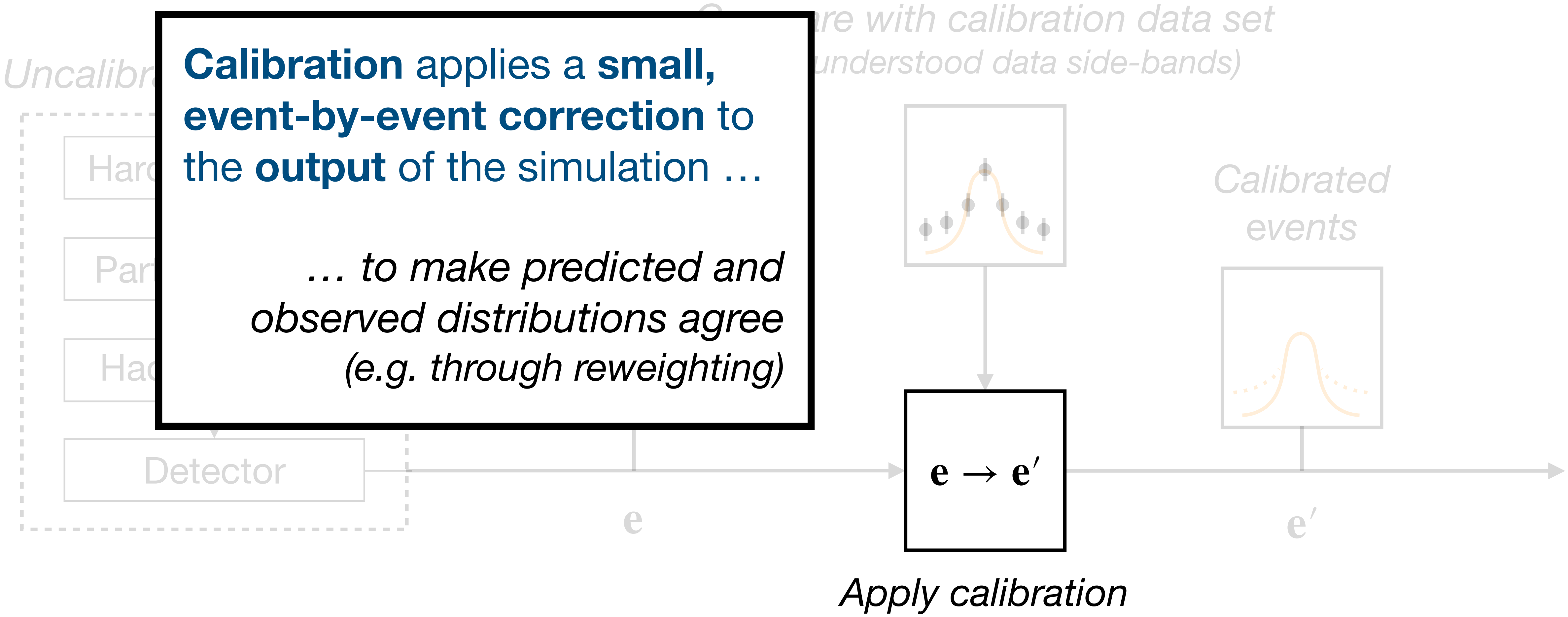
Calibrating stochastic simulation

Collider-based particle physics is in a simulation-driven era!

Detailed **simulation models** encompass collective **domain knowledge**,
from matrix elements (*TeV*) to detector signals (*eV*) … *tuned over decades!*

→ *Maximizes physics potential of our instruments*

But: simulations still need to be fine-tuned *("calibrated")*
to faithfully represent reality



Calibration applies a **small, event-by-event correction** to the **output** of the simulation …

… to make predicted and observed distributions agree (e.g. through reweighting)

Optimal transport for calibrations

Want to preserve domain knowledge in simulation!

Calibration *should result in the* ***smallest possible modification*** *of the original simulator that makes it* ***consistent with the data***

This is just the optimal transport problem!

(In Monge's form in case distributions are continuous)

$$\hat{T} = \arg\min_T \int d\mathbf{e}\ p(\mathbf{e})\ c(\mathbf{e}, T(\mathbf{e}))$$

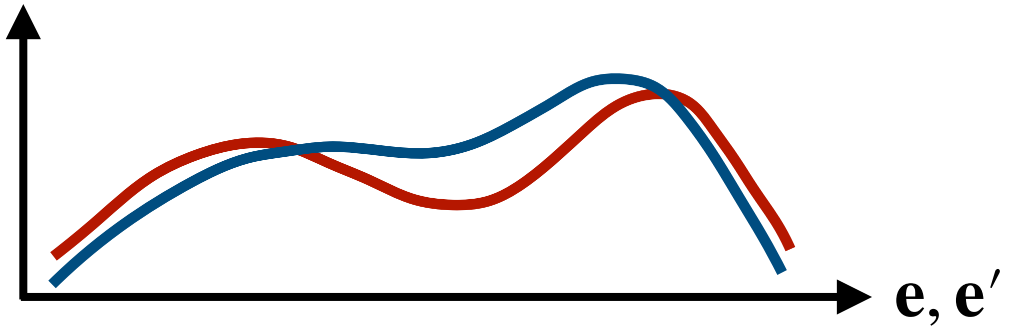
"Smallest possible modification" …

$$q(\mathbf{e}') = p(\mathbf{e})(\nabla T)^{-1}$$

… "consistent with the data"

