The Persuasion Duality*

Piotr Dworczak[†] and Anton Kolotilin[‡]

December 30, 2019

Abstract

We present a unified duality approach to Bayesian persuasion. The optimal dual variable, interpreted as a price function, is shown to be a supergradient of the concave closure of the objective function at the prior belief. Under regularity conditions, our general duality result implies known results for the case when the objective function depends only on the expected state. We apply our approach to characterize the optimal signal in the case when the state is two-dimensional.

Keywords: Bayesian persuasion, information design, duality

JEL codes: D82, D83

*We thank Isa Chavez, David Rahman, Chris Shannon, Marciano Siniscalchi, Bruno Strulovici, and Xiaoyun Qiu for helpful comments and suggestions. Kolotilin gratefully acknowledges support from the Australian Research Council Discovery Early Career Research Award DE160100964.

[†]Department of Economics, Northwestern University.

[‡]School of Economics, UNSW Business School.

1 Introduction

The past decade has witnessed an explosion of the literature on Bayesian persuasion (see Bergemann and Morris, 2019 and Kamenica, 2019 for excellent overviews). Kamenica and Gentzkow (2011) show that the optimal signal in a Bayesian persuasion problem concavifies the objective function in the space of posterior beliefs over the state.[1] Although conceptually attractive, concavification is not always a tractable approach. Thus, several recent papers (see Kolotilin, 2018, Dworczak and Martini, 2019, Galperti and Perego, 2018, and Dizdar and Kováč, 2019) used duality theory to characterize the optimal signal.

In this paper, we present a unified duality approach to the Bayesian persuasion problem. Our approach builds on and extends the geometric duality of Gale (1967). We show that the optimal dual variable is a supergradient of the concave closure of the objective function at the prior belief (Section 3). Thus, strong duality holds if and only if the concave closure is superdifferentiable at the prior. This regularity condition is always satisfied when the state space is finite and the prior has full support (because every finite-dimensional concave function is superdifferentiable at every interior point). But superdifferentiability may not hold when the state space is infinite (hence, additional assumptions are needed for strong duality to hold). We show that strong duality holds if the concave closure has bounded steepness at the prior, which in turn holds if the objective function is Lipschitz continuous. We also interpret the general duality result using an analogy to the linear production problem of Gale (1989), with the dual variable interpreted as the state price function.[2]

In the special case when the objective function depends only on a finite set of moments of the posterior distribution (for example, the expected state), we show that the price function on the space of moments is a convex function that is greater than the objective function (Section 4). This result generalizes the duality results established by Kolotilin (2018), Dworczak and Martini (2019), and Dizdar and Kováč (2019). We comment on the precise relationship to these papers in Section 7.

We illustrate the usefulness of the generalized duality approach by characterizing the optimal signal in the classical case of Rayo and Segal (2010) when the state is two-dimensional (Section 5). Moreover, relying on a characterization in Doval and Skreta (2018), we show that our duality results can be easily extended to a *constrained* persuasion problem in which the distribution of posterior beliefs must satisfy additional linear constraints (Section 6).

[1] See also Aumann and Maschler (1995) for an early version of this result.

[2] Similar interpretations have appeared in Dworczak and Martini (2019) and Galperti and Perego (2018).

2 Model

The state space Ω is a compact metric space. Let $\Delta(\Omega)$ be the space of Borel probability measures on Ω, endowed with the weak* topology. A prior belief $\mu_0 \in \Delta(\Omega)$ has full support.[3] An objective function $V : \Delta(\Omega) \to \mathbb{R}$ is upper semi-continuous.

The persuasion problem in Kamenica and Gentzkow (2011), which will be called the primal problem, is to find a *distribution of posterior beliefs* $\tau \in \Delta(\Delta(\Omega))$ to

$$\begin{aligned} &\text{maximize} \int_{\Delta(\Omega)} V(\mu)\mathrm{d}\tau(\mu) \\ &\text{subject to} \int_{\Delta(\Omega)} \mu \mathrm{d}\tau(\mu) = \mu_0. \end{aligned} \tag{P}$$

A probability measure τ that satisfies the constraint in (P) will be called feasible for (P).

Let $C(\Omega)$ be the space of continuous functions on Ω. The dual problem is to find a *price function* $P \in C(\Omega)$ to

$$\begin{aligned} &\text{minimize} \int_{\Omega} P(\omega)\mathrm{d}\mu_0(\omega) \\ &\text{subject to} \int_{\Omega} P(\omega)\mathrm{d}\mu(\omega) \geq V(\mu) \text{ for all } \mu \in \Delta(\Omega). \end{aligned} \tag{D}$$

A continuous function P that satisfies the constraint in (D) will be called feasible for (D).

