Kantorovich optimal transport problem and Shannon's optimal channel problem

Roman V. Belavkin

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

13 June 2016

In honor of
Shun-ichi Amari

Optimal transportation problems (OTPs)

Information and entropy

Optimal channel problem (OCP)

Geometry of information divergence and optimization

Dynamical OTP: Optimization of evolution

Optimal transportation problems (OTPs)

Information and entropy

Optimal channel problem (OCP)

Geometry of information divergence and optimization

Dynamical OTP: Optimization of evolution

Kantorovich's OTP

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

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

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

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

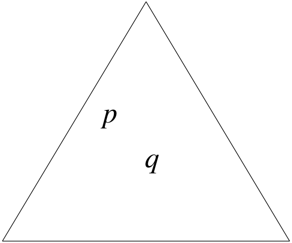
Kantorovich's OTP

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

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

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

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



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

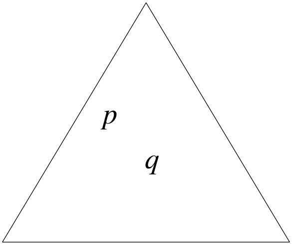
Kantorovich's OTP

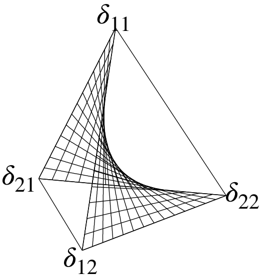
Optimal Transportation Problem (Kantorovich, 1939, 1942; Vasershtein, 1969; Dobrushin, 1970)

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

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

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



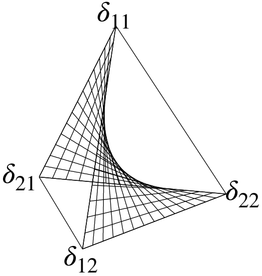


$\mathcal{P}(X) \leftrightarrow \mathcal{P}(X \otimes Y)$

Optimal Transport Plan

Optimal Transportation Problem

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



Optimal Transport Plan

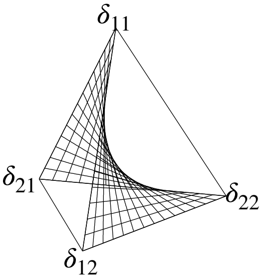
Optimal Transportation Problem

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

Optimal Transport Plan

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

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



Optimal Transport Plan

Optimal Transportation Problem

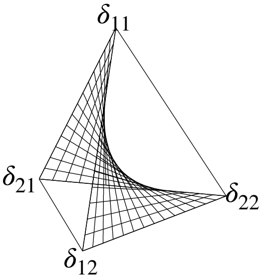
$$K_c(q,p) := \inf \left\{ \int_{X\times Y} c(x,y)\, dw \,:\, \pi_X w = q,\ \pi_Y w = p \right\}$$

Optimal Transport Plan

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

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

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



Optimal Transport Plan

Optimal Transportation Problem

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

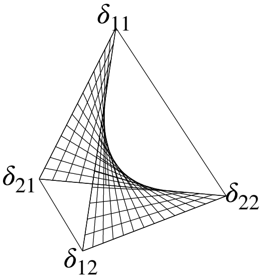
Optimal Transport Plan

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

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

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

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



Monge OTP

Optimal Transportation Problem (Monge, 1781)

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

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)

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

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

- $p(\cdot \mid x)$ 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)

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

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

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

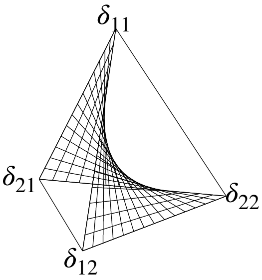
Optimal Transport

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

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

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

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



Shannon's Information and Entropy

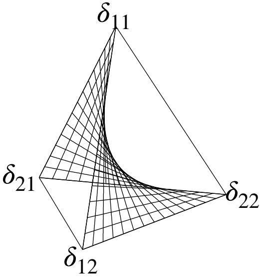
KL-divergence (Kullback & Leibler, 1951)

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

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

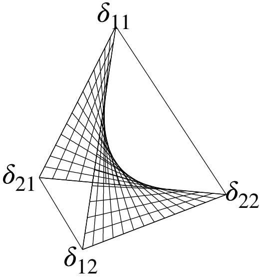
KL-divergence (Kullback & Leibler, 1951)

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

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

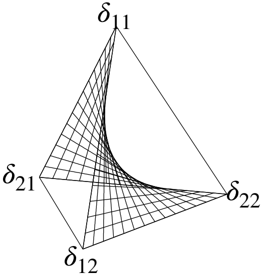
KL-divergence (Kullback & Leibler, 1951)

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

Shannon's information (Shannon, 1948)

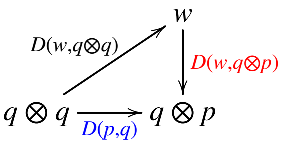
For $w \in \Gamma(q,p) \subset \mathcal{P}(X \otimes Y)$:

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



Entropy $H(p) = -\int \ln p\, dp$

$$H(p) := \sup_{w:\pi_Y w = p} I_w\{x,y\} = I_w\{y,y\}$$

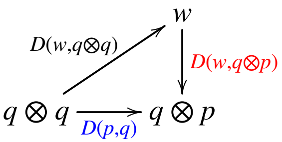

w
D(w,q⊗q)
D(w,q⊗p)
q ⊗ q
D(p,q)
q ⊗ p

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)

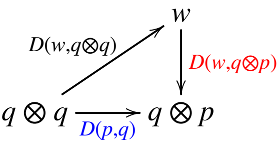


Theorem (Shannon-Pythagorean)

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

$$D_{KL}(w, q \otimes q) = {\color{red} D_{KL}(w, q \otimes p)} + {\color{blue} D_{KL}(p, q)}$$

(Belavkin, 2013a)



Proof.

$$D(w, q\otimes q) = \underbrace{D(w, q \otimes p)}_{\color{red} I_w\{x,y\}} + \underbrace{D(q \otimes p, q \otimes q)}_{\color{blue} D(p,q)} - \underbrace{\langle \ln q \otimes p - \ln q \otimes q, q \otimes p - w \rangle}_{0}$$

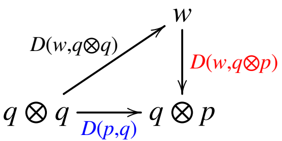
$$\ln q \otimes p - \ln q \otimes q = 1_X \otimes (\ln p - \ln q) \qquad \square$$

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.

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

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

Cross-Information (Belavkin, 2013a)

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

Shannon's OCP

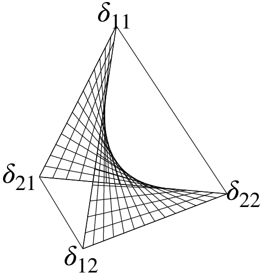
Optimal Channel Problem (Shannon, 1948)

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

Shannon's OCP

Optimal Channel Problem (Shannon, 1948)

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



Shannon's OCP

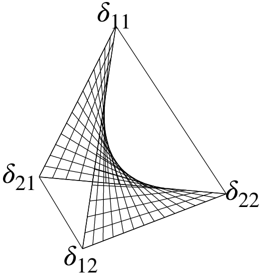
Optimal Channel Problem (Shannon, 1948)

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

Exponential family solutions

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

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



Shannon's OCP

Optimal Channel Problem (Shannon, 1948)

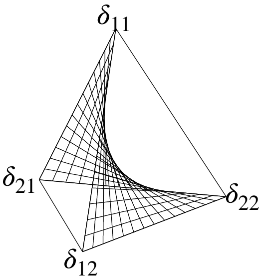
$$S_c(q,\lambda) := \inf\left\{ \int_{X\times Y} c(x,y)\,dw \,:\, \pi_X w = q,\ I_w\{x,y\} \leq \lambda \right\}$$

Exponential family solutions

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

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

- 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)

