On the relation between Kantorovich's and Shannon's optimization problems

Roman V. Belavkin

Faculty of Science and Technology
Middlesex University, London NW4 4BT, UK

March 6, 2018
Summer School on Operational Research and Applications
Higher School of Economics, Nizhny Novgorod, Russia

Optimal transportation problems (OTPs)

Information and entropy

Optimal channel problem (OCP)

Dual formulation of OTP

Geometry of information divergence and optimization

Relation between $K_c[p,q]$ and $D_{KL}[p,q]$

- $c : X \times Y \to \mathbb{R}$ — a cost function (e.g. a metric).

Relation between $K_c[p,q]$ and $D_{KL}[p,q]$

- $c : X \times Y \to \mathbb{R}$ — a cost function (e.g. a metric).
- $q \in \mathcal{P}(X)$, $p \in \mathcal{P}(Y)$, $\Gamma[q,p] \subset \mathcal{P}(X \times Y)$:

$$\Gamma[q,p] := \{w \in \mathcal{P}(X \otimes Y) : \pi_X w = q,\ \pi_Y w = p\}$$

Relation between $K_c[p,q]$ and $D_{KL}[p,q]$

- $c : X \times Y \to \mathbb{R}$ — a cost function (e.g. a metric).
- $q \in \mathcal{P}(X)$, $p \in \mathcal{P}(Y)$, $\Gamma[q,p] \subset \mathcal{P}(X \times Y)$:

$$\Gamma[q,p] := \{w \in \mathcal{P}(X \otimes Y) : \pi_X w = q,\ \pi_Y w = p\}$$

Proposition

Kantorovich (aka Wassershtain) metric (Kantorovich, 1939, 1942; Vasershtein, 1969; Dobrushin, 1970) between q, p:

$$K_c[p,q] := \inf\left\{ \int_{X\times Y} c\, dw : w \in \Gamma[q,p] \right\}$$

Relation between $K_c[p, q]$ and $D_{KL}[p, q]$

- $c : X \times Y \to \mathbb{R}$ — a cost function (e.g. a metric).
- $q \in \mathcal{P}(X)$, $p \in \mathcal{P}(Y)$, $\Gamma[q, p] \subset \mathcal{P}(X \times Y)$:

$$\Gamma[q, p] := \{w \in \mathcal{P}(X \otimes Y) : \pi_X w = q,\ \pi_Y w = p\}$$

Proposition

Kantorovich (aka Wassershtain) metric (Kantorovich, 1939, 1942; Vasershtein, 1969; Dobrushin, 1970) between q, p:

$$K_c[p, q](\lambda) := \inf\left\{ \int_{X \times Y} c\, dw : w \in \Gamma[q, p],\ D_{KL}[w, q \otimes p] \leq \lambda \right\}$$

Relation between $K_c[p,q]$ and $D_{KL}[p,q]$

- $c : X \times Y \to \mathbb{R}$ — a cost function (e.g. a metric).
- $q \in \mathcal{P}(X)$, $p \in \mathcal{P}(Y)$, $\Gamma[q,p] \subset \mathcal{P}(X \times Y)$:

$$\Gamma[q,p] := \{w \in \mathcal{P}(X \otimes Y) : \pi_X w = q,\ \pi_Y w = p\}$$

Proposition

Kantorovich (aka Wassershtain) metric (Kantorovich, 1939, 1942; Vasershtein, 1969; Dobrushin, 1970) between q, p:

$$K_c[p,q](\lambda) := \inf\left\{ \int_{X\times Y} c\, dw : w \in \Gamma[q,p]\,,\ D_{KL}[w, q\otimes q] - D_{KL}[p,q] \leq \lambda \right\}$$

Relation between $K_c[p,q]$ and $D_{KL}[p,q]$

- $c : X \times Y \to \mathbb{R}$ — a cost function (e.g. a metric).
- $q \in \mathcal{P}(X)$, $p \in \mathcal{P}(Y)$, $\Gamma[q,p] \subset \mathcal{P}(X \times Y)$:

$$\Gamma[q,p] := \{w \in \mathcal{P}(X \otimes Y) : \pi_X w = q,\ \pi_Y w = p\}$$

Proposition

Kantorovich (aka Wassershtain) metric (Kantorovich, 1939, 1942; Vasershtein, 1969; Dobrushin, 1970) between q, p:

$$K_c[p,q](\lambda) := \inf\left\{ \int_{X \times Y} c\, dw : w \in \Gamma[q,p]\,,\ \overbrace{D_{KL}[w, q \otimes q] - D_{KL}[p,q] \leq \lambda}^{\textit{Shannon's OCP}} \right\}$$

Relation between $K_c[p,q]$ and $D_{KL}[p,q]$

- $c : X \times Y \to \mathbb{R}$ — a cost function (e.g. a metric).
- $q \in \mathcal{P}(X)$, $p \in \mathcal{P}(Y)$, $\Gamma[q,p] \subset \mathcal{P}(X \times Y)$:

$$\Gamma[q,p] := \{w \in \mathcal{P}(X \otimes Y) : \pi_X w = q,\ \pi_Y w = p\}$$

Proposition

Kantorovich (aka Wassershtain) metric (Kantorovich, 1939, 1942; Vasershtein, 1969; Dobrushin, 1970) between q, p:

$$K_c[p,q](\lambda) := \inf\left\{ \int_{X\times Y} c\,dw : w \in \Gamma[q,p]\,,\ \overbrace{\underbrace{D_{KL}[w, q\otimes q] - D_{KL}[p,q]}_{\text{Shannon-Pythagor}} \le \lambda}^{\text{Shannon's OCP}} \right\}$$

Optimal transportation problems (OTPs)

Information and entropy

Optimal channel problem (OCP)

Dual formulation of OTP

Geometry of information divergence and optimization

Kantorovich's OTP

Optimal Transportation Problem (Kantorovich, 1939, 1942; Vasershtein, 1969; Dobrushin, 1970)

$$K_c[p, q] := \inf\left\{ \int_{X\times Y} c(x, y)\, dw : \pi_X w = q,\ \pi_Y w = p \right\}$$

where $c : X \times Y \to \mathbb{R}$ is a cost function (e.g. a metric).

$$\Gamma[q, p] := \{w \in \mathcal{P}(X \otimes Y) : \pi_X w = q,\ \pi_Y w = p\}$$

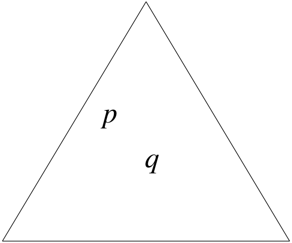
Kantorovich's OTP

Optimal Transportation Problem (Kantorovich, 1939, 1942; Vasershtein, 1969; Dobrushin, 1970)

$$K_c[p, q] := \inf \left\{ \int_{X\times Y} c(x, y)\, dw : \pi_X w = q,\ \pi_Y w = p \right\}$$

where $c : X \times Y \to \mathbb{R}$ is a cost function (e.g. a metric).

$$\Gamma[q, p] := \{w \in \mathcal{P}(X \otimes Y) : \pi_X w = q,\ \pi_Y w = p\}$$



$\mathcal{P}(X) \leftrightarrow \mathcal{P}(X \otimes Y)$

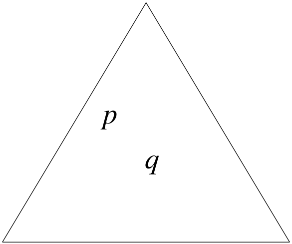
Kantorovich's OTP

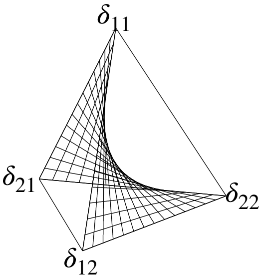
Optimal Transportation Problem (Kantorovich, 1939, 1942; Vasershtein, 1969; Dobrushin, 1970)

$$K_c[p,q] := \inf\left\{ \int_{X\times Y} c(x,y)\, dw : \pi_X w = q,\ \pi_Y w = p \right\}$$

where $c : X \times Y \to \mathbb{R}$ is a cost function (e.g. a metric).

$$\Gamma[q,p] := \{ w \in \mathcal{P}(X \otimes Y) : \pi_X w = q,\ \pi_Y w = p \}$$



