Persuasion by Dimension Reduction

Semyon Malamud, Andreas Schrimpf

Swiss Finance Institute, EPF Lausanne
Bank of International Settlements

June 24, 2022

Introduction

- Bayesian persuasion can be used to model communication and disclosure policies.
- The optimal policy for small and finite state spaces is well understood (KG 2011).
- How about the large, multi-dimensional state spaces?
- In this paper, we develop a novel, geometric approach to Bayesian persuasion when the state space is **continuous** and **multi-dimensional**.

Model

- State space Ω is open subset of $\mathbb{R}^L$.
- The prior distribution has a density $\mu_0(\omega)$ with respect to the Lebesgue measure on Ω, with $\int_\Omega \mu_0(\omega)d\omega = 1$.
- The information designer (the sender) observes ω and sends a signal to the receivers.
- The receivers use the Bayes rule to form a posterior μ after observing the signal of the sender, and then take an action $a \in \mathbb{R}^M$.
- Conditional on ω and a, the sender's utility is given by $W(a, \omega)$.
- We use D_a to denote the derivative (gradient) with respect to a and, similarly, D_{aa} is the second order derivative (Hessian).

Model

Assumption (1)

There exists a function $G : \mathbb{R}^M \times \Omega \to \mathbb{R}^M$ such that the optimal action $a = a(\mu)$ of the receivers with a posterior μ satisfies

$$\int G(a(\mu), \omega) d\mu(\omega) = 0 \tag{1}$$

Furthermore, G satisfies the following conditions:

- G is continuously differentiable in a.
- G is uniformly monotone in a for each ω so that $\varepsilon \|z\|^2 \leq -z^\top D_a G(a, \omega) z \leq \varepsilon^{-1} \|z\|^2$ for some $\varepsilon > 0$ and all $z \in \mathbb{R}^M$.
- the unique solution $a_*(\omega)$ to $G(a_*(\omega), \omega) = 0$ is square integrable: $E\left[\|a_*(\omega)\|^2\right] < \infty$.

Lemma (1)

For any posterior μ, there exists a unique action $a = a(\mu)$ satisfying (1) and

$$\|a(\mu)\|^2 \leq \kappa \int_\Omega \|a_*(\omega)\|^2 \, d\mu(\omega),$$

for some universal $\kappa > 0$.

Problem Formulation

Definition (1)

Let

$$\bar{W}(\mu) = \int_\Omega W(a(\mu), \omega) d\mu(\omega)$$

be the expected utility of the sender conditional on a posterior μ, with $a(\mu)$ defined in (1). The optimal Bayesian persuasion (optimal information design) problem is to maximize

$$\int_{\Delta(\Omega)} \bar{W}(\mu) d\tau(\mu)$$

over all distributions of posterior beliefs $\tau \in \Delta(\Delta(\Omega))$ satisfying

$$\int_{\Delta(\Omega)} \mu d\tau(\mu) = \mu_0$$

Problem Formulation

- We denote the value of this problem by $V(\mu_0)$.
- A solution τ (a distribution of posterior beliefs) to this problem is called an optimal information design.
- We say that an information design is **pure** (does not involve randomization) if there exists a map $a : \Omega \to \mathbb{R}^M$ such that τ is induced by this map.
- That is, the distribution of τ coincides with that of $\left\{\mu_a : a \in \mathbb{R}^M\right\}$ where

$$\mu_a(\omega) = \text{Prob}(\omega \mid a(\omega) = a), a \in \mathbb{R}^M$$

  is the posterior after observing the realization of the signal $a(\omega)$.
- A pure information design where the signal $a(\omega)$ coincides with the optimal action of the receivers, such that

$$\int G(a, \omega) d\mu_a(\omega) = E[G(a, \omega) \mid a(\omega) = a] = 0$$

  for all $a \in \mathbb{R}^M$, will be referred to as an optimal policy.

Pure Optimal Policies

- We will use Supp(a) to denote the support of any map $a : \mathbb{R}^L \to \mathbb{R}^M$:

$$\text{Supp}(a) = \left\{ x \in \mathbb{R}^M : \mu_0(\{\omega : \|a(\omega) - x\| < \varepsilon\}) > 0, \forall \varepsilon > 0 \right\}.$$

Definition (2)

Recall that $a_*(\omega)$ is the unique solution to $G\left(a_*(\omega), \omega\right) = 0$ (see Assumption 1). For any map $x : \mathbb{R}^M \to \mathbb{R}^M$, we define

$$c(a, \omega; x) \equiv \underbrace{W\left(a_*(\omega), \omega\right) - W(a, \omega)}_{\text{utility gain from inducing a different action}} + \underbrace{x(a)^\top G(a, \omega)}_{\text{shadow cost of agents' PC}}$$

Everywhere in the sequel, we refer to c as the cost of **information transport**.