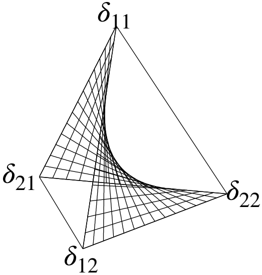
$$S_c(q, \lambda) := \inf \left\{ \int_{X \times Y} c(x, y)\, dw \;:\; \pi_X w = q,\; I_w\{x, y\} \leq \lambda \right\}$$

Exponential family solutions

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

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

- 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)

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

Relation to Kantorovich OTP

Optimal Channel Problem

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

Relation to Kantorovich OTP

Optimal Channel Problem

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

Optimal Transportation Problem

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

Relation to Kantorovich OTP

Optimal Channel Problem

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

Optimal Transportation Problem

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

- 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

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

Optimal Transportation Problem

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

- 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

$$S_c(q, \lambda) \leq K_c(q,p)$$

Inverse Optimal Values

Inverse of the OCP Value

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

Inverse Optimal Values

Inverse of the OCP Value

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

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

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

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

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

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 $\upsilon = K_c(q, p)$, then

$$S_c^{-1}(q, \upsilon) \leq K_c^{-1}(q, p, \upsilon)$$

Optimal transportation problems (OTPs)

Information and entropy

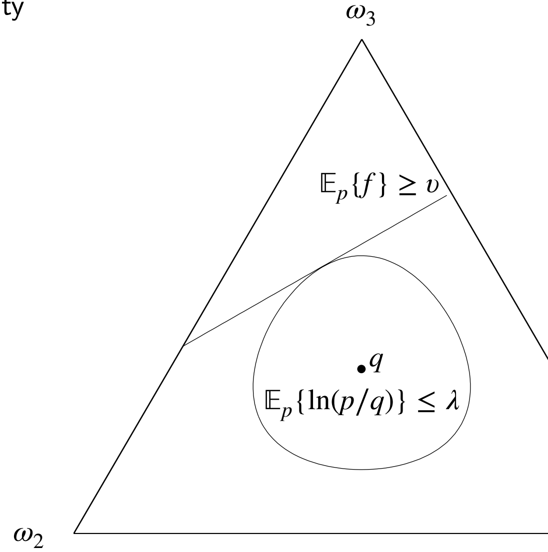
Optimal channel problem (OCP)

Geometry of information divergence and optimization

Dynamical OTP: Optimization of evolution

Problems on Conditional Extremum

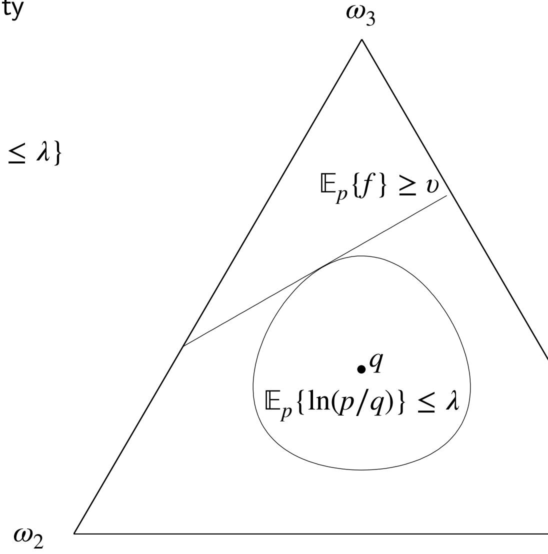
- $\mathbb{E}_p\{u\} = \langle u, p \rangle$ expected utility



Problems on Conditional Extremum

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

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



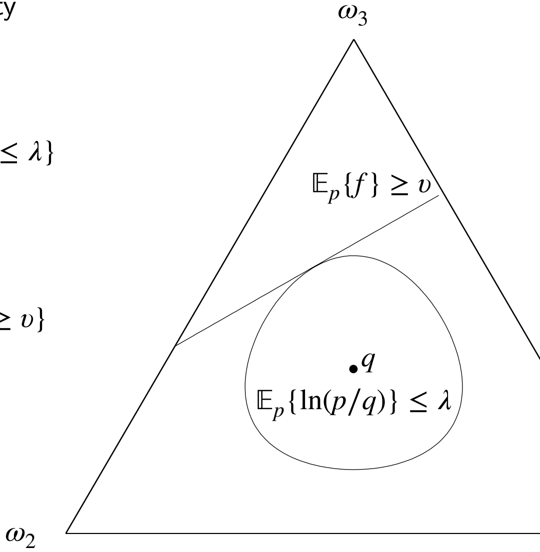
Problems on Conditional Extremum

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

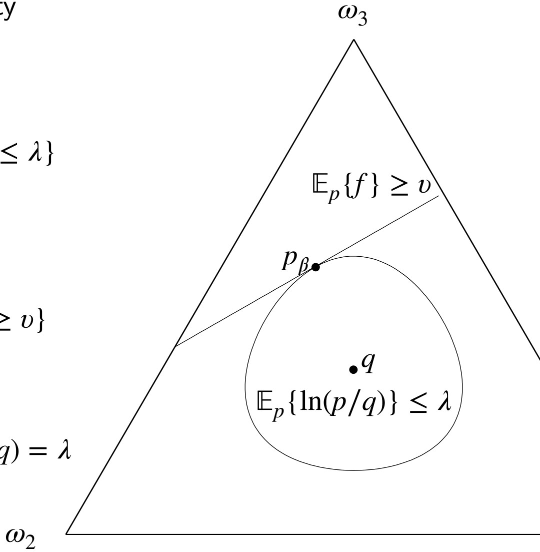
$$\upsilon_u(\lambda) := \sup\{\langle u, p\rangle : F(p, q) \leq \lambda\}$$

- $\lambda_u(\upsilon) = \upsilon_u^{-1}(\upsilon)$ 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
- $\upsilon_u(\lambda) = -S_{-u}(q,\lambda)$ utility of information λ:

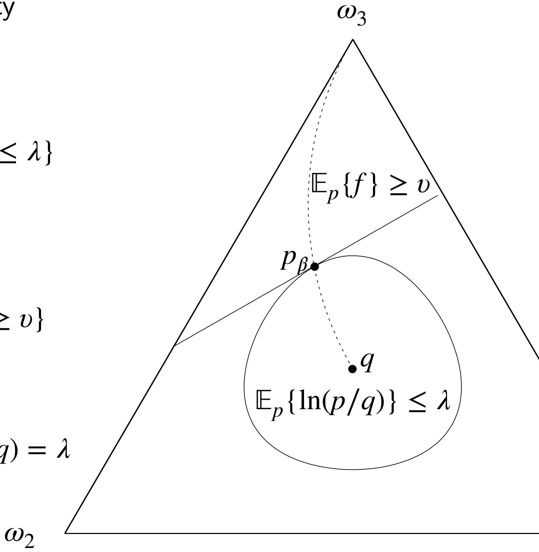
$$\upsilon_u(\lambda) := \sup\{\langle u, p\rangle : F(p,q) \leq \lambda\}$$

- $\lambda_u(\upsilon) = \upsilon_u^{-1}(\upsilon)$ 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
- $\upsilon_u(\lambda) = -S_{-u}(q, \lambda)$ utility of information λ:

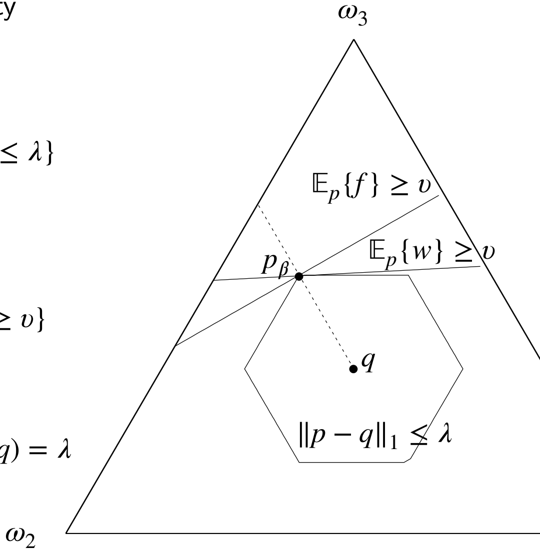
$$\upsilon_u(\lambda) := \sup\{\langle u, p \rangle : F(p, q) \leq \lambda\}$$