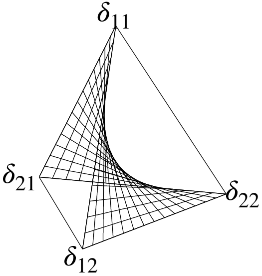


$\mathcal{P}(X) \leftrightarrow \mathcal{P}(X \otimes Y)$

Optimal Transport Plan

Optimal Transportation Problem

$$K_c[p,q] := \inf \left\{ \int_{X\times Y} c(x,y)\, dw \,:\, \pi_X w = q,\ \pi_Y w = p \right\}$$



Optimal Transport Plan

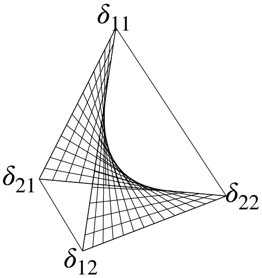
Optimal Transportation Problem

$$K_c[p,q] := \inf \left\{ \int_{X\times Y} c(x,y)\,dw \,:\, \pi_X w = q,\ \pi_Y w = p \right\}$$

Optimal Transport Plan

- Linear operator (Markov morphism) $T : \mathcal{P}(X) \to \mathcal{P}(Y)$:

$$q \mapsto Tq = p = \int_X p(y \mid x)\,dq$$



Optimal Transport Plan

Optimal Transportation Problem

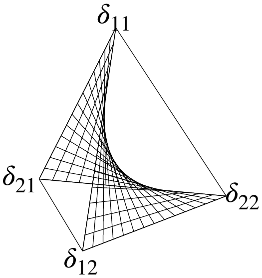
$$K_c[p,q] := \inf \left\{ \int_{X\times Y} c(x,y)\,dw \,:\, \pi_X w = q,\ \pi_Y w = p \right\}$$

Optimal Transport Plan

- Linear operator (Markov morphism) $T : \mathcal{P}(X) \to \mathcal{P}(Y)$:

$$q \mapsto Tq = p = \int_X p(y \mid x)\,dq$$

- $p(y \mid x)$ — Markov transition kernel



Optimal Transport Plan

Optimal Transportation Problem

$$K_c[p,q] := \inf \left\{ \int_{X\times Y} c(x,y)\,dw \,:\, \pi_X w = q,\ \pi_Y w = p \right\}$$

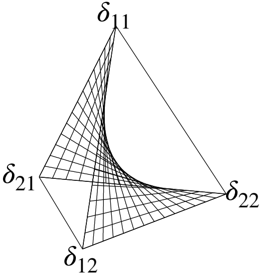
Optimal Transport Plan

- Linear operator (Markov morphism) $T : \mathcal{P}(X) \to \mathcal{P}(Y)$:

$$q \mapsto Tq = p = \int_X p(y \mid x)\,dq$$

- $p(y \mid x)$ — Markov transition kernel
- T is determined by $w \in \mathcal{P}(X \otimes Y)$:

$$w = p(y \mid x) \otimes q$$



Monge OTP

Optimal Transportation Problem (Monge, 1781)

$$K_c[p,q] := \inf \left\{ \int_X c(x, f(x))\, dq : f : p = q \circ f^{-1} \right\}$$

where $p = q \circ f^{-1}$ is push-forward under measurable mapping $f : X \to Y$:

$$p(E) = q \circ f^{-1}(E) = q\{x : f(x) \in E\}$$

Monge OTP

Optimal Transportation Problem (Monge, 1781)

$$K_c[p,q] := \inf \left\{ \int_X c(x,f(x))\, dq : f : p = q \circ f^{-1} \right\}$$

where $p = q \circ f^{-1}$ is push-forward under measurable mapping $f : X \to Y$:

$$p(E) = q \circ f^{-1}(E) = q\{x : f(x) \in E\}$$

Optimal Transport

- $p(E \mid x)$ has the form:

$$\delta_{f(x)}(E) = \begin{cases} 1 & \text{if } f(x) \in E \\ 0 & \text{otherwise} \end{cases}$$

Monge OTP

Optimal Transportation Problem (Monge, 1781)

$$K_c[p,q] := \inf \left\{ \int_X c(x,f(x))\, dq : f : p = q \circ f^{-1} \right\}$$

where $p = q \circ f^{-1}$ is push-forward under measurable mapping $f : X \to Y$:

$$p(E) = q \circ f^{-1}(E) = q\{x : f(x) \in E\}$$

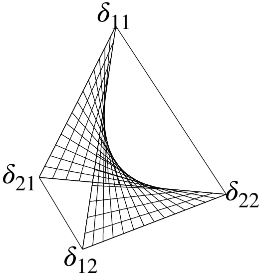
Optimal Transport

- $p(E \mid x)$ has the form:

$$\delta_{f(x)}(E) = \begin{cases} 1 & \text{if } f(x) \in E \\ 0 & \text{otherwise} \end{cases}$$

- $w_f \in \partial \mathcal{P}(X \otimes Y)$:

$$w_f(X, Y \setminus f(X)) = 0$$



Optimal transportation problems (OTPs)

Information and entropy

Optimal channel problem (OCP)

Dual formulation of OTP

Geometry of information divergence and optimization

Shannon's Information and Entropy

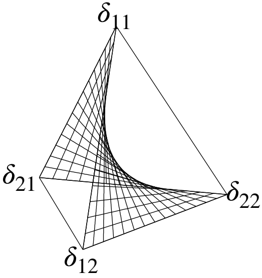
KL-divergence (Kullback & Leibler, 1951)

$$D_{KL}[p, q] = \int [\ln p - \ln q] \, dp$$

Shannon's Information and Entropy

KL-divergence (Kullback & Leibler, 1951)

$$D_{KL}[p,q] = \int [\ln p - \ln q]\, dp$$



Shannon's Information and Entropy

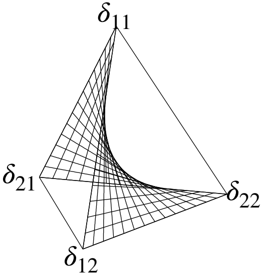
KL-divergence (Kullback & Leibler, 1951)

$$D_{KL}[p, q] = \int [\ln p - \ln q] \, dp$$

Shannon's information (Shannon, 1948)

For $w \in \Gamma[q, p] \subset \mathcal{P}(X \otimes Y)$:

$$\begin{aligned} I_w\{x, y\} :=& D_{KL}[w, q \otimes p] \\ =& H[q] - H[q(x \mid y)] \\ =& H[p] - H[p(y \mid x] \end{aligned}$$



Shannon's Information and Entropy

KL-divergence (Kullback & Leibler, 1951)

$$D_{KL}[p,q] = \int [\ln p - \ln q]\, dp$$

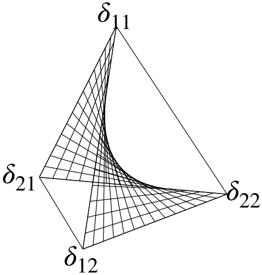
Shannon's information (Shannon, 1948)

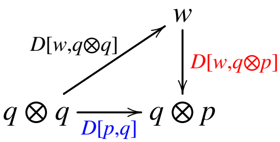
For $w \in \Gamma[q,p] \subset \mathcal{P}(X \otimes Y)$:

$$\begin{aligned} I_w\{x,y\} :=& D_{KL}[w, q \otimes p] \\ =& H[q] - H[q(x \mid y)] \\ =& H[p] - H[p(y \mid x] \end{aligned}$$

Entropy $H[p] = -\int \ln p \, dp$

$$H[p] := I_w\{y,y\} = \sup_{w : \pi_Y w = p} I_w\{x,y\}$$



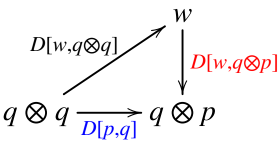

w
$D[w,q\otimes q]$
$D[w,q\otimes p]$
$q \otimes q$
$q \otimes p$
$D[p,q]$

Theorem (Shannon-Pythagorean)

- $w \in \mathcal{P}(X \otimes Y)$, $\pi_X w = q$, $\pi_Y w = p$

$$D_{KL}[w, q \otimes q] = D_{KL}[w, q \otimes p] + D_{KL}[p, q]$$

(Belavkin, 2013a)

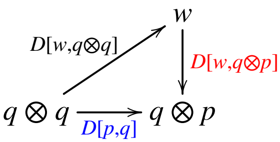


Theorem (Shannon-Pythagorean)