- Note that, for any policy a satisfying $E[G(a(\omega), \omega) \mid a(\omega) = a] = 0$ and any well-behaved x, we always have

$$\underbrace{E\left[W(a(\omega), \omega) - W\left(a_*(\omega), \omega\right)\right]}_{\text{gain from concealing information}} = -\underbrace{E[c(a(\omega), \omega; x)]}_{\text{expected cost}}$$

Pure Optimal Policies

Theorem (1)

There always exists a Borel-measurable pure optimal policy $a(\omega)$ solving the problem of Definition 1. Furthermore, if we define the optimal information manifold $\Xi = \text{Supp}(a)$ and

$$x(a)^\top = E\left[D_a W(a,\omega) \mid a\right] E\left[D_a G(a,\omega) \mid a\right]^{-1}$$

then

$$c(a(\omega),\omega;x) \leq 0 \quad \text{(transporting information from } \omega \text{ to a } (\omega) \text{ has a negative cost)} \tag{4}$$

and

$$a(\omega) \in \arg\min_{b\in\Xi} c(b,\omega;x) \quad \text{(transporting information to a}(\omega)\text{ is optimal)} \tag{5}$$

and the function $c(a(\omega),\omega;x)$ is Lipschitz continuous in ω. Furthermore, any optimal information design satisfies (4) and (5).

- There is an interesting connection between Theorem 1 and optimal transport theory.

Optimal Transport Theory: Background

- Consider N mines with position $(x_i)_{i=1}^N$ and resource $(\mu_i)_{i=1}^N$.
- M factories with position $(y_i)_{i=1}^M$ and demand $(\nu_i)_{i=1}^M$.
- The distance $c(x_i, y_j), i = 1, \dots, N, j = 1, \dots, M$.
- Task is to find a **transportation plan**: Γ_{ij}, how many resources to move from x_i to y_j.
- The total cost (Kantorovich problem):

$$\min_{\Gamma} \sum_{i=1}^{N} \Gamma_{ij} c(x_i, y_j), \text{ s.t. } \sum_{i}^{N} \Gamma_{ij} = \nu_j \text{ and } \sum_{j=1}^{M} \Gamma_{ij} = \mu_i$$

- Monge problem **transportation map**: one mine only ships to one resource.

Optimal Transport Theory: Background

Theorem (Monge-Kantorovich duality)

Let $\mathcal{X}$ and $\mathcal{Y}$ be two Banach spaces, and let P and Q be two probability measures on $\mathcal{X}$ and $\mathcal{Y}$ respectively. Let $\Phi : \mathcal{X} \times \mathcal{Y} \to \mathbb{R} \cup \{-\infty\}$ be an upper semicontinuous surplus function bounded from above. Then,

- The value of the primal Monge-Kantorovich problem

  $$\sup_{\pi \in \mathcal{M}(P,Q)} \mathbb{E}_\pi[\Phi(X,Y)]$$

  coincides with the value of the dual

  $$\inf_{u,v} \mathbb{E}_P[u(X)] + \mathbb{E}_Q[v(Y)]$$
  $$s.t. \;\; u(x) + v(y) \geq \Phi(x,y)$$

  where the infimum is over measurable and integrable functions u and v, and the inequality constraint should be satisfied for almost every x and almost every y (all these statements are respective to measures P and Q);
- An optimal solution π to primal problem exists.

Optimal Transport Theory: Background

Example (Principal-agent problems)

We consider a basic principal-agent model with possibly multivariate characteristics. Consider an agent of type $x \in \mathcal{X}$, where $\mathcal{X}$ is a subset of $\mathbb{R}^d$. Assume that the types of agents follow a probability distribution P over $\mathcal{X}$. Based on the type x' announced by the agent, the principal decides on an outcome $y = T(x') \in \mathcal{Y}$, where $\mathcal{Y}$ is also a subset of $\mathbb{R}^d$, and on a payment $v(y)$ made by the agent, so that agent x 's utility of announcing type x' is $\Phi(x, T(x')) - v(T(x'))$. Then T is implementable in dominant strategy if and only if there exists a payment schedule $v(\cdot)$ such that

$$T(x) \in \arg\max_y \{\Phi(x, y) - v(y)\}.$$

Optimal Transport Theory: Background

Theorem (Carlier)

The following statements are equivalent:

- The map $T : \mathcal{X} \to \mathcal{Y}$ is implementable (IC).
- The map T solves the Monge problem associated with Φ between measures P and $Q = T\#P$, that is,

$$\mathbb{E}_P[\Phi(X, T(X))] = \max_{\pi \in \mathcal{M}(P,Q)} \mathbb{E}_\pi[\Phi(X, Y)].$$

Further, when this is the case, the payment rules such that (T, v) is implementable are such that there is some function u such that (u, v) is a solution to the dual problem

$$\begin{aligned} \min_{u,v} \mathbb{E}_P[u(X)] + \mathbb{E}_P[v(T(X))] \\ s.t.\ \ u(x) + v(y) \geq \Phi(x, y) \end{aligned}$$

Optimal Transport Theory: Background

- From this equivalence we see that testing whether a given map T is implementable is equivalent to an optimal transportation problem. When Φ satisfies the conditions in chapter 7 , then for P-almost every x,

$$\nabla u(x) = \frac{\partial \Phi(x, T(x))}{\partial x},$$

where u appears in the solution to dual problem. This suggests that the payment v is determined from the knowledge of T. Indeed, the condition implies that u is determined up to a constant; next, v is determined on $\mathcal{Y} = T(\mathcal{X})$ by

$$v(T(x)) = \Phi(x, T(x)) - u(x).$$

Optimal Transport Theory

Definition (3)

Consider two probability measures, $\mu_0(\omega)d\omega$ (distribution of mines) on Ω and ν on Ξ (distribution of factories).

- The optimal map problem (the Monge problem) is to find a map $X : \Omega \to \Xi$ that minimizes $\int c(X(\omega), \omega)\mu_0(\omega)d\omega$ under the constraint that the random variable $\chi = X(\omega)$ is distributed according to ν.
- The Kantorovich problem is to find a probability measure γ on $\Xi \times \Omega$ that minimizes $\int c(\chi, \omega)d\gamma(\chi, \omega)$ over all γ whose marginals coincide with $\mu_0(\omega)d\omega$ and ν, respectively.

- It is known that, under very general conditions, the Monge problem and its Kantorovich relaxation have identical values, and an optimal map exists.
- It turns out that any optimal policy of Theorem 1 solves the Monge problem.

Optimal Transport Theory

Theorem (2)

Any optimal policy $a(\omega)$ solves the Monge problem with ν being the distribution of the random vector $a(\omega) \in \mathbb{R}^M$.

- Proof: Let $(a(\omega), x(a))$ be an optimal policy and let

$$\phi^c(a) = \inf_{\omega} \left(c(a, \omega; x) - \phi_{\Xi}(\omega; x)\right)$$

- Pick an $a \in \Xi$. Since $a \in \Xi$, there exists a $\tilde{\omega}$ such that $a = a(\tilde{\omega})$ and hence

$$\phi^c(a) = \inf_{\omega} \left(c(a, \omega; x) - \phi_{\Xi}(\omega; x)\right) \leq c(a, \tilde{\omega}; x) - \phi_{\Xi}(\tilde{\omega}; x) = 0.$$

- Thus,

$$\int \phi^c(a(\omega))\mu_0(\omega)d\omega = 0$$

- At the same time,

$$c(a, \omega; x) - \phi_{\Xi}(\omega; x) = c(a, \omega; x) - \inf_{b \in \Xi} c(b, \omega; x) \geq 0.$$

- Thus, $\phi^c(a) = 0$ for all $a \in \Xi$. Now, by the definition of ϕ^c, we always have

$$\phi^c(a) + \phi_{\Xi}(\omega; x) \leq c(a, \omega)$$

for an optimal policy.

Optimal Transport Theory

Theorem (2)

Any optimal policy $a(\omega)$ solves the Monge problem with ν being the distribution of the random vector $a(\omega) \in \mathbb{R}^M$.

- Let γ be the measure on $\Xi \times \Omega$ describing the joint distribution of $\chi = a(\omega)$ and ω. Then,

$$\int c(a,\omega)d\gamma(a,\omega) = \int c(a(\omega),\omega)\mu_0(\omega)d\omega = \int \phi_\Xi(\omega;x)\mu_0(\omega)d\omega = \int \phi_\Xi(a)d\nu(a)$$

- Pick any measure π from the Kantorovich problem. Then,

$$\begin{aligned}
&\int c(a,\omega)d\gamma(a,\omega) = \int \left(\phi_\Xi(\omega;x)\right) d\gamma(a,\omega) \\
&= \int \phi_\Xi(\omega;x)\mu_0(\omega)d\omega + \int \phi^c(a(\omega))\mu_0(\omega)d\omega \\
&= \int \left(\phi_\Xi(\omega;x) + \phi^c(a(\omega))\right)\mu_0(\omega)d\omega \\
&= \int \left(\phi_\Xi(\omega;x) + \phi^c(a)\right) d\pi(a,\omega) \le \int c(a,\omega)d\pi(a,\omega)
\end{aligned}$$

Thus, γ minimizes the cost in the Kantorovich problem.