- $\lambda_u(\upsilon) = \upsilon_u^{-1}(\upsilon)$ 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 $\upsilon_u(\lambda) := \sup\{\langle u, p\rangle : F(p,q) \leq \lambda\}$ $\big(\lambda_u(\upsilon) := \inf\{F(p,q) : \langle u, p\rangle \geq \upsilon\}\big)$:

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

General Solution

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

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

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

$$\begin{aligned} \partial_p L(p, \beta^{-1}) &= \{\beta u\} - \partial_p F(p,q) \ni 0 \\ \partial_{\beta^{-1}} L(p, \beta^{-1}) &= \lambda - F(p,q) = 0 \end{aligned}$$

General Solution

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

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

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

$$\begin{aligned} \partial_p L(p, \beta^{-1}) &= \{\beta u\} - \partial_p F(p,q) \ni 0 \\ \partial_{\beta^{-1}} L(p, \beta^{-1}) &= \lambda - F(p,q) = 0 \end{aligned}$$

- 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 $D_{KL}(p,q) = \langle \ln(p/q), p\rangle - \langle 1, p-q\rangle$:

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

Example: Exponential Solution

- For $D_{KL}(p,q) = \langle \ln(p/q), p\rangle - \langle 1, p-q\rangle$:

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

- 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 $D_{KL}(p,q) = \langle \ln(p/q), p\rangle - \langle 1, p-q\rangle$:

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

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

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

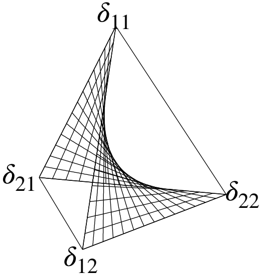
- Optimal solutions are gradients of $D^*_{KL}(u,q) = \ln\langle e^u, q\rangle$:

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

Solution to Shannon's OCP

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

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

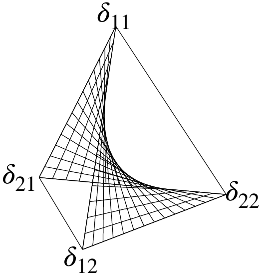


Solution to Shannon's OCP

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

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

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

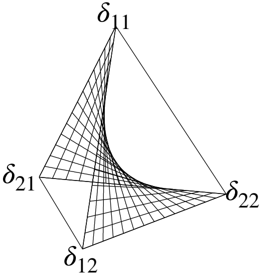


Solution to Shannon's OCP

- 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 $\delta_{f(x)}(\cdot)$.



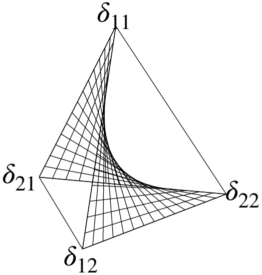
Solution to Shannon's OCP

- 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:

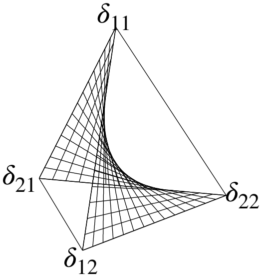
$$D^*_{KL}(u, q \otimes p) = \ln \int e^u\, q \otimes p$$



Solution to Kantorovich's OTP

- $\Gamma(q, p)$ is convex:

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



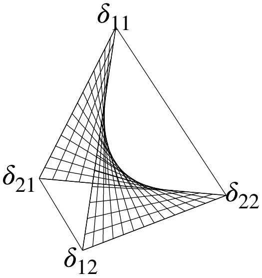
Solution to Kantorovich's OTP

- $\Gamma(q, p)$ 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:

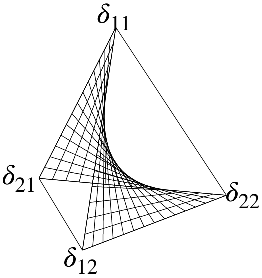
$$\pi_X((1-t)w_1 + tw_2) = (1-t)q + tq = q$$

- 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:

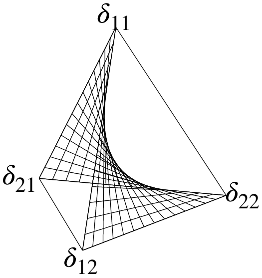
$$\pi_X((1-t)w_1 + tw_2) = (1-t)q + tq = q$$

- 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* $w(\beta) \in \mathcal{P}(X \otimes Y)$

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

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

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.* $\beta \to \infty$*).*

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* $v \in \partial\mathcal{P}(X \otimes Y)$ *with* $F(v) = F(w(\beta)) = \lambda$

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

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* $v \in \partial\mathcal{P}(X \otimes Y)$ *with* $F(v) = F(w(\beta)) = \lambda$

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

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

$$F(v) > F(w(\beta))$$

Strict Bounds for Monge OTP

Corollary

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

Strict Bounds for Monge OTP

Corollary

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

Strict Bounds for Monge OTP

Corollary

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

$$K_c(q,p) > S_c(q,\lambda) > 0$$

Strict Bounds for Monge OTP

Corollary

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

$$K_c(q, p) > S_c(q, \lambda) > 0$$

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

$$K_c^{-1}(q, p, \upsilon) > S_c^{-1}(q, \upsilon)$$

Optimal Transport and the Expected Utility Principle

- Let $(X \times Y, \lesssim)$ be a set with preference relation (total pre-order).

Optimal Transport and the Expected Utility Principle

- Let $(X \times Y, \lesssim)$ be a set with preference relation (total pre-order).
- The Neumann and Morgenstern (1944) EU principle states that for any $v, w \in \mathcal{P}(X \otimes Y)$:

$$v \lesssim w \quad \iff \quad \mathbb{E}_v\{u\} \leq \mathbb{E}_w\{u\} \quad \exists\, u : X \times Y \to \mathbb{R}$$

Optimal Transport and the Expected Utility Principle

- Let $(X \times Y, \lesssim)$ be a set with preference relation (total pre-order).
- The Neumann and Morgenstern (1944) EU principle states that for any $v, w \in \mathcal{P}(X \otimes Y)$:

$$v \lesssim w \quad \iff \quad \mathbb{E}_v\{u\} \leq \mathbb{E}_w\{u\} \quad \exists\, u : X \times Y \to \mathbb{R}$$

- Based on linear and Archimedian axioms:

$$v \lesssim w \quad \iff \quad \lambda v \lesssim \lambda w\,, \qquad \forall\, \lambda > 0 \tag{1}$$

$$v \lesssim w \quad \iff \quad v + r \lesssim w + r\,, \qquad \forall\, r \in L \tag{2}$$

$$nv \lesssim w\,, \qquad \forall\, n \in \mathbb{N} \quad \Rightarrow \quad v \lesssim 0 \tag{3}$$

Optimal Transport and the Expected Utility Principle

- Let $(X \times Y, \lesssim)$ be a set with preference relation (total pre-order).
- The Neumann and Morgenstern (1944) EU principle states that for any $v, w \in \mathcal{P}(X \otimes Y)$:

$$v \lesssim w \quad \iff \quad \mathbb{E}_v\{u\} \leq \mathbb{E}_w\{u\} \quad \exists\, u : X \times Y \to \mathbb{R}$$

- Based on linear and Archimedian axioms:

$$v \lesssim w \quad \iff \quad \lambda v \lesssim \lambda w\,, \qquad \forall\, \lambda > 0 \tag{1}$$

$$v \lesssim w \quad \iff \quad v + r \lesssim w + r\,, \qquad \forall\, r \in L \tag{2}$$

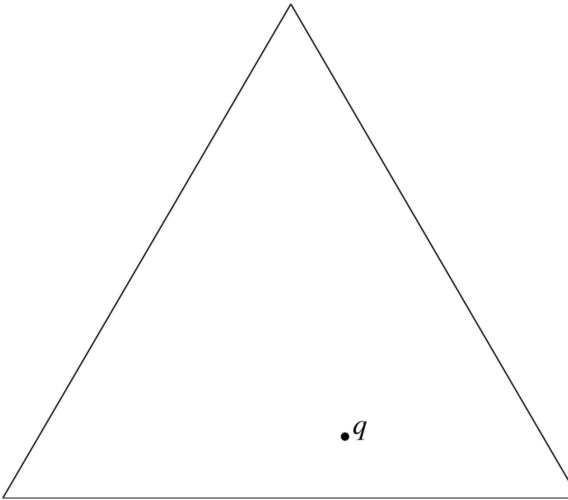
$$nv \lesssim w\,, \qquad \forall\, n \in \mathbb{N} \quad \Rightarrow \quad v \lesssim 0 \tag{3}$$

Theorem

Pre-ordered linear space $(L, \lesssim)$ satisfies (1), (2) and (3) if and only if $(L, \lesssim)$ has a utility representation by a closed linear functional $u : L \to \mathbb{R}$.