- $w \in \mathcal{P}(X \otimes Y)$, $\pi_X w = q$, $\pi_Y w = p$

$$D_{KL}[w, q \otimes q] = D_{KL}[w, q \otimes p] + D_{KL}[p, q]$$

(Belavkin, 2013a)



Proof.

$$D[w, q\otimes q] = \underbrace{D[w, q \otimes p]}_{I_w\{x,y\}} + \underbrace{D[q \otimes p, q \otimes q]}_{D[p,q]} - \underbrace{\langle \ln q \otimes p - \ln q \otimes q, q \otimes p - w\rangle}_{0}$$

$$\ln q \otimes p - \ln q \otimes q = 1_X \otimes (\ln p - \ln q)$$

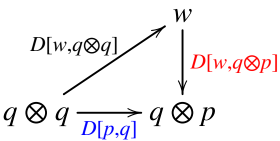
□

Theorem (Shannon-Pythagorean)

- $w \in \mathcal{P}(X \otimes Y)$, $\pi_X w = q$, $\pi_Y w = p$

$$D_{KL}[w, q \otimes q] = D_{KL}[w, q \otimes p] + D_{KL}[p, q]$$

(Belavkin, 2013a)



Proof.

$$D[w, q{\otimes}q] = \underbrace{D[w, q \otimes p]}_{I_w\{x,y\}} + \underbrace{D[q \otimes p, q \otimes q]}_{D[p,q]} - \underbrace{\langle \ln q \otimes p - \ln q \otimes q, q \otimes p - w \rangle}_{0}$$

$$\ln q \otimes p - \ln q \otimes q = 1_X \otimes (\ln p - \ln q)$$

□

Cross-Information (Belavkin, 2013a)

$$D_{KL}[w, q \otimes q] = \underbrace{-\langle \ln q, p \rangle}_{\text{Cross-entropy}} - \underbrace{\Big(H[p] - D_{KL}[w, q \otimes p]\Big)}_{H[p(y|x)]}$$

Optimal transportation problems (OTPs)

Information and entropy

Optimal channel problem (OCP)

Dual formulation of OTP

Geometry of information divergence and optimization

Shannon's OCP

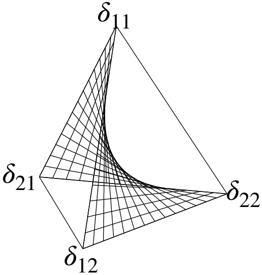
Optimal Channel Problem (Shannon, 1948)

$$R_c[q](\lambda) := \inf\left\{ \int_{X\times Y} c(x,y)\,dw \,:\, \pi_X w = q,\ I_w\{x,y\} \leq \lambda \right\}$$

Shannon's OCP

Optimal Channel Problem (Shannon, 1948)

$$R_c[q](\lambda) := \inf \left\{ \int_{X\times Y} c(x,y)\, dw \,:\, \pi_X w = q,\ I_w\{x,y\} \leq \lambda \right\}$$



Shannon's OCP

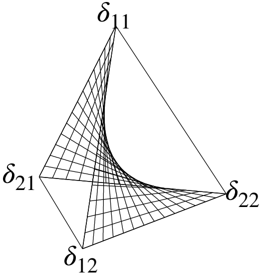
Optimal Channel Problem (Shannon, 1948)

$$R_c[q](\lambda) := \inf\left\{ \int_{X\times Y} c(x,y)\,dw \,:\, \pi_X w = q,\ I_w\{x,y\} \le \lambda \right\}$$

Exponential family solutions

- Optimal $T : \mathcal{P}(X) \to \mathcal{P}(Y)$ is defined by

$$w = e^{-\beta c - \ln Z}\, q \otimes p\,, \quad \beta^{-1} = -dR_c[q](\lambda)/d\lambda$$



Shannon's OCP

Optimal Channel Problem (Shannon, 1948)

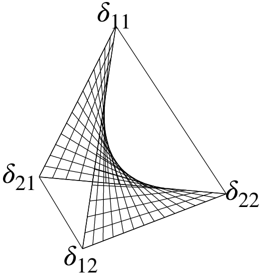
$$R_c[q](\lambda) := \inf\left\{ \int_{X\times Y} c(x,y)\,dw \,:\, \pi_X w = q,\ I_w\{x,y\} \leq \lambda \right\}$$

Exponential family solutions

- Optimal $T : \mathcal{P}(X) \to \mathcal{P}(Y)$ is defined by

$$w = e^{-\beta c - \ln Z}\, q \otimes p\,, \quad \beta^{-1} = -dR_c[q](\lambda)/d\lambda$$

- Observe that $w \notin \partial\mathcal{P}(X \otimes Y)$, unless $q \otimes p \in \partial\mathcal{P}(X \otimes Y)$ or $\beta \to \infty$.



Shannon's OCP

Optimal Channel Problem (Shannon, 1948)

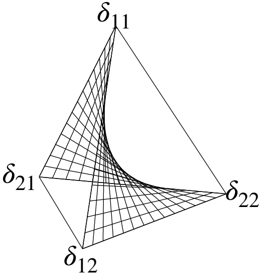
$$R_c[q](\lambda) := \inf\left\{ \int_{X\times Y} c(x,y)\,dw \,:\, \pi_X w = q,\ I_w\{x,y\} \leq \lambda \right\}$$

Exponential family solutions

- Optimal $T : \mathcal{P}(X) \to \mathcal{P}(Y)$ is defined by

  $$w = e^{-\beta c - \ln Z}\, q \otimes p\,, \quad \beta^{-1} = -dR_c[q](\lambda)/d\lambda$$

- Observe that $w \notin \partial\mathcal{P}(X \otimes Y)$, unless $q \otimes p \in \partial\mathcal{P}(X \otimes Y)$ or $\beta \to \infty$.



Value of Information (Stratonovich, 1965)

$$V(\lambda) := R_c[q](0) - R_c[q](\lambda) = \sup\{\mathbb{E}_w\{u\} \,:\, I_w\{x,y\} \leq \lambda\}$$

Relation to Kantorovich OTP

Optimal Channel Problem

$$R_c[q](\lambda) := \inf\left\{\int_{X\times Y} c(x,y)\,dw \,:\, \pi_X w = q,\ I_w\{x,y\} \leq \lambda\right\}$$

Relation to Kantorovich OTP

Optimal Channel Problem

$$R_c[q](\lambda) := \inf \left\{ \int_{X\times Y} c(x,y)\, dw \,:\, \pi_X w = q,\ I_w\{x,y\} \leq \lambda \right\}$$

Optimal Transportation Problem

$$K_c[q,p](\lambda) := \inf \left\{ \int_{X\times Y} c(x,y)\, dw \,:\, \pi_X w = q,\ \pi_Y w = p, \qquad \right\}$$

Relation to Kantorovich OTP

Optimal Channel Problem

$$R_c[q](\lambda) := \inf \left\{ \int_{X\times Y} c(x,y)\, dw \,:\, \pi_X w = q,\ I_w\{x,y\} \leq \lambda \right\}$$

Optimal Transportation Problem

$$K_c[q,p](\lambda) := \inf \left\{ \int_{X\times Y} c(x,y)\, dw \,:\, \pi_X w = q,\ \pi_Y w = p,\ I_w\{x,y\} \leq \lambda \right\}$$

Relation to Kantorovich OTP

Optimal Channel Problem

$$R_c[q](\lambda) := \inf\left\{\int_{X\times Y} c(x,y)\,dw \,:\, \pi_X w = q,\ I_w\{x,y\} \leq \lambda\right\}$$

Optimal Transportation Problem

$$K_c[q,p](\lambda) := \inf\left\{\int_{X\times Y} c(x,y)\,dw \,:\, \pi_X w = q,\ \pi_Y w = p,\ I_w\{x,y\} \leq \lambda\right\}$$

- q and p have entropies $H[q]$ and $H[p]$ and

$$0 \leq I_w\{x,y\} \leq \min\{H[q], H[p]\}$$

Relation to Kantorovich OTP

Optimal Channel Problem

$$R_c[q](\lambda) := \inf \left\{ \int_{X \times Y} c(x,y)\, dw \;:\; \pi_X w = q,\; I_w\{x,y\} \leq \lambda \right\}$$

Optimal Transportation Problem

$$K_c[q,p](\lambda) := \inf \left\{ \int_{X \times Y} c(x,y)\, dw \;:\; \pi_X w = q,\; \pi_Y w = p,\; I_w\{x,y\} \leq \lambda \right\}$$