We interpret the persuasion problem as a linear production problem of Gale (1989). The states are economic resources, and the probability measure μ_0 is a producer's endowment of resources. The set $\Delta(\Omega)$ is the set of linear production processes available to the producer. A process $\mu \in \Delta(\Omega)$ operated at unit level consumes the measure μ of resources and generates income $V(\mu)$. A production plan $\tau \in \Delta(\Delta(\Omega))$ describes the level at which each process μ is operated. The primal problem is for the producer to find a production plan that exhausts the endowment μ_0 and maximizes the total income.

To interpret the dual problem, imagine that there is a wholesaler who wants to buy out the producer. The wholesaler sets a unit price $P(\omega)$ for each resource ω. The producer's (opportunity) cost of operating a process μ at unit level is thus $\int_\Omega P(\omega)\mathrm{d}\mu(\omega)$. A price function $P \in C(\Omega)$ is feasible for the wholesaler if the income generated by each process of the producer is not greater than the cost of operating the process, which makes the producer willing to sell all the resources. The dual problem is for the wholesaler to find feasible prices that minimize the total cost of buying up all the resources.[4]

[3] This is effectively without loss of generality because we can always redefine Ω to be $\mathrm{supp}(\mu_0)$.

[4] Dworczak and Martini (2019) offer a related intuition with the producer replaced by a consumer, pro-

3 Duality

In this section, we establish weak and strong duality for the persuasion problem. Within our interpretation, weak duality states that the total income generated by the producer cannot be greater than the total cost of the resources under prices that make the producer willing to sell the resources. Moreover, if there exists a plan for the producer and prices for the wholesaler that equalize the total income with the total cost, then this plan must be optimal for the producer, and the prices must be optimal for the wholesaler. Strong duality establishes the existence of such a plan and prices for any endowment of resources.

Theorem 1 (Weak duality). *If $\tau \in \Delta(\Delta(\Omega))$ is feasible for (P) and $P \in C(\Omega)$ is feasible for (D), then*

$$\int_{\Omega} P(\omega)\mathrm{d}\mu_0(\omega) \geq \int_{\Delta(\Omega)} V(\mu)\mathrm{d}\tau(\mu). \tag{G}$$

Moreover, if (G) holds with equality, that is,

$$\int_{\Omega} P(\omega)\mathrm{d}\mu_0(\omega) = \int_{\Delta(\Omega)} V(\mu)\mathrm{d}\tau(\mu), \tag{O}$$

then τ and P are optimal solutions to (P) and (D), respectively.

Proof. The proof is standard and hence relegated to Appendix B.1. □

We can also show that there is never a duality gap. The value of a problem is defined as the extremum (supremum or infimum) of the objective function over the feasible set.

Proposition 1 (No duality gap). *The problems (P) and (D) have the same value.*

Proof. See Appendix B.2. □

In the light of Proposition 1, strong duality will hold if the problems (P) and (D) have optimal solutions. Kamenica and Gentzkow (2011) show that the problem (P) always has an optimal solution and its value is equal to the concave closure of the objective function at the prior.[5] The concave closure of V at μ is defined as

$$\widehat{V}(\mu) = \inf \left\{ H(\mu) \middle| \, H : \Delta(\Omega) \to \mathbb{R}, \, H \geq V, \, H \text{ is affine and continuous} \right\}.$$

duction plans by consumption bundles, and the wholesaler by a Walrasian auctioneer who sets prices in a "Persuasion economy" to clear the market.

[5] See the Online Appendix of Kamenica and Gentzkow (2011) for the result when Ω is infinite. Kamenica and Gentzkow use a different definition of a concave closure but it can be shown to be equivalent to our definition when V is upper semi-continuous (we adopt Definition 7.4 from Aliprantis and Border, 2006).

The concave closure is a concave upper semi-continuous function. However, the problem (D) does not always have an optimal solution. So we need to impose a regularity condition.

Definition 1 (Regularity)**.** The persuasion problem is regular if $\widehat{V}$ is superdifferentiable at μ_0: There exists an affine continuous function $\overline{V} : \Delta(\Omega) \to \mathbb{R}$ such that $\overline{V}(\mu_0) = \widehat{V}(\mu_0)$ and $\overline{V}(\mu) \geq \widehat{V}(\mu)$ for all $\mu \in \Delta(\Omega)$.