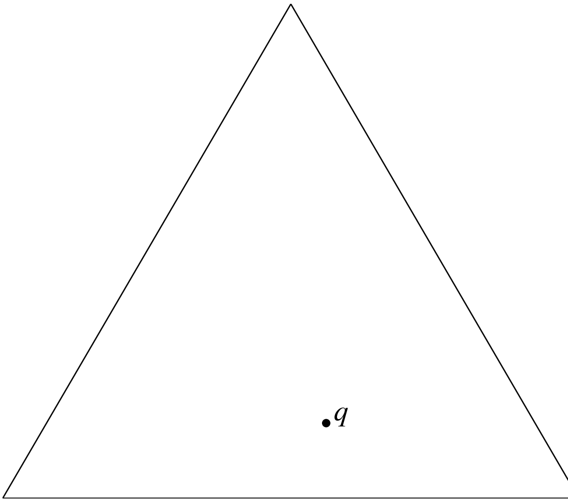
Dynamical Problems

- $q(t) \mapsto Tq(t) = q(t+1)$



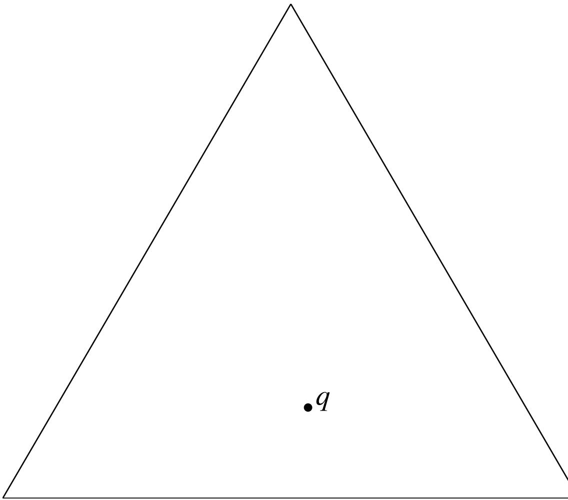
Dynamical Problems

- $q(t) \mapsto Tq(t) = q(t+1)$



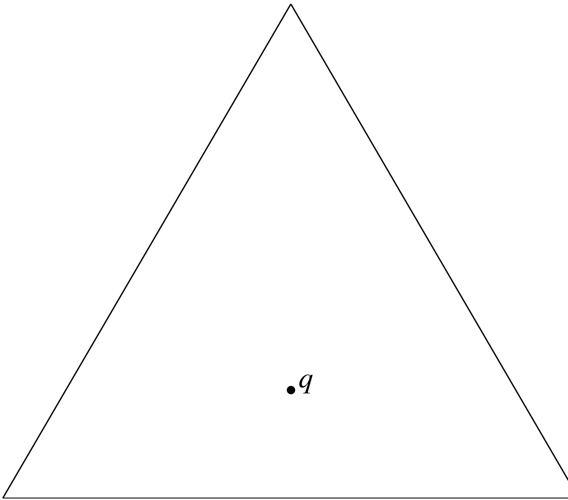
Dynamical Problems

- $q(t) \mapsto Tq(t) = q(t+1)$



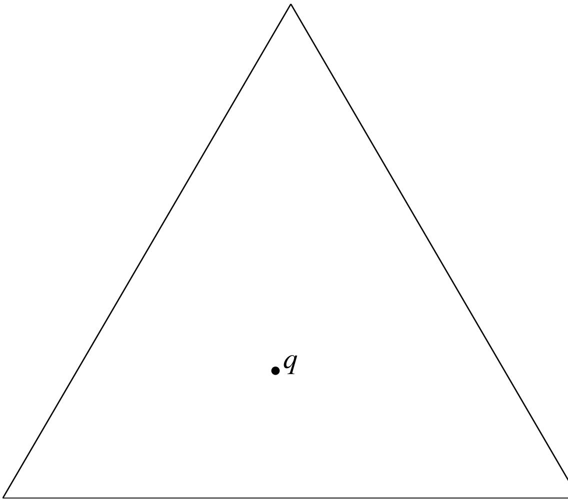
Dynamical Problems

- $q(t) \mapsto Tq(t) = q(t+1)$



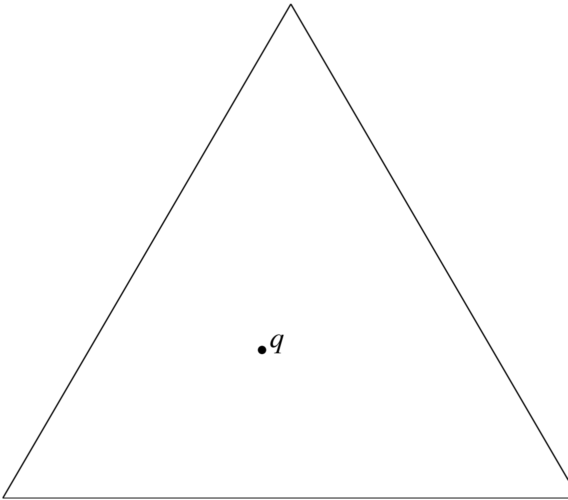
Dynamical Problems

- $q(t) \mapsto Tq(t) = q(t+1)$



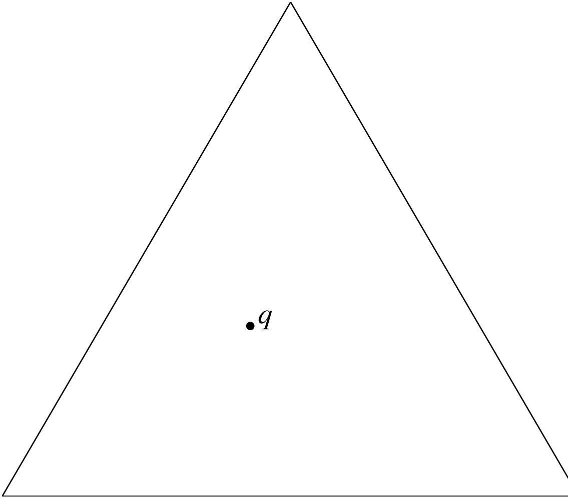
Dynamical Problems

- $q(t) \mapsto Tq(t) = q(t+1)$



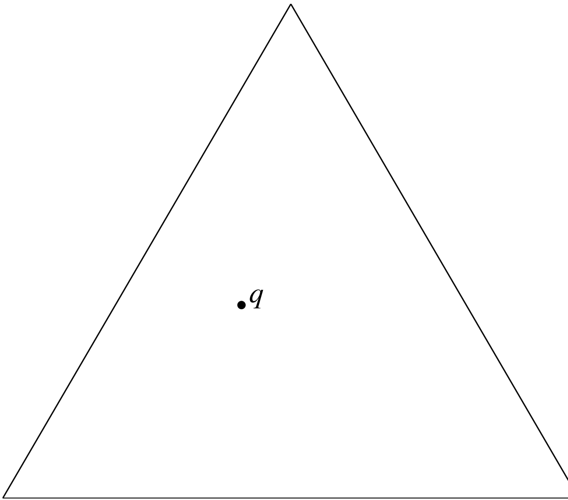
Dynamical Problems

- $q(t) \mapsto Tq(t) = q(t+1)$



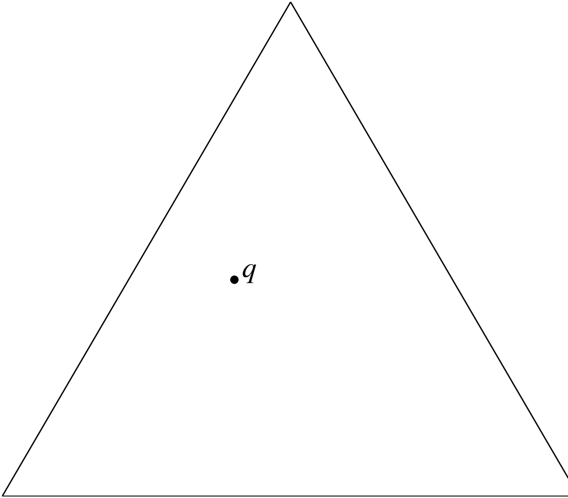
Dynamical Problems

- $q(t) \mapsto Tq(t) = q(t+1)$



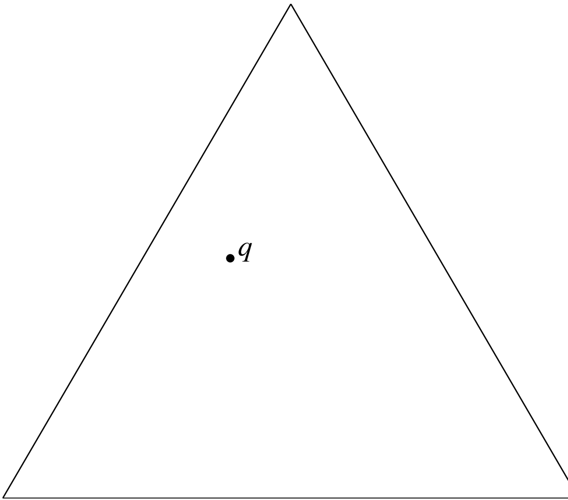
Dynamical Problems

- $q(t) \mapsto Tq(t) = q(t+1)$