$p(\mathbf{e})$ **… uncalibrated distribution**

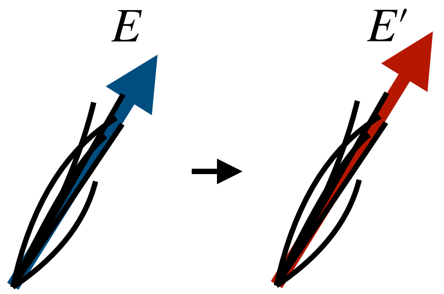
$q(\mathbf{e}')$ **… calibration data**



Calibration: ***unbinned***

per-event modification

$$\mathbf{e} \mapsto \mathbf{e}' = \hat{T}(\mathbf{e})$$



[2107.08648]

Optimal transport for calibrations

Want to preserve domain knowledge in simulation!

Calibration** should resul t in the **smalles

ulator

*that makes it **c***

This is just the o

(In Monge's form in ca

Which cost function to use?

Part of the problem specification!

Encodes degree of confidence in different aspects of the simulation

$$\hat{T} = \arg\min_{T} \int d\mathbf{e}\; p(\mathbf{e})\; c(\mathbf{e}, T(\mathbf{e}))$$

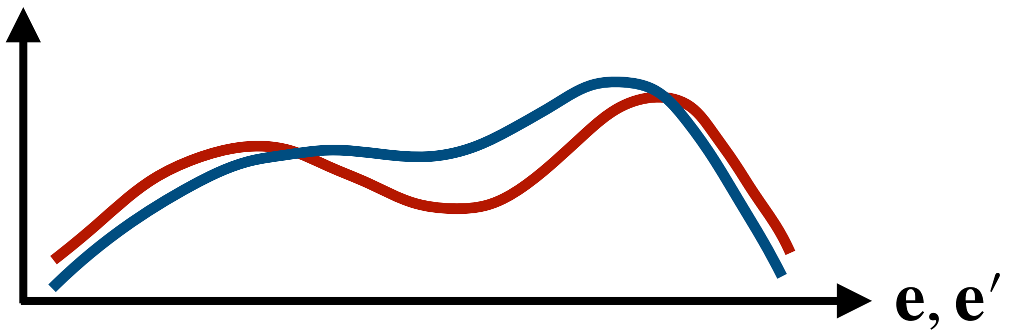
"Smallest possible modification" …

$$q(\mathbf{e}') = p(\mathbf{e})(\nabla T)^{-1}$$

… "consistent with the data"