We will refer to the function $\overline{V}$ defined above as the *supporting hyperplane* of $\widehat{V}$ at μ_0. By the Riesz-Markov-Kakutani representation theorem, every linear continuous function H on $\Delta(\Omega)$ can be represented as $H(\mu) = \int_\Omega P(\omega)\mathrm{d}\mu(\omega)$ for some function $P \in C(\Omega)$. Thus, $\widehat{V}$ is superdifferentiable at μ_0 if and only if $\widehat{V}$ has a *supergradient* at μ_0 defined as a function $P \in C(\Omega)$ such that $\widehat{V}(\mu) - \widehat{V}(\mu_0) \leq \int_\Omega P(\omega)\mathrm{d}(\mu - \mu_0)(\omega)$ for all $\mu \in \Delta(\Omega)$.

When Ω is finite, the persuasion problem is regular. This is because every finite-dimensional concave function is superdifferentiable at every interior point of its domain (the full-support assumption makes sure that μ_0 lies in the interior of the set of probability measures). However, when Ω is infinite, the set of probability measures has an empty (relative) interior in the weak* topology – any μ_0 is a boundary point. As a result, the hyperplane separating $(\mu_0, \widehat{V}(\mu_0))$ from the graph of $\widehat{V}$ may be vertical, and hence a supporting hyperplane of $\widehat{V}$ may fail to exist.[6] Moreover, in infinite-dimensional spaces concavity and upper semi-continuity of a function is not in general sufficient for existence of a supporting hyperplane even at interior points.[7]

In Appendix A, we show that a sufficient condition for regularity to hold is that $\widehat{V}$ has "bounded steepness" at μ_0 – the marginal increase in the value of the persuasion problem is bounded above for a small perturbation of the prior. A sufficient condition for this property is, for example, that V is Lipschitz continuous.[8]

Theorem 2 (Strong duality)**.** *The problems* (P) *and* (D) *have optimal solutions that satisfy* (O) *(that is, strong duality holds) if and only if the persuasion problem is regular.*

The proof below shows that the optimal price function can be identified with a supergradient of the concave closure of V at the prior μ_0. Geometrically, any price function P defines a hyperplane on $\Delta(\Omega)$ that passes through each extreme point $(\delta_\omega, P(\omega))$, where δ_ω is the Dirac measure at ω. The price function P is feasible for (D) if the corresponding hyperplane lies above V on $\Delta(\Omega)$. The dual problem reduces to finding a hyperplane that lies above V

[6] For example, consider a concave continuous function $f(x) = \sqrt{x}$ on $[0, 1]$. That function is not superdifferentiable at the boundary point $x = 0$ because the supporting hyperplane would have to be vertical.

[7] For example, Brøndsted and Rockafellar (1965) construct a concave upper semi-continuous function that is nowhere superdifferentiable.

[8] See Appendix A. That last condition can be significantly relaxed in more specific instances of the problem, as in Dworczak and Martini (2019) or Dizdar and Kováč (2019).

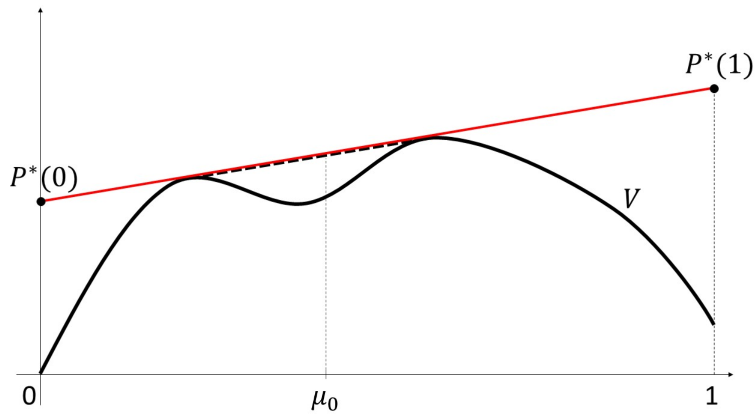

Figure 3.1: An objective function V and the optimal price function $P^\star$ in the case of a binary state, $\Omega = \{0, 1\}$.

and whose value at the prior μ_0 is minimized. Thus, the optimal hyperplane is a hyperplane that supports the concave closure of V at μ_0, and the optimal price $P^\star(\omega)$ of state ω is the value of the supporting hyperplane at δ_ω (see Figure 3.1).