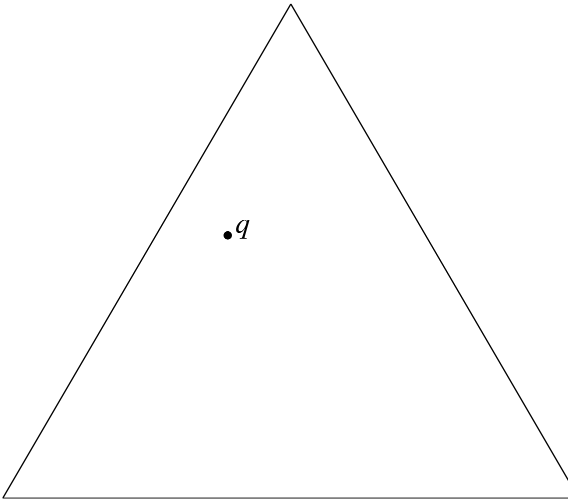
Dynamical Problems

- $q(t) \mapsto Tq(t) = q(t+1)$



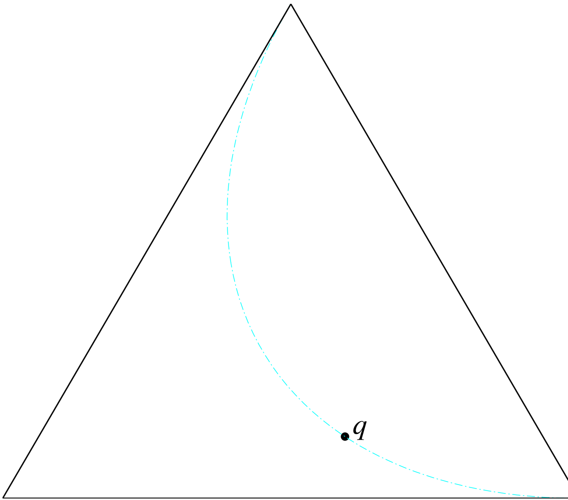
Dynamical Problems

- $q(t) \mapsto Tq(t) = q(t+1)$
- $q(t) \mapsto T^s q(t) = q(t+s)$



Dynamical Problems

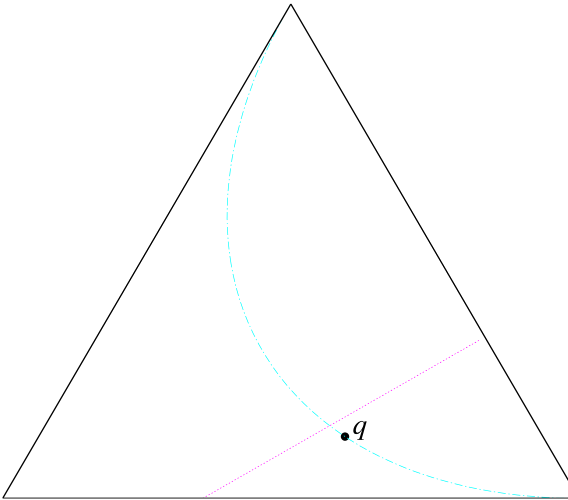
- $q(t) \mapsto Tq(t) = q(t+1)$
- $q(t) \mapsto T^s q(t) = q(t+s)$
- $q(t) \to \delta_\top$



Dynamical Problems

- $q(t) \mapsto Tq(t) = q(t+1)$
- $q(t) \mapsto T^s q(t) = q(t+s)$
- $q(t) \to \delta_\top$
- Expected utility:

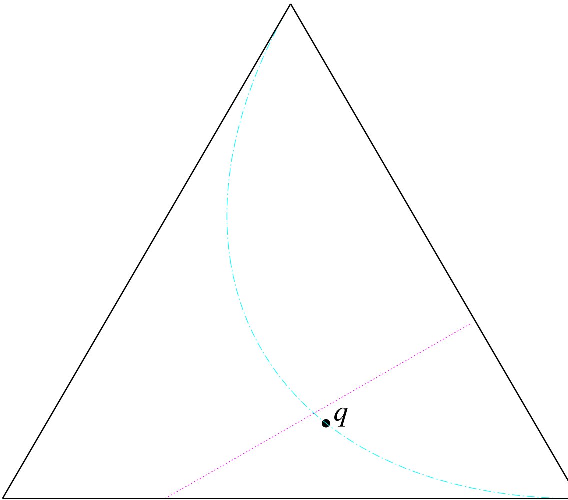
$$\mathbb{E}_{q(t)}\{u\} = \int_X u \, dq(t)$$



Dynamical Problems

- $q(t) \mapsto Tq(t) = q(t+1)$
- $q(t) \mapsto T^s q(t) = q(t+s)$
- $q(t) \to \delta_{\top}$
- Expected utility:

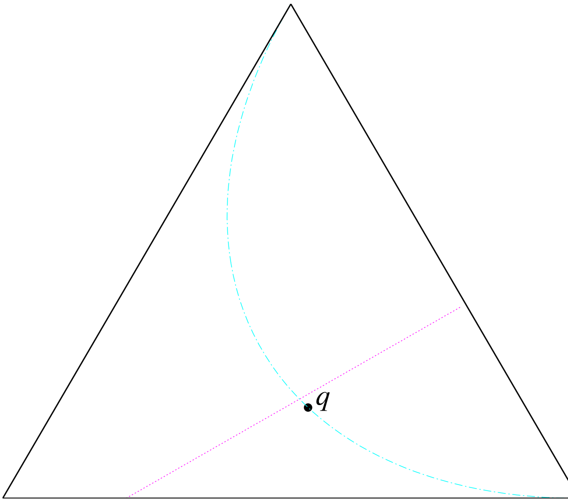
$$\mathbb{E}_{q(t)}\{u\} = \int_X u \, dq(t)$$