$p(\mathbf{e})$ **… uncalibrated distribution**

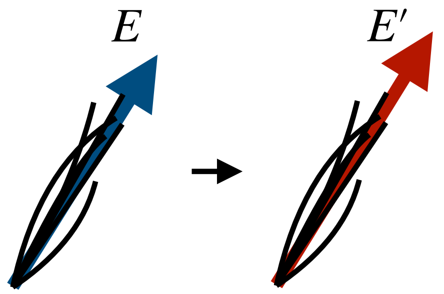
$q(\mathbf{e}')$ **… calibration data**



Calibration: ***unbinned***

per-event modification

$$\mathbf{e} \mapsto \mathbf{e}' = \hat{T}(\mathbf{e})$$

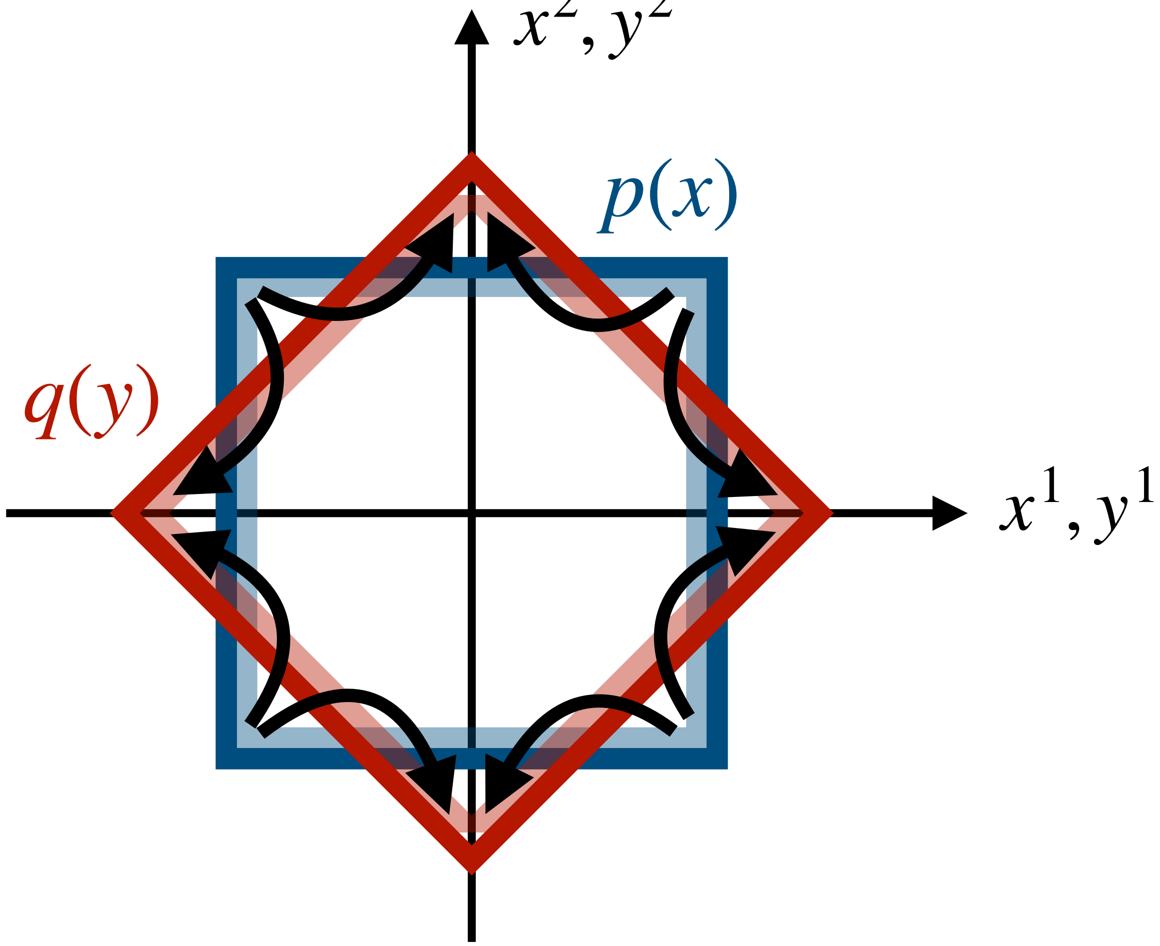


[2107.08648]

Calibrating simulations: the right cost function

Example from before: simulation of a square, but rotation angle incorrectly modeled