Optimal Transport Theory

Theorem (2)

Any optimal policy $a(\omega)$ solves the Monge problem with ν being the distribution of the random vector $a(\omega) \in \mathbb{R}^M$.

- In our setting, this result has a similar flavour: In order to induce an optimal action, the sender optimally aligns actions a with the state ω to minimize the cost of information transport, c.
- However, there is a major difference between Bayesian persuasion and classic optimal transport.
- In the Monge-Kantorovich problem, factories are in fixed locations and we need to design the transport plan. In contrast, in Bayesian persuasion the "location of factories" is endogenous: It is the support Ξ of the map $a(\omega)$.

Moment Persuasion

- In this section, we consider a setup where $G(a, \omega) = a - g(\omega)$ for some continuous functions $g(\omega) = (g_i(\omega))_{i=1}^M : \Omega \to \mathbb{R}^M$.
- Dworczak and Kolotilin (2019) refer to this setup as "moment persuasion".

Assumption (3)

We have $G(a, \omega) = a - g(\omega), a_*(\omega) = g(\omega), W(a, \omega) = W(a)$ and $|W(a)| + \|D_a W\| \le f\left(\|a\|^2\right)$ for some convex function f satisfying $E\left[\|g(\omega)\|^2 f\left(\|g(\omega)\|^2\right)\right] < \infty$.

- Eq.(3) reduces to

$$x(a)^\top = E\left[D_a W(a, \omega) \mid a\right] E\left[D_a G(a, \omega) \mid a\right]^{-1} = D_a W(a).$$

- We define

$$c(a, b) = W(b) - W(a) + D_a W(a)(a - b).$$

- The cost of information transport, c, coincides with the classic Bregman divergence that plays an important role in convex analysis (see, e.g., Rockafellar (1970)).
- In particular, as the graph of a convex function always lies above a tangent hyperplane, $c(a, b) \ge 0$ when W is convex, and hence $c(a, b)$ can be interpreted as "distance".
- However, in our setting W is generally not convex and hence c can take negative values.

Bregman Projection

- We define the Bregman Projection $\mathcal{P}_\Xi$ onto a set Ξ via

$$\mathcal{P}_\Xi(b) = \arg\min_{a \in \Xi} c(a, b) \quad \text{(definition of a Bregman Projection).}$$

- In other words, $\mathcal{P}_\Xi$ projects b onto the point $a \in \Xi$ that attains the lowest Bregman divergence.
- As neither Ξ nor W are convex, standard results about Bregman projections do not apply.
- Our key objective here is to understand the structure of the support set Ξ of an optimal policy.

Bregman Projection

Definition (4)

Let conv(X) be the closed convex hull of a set $X \subset \mathbb{R}^M$.

- A set $\Xi \subset \mathbb{R}^M$ is X-maximal if $\inf_{a \in \Xi} c(a, b) \leq 0$ for all $b \in X$.
- A set Ξ is W-monotone if $c\left(a_1, a_2\right) \geq 0$ for all $a_1, a_2 \in \Xi$.
- A set Ξ is W-convex if $W\left(ta_1 + (1-t)a_2\right) \leq tW\left(a_1\right) + (1-t)W\left(a_2\right)$ for all $a_1, a_2 \in \Xi, t \in [0, 1]$.
- We also define $\phi_{\Xi}(x) = \inf_{a \in \Xi} c(a, x)$.

Bregman Projection

Theorem (Optimal Policies are Projections)

There always exists a pure optimal policy. Furthermore:

- Each such policy is a Bregman projection onto an optimal information manifold Ξ

$$a(\omega) \in \mathcal{P}_{\Xi}(g(\omega)) \equiv \arg\min_{a \in \Xi} c(a, g(\omega)),$$

  for Lebesgue-almost every ω.
- Any optimal information manifold is conv $(g(\Omega))$-maximal, W-convex, and W-monotone.
- The pool of every signal value a_1 is $\text{Pool}\,(a_1) = \{\omega \in \Omega : a(\omega) = a_1\} = a^{-1}\,(a_1)$.