Dynamical Problems

- $q(t) \mapsto Tq(t) = q(t+1)$
- $q(t) \mapsto T^s q(t) = q(t+s)$
- $q(t) \to \delta_{\top}$
- Expected utility:

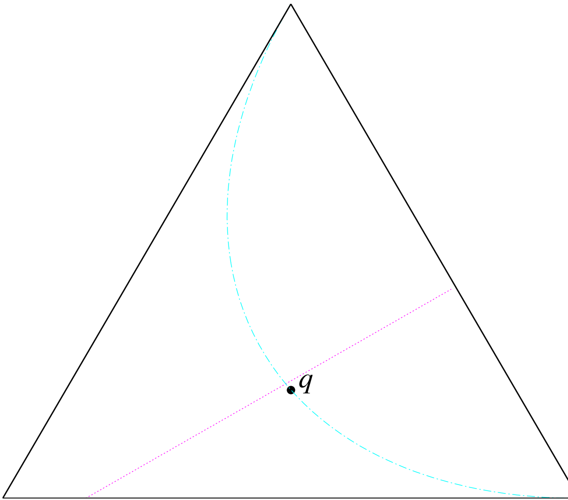
$$\mathbb{E}_{q(t)}\{u\} = \int_X u \, dq(t)$$



Dynamical Problems

- $q(t) \mapsto Tq(t) = q(t+1)$
- $q(t) \mapsto T^s q(t) = q(t+s)$
- $q(t) \to \delta_\top$
- Expected utility:

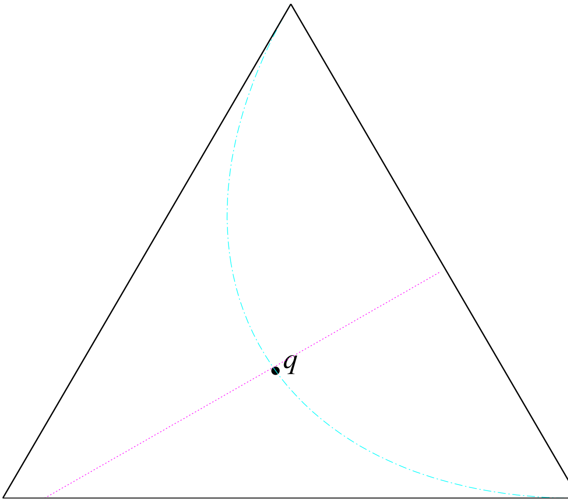
$$\mathbb{E}_{q(t)}\{u\} = \int_X u \, dq(t)$$



Dynamical Problems

- $q(t) \mapsto Tq(t) = q(t+1)$
- $q(t) \mapsto T^s q(t) = q(t+s)$
- $q(t) \to \delta_\top$
- Expected utility:

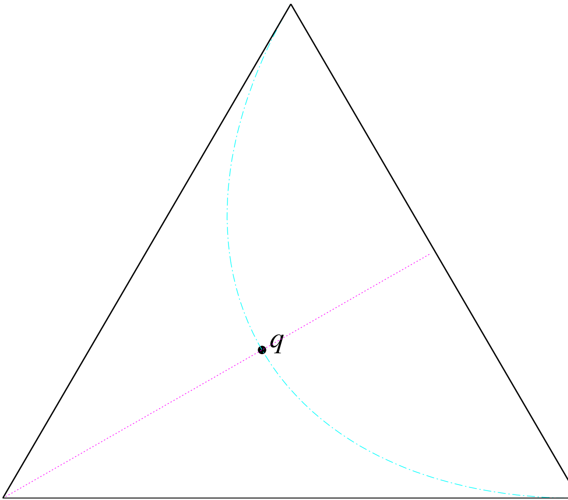
$$\mathbb{E}_{q(t)}\{u\} = \int_X u \, dq(t)$$



Dynamical Problems

- $q(t) \mapsto Tq(t) = q(t+1)$
- $q(t) \mapsto T^s q(t) = q(t+s)$
- $q(t) \to \delta_{\top}$
- Expected utility:

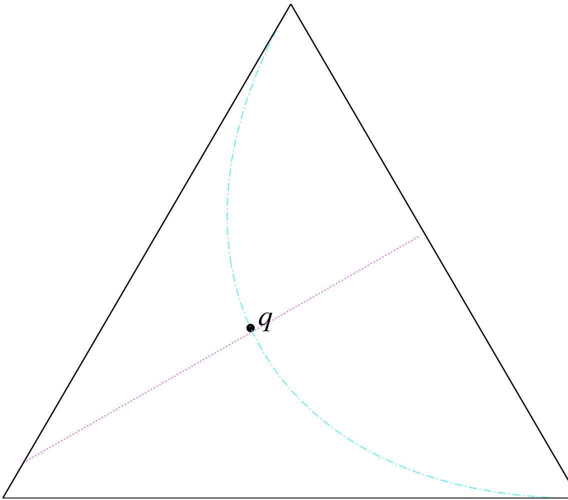
$$\mathbb{E}_{q(t)}\{u\} = \int_X u\, dq(t)$$



Dynamical Problems

- $q(t) \mapsto Tq(t) = q(t+1)$
- $q(t) \mapsto T^s q(t) = q(t+s)$
- $q(t) \to \delta_{\top}$
- Expected utility:

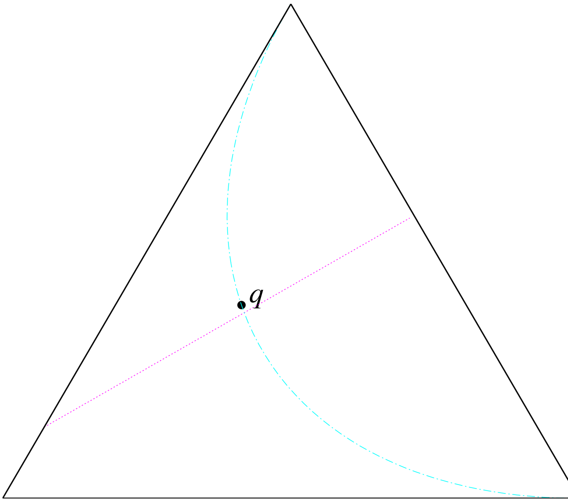
$$\mathbb{E}_{q(t)}\{u\} = \int_X u \, dq(t)$$



Dynamical Problems

- $q(t) \mapsto Tq(t) = q(t+1)$
- $q(t) \mapsto T^s q(t) = q(t+s)$
- $q(t) \to \delta_{\top}$
- Expected utility:

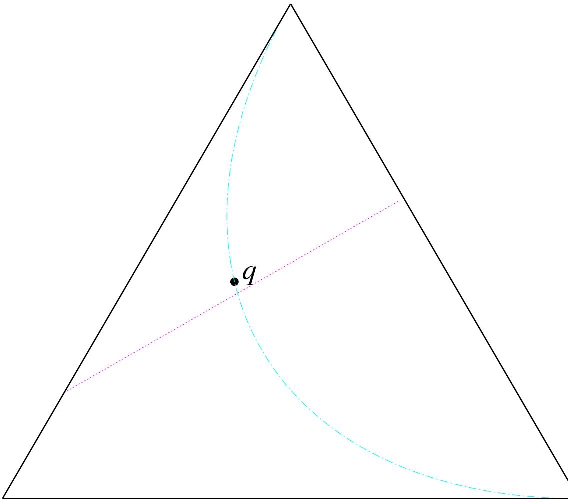
$$\mathbb{E}_{q(t)}\{u\} = \int_X u \, dq(t)$$