Uncalibrated simulation **Calibration data**

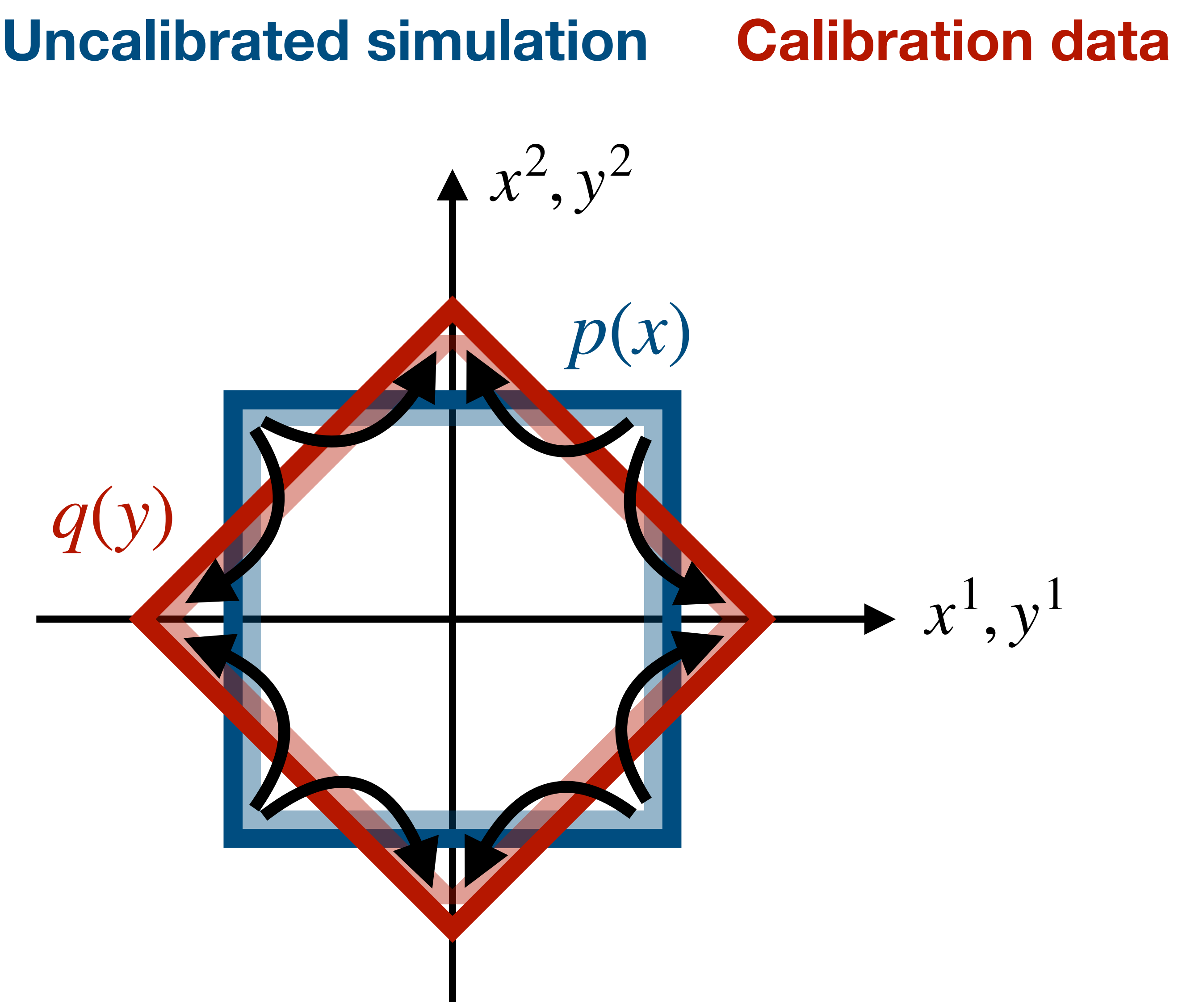


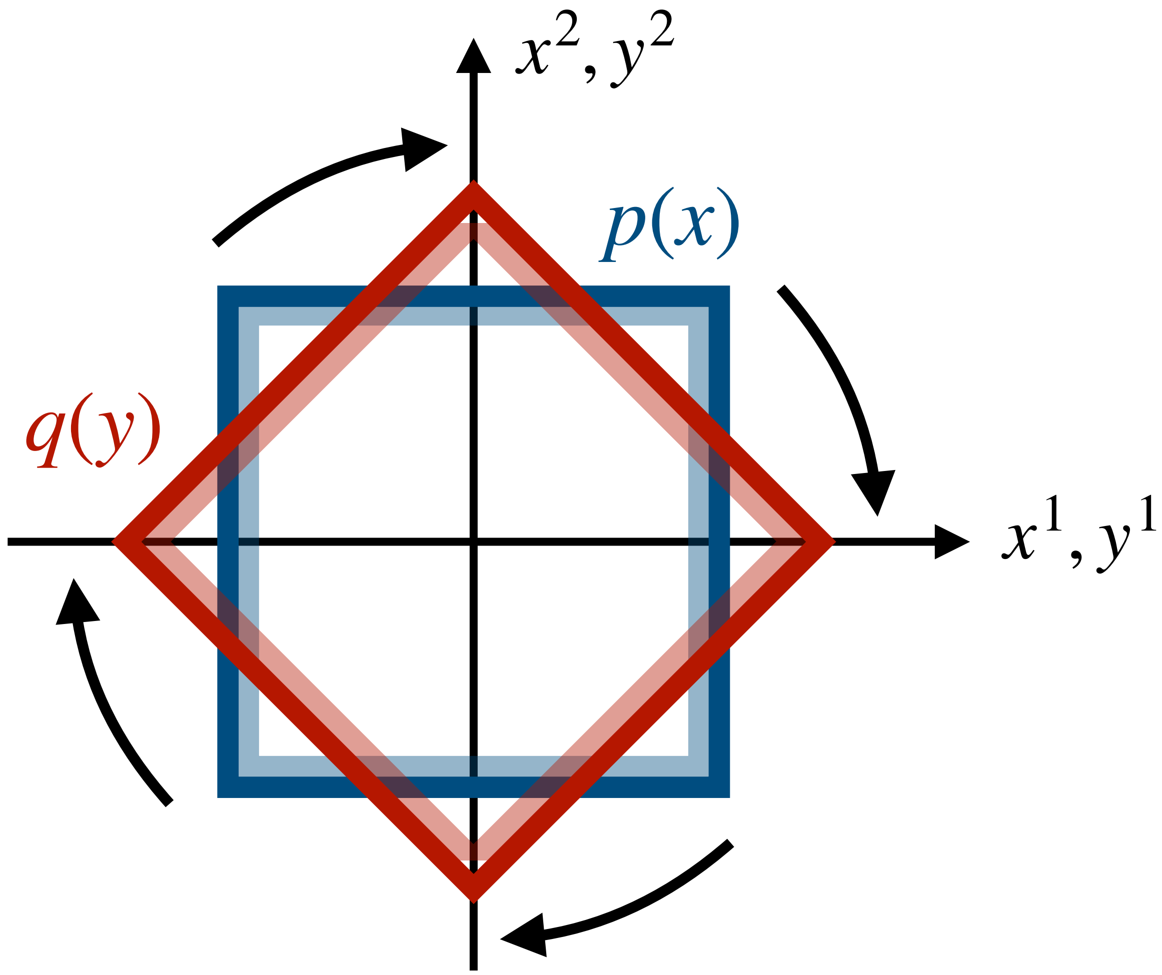
Optimal in Euclidean plane

$$ds^2 = dr^2 + r^2 d\phi^2$$

Calibrating simulations: the right cost function

Example from before: simulation of a square, but rotation angle incorrectly modeled