- q and p have entropies $H[q]$ and $H[p]$ and

$$0 \leq I_w\{x,y\} \leq \min\{H[q], H[p]\}$$

- $K_c[q,p]$ has implicit constraint $I_w\{x,y\} \leq \lambda = \min\{H[q], H[p]\}$ and

$$R_c[q](\lambda) \leq K_c[q,p](\lambda)$$

Inverse Optimal Values

Inverse of the OCP Value

$$R_c^{-1}[q](\upsilon) := \inf\left\{ I_w\{x,y\} : \pi_X w = q,\ \int c\, dw \le \upsilon \right\}$$

Inverse Optimal Values

Inverse of the OCP Value

$$R_c^{-1}[q](\upsilon) := \inf \left\{ I_w\{x,y\} : \pi_X w = q,\ \int c\, dw \leq \upsilon \right\}$$

Inverse of the OTP Value

$$K_c^{-1}[q,p](\upsilon) := \inf \left\{ I_w\{x,y\} : \pi_X w = q,\ \pi_Y w = p,\ \int c\, dw \leq \upsilon \right\}$$

Inverse Optimal Values

Inverse of the OCP Value

$$R_c^{-1}[q](\upsilon) := \inf\left\{ I_w\{x,y\} : \pi_X w = q,\ \int c\,dw \leq \upsilon \right\}$$

Inverse of the OTP Value

$$K_c^{-1}[q,p](\upsilon) := \inf\left\{ I_w\{x,y\} : \pi_X w = q,\ \pi_Y w = p,\ \int c\,dw \leq \upsilon \right\}$$

- These inverse values represent the smallest amount of Shannon's information required to achieve expected cost $\int c\,dw = \upsilon$.

Inverse Optimal Values

Inverse of the OCP Value

$$R_c^{-1}[q](\upsilon) := \inf\left\{I_w\{x,y\} : \pi_X w = q,\ \int c\,dw \leq \upsilon\right\}$$

Inverse of the OTP Value

$$K_c^{-1}[q,p](\upsilon) := \inf\left\{I_w\{x,y\} : \pi_X w = q,\ \pi_Y w = p,\ \int c\,dw \leq \upsilon\right\}$$

- These inverse values represent the smallest amount of Shannon's information required to achieve expected cost $\int c\,dw = \upsilon$.
- If $\upsilon = K_c[q,p](\lambda)$, then

$$R_c^{-1}[q](\upsilon) \leq K_c^{-1}[q,p](\upsilon)$$

Common Solution

Theorem

Let w_{OCP} *and* $w_{OTP} \in \mathcal{P}(X \times Y)$ *be optimal solutions to OCP and OTP problems with the same constraint* $I(x,y) \leq \lambda$. *Then* $R_c[q](\lambda) = K_c[p,q](\lambda)$ *if and only if* $w_{OCP} = w_{OTP} \in \Gamma[p,q]$.

Common Solution

Theorem

Let w_{OCP} and $w_{OTP} \in \mathcal{P}(X \times Y)$ be optimal solutions to OCP and OTP problems with the same constraint $I(x, y) \leq \lambda$. Then $R_c[q](\lambda) = K_c[p, q](\lambda)$ if and only if $w_{OCP} = w_{OTP} \in \Gamma[p, q]$.

Proof.

- $w_{OCP} \in \partial D^*[-\beta c, q \otimes p]$ of subdifferential of D^* at $u = -\beta\, c$:

$$D^*[u, q \otimes p] = \ln \int_{X \times Y} e^{u(x,y)}\, dq(x)\, dp(y)$$

- $w_{OTP} \in \partial D^*[-\beta c, q \otimes p]$ implies

$$(1 - t)w_{OCP} + tw_{OTP} \in \partial D^*[-\beta c, q \otimes p]\,, \quad t \in [0, 1]$$

Common Solution

Theorem

Let w_{OCP} and $w_{OTP} \in \mathcal{P}(X \times Y)$ be optimal solutions to OCP and OTP problems with the same constraint $I(x,y) \leq \lambda$. Then $R_c[q](\lambda) = K_c[p,q](\lambda)$ if and only if $w_{OCP} = w_{OTP} \in \Gamma[p,q]$.

Proof.

- $w_{OCP} \in \partial D^*[-\beta c, q \otimes p]$ of subdifferential of D^* at $u = -\beta\, c$:

$$D^*[u, q \otimes p] = \ln \int_{X \times Y} e^{u(x,y)}\, dq(x)\, dp(y)$$

- $w_{OTP} \in \partial D^*[-\beta c, q \otimes p]$ implies

$$(1-t)w_{OCP} + t w_{OTP} \in \partial D^*[-\beta c, q \otimes p]\,, \quad t \in [0,1]$$

- and the KL-divergence $D[w, q \otimes p]$, the dual of $D^*[u, q \otimes p]$, not strictly convex.

Optimal transportation problems (OTPs)

Information and entropy

Optimal channel problem (OCP)

Dual formulation of OTP

Geometry of information divergence and optimization

Dual formulation

- Consider $f : X \to \mathbb{R}$ and $g : Y \to \mathbb{R}$ such that

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

Dual formulation

- Consider $f : X \to \mathbb{R}$ and $g : Y \to \mathbb{R}$ such that

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

- Define

$$J_c[p, q] := \sup \left\{ \mathbb{E}_p\{f\} - \mathbb{E}_q\{g\} : f(x) - g(y) \leq c(x, y) \right\}$$

Dual formulation

- Consider $f : X \to \mathbb{R}$ and $g : Y \to \mathbb{R}$ such that

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

- Define

$$J_c[p, q] := \sup \left\{ \mathbb{E}_p\{f\} - \mathbb{E}_q\{g\} : f(x) - g(y) \leq c(x, y) \right\}$$

- Clearly

$$J_c[p, q] \leq K_c[p, q]$$

Dual formulation

- Consider $f : X \to \mathbb{R}$ and $g : Y \to \mathbb{R}$ such that

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

- Define

$$J_c[p, q] := \sup \left\{ \mathbb{E}_p\{f\} - \mathbb{E}_q\{g\} : f(x) - g(y) \leq c(x, y) \right\}$$

- Clearly

$$J_c[p, q] \leq K_c[p, q]$$

- Can $J_c[p, q]$ be related to $D_{KL}[p, q]$?

KL-divergence

- $$\begin{aligned} D[p,q] &= D[p,r] + D[r,q] - \int_X \ln \frac{dq(x)}{dr(x)} \, [dp(x) - dr(x)] \\ &= D[p,r] - D[q,r] - \int_X \ln \frac{dq(x)}{dr(x)} \, [dp(x) - dq(x)] \end{aligned}$$

KL-divergence

- $$D[p,q] = D[p,r] + D[r,q] - \int_X \ln \frac{dq(x)}{dr(x)}\,[dp(x) - dr(x)]$$
  $$= D[p,r] - D[q,r] - \int_X \ln \frac{dq(x)}{dr(x)}\,[dp(x) - dq(x)]$$
- Let $f(x) - g(y) \leq c(x,y)$ satisfy also:
  $$\beta f(x) = \nabla D[p,r] = \ln \frac{dp(x)}{dr(x)}\,, \qquad \beta \geq 0$$
  $$\alpha g(x) = \nabla D[q,r] = \ln \frac{dq(x)}{dr(x)}\,, \qquad \alpha \geq 0$$

KL-divergence

- $$D[p,q] = D[p,r] + D[r,q] - \int_X \ln \frac{dq(x)}{dr(x)} \, [dp(x) - dr(x)]$$
  $$= D[p,r] - D[q,r] - \int_X \ln \frac{dq(x)}{dr(x)} \, [dp(x) - dq(x)]$$
- Let $f(x) - g(y) \leq c(x,y)$ satisfy also:
  $$\beta f(x) = \nabla D[p,r] = \ln \frac{dp(x)}{dr(x)} \,, \qquad \beta \geq 0$$
  $$\alpha g(x) = \nabla D[q,r] = \ln \frac{dq(x)}{dr(x)} \,, \qquad \alpha \geq 0$$
- Thus $dp(x) = e^{\beta f(x) - \kappa[\beta f]} \, dr(x)$, $dq(x) = e^{\alpha \, g(x) - \kappa[\alpha g]} \, dr(x)$ and
  $$D[p,r] = \beta \, \mathbb{E}_p\{f\} - \kappa[\beta f] \,, \quad D[q,r] = \alpha \, \mathbb{E}_q\{g\} - \kappa[\alpha \, g]$$