- Formally, it means that if $g(\omega) \in \Xi$, then it is optimal to reveal the true value of $g(\omega)$ instead of sending a signal corresponding to a different point on the manifold.
- As $c(a,b) \geq 0$ for all $a, b \in \Xi$, we have $c\,(\mathcal{P}_{\Xi}(b), a) \geq 0$ for all a and, by direct calculation, $c\,(\mathcal{P}_{\Xi}(b), \mathcal{P}_{\Xi}(b)) = 0$.
- Thus, we have the projection property: for any $x \in \mathcal{P}_{\Xi}(b)$, we have $x \in \mathcal{P}_{\Xi}(x)$.

Bregman Projection

- We now discuss the two key properties of an optimal information manifold: monotonicity and maximality.
- Suppose for simplicity that $M = 2, W(a) = a_1 a_2$, and $g(\omega) = \begin{pmatrix} \omega_1 \\ \omega_2 \end{pmatrix}$, as in Rochet and Vila (1994) and Rayo and Segal (2010).
- Then, $c(a, b) = (a_1 - a_2)(b_1 - b_2)$, and W-monotonicity means that

$$(a_1 - a_2)(b_1 - b_2) \geq 0.$$

- Thus, in the language of Rayo and Segal (2010), for any two signals $s_1 \neq s_2$, expected prospects $a(i) = E[\omega \mid s_i] = \begin{pmatrix} E[\omega_1 \mid s_i] \\ E[\omega_2 \mid s_i] \end{pmatrix}$ are ordered: A better signal reveals that both expected dimensions of the prospect are better.
- This ordering immediately implies the existence of a monotone increasing function $f : \mathbb{R} \to \mathbb{R}$ such that

$$\Xi \subset \text{Graph}(f) = \left\{ \begin{pmatrix} f(a_2) \\ a_2 \end{pmatrix} : a_2 \in \mathbb{R} \right\}.$$

- Clearly, this graph is a one-dimensional object (a curve) and, thus, so is the optimal information manifold, Ξ.
- Therefore, optimal persuasion is achieved by dimension reduction and W-monotonicity imposes a lower-dimensional structure on Ξ.

Convexity of Pools

- For a discrete state space, Rayo and Segal (2010) show that the pool of every signal is a discrete subset of a line segment.
- Therefore, we intuitively expect that, in the continuous case, the pool of every signal is convex (i.e., there are no gaps in the segment).
- Convex pools correspond to monotone partitions when $M = 1$ When $g(\omega)$ is non-linear, the convexity of pools cannot be guaranteed.
- However, as Dworczak and Martini (2019) show, when $g(\omega) = \omega$ and $M = 1$, monotone partitions are indeed optimal when $W(a)$ is affine-closed.

Proposition (1 Monotone partitions (Convexity of pools))

Suppose that $g(\omega) = \omega$ and Ω is convex. Then, there exists a pure optimal policy $a(\omega)$ such that the map $\omega \to D_a W(a(\omega))$ is monotone increasing on Ω and the set

$$\{\omega \in \Omega : D_a W(a(\omega)) = a\} = \cup_{b \in (D_a W)^{-1}(a)} \mathrm{Pool}(b)$$

is always convex. If the map $a \to D_a W(a)$ is injective, then the pool of every signal is convex (up to a set of measure zero [29] and $a(\omega)$ is an idempotent: $a(a(\omega)) = a(\omega)$.

Convexity of Pools

Theorem (4 Maximality is both necessary and sufficient)

Let Ξ be a conv $(g(\Omega))$-maximal subset of $\mathbb{R}^M$. Suppose that a map a (ω) satisfies $a(\omega) \in \mathcal{P}_\Xi(g(\omega))$ and $a(\omega) = E[g(\omega) \mid a(\omega)]$ for Lebesgue-almost all ω. Then, a is an optimal policy.

- This fact drastically simplifies the search for optimal information designs.
- It implies that finding an optimal policy reduces to two steps:
  - step 1: find a set of candidates Ξ that are conv$(g(\Omega))$-maximal
  - step 2: solve the following integro-differential equation:

$$\mathcal{P}_\Xi(g(\omega)) = E\left[g(\omega) \mid \mathcal{P}_\Xi(g(\omega))\right].$$

Optimal Information Manifold

- For any symmetric matrix H, let $\nu_+(H)$ be the number of strictly positive eigenvalues and $\nu(H) \geq \nu_+(H)$ the number of nonnegative eigenvalues.
- Recall that the dimension, $\dim(X)$, of a convex set X is the dimension of the smallest linear manifold containing it.