Optimal in Euclidean plane

$$ds^2 = dr^2 + r^2 d\phi^2$$

Optimal on a cone manifold

$$ds^2 = \alpha^2 dr^2 + r^2 d\phi^2, \alpha > 1$$

Use this if rotational degree of freedom is known to be poorly modeled

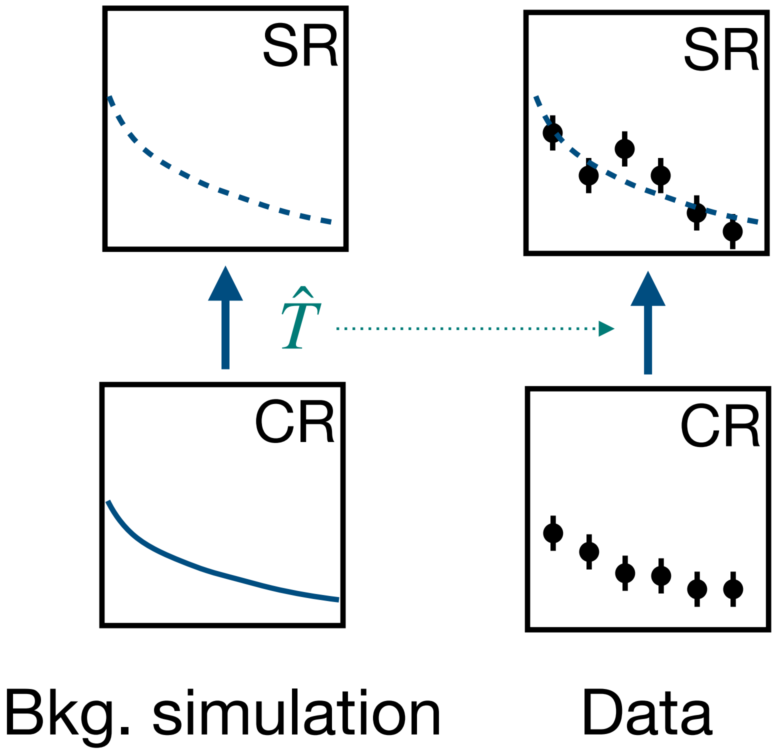
Extrapolation models

The calibration transfers information about one distribution *(calibration data)* **onto another** *(uncalibrated simulation)*

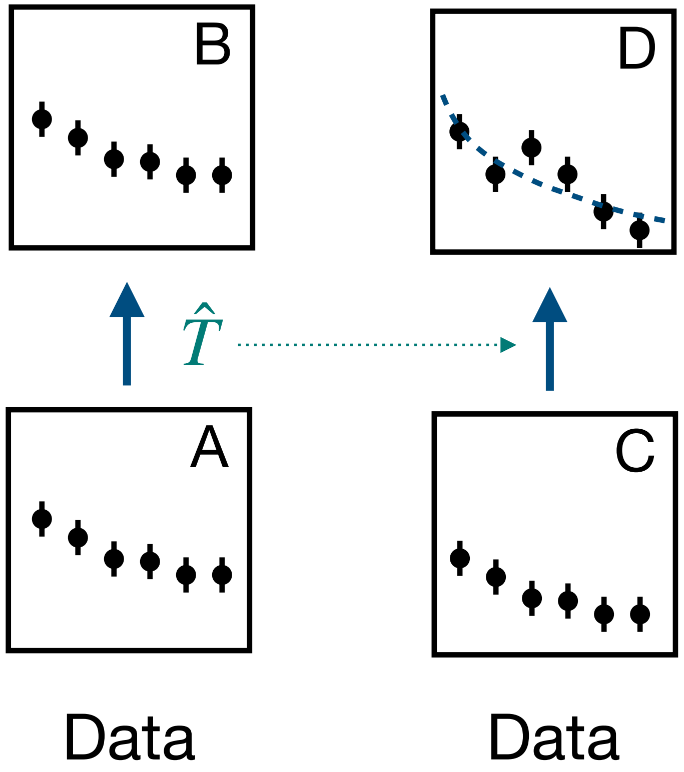
Very similar setting: extrapolation of backgrounds from control region into signal region

[2208.02807]

[2203.09470]



“Derive on simulation, apply to data”

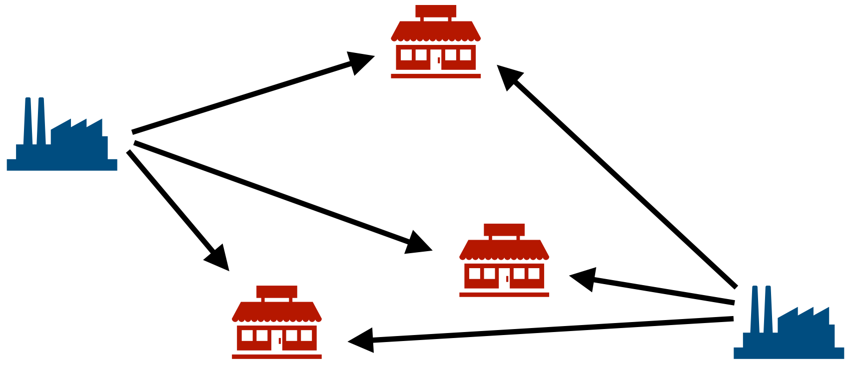


“ABCD method”