KL-divergence

- $$D[p,q] = D[p,r] + D[r,q] - \int_X \ln \frac{dq(x)}{dr(x)} \, [dp(x) - dr(x)]$$
  $$= D[p,r] - D[q,r] - \int_X \ln \frac{dq(x)}{dr(x)} \, [dp(x) - dq(x)]$$
- Let $f(x) - g(y) \leq c(x,y)$ satisfy also:
  $$\beta f(x) = \nabla D[p,r] = \ln \frac{dp(x)}{dr(x)}\,, \qquad \beta \geq 0$$
  $$\alpha g(x) = \nabla D[q,r] = \ln \frac{dq(x)}{dr(x)}\,, \qquad \alpha \geq 0$$
- Thus $dp(x) = e^{\beta f(x) - \kappa[\beta f]} \, dr(x)$, $dq(x) = e^{\alpha\, g(x) - \kappa[\alpha g]} \, dr(x)$ and
  $$D[p,r] = \beta \, \mathbb{E}_p\{f\} - \kappa[\beta f]\,, \quad D[q,r] = \alpha \, \mathbb{E}_q\{g\} - \kappa[\alpha\, g]$$
- Therefore
  $$D[p,q] = \beta \mathbb{E}_p\{f\} - \alpha \mathbb{E}_q\{g\} - (\kappa[\beta f] - \kappa[\alpha g]) - \alpha \int_X g(x) \, [dp(x) - dq(x)]$$

KL-divergence (cont.)

- Therefore

$$D[p,q] = \beta \mathbb{E}_p\{f\} - \alpha \mathbb{E}_q\{g\} - (\kappa[\beta f] - \kappa[\alpha g]) - \alpha \int_X g(x)\,[dp(x) - dq(x)]$$

KL-divergence (cont.)

- Therefore

$$D[p,q] = \beta\mathbb{E}_p\{f\} - \alpha\mathbb{E}_q\{g\} - (\kappa[\beta f] - \kappa[\alpha g]) - \alpha\int_X g(x)\,[dp(x) - dq(x)]$$

- Let us denote

$$J_{c,\varepsilon}[p,q] := \frac{1}{\varepsilon}\left[\beta\mathbb{E}_p\{f\} - \alpha\mathbb{E}_q\{g\}\right]$$

where $\varepsilon = \inf\{\epsilon \geq 0 : \beta f(x) - \alpha g(y) \leq \epsilon\, c(x,y)\}$.

KL-divergence (cont.)

- Therefore

$$D[p,q] = \beta\mathbb{E}_p\{f\} - \alpha\mathbb{E}_q\{g\} - (\kappa[\beta f] - \kappa[\alpha g]) - \alpha\int_X g(x)\,[dp(x) - dq(x)]$$

- Let us denote

$$J_{c,\varepsilon}[p,q] := \frac{1}{\varepsilon}\left[\beta\mathbb{E}_p\{f\} - \alpha\mathbb{E}_q\{g\}\right]$$

where $\varepsilon = \inf\{\epsilon \geq 0 \,:\, \beta f(x) - \alpha g(y) \leq \epsilon\, c(x,y)\}$.

- We have the following inequality:

$$J_{c,\varepsilon}[p,q] \leq J_c[p,q] \leq K_c[p,q]$$

KL-divergence (cont.)

- Therefore

$$D[p,q] = \beta\mathbb{E}_p\{f\} - \alpha\mathbb{E}_q\{g\} - (\kappa[\beta f] - \kappa[\alpha g]) - \alpha \int_X g(x)\,[dp(x) - dq(x)]$$

- Let us denote

$$J_{c,\varepsilon}[p,q] := \frac{1}{\varepsilon}\left[\beta\mathbb{E}_p\{f\} - \alpha\mathbb{E}_q\{g\}\right]$$

where $\varepsilon = \inf\{\epsilon \geq 0 : \beta f(x) - \alpha g(y) \leq \epsilon\, c(x,y)\}$.

- We have the following inequality:

$$J_{c,\varepsilon}[p,q] \leq J_c[p,q] \leq K_c[p,q]$$

- and

$$D[p,q] \leq \varepsilon\, K_c[p,q] - (\kappa[\beta f] - \kappa[\alpha g]) - \alpha \int g(x)\,[dp(x) - dq(x)]$$

Common solution for Dual OTP

Theorem

Let (f, g) be the solution to the dual OTP. If there exists a reference measure $r \in \mathcal{P}(X)$ such that $f = \nabla D[p, r]$ and $g = \nabla D[q, r]$, then

$$D[p, q] = K_c[p, q] - (\kappa[f] - \kappa[g]) - \int g(x)\,[dp(x) - dq(x)]$$

Common solution for Dual OTP

Theorem

Let (f, g) be the solution to the dual OTP. If there exists a reference measure $r \in \mathcal{P}(X)$ such that $f = \nabla D[p, r]$ and $g = \nabla D[q, r]$, then

$$D[p, q] = K_c[p, q] - (\kappa[f] - \kappa[g]) - \int g(x)\,[dp(x) - dq(x)]$$

Proof.

- $f = \nabla D[p, r]$ and $g = \nabla D[q, r]$ imply that $\alpha = \beta = 1$, and

$$p = \exp(f - \kappa[f])\,r\,, \quad q = \exp(g - \kappa[g])\,r$$

- The result follows, because

$$\mathbb{E}_p\{f\} - \mathbb{E}_q\{g\} = J_c[p, q] = K_c[p, q]$$

Optimal transportation problems (OTPs)

Information and entropy

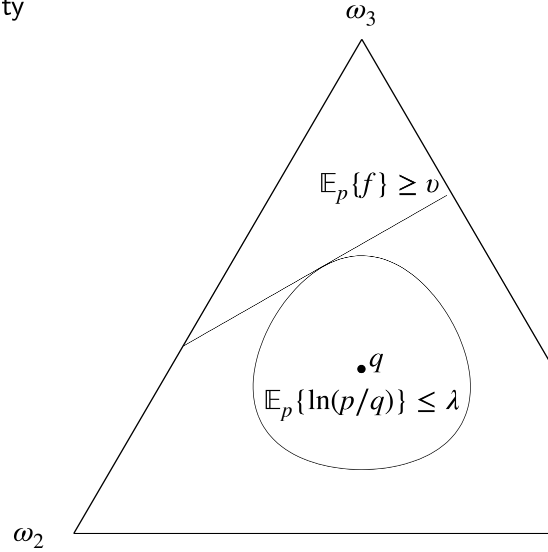
Optimal channel problem (OCP)

Dual formulation of OTP

Geometry of information divergence and optimization

Problems on Conditional Extremum

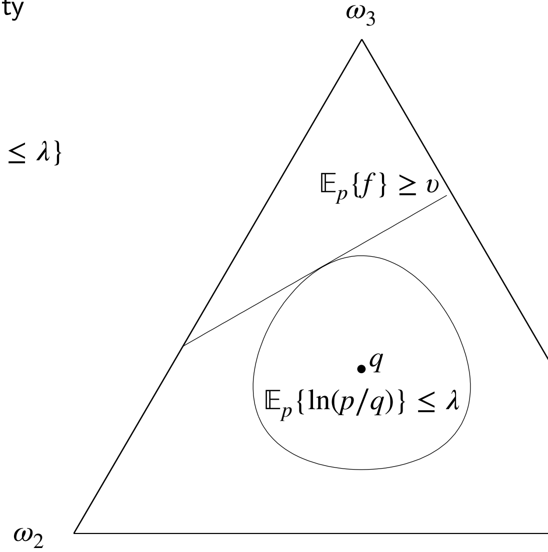
- $\mathbb{E}_p\{u\} = \langle u, p \rangle$ expected utility