Proof. If. Given that Ω is a compact metric space and V is upper semi-continuous, there exists an optimal solution $\tau^\star$ for each prior $\mu_0 \in \Delta(\Omega)$ (see the Online Appendix of Kamenica and Gentzkow, 2011, or our Appendix B.3). Moreover, Kamenica and Gentzkow show that

$$\widehat{V}(\mu_0) = \int_{\Delta(\Omega)} V(\mu) \mathrm{d}\tau^\star(\mu).$$

If the persuasion problem is regular, $\widehat{V}$ has a supporting hyperplane $\overline{V}$ at μ_0. Consider $P^\star(\omega) = \overline{V}(\delta_\omega)$. Because $\overline{V}$ is affine, for all $\mu \in \Delta(\Omega)$,

$$\overline{V}(\mu) = \int_\Omega \overline{V}(\delta_\omega) \mathrm{d}\mu(\omega) = \int_\Omega P^\star(\omega) \mathrm{d}\mu(\omega).$$

Moreover, $P^\star \in C(\Omega)$, because $\overline{V}$ is continuous in the weak* topology. Finally, because $\overline{V}$ is a supporting hyperplane of $\widehat{V}$ at μ_0,

$$\int_\Omega P^\star(\omega) \mathrm{d}\mu_0(\omega) = \widehat{V}(\mu_0) \quad \text{and} \quad \int_\Omega P^\star(\omega) \mathrm{d}\mu(\omega) \geq \widehat{V}(\mu) \text{ for all } \mu \in \Delta(\Omega),$$

showing that $P^\star$ satisfies (O) and is feasible for (D). Thus, strong duality holds by Theorem 1.

Only if. Suppose that strong duality holds and let $P^\star \in C(\Omega)$ be an optimal solution to (D). Define an affine continuous function $\overline{V} : \Delta(\Omega) \to \mathbb{R}$ by $\overline{V}(\mu) = \int_\Omega P^\star(\omega) \mathrm{d}\mu(\omega)$. Then, we know from (O) that $\overline{V}(\mu_0) = \widehat{V}(\mu_0)$. Moreover, for all $\mu \in \Delta(\Omega)$, we have

$$\overline{V}(\mu) = \int_\Omega P^\star(\omega)\mathrm{d}\mu(\omega) \geq V(\mu).$$

Because $\overline{V}$ is an affine continuous function such that $\overline{V} \geq V$, it follows from the definition of $\widehat{V}$ that $\widehat{V}(\mu) \leq \overline{V}(\mu)$ for all $\mu \in \Delta(\Omega)$. But this shows that $\overline{V}$ is a supporting hyperplane of $\widehat{V}$ at μ_0. Thus, the persuasion problem is regular. □

An important implication of duality theory is complementary slackness, which gives a simple way to verify whether a candidate solution to the primal problem is optimal. Within our interpretation, complementary slackness states that if the cost of operating a process exceeds the income it generates, then this process will not be operated.

Corollary 1 (Complementary slackness)**.** *The feasible distribution τ and the feasible price function P are optimal solutions to* (P) *and* (D)*, respectively, if and only if*

$$supp(\tau) \subseteq \left\{\mu \in \Delta(\Omega) : \int_\Omega P(\omega)\mathrm{d}\mu(\omega) = V(\mu)\right\}. \tag{C}$$

Proof. The proof is standard and hence relegated to Appendix B.4. □

Corollary 1 does not require regularity of the persuasion problem. To illustrate how it can be used, we derive the well-known condition in Kamenica and Gentzkow (2011) for the optimality of full disclosure. Under full disclosure, the support of τ is the set of all δ_ω. Condition (C) thus simplifies to the requirement that $P(\omega) = V(\delta_\omega)$ for all ω. Substituting P in the constraint of (D) yields that full disclosure is optimal if and only if $\int_\Omega V(\delta_\omega)\mathrm{d}\mu(\omega) \geq V(\mu)$ for all $\mu \in \Delta(\Omega)$.[9]

4 Moment Persuasion

We now show how our approach specializes to the case of a persuasion problem in which the objective function depends only on certain moments of the posterior.