Corollary (1 Pools are low-dimensional sets)

We have $\dim(\text{conv}(g(\text{Pool}(a)))) \leq M - \nu_+ (D_{aa}W(a))$. In particular, if $D_{aa}W(a)$ has at least one strictly positive eigenvalue for any a, then the convex hull of the image of the pool of any signal, $\text{conv}(g(\text{Pool}(a)))$, has measure zero. If g is locally injective and bi-Lipschitz, then Pool (a) also has Lebesgue measure zero for each a.

- This implies that the observations in Rochet and Vila (1994) and Rayo and Segal (2010) are typical for $M > 1$: Pools have measure zero, and can only have positive measure if a is a local maximum, so that $D_{aa}W(a)$ is negative semi-definite.
- In the Rayo and Segal (2010) setting, $W(a) = a_1 G(a_2)$ and

$$D_{aa}W(a) = \begin{pmatrix} 0 & G'(a_2) \\ G'(a_2) & a_1 G''(a_2) \end{pmatrix}$$

  with $G'(a_2) > 0$
- Hence $\det(D_{aa}W(a)) = -(G'(a_2))^2 < 0$, so that $D_{aa}W(a)$ always has exactly one positive and one negative eigenvalue.

Optimal Information Manifold

Definition (5)

A ν-dimensional (topological) manifold $\Xi \subset \mathbb{R}^M$ is a set such that every point $a \in \Xi$ has a neighborhood homeomorphic to $\mathbb{R}^\nu$. The respective homeomorphism is called a (local) coordinate map. A ν-dimensional Lipschitz (respectively, C^k)-manifold is such that the respective homeomorphism and its inverse are Lipschitz-continuous (respectively, k-times continuously differentiable).

- A manifold can typically be defined in two ways: through a coordinate map or through a system of equations.
- For example, the unit circle $\Xi = \left\{(a_1, a_2) \in \mathbb{R}^2 : a_1^2 + a_2^2 = 1\right\}$ is a smooth 1-dimensional manifold defined by one equation $\Psi(a) = 0$ with $\Psi(a) = a_1^2 + a_2^2 - 1$.

Optimal Information Manifold

- Our goal is to show that Ξ is a lower-dimensional set. In order to gain intuition for the origins of lower-dimensionality of an optimal information manifold, consider first the case when W is linear-quadratic,

$$W(a) = 0.5a^\top Ha + h^\top a.$$

- In this case, by direct calculation, $c(a,b) = (a-b)^\top H(a-b)$ and hence W-monotonicity of Ξ implies that for any $a_1, a_2 \in \Xi$, we have $(a_1 - a_2)^\top H(a_1 - a_2) \geq 0$. As we will now show, this imposes a low-dimensional structure on Ξ.
- Let $H = VDV^\top$ be the eigenvalue decomposition of H where $D = \mathrm{diag}(\lambda_1, \cdots, \lambda_M)$ with $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_M$. Let ν be the number of nonnegative eigenvalues of H.
- Let also $\tilde{a}_i = |D|^{1/2} V^\top a_i = \begin{pmatrix} \tilde{a}_{i,-\nu} \\ \tilde{a}_{i,\nu+} \end{pmatrix}$, where we have split the vector into two components corresponding to nonnegative and negative eigenvalues, respectively.
- The monotonicity condition $(a_1 - a_2)^\top H(a_1 - a_2) \geq 0$ now takes the form

$$\underbrace{\|\tilde{a}_{1,-\nu} - \tilde{a}_{2,-\nu}\|^2}_{\text{change in good information}} \geq \underbrace{\|\tilde{a}_{1,\nu+} - \tilde{a}_{2,\nu+}\|^2}_{\text{change in bad information}} \quad \text{for all } a_1, a_2 \in \Xi.$$

- This immediately implies the existence of a map $f : \mathbb{R}^\nu \to \mathbb{R}^{M-\nu}$ such that $\tilde{a}_{\nu+} = f(\tilde{a}_{-\nu})$ because the coincidence of $\tilde{a}_{1,-\nu}$ with $\tilde{a}_{2,-\nu}$ always implies the coincidence of $\tilde{a}_{1,\nu+}$ with $\tilde{a}_{2,\nu+}$.

Optimal Information Manifold

Theorem (5 Ξ is a lower-dimensional manifold)

Let Ξ be an optimal information manifold (the support of an optimal policy) and $\nu(a) = \nu\left(D_{aa}W(a)\right)$ be the local degree of convexity of W. Then, for any open set $B, \Xi \cap B$ is a subset of a Lipschitz manifold of dimension at most $\sup_{a \in B} \nu(a)$.