Optimal transport: ∃ *unbinned, high-dimensional equivalents to many established analysis techniques (but also no panacea!)*

Summary and outlook

Optimal transport: from a question in mathematics …



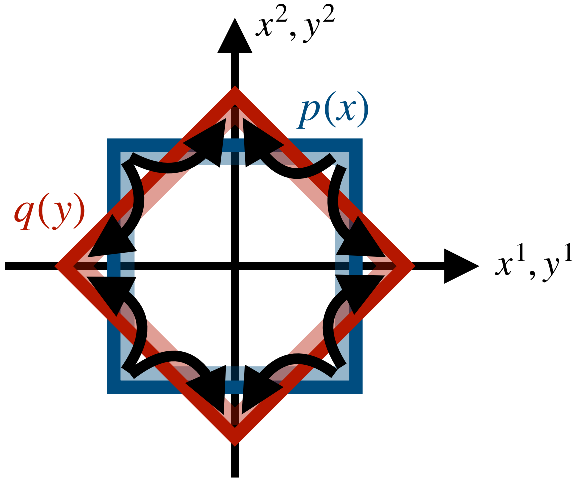
$$\hat{\pi} = \arg\min_{\pi} \int dx\, dy\ \pi(x,y)\ c(x,y)$$

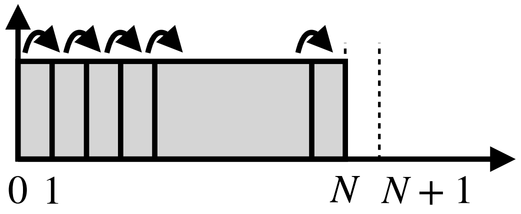
Kantorovich

$$\hat{T} = \arg\min_{T} \int dx\ p(x)\ c(x, T(x))$$

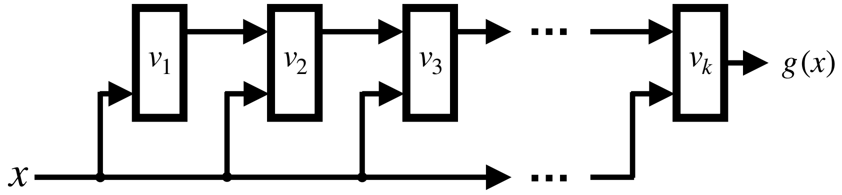
Monge

… with deep and intriguing solutions …

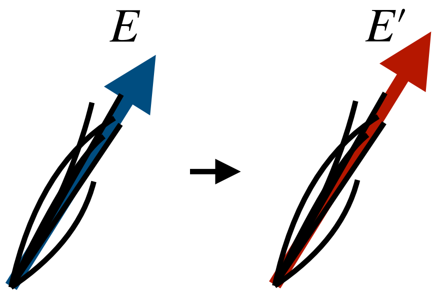




… made accessible through modern machine learning



… starting to enter high-energy physics!