Problems on Conditional Extremum

- $\mathbb{E}_p\{u\} = \langle u, p \rangle$ expected utility
- $v_u(\lambda) = -R_{-u}[q](\lambda)$ utility of information λ:

$$v_u(\lambda) := \sup\{\langle u, p \rangle : F[p, q] \leq \lambda\}$$



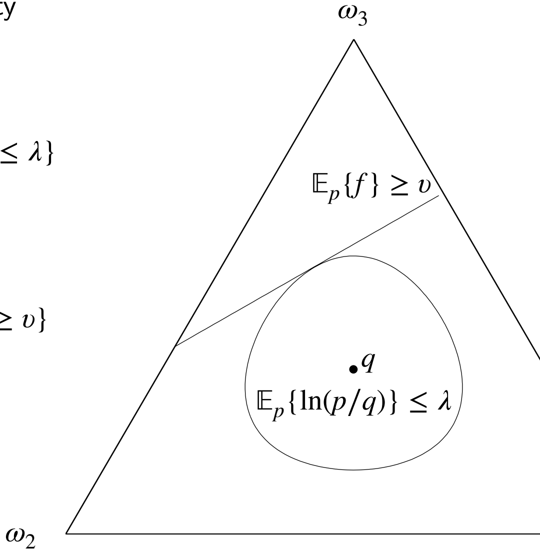
Problems on Conditional Extremum

- $\mathbb{E}_p\{u\} = \langle u, p\rangle$ expected utility
- $\upsilon_u(\lambda) = -R_{-u}[q](\lambda)$ utility of information λ:

$$\upsilon_u(\lambda) := \sup\{\langle u, p\rangle : F[p, q] \leq \lambda\}$$

- $\lambda_u(\upsilon) = \upsilon_u^{-1}(\upsilon)$ information of utility υ:

$$\lambda_u(\upsilon) := \inf\{F[p, q] : \langle u, p\rangle \geq \upsilon\}$$



Problems on Conditional Extremum

- $\mathbb{E}_p\{u\} = \langle u, p \rangle$ expected utility
- $\upsilon_u(\lambda) = -R_{-u}[q](\lambda)$ utility of information λ:

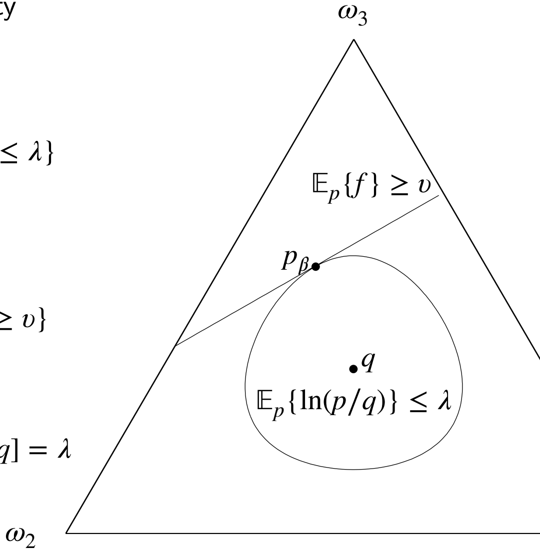
$$\upsilon_u(\lambda) := \sup\{\langle u, p\rangle : F[p, q] \leq \lambda\}$$

- $\lambda_u(\upsilon) = \upsilon_u^{-1}(\upsilon)$ information of utility υ:

$$\lambda_u(\upsilon) := \inf\{F[p, q] : \langle u, p\rangle \geq \upsilon\}$$

- $p(\beta)$ optimal solutions:

$$p(\beta) \in \partial F^*[\beta u, q]\,, \quad F[p(\beta), q] = \lambda$$



Problems on Conditional Extremum

- $\mathbb{E}_p\{u\} = \langle u, p\rangle$ expected utility
- $\upsilon_u(\lambda) = -R_{-u}[q](\lambda)$ utility of information λ:

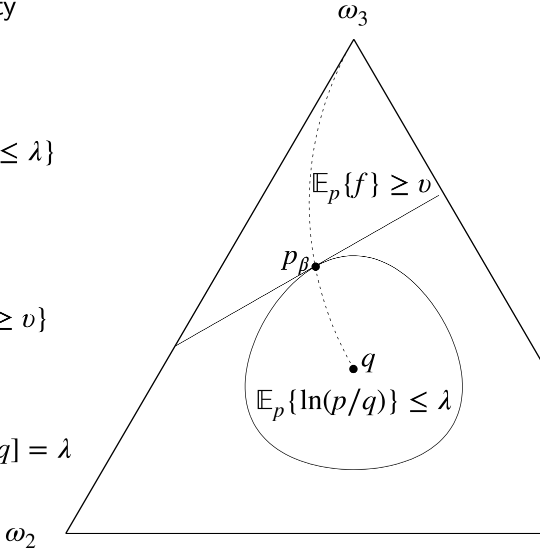
$$\upsilon_u(\lambda) := \sup\{\langle u, p\rangle : F[p,q] \le \lambda\}$$

- $\lambda_u(\upsilon) = \upsilon_u^{-1}(\upsilon)$ information of utility υ:

$$\lambda_u(\upsilon) := \inf\{F[p,q] : \langle u, p\rangle \ge \upsilon\}$$

- $p(\beta)$ optimal solutions:

$$p(\beta) \in \partial F^*[\beta u, q]\,, \quad F[p(\beta), q] = \lambda$$



Problems on Conditional Extremum

- $\mathbb{E}_p\{u\} = \langle u, p\rangle$ expected utility
- $\upsilon_u(\lambda) = -R_{-u}[q](\lambda)$ utility of information λ:

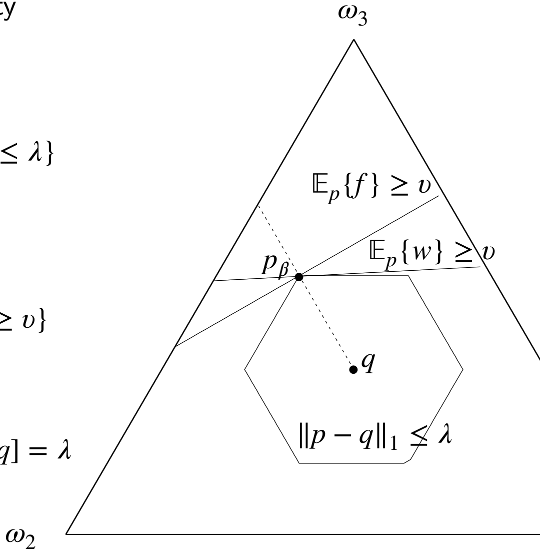
$$\upsilon_u(\lambda) := \sup\{\langle u, p\rangle : F[p,q] \leq \lambda\}$$

- $\lambda_u(\upsilon) = \upsilon_u^{-1}(\upsilon)$ information of utility υ:

$$\lambda_u(\upsilon) := \inf\{F[p,q] : \langle u, p\rangle \geq \upsilon\}$$

- $p(\beta)$ optimal solutions:

$$p(\beta) \in \partial F^*[\beta u, q]\,, \quad F[p(\beta), q] = \lambda$$



General Solution

- Lagrange function for $\upsilon_u(\lambda) := \sup\{\langle u, p\rangle : F[p,q] \leq \lambda\}$ $\big(\lambda_u(\upsilon) := \inf\{F[p,q] : \langle u, p\rangle \geq \upsilon\}\big)$:

$$L(p, \beta^{-1}) = \langle u, p\rangle + \beta^{-1}(\lambda - F[p,q])$$

$$\Big(L(p, \beta) = F[p,q] + \beta(\upsilon - \langle u, p\rangle)\Big)$$

General Solution

- Lagrange function for $\upsilon_u(\lambda) := \sup\{\langle u, p\rangle : F[p,q] \leq \lambda\}$ $(\lambda_u(\upsilon) := \inf\{F[p,q] : \langle u, p\rangle \geq \upsilon\})$:

$$L(p, \beta^{-1}) = \langle u, p\rangle + \beta^{-1}(\lambda - F[p,q])$$
$$\Big(L(p, \beta) = F[p,q] + \beta(\upsilon - \langle u, p\rangle)\Big)$$

- Necessary and sufficient conditions $\partial L \ni 0$:

$$\partial_p L(p, \beta^{-1}) = \{\beta u\} - \partial_p F[p,q] \ni 0$$
$$\partial_{\beta^{-1}} L(p, \beta^{-1}) \ni \lambda - F[p,q] = 0$$

General Solution

- Lagrange function for $\upsilon_u(\lambda) := \sup\{\langle u, p\rangle : F[p,q] \leq \lambda\}$ $\big(\lambda_u(\upsilon) := \inf\{F[p,q] : \langle u, p\rangle \geq \upsilon\}\big)$:

$$L(p, \beta^{-1}) = \langle u, p\rangle + \beta^{-1}(\lambda - F[p,q])$$
$$\Big(L(p, \beta) = F[p,q] + \beta(\upsilon - \langle u, p\rangle)\Big)$$