Dynamical Problems

- $q(t) \mapsto Tq(t) = q(t+1)$
- $q(t) \mapsto T^s q(t) = q(t+s)$
- $q(t) \to \delta_{\top}$
- Expected utility:

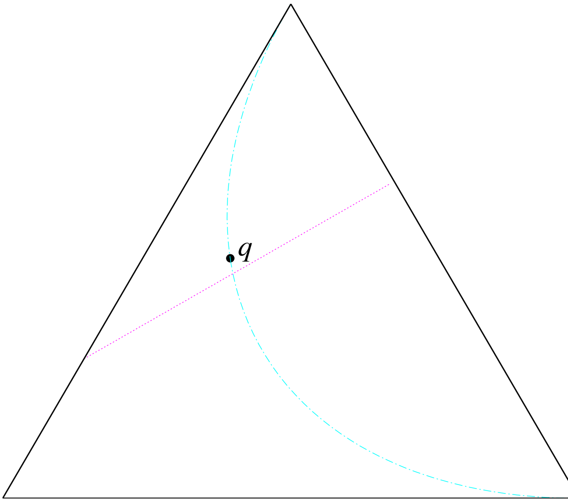
$$\mathbb{E}_{q(t)}\{u\} = \int_X u \, dq(t)$$



Dynamical Problems

- $q(t) \mapsto Tq(t) = q(t+1)$
- $q(t) \mapsto T^s q(t) = q(t+s)$
- $q(t) \to \delta_\top$
- Expected utility:

$$\mathbb{E}_{q(t)}\{u\} = \int_X u\, dq(t)$$

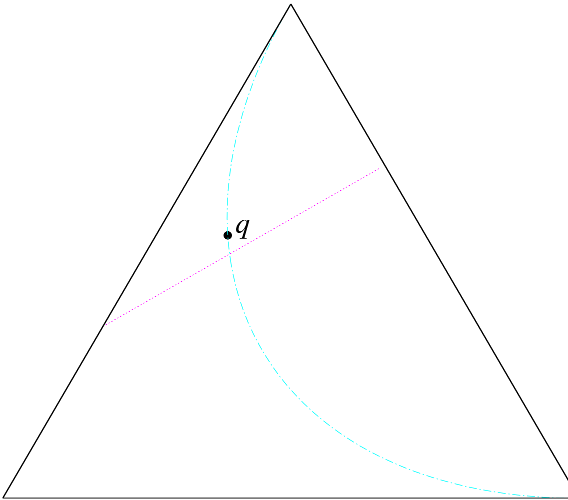


Dynamical Problems

- $q(t) \mapsto Tq(t) = q(t+1)$
- $q(t) \mapsto T^s q(t) = q(t+s)$
- $q(t) \to \delta_\top$
- Expected utility:

$$\mathbb{E}_{q(t)}\{u\} = \int_X u\, dq(t)$$

- Why not directly $q \mapsto Tq = \delta_\top$?



Example: Search in Hamming space $\{1, \dots, \alpha\}^l$

- Find $\mathsf{T} \in \{1, \dots, \alpha\}^l$.

Example: Search in Hamming space $\{1,\dots,\alpha\}^l$

- Find $\top \in \{1,\dots,\alpha\}^l$.
- Hamming metric $d_H(x,y) = \|y - x\|_H = l - \sum_{i=1}^{l} \delta_{x_i y_i}$

Example: Search in Hamming space $\{1, \dots, \alpha\}^l$

- Find $\top \in \{1, \dots, \alpha\}^l$.
- Hamming metric $d_H(x, y) = \|y - x\|_H = l - \sum_{i=1}^{l} \delta_{x_i y_i}$
- $|\{1, \dots, \alpha\}^l| = \alpha^l$

Example: Search in Hamming space $\{1, \dots, \alpha\}^l$

- Find $\top \in \{1, \dots, \alpha\}^l$.
- Hamming metric $d_H(x, y) = \|y - x\|_H = l - \sum_{i=1}^{l} \delta_{x_i y_i}$
- $|\{1, \dots, \alpha\}^l| = \alpha^l$
- To find $\top$ in one step we need $l \log_2 \alpha$ bits.

Example: Search in Hamming space $\{1, \dots, \alpha\}^l$

- Find $\top \in \{1, \dots, \alpha\}^l$.
- Hamming metric $d_H(x, y) = \|y - x\|_H = l - \sum_{i=1}^{l} \delta_{x_i y_i}$
- $|\{1, \dots, \alpha\}^l| = \alpha^l$
- To find $\top$ in one step we need $l \log_2 \alpha$ bits.
- $d_H(\top, x)$ communicates no more than $\log_2(l + 1)$ bits.

Example: Search in Hamming space $\{1, \dots, \alpha\}^l$

- Find $\top \in \{1, \dots, \alpha\}^l$.
- Hamming metric $d_H(x, y) = \|y - x\|_H = l - \sum_{i=1}^{l} \delta_{x_i y_i}$
- $|\{1, \dots, \alpha\}^l| = \alpha^l$
- To find $\top$ in one step we need $l \log_2 \alpha$ bits.
- $d_H(\top, x)$ communicates no more than $\log_2(l + 1)$ bits.
- For $l \geq 2$

$$\log_2(l + 1) < l \log_2 \alpha$$

Optimal Control of Mutation Rate

- Optimal transition kernels are

$$P_\beta(y \mid x) = \frac{e^{-\beta \|y-x\|_H}}{[1+(\alpha-1)e^{-\beta}]^l} = \prod_{i=1}^{l} \frac{e^{-\beta(1-\delta_{x_i y_i})}}{1+(\alpha-1)e^{-\beta}}$$

Optimal Control of Mutation Rate

- Optimal transition kernels are

$$P_\beta(y \mid x) = \frac{e^{-\beta\,\|y-x\|_H}}{[1+(\alpha-1)e^{-\beta}]^l} = \prod_{i=1}^{l} \frac{e^{-\beta\,(1-\delta_{x_i y_i})}}{1+(\alpha-1)e^{-\beta}}$$

- $\beta = \ln\left(\mu^{-1}-1\right) + \ln(\alpha-1)$, where $\mu = \upsilon/l$ is the mutation rate, defined by the constraint $\mathbb{E}\{\|y-x\|_H\} \leq \upsilon$.

Optimal Control of Mutation Rate

- Optimal transition kernels are

$$P_\beta(y \mid x) = \frac{e^{-\beta \|y-x\|_H}}{[1+(\alpha-1)e^{-\beta}]^l} = \prod_{i=1}^{l} \frac{e^{-\beta(1-\delta_{x_i y_i})}}{1+(\alpha-1)e^{-\beta}}$$

- $\beta = \ln\left(\mu^{-1} - 1\right) + \ln(\alpha - 1)$, where $\mu = \upsilon/l$ is the mutation rate, defined by the constraint $\mathbb{E}\{\|y-x\|_H\} \leq \upsilon$.
- Optimal randomization corresponds to independent substitution of letters with probability μ (mutation rate):

$$P_\mu(r \mid n) = \binom{l}{r} \mu(n)^r (1-\mu(n))^{l-r}$$

Optimal Control of Mutation Rate

- Optimal transition kernels are

$$P_\beta(y \mid x) = \frac{e^{-\beta \|y-x\|_H}}{[1+(\alpha-1)e^{-\beta}]^l} = \prod_{i=1}^{l} \frac{e^{-\beta(1-\delta_{x_i y_i})}}{1+(\alpha-1)e^{-\beta}}$$

- $\beta = \ln\left(\mu^{-1}-1\right) + \ln(\alpha-1)$, where $\mu = \upsilon/l$ is the mutation rate, defined by the constraint $\mathbb{E}\{\|y-x\|_H\} \leq \upsilon$.
- Optimal randomization corresponds to independent substitution of letters with probability μ (mutation rate):

$$P_\mu(r \mid n) = \binom{l}{r} \mu(n)^r (1-\mu(n))^{l-r}$$

- Which $\mu(n)$?