Assume that $V(\mu) = v\left(\int_\Omega m(\omega)\mathrm{d}\mu(\omega)\right)$ for all $\mu \in \Delta(\Omega)$, some continuous $m : \Omega \to \mathbb{R}^N$, and some upper semi-continuous $v : \mathbb{R}^N \to \mathbb{R}$. Let X be the space of moments, that is, the convex hull of the set $m(\Omega)$, and let F_0 be the distribution of $x = m(\omega)$ induced by the prior μ_0 over ω.

[9]This condition, referred to as outer-convexity of V, also appears in Dworczak and Martini (2019).

Under these assumptions, a distribution τ of posterior beliefs μ matters only through the induced distribution G of moments $\mathbb{E}_\mu[m(\omega)]$. By Strassen's Theorem (see Gentzkow and Kamenica, 2016 and Kolotilin, 2018, for details), $G \in \Delta(X)$ is induced by some $\tau \in \Delta(\Delta(\Omega))$ that is feasible for (P) if and only if G is smaller than the prior F_0 in the convex order (denoted by $G \leq_{cx} F_0$). Summing up, the primal problem (P) reduces to the problem of finding a distribution of moments $G \in \Delta(X)$ to

$$\begin{aligned} &\text{maximize} \int_\Omega v(x)\mathrm{d}G(x) \\ &\text{subject to } G \leq_{cx} F_0. \end{aligned} \tag{P$_\mathrm{m}$}$$

Given that the primal problem only depends on the distribution of moments, it is natural to conjecture that it is also sufficient to find prices of moments. The key insight of this section is that prices for moments can be derived from the prices of states P by finding the "cheapest" way in which a given moment x can be generated by pooling the states:

$$p(x) = \min\left\{ \int_\Omega P(\omega)\mathrm{d}\mu(\omega) : \mu \in \Delta(\Omega) \text{ and } \int_\Omega m(\omega)\mathrm{d}\mu(\omega) = x \right\}. \tag{M}$$

In particular, we will show that the dual problem (D) reduces to the problem of finding a price function $p \in C(X)$ to

$$\begin{aligned} &\text{minimize} \int_\Omega p(x)\mathrm{d}F_0(x) \\ &\text{subject to } p(x) \geq v(x) \text{ for all } x \in X, \\ &p \text{ is convex on } X. \end{aligned} \tag{D$_\mathrm{m}$}$$

The primal and the dual problems have similar interpretations as before, with the difference that the producer operates a linear production technology with X being both the set of resources and the set of available production processes. To see why we obtain convex prices in the dual, note that the producer can always transform a measure $G \in \Delta(X)$ of resources into one unit of resource $x' = \mathbb{E}_G[x]$. If prices failed to be convex, the producer could sell at effectively higher prices by engaging in such transformations. Thus, the wholesaler may as well offer convex prices from the beginning.

Theorem 3. *If $G \in \Delta(X)$ is feasible for* (P$_\mathrm{m}$)*, $p \in C(X)$ is feasible for* (D$_\mathrm{m}$)*, and*

$$\int_X p(x)\mathrm{d}F_0(x) = \int_X v(x)\mathrm{d}G(x), \tag{O$_\mathrm{m}$}$$

then G and p are optimal solutions to ($\mathrm{P_m}$) *and* ($\mathrm{D_m}$)*, respectively.*

The problems ($\mathrm{P_m}$) *and* ($\mathrm{D_m}$) *have optimal solutions that satisfy* ($\mathrm{O_m}$) *if and only if the persuasion problem is regular.*

Proof. We prove two lemmas first.

Lemma 1. *If $p \in C(X)$ is feasible for* ($\mathrm{D_m}$)*, then $P \in C(\Omega)$ defined by $P(\omega) = p(m(\omega))$ is feasible for* (D)*.*

Proof. If $p \in C(X)$ is feasible for ($\mathrm{D_m}$), then $P \in C(\Omega)$ is feasible for (D) because

$$\int_\Omega P(\omega)\mathrm{d}\mu(\omega) = \int_\Omega p(m(\omega))\mathrm{d}\mu(\omega) \geq p\left(\int_\Omega m(\omega)\mathrm{d}\mu(\omega)\right) \geq v\left(\int_\Omega m(\omega)\mathrm{d}\mu(\omega)\right) = V(\mu). \quad \square$$

Lemma 2. *If $P \in C(\Omega)$ is feasible for* (D)*, then $p \in C(X)$ defined by* (M) *is feasible for* ($\mathrm{D_m}$)*. Moreover, $p(m(\omega)) \leq P(\omega)$ for all $\omega \in \Omega$.*