- Necessary and sufficient conditions $\partial L \ni 0$:

$$\partial_p L(p, \beta^{-1}) = \{\beta u\} - \partial_p F[p,q] \ni 0$$
$$\partial_{\beta^{-1}} L(p, \beta^{-1}) \ni \lambda - F[p,q] = 0$$

- Optimal solutions are subgradients of $F^*[u,q] = \sup\{\langle u, p\rangle - F[p,q]\}$:

$$p(\beta) \in \partial F^*[\beta u]\,, \qquad F[p,q] = \lambda \quad \Big(\langle u, p(\beta)\rangle = \upsilon\Big)$$

Example: Exponential Solution

- For $F[p,q] = D_{KL}[p,q]$:

$$L(p,\beta^{-1}) = \langle u,p\rangle + \beta^{-1}(\lambda - \langle \ln(p/q),p\rangle + \langle 1,p-q\rangle)$$

Example: Exponential Solution

- For $F[p,q] = D_{KL}[p,q]$:

$$L(p,\beta^{-1}) = \langle u,p\rangle + \beta^{-1}(\lambda - \langle \ln(p/q),p\rangle + \langle 1,p-q\rangle)$$

- Necessary and sufficient conditions $\nabla L(p,\beta^{-1}) = 0$:

$$\begin{aligned} \nabla_p L(p,\beta^{-1}) &= u - \beta^{-1}\ln(p/q) = 0 \\ \partial_{\beta^{-1}} L(p,\beta^{-1}) &= \lambda - D_{KL}[p,q] = 0 \end{aligned}$$

Example: Exponential Solution

- For $F[p,q] = D_{KL}[p,q]$:

$$L(p,\beta^{-1}) = \langle u,p\rangle + \beta^{-1}(\lambda - \langle \ln(p/q),p\rangle + \langle 1,p-q\rangle)$$

- Necessary and sufficient conditions $\nabla L(p,\beta^{-1}) = 0$:

$$\begin{aligned} \nabla_p L(p,\beta^{-1}) &= u - \beta^{-1}\ln(p/q) = 0 \\ \partial_{\beta^{-1}} L(p,\beta^{-1}) &= \lambda - D_{KL}[p,q] = 0 \end{aligned}$$

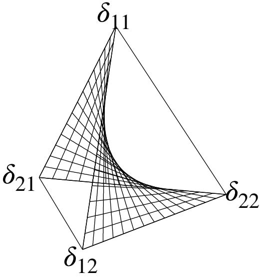
- Optimal solutions are gradients of $D^*_{KL}[u,q] = \ln\langle e^u, q\rangle$:

$$p(\beta) = e^{\beta\, u - \ln Z[\beta u]}\, q\,, \qquad D_{KL}[p(\beta),q] = \lambda$$

Solution to Shannon's OCP

- The solution for $I_w\{x,y\} = D_{KL}[w, q \otimes p] \leq \lambda$:

$$w(\beta) = e^{\beta u - \ln Z[\beta u]}\, q \otimes p$$

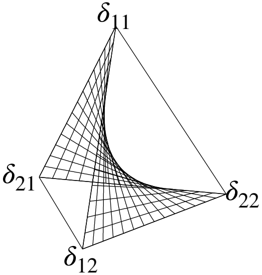


Solution to Shannon's OCP

- The solution for $I_w\{x, y\} = D_{KL}[w, q \otimes p] \leq \lambda$:

$$w(\beta) = e^{\beta u - \ln Z[\beta u]}\, q \otimes p$$

- $w \notin \partial \mathcal{P}(X \otimes Y)$.

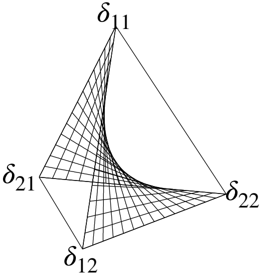


Solution to Shannon's OCP

- The solution for $I_w\{x,y\} = D_{KL}[w, q \otimes p] \leq \lambda$:

$$w(\beta) = e^{\beta u - \ln Z[\beta u]}\, q \otimes p$$

- $w \notin \partial \mathcal{P}(X \otimes Y)$.
- $T : \mathcal{P}(X) \to \mathcal{P}(Y)$ cannot have kernel $\delta_{f(x)}(\cdot)$.



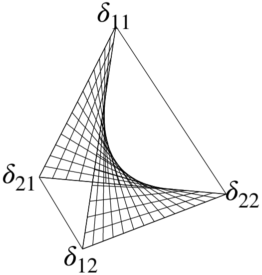
Solution to Shannon's OCP

- The solution for $I_w\{x,y\} = D_{KL}[w, q \otimes p] \leq \lambda$:

$$w(\beta) = e^{\beta u - \ln Z[\beta u]}\, q \otimes p$$

- $w \notin \partial \mathcal{P}(X \otimes Y)$.
- $T : \mathcal{P}(X) \to \mathcal{P}(Y)$ cannot have kernel $\delta_{f(x)}(\cdot)$.
- The dual is strictly convex:

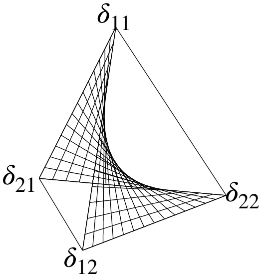
$$D^*_{KL}[u, q \otimes p] = \ln \int e^u\, q \otimes p$$



Solution to Kantorovich's OTP

- $\Gamma[q, p]$ is convex:

$$\pi_X[(1-t)w_1 + tw_2] = (1-t)q + tq = q$$



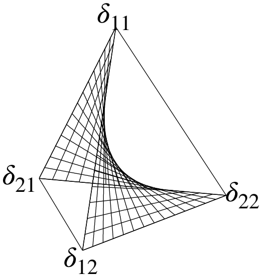
Solution to Kantorovich's OTP

- $\Gamma[q,p]$ is convex:

$$\pi_X[(1-t)w_1 + tw_2] = (1-t)q + tq = q$$

- There exists closed convex functional F:

$$\Gamma[q,p] = \{w : F[w, q \otimes p] \leq 1\}$$



Solution to Kantorovich's OTP

- $\Gamma[q,p]$ is convex:

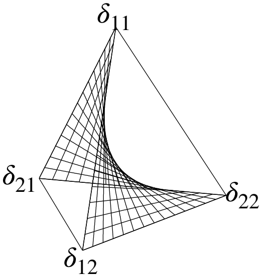
$$\pi_X[(1-t)w_1 + tw_2] = (1-t)q + tq = q$$

- There exists closed convex functional F:

$$\Gamma[q,p] = \{w : F[w, q \otimes p] \leq 1\}$$

- Then the solution to OTP is:

$$w(\beta) \in \partial F^*[-\beta c, q \otimes p]$$



Solution to Kantorovich's OTP

- $\Gamma[q,p]$ is convex:

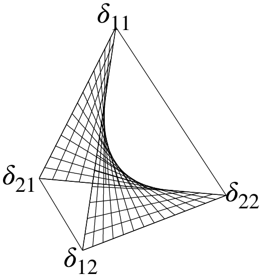
$$\pi_X[(1-t)w_1 + tw_2] = (1-t)q + tq = q$$

- There exists closed convex functional F:

$$\Gamma[q,p] = \{w : F[w, q \otimes p] \leq 1\}$$

- Then the solution to OTP is:

$$w(\beta) \in \partial F^*[-\beta c, q \otimes p]$$



Monge-Amper equation

$$q = p \circ \nabla\varphi |\nabla^2 \varphi|$$

where $\varphi : X \to \mathbb{R} \cup \{\infty\}$ is convex, and $\nabla\varphi : X \to Y$ is such that $p = q \circ (\nabla\varphi)^{-1}$ (McCann, 1995; Villani, 2009).