Optimal Control of Mutation Rate

- Optimal transition kernels are

$$P_\beta(y \mid x) = \frac{e^{-\beta \|y-x\|_H}}{[1+(\alpha-1)e^{-\beta}]^l} = \prod_{i=1}^{l} \frac{e^{-\beta(1-\delta_{x_i y_i})}}{1+(\alpha-1)e^{-\beta}}$$

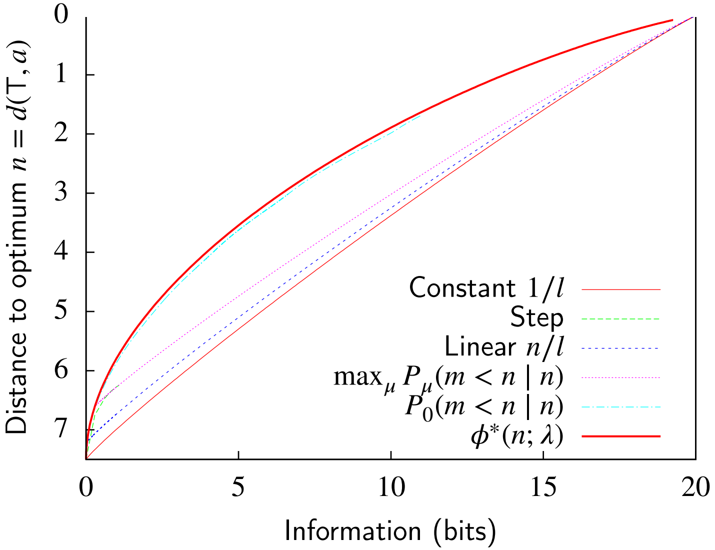
- $\beta = \ln\left(\mu^{-1} - 1\right) + \ln(\alpha - 1)$, where $\mu = \upsilon/l$ is the mutation rate, defined by the constraint $\mathbb{E}\{\|y-x\|_H\} \leq \upsilon$.
- Optimal randomization corresponds to independent substitution of letters with probability μ (mutation rate):

$$P_\mu(r \mid n) = \binom{l}{r} \mu(n)^r (1-\mu(n))^{l-r}$$

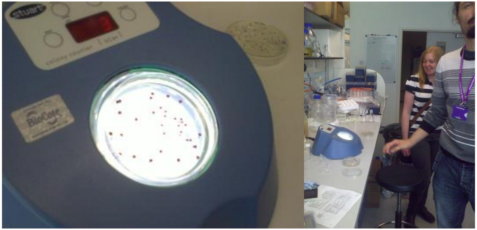
- Which $\mu(n)$?
- The CDF method:

$$\mu(n) = \sum_{m=0}^{n-1} \binom{l}{m} \frac{(\alpha-1)^m}{\alpha^l}$$

Evolution of Fitness in Information

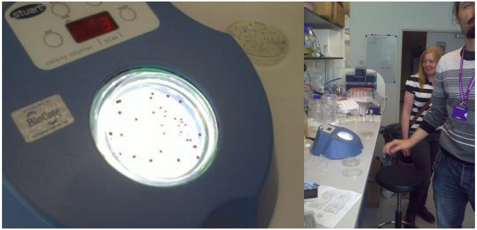


Mutation Rate Control in *E. coli*



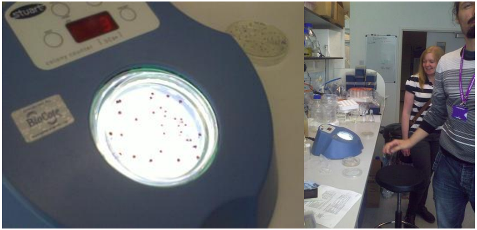
- Used strains of *Escherichia coli* K-12 MG1665

Mutation Rate Control in *E. coli*



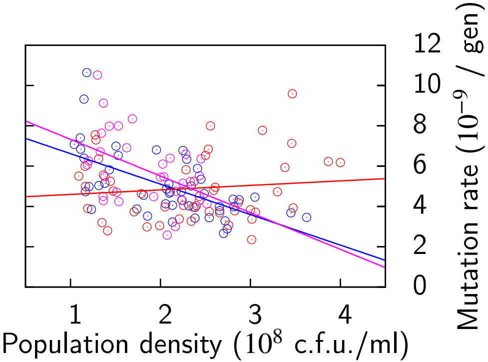
- Used strains of *Escherichia coli* K-12 MG1665
- Fluctuation test using media 50μg/ml of Rifamipicin

Mutation Rate Control in *E. coli*



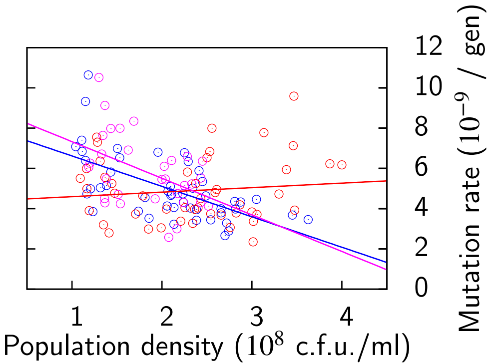
- Used strains of *Escherichia coli* K-12 MG1665
- Fluctuation test using media 50μg/ml of Rifamipicin
- Estimated mutation rates μ in *E.coli* strains grown in Davis minimal medium with different amount of glucose.

Experimental Results (Krašovec et al., 2014)



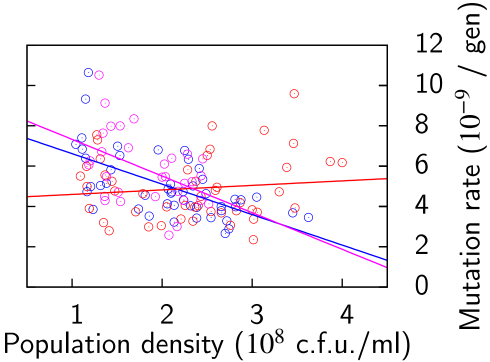
- Strong relationship between μ and density of cells ($p < .0001$).

Experimental Results (Krašovec et al., 2014)



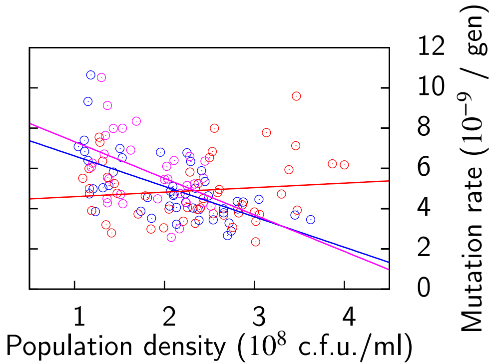
- Strong relationship between μ and density of cells ($p < .0001$).
- No such relationship in the *luxS* quorum sensing mutant ($p = .0234$).

Experimental Results (Krašovec et al., 2014)



- Strong relationship between μ and density of cells ($p < .0001$).
- No such relationship in the *luxS* quorum sensing mutant ($p = .0234$).

Experimental Results (Krašovec et al., 2014)



- Strong relationship between μ and density of cells ($p < .0001$).
- No such relationship in the *luxS* quorum sensing mutant ($p = .0234$).

Krašovec, R., Belavkin, R., Aston, J., Channon, A., Aston, E., Rash, B., Kadirvel, M., Forbes. S., Knight, C. G. (2014, April). Mutation-rate-plasticity in rifampicin resistance depends on Escherichia coli cell-cell interactions. *Nature Communications*, Vol. 5 (3742).

Optimal transportation problems (OTPs)

Information and entropy

Optimal channel problem (OCP)

Geometry of information divergence and optimization

Dynamical OTP: Optimization of evolution

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)

Krašovec, R., Belavkin, R. V., Aston, J. A. D., Channon, A., Aston, E., Rash, B. M., et al. (2014, April). Mutation rate plasticity in rifampicin resistance depends on escherichia coli cell-cell interactions. *Nature Communications*, *5*(3742).

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.

Neumann, J. von, & Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton, NJ: Princeton University Press.

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.