A (non-exhaustive) set of references

Optimal transport

C. Villani, *“Topics in optimal transportation”*, Am. Math. Soc. (2003) [link]

W. Gangbo, R. J. McCann, *“The geometry of optimal transportation”*, Acta Math., 177, 113–161 (1996) [link]

Machine learning for optimal transport

A. V. Makkuva et al, *“Optimal transport mapping via input convex neural networks”*, PMLR 119:6672-6681 (2020) [link]

B. Amos et al, *“Input Convex Neural Networks”*, PMLR 70:146-155 (2017) [link]

S. Cohen et al, *“Riemannian Convex Potential Maps”*, PMLR 139:2028-2038 (2021) [link]

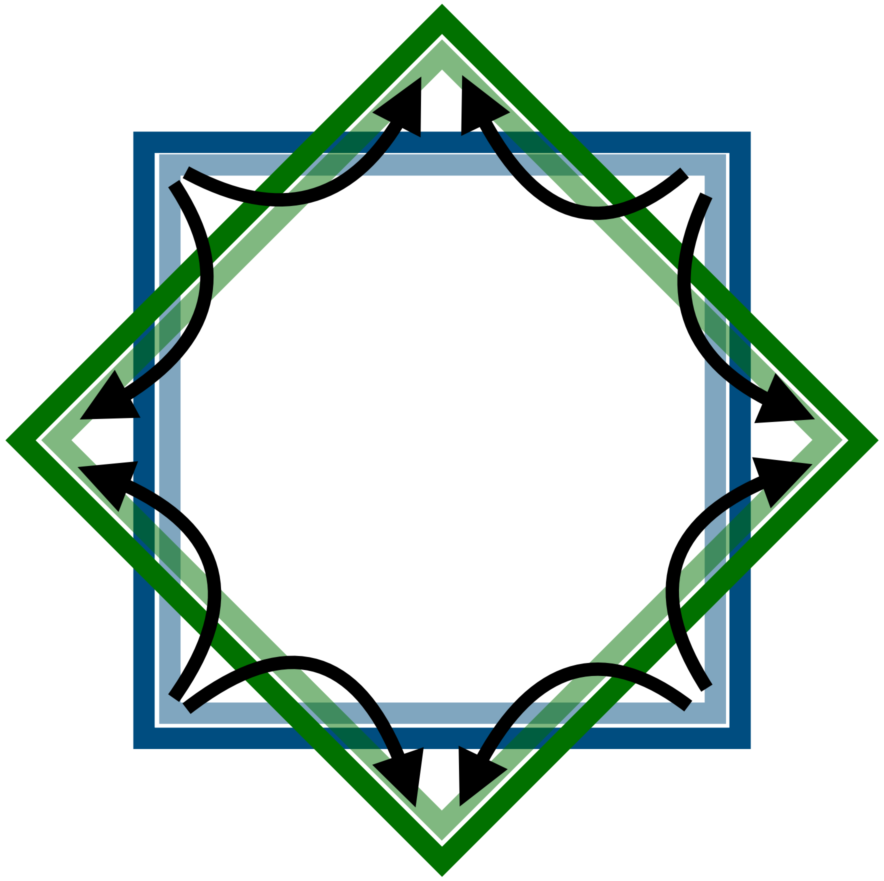
Applications in high-energy physics

P. T. Komiske et al, “*The hidden geometry of particle collisions*”, JHEP 2020, 6 (2020) [link]

T. Manoule et al, *“Background modeling for double Higgs boson production: density ratios and optimal transport”* [arXiv:2208.02807] (2022)

J. Raine et al, *“CURTAINs for your Sliding Window: Constructing Unobserved Regions by Transforming Adjacent Intervals”* [arXiv:2203.09470] (2022)

C. Pollard, PW, *“Transport away your problems: calibrating stochastic simulations with optimal transport”*, Nucl. Instrum. Meth. A 1027, 166119 (2022) [link]



Optimal transport in high-energy physics

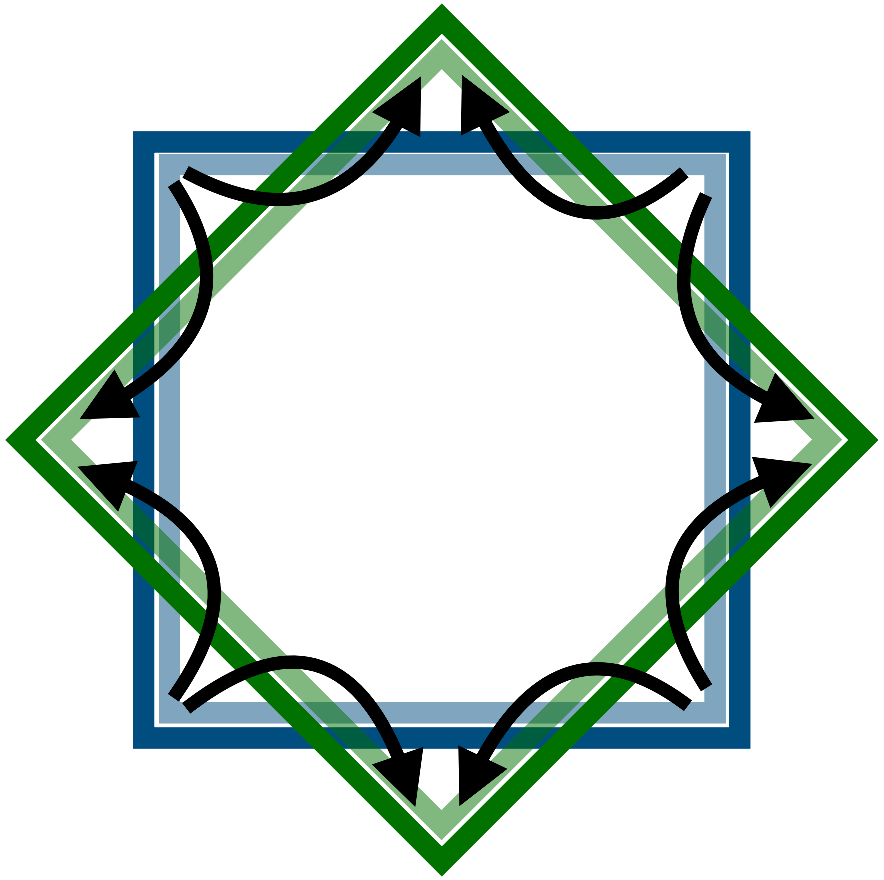
Theory and applications

October 12, 2022

philipp.windischhofer@uchicago.edu

Philipp Windischhofer
University of Chicago

THE UNIVERSITY OF CHICAGO



Backup

How to compute the Legendre transform?