- Theorem 5 only implies that we can characterize Ξ as $\Xi = \left\{a \in \mathbb{R}^M : a = f(\theta), \theta \in \Theta\right\}$, where $\Theta \subset \mathbb{R}^\nu$ is a lower-dimensional subset with unknown properties.
- Rewriting the projection as

$$a(\omega) = \arg\min_{\theta \in \Theta} c(f(\theta), g(\omega))$$

- FOC w.r.t θ: $D_\theta c(f(\theta), g(\omega)) = 0$, where

$$c(a, b) = W(b) - W(a) + D_a W(a)(a - b).$$

Optimal Information Manifold

Corollary (2 Characterization of Pools)

Let $a(\omega)$ be a pure optimal policy and Ξ the corresponding optimal information manifold. Suppose that $D_{aa}W(a)$ is non-degenerate and that, for any $\varepsilon > 0$, $\text{conv}\,(g\,(\text{Pool}\,(\Xi \cap B_\varepsilon(a)))) \subset \mathbb{R}^M$ has positive Lebesgue measure.
Let $f : \mathbb{R}^\nu \to \Xi$ be local coordinates from Theorem 5 in a small neighborhood of a, and let $\Theta = f^{-1}\,(\Xi \cap B_\varepsilon(a))$. Then, for Lebesgue-almost every $\theta \in \Theta, f$ is differentiable, with a Jacobian $\text{Df}(\theta) \in \mathbb{R}^{M\times\nu}$, and we have

- The matrix $Df(\theta)^\top D_{aa}W(f(\theta))Df(\theta) \in \mathbb{R}^{\nu\times\nu}$ is Lebesgue-almost surely symmetric and positive semi-definite.
- Lebesgue-almost every ω satisfies

$$Df(\theta)^\top D_{aa}W(f(\theta))(f(\theta) - g(\omega)) = 0$$

  when $\omega \in \text{Pool}(f(\theta))$.

Optimal Information Manifold

- As an illustration, consider the case $W(a) = a_1 a_2, g(\omega) = \begin{pmatrix} \omega_1 \\ \omega_2 \end{pmatrix}$ of Rochet and Vila (1994) and Rayo and Segal (2010).
- In this case, $f(\theta) = \begin{pmatrix} \varphi(\theta) \\ \theta \end{pmatrix}$ and item (1) takes the form

$$Df(\theta)^\top D_{aa} W(f(\theta)) Df(\theta) = (\varphi'(\theta), 1) \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \varphi'(\theta) \\ 1 \end{pmatrix} = 2\varphi'(\theta) \geq 0,$$

  confirming that Ξ is the graph of a monotone increasing function.
- While the pool equation takes the form

$$\omega_1 = \varphi'(\theta)\,(\theta - \omega_2) + \varphi(\theta).$$

- Thus, pools are (subsets of) lines orthogonal to the manifold Ξ. Furthermore, the slope of each line is determined by the slope of the manifold: The steeper the slope, the stronger the separation is between different signals on Ξ. Thus, already in this simple setting, formula (10) provides a novel insight: an equality between the degree of separation and the steepness of pools.

Concealing the Tails

- In the preceding example, Ξ is the graph of a monotone function. If $\Omega = \mathbb{R}$ is unbounded, the maximality of Ξ implies that Ξ extends all the way to infinity.
- This has important implications for the nature of signals for extreme (tail) state realizations-specifically, the optimal policy always reveals (some) information about the tails.
- In particular, there will always be states $\omega = \begin{pmatrix} \varphi(\theta) \\ \theta \end{pmatrix}$ with arbitrarily large θ that are revealed.
- However, this property cannot be true in a general persuasion problem.

Definition (6)

Let $\mathcal{C}(a,\varepsilon) = \left\{ b \in \mathbb{R}^M : b^\top a/(\|a\| \cdot \|b\|) > 1 - \varepsilon \right\}$ be the ε-cone around a : the set of vectors b that point in **approximately the same direction** as a. We say that the value function W is concave along rays for large a if there exists a small $\varepsilon > 0$ and a large $K > 0$ such that $b^\top D_{aa} W(a) b < 0$ for all a with $\|a\| > K$ and all $b \in \mathcal{C}(a,\varepsilon)$).

Concealing the Tails

Proposition (2 Concealing Tail Information)

Suppose that $\text{conv}(g(\Omega)) = \mathbb{R}^M$ and let Ξ be an optimal information manifold.