Strict Inequalities

Theorem (Belavkin, 2013b)

- *Let* $\{w(\beta)\}_u$ *be a family of* $w(\beta) \in \mathcal{P}(X \otimes Y)$

$$\textit{maximizing } \mathbb{E}_w\{u\} \textit{ on sets } \{w : F[w] \leq \lambda\}\,,\ \forall\, \lambda = F[w]$$

Strict Inequalities

Theorem (Belavkin, 2013b)

- *Let* $\{w(\beta)\}_u$ *be a family of* $w(\beta) \in \mathcal{P}(X \otimes Y)$

$$\textit{maximizing } \mathbb{E}_w\{u\} \textit{ on sets } \{w : F[w] \leq \lambda\}, \ \forall \lambda = F[w]$$

- $F : \mathcal{P}(X \otimes Y) \to \mathbb{R} \cup \{\infty\}$ *closed convex and minimized at*

$$q \otimes p \in \partial F^*(0) \subset Int(\mathcal{P}(X \otimes Y))$$

Strict Inequalities

Theorem (Belavkin, 2013b)

- *Let* $\{w(\beta)\}_u$ *be a family of* $w(\beta) \in \mathcal{P}(X \otimes Y)$

$$\text{maximizing } \mathbb{E}_w\{u\} \text{ on sets } \{w : F[w] \leq \lambda\}, \ \forall \lambda = F[w]$$

- $F : \mathcal{P}(X \otimes Y) \to \mathbb{R} \cup \{\infty\}$ *closed convex and minimized at*

$$q \otimes p \in \partial F^*(0) \subset Int(\mathcal{P}(X \otimes Y))$$

- *If* F^* *is strictly convex, then*

Strict Inequalities

Theorem (Belavkin, 2013b)

- *Let* $\{w(\beta)\}_u$ *be a family of* $w(\beta) \in \mathcal{P}(X \otimes Y)$

$$\textit{maximizing } \mathbb{E}_w\{u\} \textit{ on sets } \{w : F[w] \leq \lambda\}, \ \forall \lambda = F[w]$$

- $F : \mathcal{P}(X \otimes Y) \to \mathbb{R} \cup \{\infty\}$ *closed convex and minimized at*

$$q \otimes p \in \partial F^*(0) \subset Int(\mathcal{P}(X \otimes Y))$$

- *If* F^* *is strictly convex, then*
  1. $w(\beta) \in \partial\mathcal{P}(X \otimes Y)$ *iff* $\lambda \geq \sup F$ *(i.e.* $\beta \to \infty$*).*

Strict Inequalities

Theorem (Belavkin, 2013b)

- *Let* $\{w(\beta)\}_u$ *be a family of* $w(\beta) \in \mathcal{P}(X \otimes Y)$

$$\textit{maximizing } \mathbb{E}_w\{u\} \textit{ on sets } \{w : F[w] \leq \lambda\}, \ \forall \lambda = F[w]$$

- $F : \mathcal{P}(X \otimes Y) \to \mathbb{R} \cup \{\infty\}$ *closed convex and minimized at*

$$q \otimes p \in \partial F^*(0) \subset Int(\mathcal{P}(X \otimes Y))$$

- *If* F^* *is strictly convex, then*
  1. $w(\beta) \in \partial \mathcal{P}(X \otimes Y)$ *iff* $\lambda \geq \sup F$ *(i.e.* $\beta \to \infty$*).*
  2. *For any* $\sigma \in \partial \mathcal{P}(X \otimes Y)$ *with* $F[\sigma] = F[w(\beta)] = \lambda$

$$\mathbb{E}_\sigma\{u\} < \mathbb{E}_{w(\beta)}\{u\}$$

Strict Inequalities

Theorem (Belavkin, 2013b)

- *Let* $\{w(\beta)\}_u$ *be a family of* $w(\beta) \in \mathcal{P}(X \otimes Y)$

$$\text{maximizing } \mathbb{E}_w\{u\} \text{ on sets } \{w : F[w] \leq \lambda\}, \ \forall \lambda = F[w]$$

- $F : \mathcal{P}(X \otimes Y) \to \mathbb{R} \cup \{\infty\}$ *closed convex and minimized at*

$$q \otimes p \in \partial F^*(0) \subset Int(\mathcal{P}(X \otimes Y))$$

- *If* F^* *is strictly convex, then*
  1. $w(\beta) \in \partial\mathcal{P}(X \otimes Y)$ *iff* $\lambda \geq \sup F$ *(i.e.* $\beta \to \infty$*).*
  2. *For any* $\sigma \in \partial\mathcal{P}(X \otimes Y)$ *with* $F[\sigma] = F[w(\beta)] = \lambda$

$$\mathbb{E}_\sigma\{u\} < \mathbb{E}_{w(\beta)}\{u\}$$

  3. *For any* $\sigma \in \partial\mathcal{P}(X \otimes Y)$ *with* $\mathbb{E}_\sigma\{u\} = \mathbb{E}_{w(\beta)}\{u\} = \upsilon$

$$F[\sigma] > F[w(\beta)]$$

Strict Bounds for Monge OTP

Corollary

- *Let* $w_f \in \Gamma[q, p]$ *be a solution to Monge OTP* $K_c[p, q]$.

Strict Bounds for Monge OTP

Corollary

- *Let* $w_f \in \Gamma[q,p]$ *be a solution to Monge OTP* $K_c[p,q]$*.*
- *Let* $w(\beta)$ *is a solution to Shannon's OCP* $R_c[q](\lambda)$*.*

Strict Bounds for Monge OTP

Corollary

- *Let* $w_f \in \Gamma[q,p]$ *be a solution to Monge OTP* $K_c[p,q]$*.*
- *Let* $w(\beta)$ *is a solution to Shannon's OCP* $R_c[q](\lambda)$*.*
- *If* w_f *and* $w(\beta)$ *have equal* $I_{w_f}\{x,y\} = I_{w(\beta)}\{x,y\} = \lambda < \sup I_w\{x,y\}$*, then*

$$K_c[p,q](\lambda) > R_c[q](\lambda) > 0$$

Strict Bounds for Monge OTP

Corollary

- *Let* $w_f \in \Gamma[q,p]$ *be a solution to Monge OTP* $K_c[p,q]$.
- *Let* $w(\beta)$ *is a solution to Shannon's OCP* $R_c[q](\lambda)$.
- *If* w_f *and* $w(\beta)$ *have equal* $I_{w_f}\{x,y\} = I_{w(\beta)}\{x,y\} = \lambda < \sup I_w\{x,y\}$, *then*

$$K_c[p,q](\lambda) > R_c[q](\lambda) > 0$$

- *If* w_f *and* $w(\beta)$ *achieve equal values* $K_c[p,q] = R_c[q](\lambda) = \upsilon > 0$, *then*

$$K_c^{-1}[p,q](\upsilon) > R_c^{-1}[q](\upsilon)$$

Optimal transportation problems (OTPs)

Information and entropy

Optimal channel problem (OCP)

Dual formulation of OTP

Geometry of information divergence and optimization

Belavkin, R. V. (2013a). Law of cosines and Shannon-Pythagorean theorem for quantum information. In F. Nielsen & F. Barbaresco (Eds.), *Geometric science of information* (Vol. 8085, pp. 369–376). Heidelberg: Springer.

Belavkin, R. V. (2013b). Optimal measures and Markov transition kernels. *Journal of Global Optimization*, *55*, 387–416.

Dobrushin, R. L. (1970). Prescribing a system of random variables by conditional distributions. *Theory Probab. Appl.*, *15*(3), 458âĂŞ-486.

Kantorovich, L. V. (1939). *Mathematical methods in the organization and planning of production* (Tech. Rep.). Leningrad Univ. (English translation: Management Science, 6, 4 (1960), 363âĂŞ422)

Kantorovich, L. V. (1942). On translocation of masses. *USSR AS Doklady*, *37*(7–8), 227–229. ((in Russian). English translation: J. Math. Sci., 133, 4 (2006), 1381âĂŞ1382)

Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, *22*(1), 79–86.

McCann, R. J. (1995). Existence and uniqueness of monotone measure-preserving maps. *Duke Math. J.*, *80*(2), 309âĂŞ-323.

Monge, G. (1781). *Mémoire sur la théorie des déblais et de remblais*. Paris: Histoire de l'AcadÂťemie Royale des Sciences avec les Mémoires de Mathématique & de Physique.

Shannon, C. E. (1948, July and October). A mathematical theory of communication. *Bell System Technical Journal*, *27*, 379–423 and 623–656.

Stratonovich, R. L. (1965). On value of information. *Izvestiya of USSR Academy of Sciences, Technical Cybernetics*, *5*, 3–12. (In Russian)

Vasershtein, L. N. (1969). Markov processes over denumerable products of spaces describing large system of automata. *Problems Inform. Transmission*, *5*(3), 47âĂŞ-52.

Villani, C. (2009). *Optimal transport. old and new* (Vol. 338). Springer-Verlag.