Original form:

$$\hat{f}, \hat{g} = \arg\max_{f,g} \int dy\, q(y)\, f(y) + \int dx\, p(x) g(x)$$

$$g(x) + f(y) \leq c(x, y)$$

Turns out: maximum attained when $g(x)$ and $f(y)$ are Legendre-conjugates

$$\begin{aligned} g(x) \sim f^*(x) &= \max_y \left[x \cdot y - f(y)\right] \\ &= \max_h \left[x \cdot \nabla h(x) - f(\nabla h(x))\right] \end{aligned}$$

(h is an auxiliary function)

The problem then becomes

$$\hat{f} = \arg\min_{f \in \mathrm{cvx}} \max_{h \in \mathrm{cvx}} \int dy\, q(y) f(y) + \int dx\, p(x) \; \left[x \cdot \nabla h(x) - f(\nabla h(x))\right]$$

… where both f and h are convex functions.

General convex cost functions

For general convex cost functions $c(\mathbf{x}, \mathbf{y})$:

Transport potential

Cost-minimising transport function is of the form $\mathbf{x} \mapsto \mathbf{x} - \underbrace{\nabla c^{-1}}(\nabla \hat{g}(\mathbf{x}))$

Becomes the identity for

$$c(\mathbf{x}, \mathbf{y}) = \frac{1}{2} | \mathbf{x} - \mathbf{y} |^2$$

(as used in main body)

The transport potential is a c-concave function …

$$\hat{g}(\mathbf{x}) = \inf_{\mathbf{x}', \lambda} c(\mathbf{x}, \mathbf{x}') + \lambda$$

… it can be written as the superposition of shifted copies of the cost function

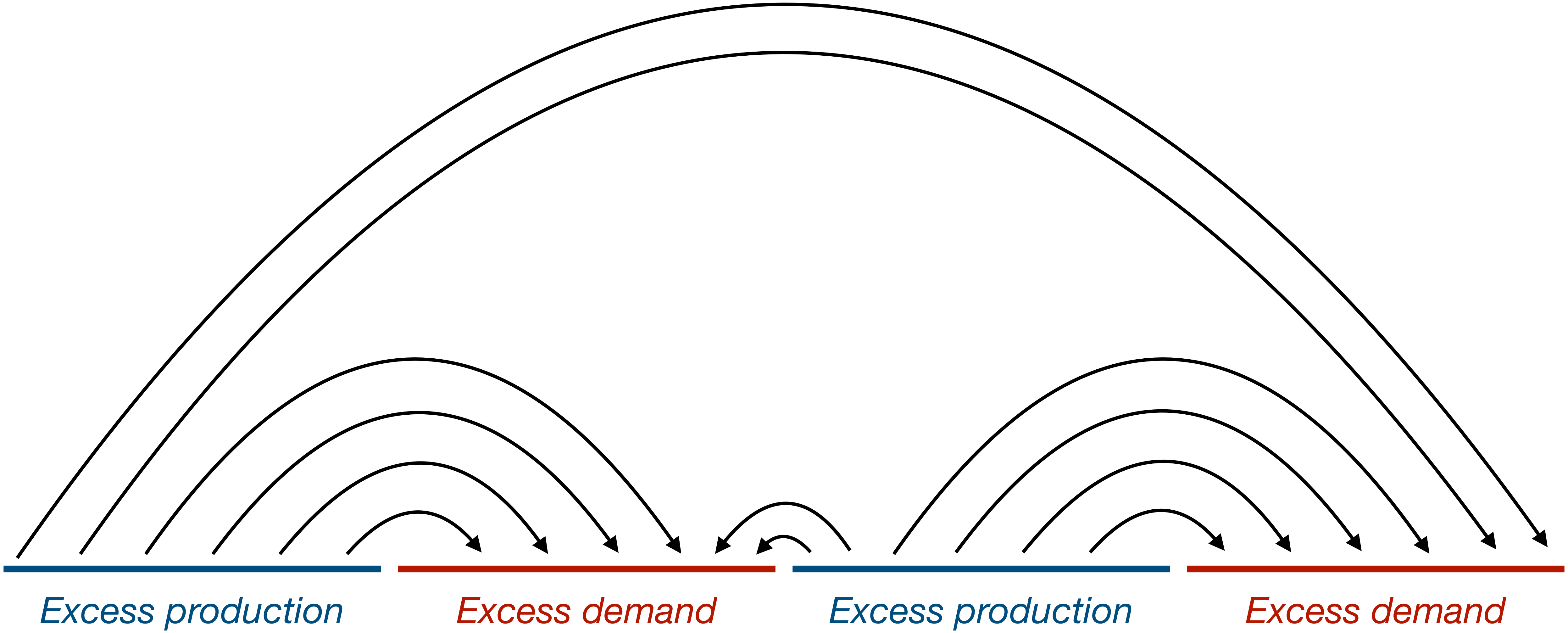
For $c(\mathbf{x}, \mathbf{y}) = \frac{1}{2} | \mathbf{x} - \mathbf{y} |^2$, this specializes to the standard definition of convexity for the potential on slide 38

Concave cost functions

What about concave $c(\mathbf{x}, \mathbf{y}) = h(|\mathbf{x} - \mathbf{y}|)$:

Useful in economics: **absolute cost** for transport **increases with distance**, but **cost per distance decreases** for longer legs

Solution has intricate structure: long-distance legs interspersed with local transport



R. McCann, *"Exact solutions to the transportation problem on the line"* [link]