- If $W(a) = a^\top Ha + h^\top a$ with $\det H \neq 0$, then $\Xi = f(\Theta)$ for some Lipschitz $f : \mathbb{R}^{\nu(H)} \to \mathbb{R}^M$, where Θ extends indefinitely in all directions.
- If W is concave along rays for large a, then there exists a constant K independent of the prior μ_0, such that any optimal information manifold satisfies $\Xi \subset B_K(0)$.

- Proposition 2 shows how a weak form of the sender's aversion of large risks makes it optimal to conceal information about large state realizations.
- Instead of assuming concavity (risk aversion) occurs in all directions, it is enough to assume it exists along rays according to Definition 6.

Concealing the Tails

Corollary (3 Concave Marginal Utility Implies Concealing the Tails)

- Let H be a non-degenerate, $M \times M$ positive-definite matrix. Suppose that $W(a) = \varphi\,(a'Ha)$ for some φ with $-\varphi''(x)/\,|\varphi'(x)| > \varepsilon$ for some $\varepsilon > 0$ and all sufficiently large x. Then, $\nu\,(D_{aa}(W(a))) \geq M - 1$ for all a.
- Yet, W is concave along rays for large a and, hence, any optimal information manifold is bounded, contained in a ball of radius K that is independent of the prior μ_0.

- This shows explicitly how non-linear marginal utility alters the nature of optimal information manifolds, leading to a phenomenon that we call "information compression", whereby potentially unbounded information is compressed into a bounded signal.

Supplying Product Information

- In this model, the sender is endowed with a prospect randomly drawn from $\mu_0(\pi, v)$. Each prospect is characterized by $\omega = \begin{pmatrix} \pi \\ v \end{pmatrix}$ where π is the prospect's profitability to the sender and v is its value to the receiver.
- After observing the signal of the sender, the receiver decides whether to accept the prospect.
- Whenever the receiver accepts the prospect, she forgoes an outside option worth r, which is a random variable independent of ω and drawn from a c.d.f. G over $\mathbb{R}$.
- Thus, the sender and receiver obtain payoffs, respectively, equal to $q\pi$ and $q(v-r)$ where $q = 1$ if the prospect is accepted and zero otherwise.
- Defining $g(\omega) = \begin{pmatrix} \pi \\ v \end{pmatrix}$, we get by direct calculation that the sender's and receiver's expected utilities are respectively given by

$$W(a) = a_1 G(a_2), U(a) = \int_{\mathbb{R}} \max\{a_2 - r, 0\} \, dG(r) = \int_{-\infty}^{a_2} (a_2 - r) \, dG(r).$$

- We can calculate

$$c(a, b) = b_1 (G(b_2) - G(a_2)) - a_1 G'(a_2)(b_2 - a_2).$$

Supplying Product Information

- Consider first the case when $G(b) = b$ (uniform acceptance rate). As Rayo and Segal (2010) show in a discrete state space setting, the set of possible signals' payoffs $\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} E[\pi \mid a] \\ E[v \mid a] \end{pmatrix}$ (that is, the optimal information manifold Ξ) is ordered: For any two possible signals' payoffs $a, \tilde{a} \in \Xi$, we always have $(a_1 - \tilde{a}_1)(a_2 - \tilde{a}_2) \geq 0$.
- As we explain above, this is a direct consequence of W-monotonicity because $c(a_1, \tilde{a}_1) = (a_1 - \tilde{a}_1)(a_2 - \tilde{a}_2)$.
- Ξ is in fact a graph of a monotone increasing function $a_1 = f(a_2)$.

Proposition (3)

There always exists a pure optimal policy $a(\omega) = \begin{pmatrix} a_1(\omega) \\ a_2(\omega) \end{pmatrix}$. For each such policy, there exists a function $f(a_2)$ such that $a_1(\omega) = f(a_2(\omega))$ for all ω and, hence, the optimal information Ξ is the graph $\{(f(a_2), a_2)\}$. The function $f(a_2)(G'(a_2))^{1/2}$ is monotone increasing in a_2. For each a_2 , Pool (a_2) is a convex segment of the line $\pi = \kappa_1(a_2)v + \kappa_2(a_2)$ with

$$\kappa_1(a_2) = -\left(f(a_2)G'(a_2)\right)' / G'(a_2), \kappa_2(a_2) = f(a_2) - a_2\kappa_1(a_2).$$

Supplying Product Information

- We show that the optimal way to communicate a high-dimensional signal is through dimension reduction, achieved by projecting the signal onto a lower-dimensional optimal information manifold.
- We derive several analytical results regarding the shape and the geometry of the optimal information manifold and the corresponding optimal pools.
- In particular, we show when it is optimal for the sender to conceal the tails and project the signal onto a compact manifold.

Thanks!