Proof. Suppose that $P \in C(\Omega)$ is feasible for (D). The minimum in the definition (M) of p is attained at some μ_x and $p \in C(X)$ because the objective is continuous in the weak* topology and the feasible set is a compact-valued continuous correspondence. For each $x \in X$,

$$p(x) = \int_\Omega P(\omega)\mathrm{d}\mu_x(\omega) \geq V(\mu_x) = v(x).$$

Moreover, for each $\lambda \in (0, 1)$, x_1, $x_2 \in X$,

$$\begin{aligned}&\lambda p(x_1) + (1-\lambda)p(x_2) = \lambda \int_\Omega P(\omega)\mathrm{d}\mu_{x_1}(\omega) + (1-\lambda)\int_\Omega P(\omega)\mathrm{d}\mu_{x_2}(\omega)\\ &= \int_\Omega P(\omega)\mathrm{d}\left(\lambda\mu_{x_1}(\omega) + (1-\lambda)\mu_{x_2}(\omega)\right) \geq \int_\Omega P(\omega)\mathrm{d}\mu_{\lambda x_1 + (1-\lambda)x_2}(\omega) = p(\lambda x_1 + (1-\lambda)x_2),\end{aligned}$$

where the inequality follows from the definition of μ_x. Thus, $p \in C(X)$ is feasible for ($\mathrm{D_m}$). Finally, by the definition of p and δ_ω, for each $\omega \in \Omega$,

$$p(m(\omega)) = \int_\Omega P(\omega')\mathrm{d}\mu_{m(\omega)}(\omega') \leq \int_\Omega P(\omega')\mathrm{d}\delta_\omega(\omega') = P(\omega). \quad \square$$

Since (P) reduces to ($\mathrm{P_m}$) under the assumptions of this section, Theorems 1 and 2 imply a version of Theorem 3 in which $p(x)$, $F_0(x)$, and ($\mathrm{D_m}$) are replaced with $P(\omega)$, $\mu_0(\omega)$, and (D) in the statement. Thus, to prove Theorem 3, it suffices to show that the problems ($\mathrm{D_m}$) and (D) have the same value and that ($\mathrm{D_m}$) has an optimal solution if and only if (D) does.

If $p \in C(X)$ is feasible for (D$_\text{m}$), then $P \in C(\Omega)$ given by Lemma 1 is feasible for (D) and

$$\int_\Omega P(\omega)\mathrm{d}\mu_0(\omega) = \int_\Omega p(m(\omega))\mathrm{d}\mu_0(\omega) = \int_X p(x)\mathrm{d}F_0(x),$$

where the first equality is by definition of P and the second by definition of F_0. Conversely, if $P \in C(\Omega)$ is feasible for (D), then $p \in C(X)$ given by Lemma 2 is feasible for (D$_\text{m}$) and

$$\int_X p(x)\mathrm{d}F_0(x) = \int_\Omega p(m(\omega))\mathrm{d}\mu_0(\omega) \leq \int_\Omega P(\omega)\mathrm{d}\mu_0(\omega).$$

Thus, the problems (D$_\text{m}$) and (D) have the same value. Moreover, if p is an optimal solution to (D$_\text{m}$), then P given by Lemma 1 is an optimal solution to (D). Finally, if P is an optimal solution to (D), then p given by Lemma 2 is an optimal solution to (D$_\text{m}$). □

In Appendix A, we show that the moment persuasion problem is regular if v is Lipschitz continuous. Other authors have identified weaker sufficient conditions on v in the case when $\omega \in \mathbb{R}$ and $m(\omega) = \omega$ (see especially Dizdar and Kováč, 2019).

5 Application: Two-Dimensional Moment Persuasion

Consider the following special case of moment persuasion. The set Ω is a compact set in $\mathbb{R}^2_+$, with a generic element denoted by $\omega = (\omega_1, \omega_2)$. The objective function is $V(\mu) = \left(\int_\Omega \omega_1 \mathrm{d}\mu(\omega)\right)\left(\int_\Omega \omega_2 \mathrm{d}\mu(\omega)\right)$, so that $m(\omega) = \omega$ and $v(x_1, x_2) = x_1 x_2$. Thus, the vector of moments $x = \mathbb{E}_\mu[m(\omega)]$ is the posterior mean $\mathbb{E}_\mu[\omega]$, the space of posterior means X is the convex hull of Ω, and the prior F_0 over X coincides with the prior μ_0 over Ω. Because the above function v is Lipschitz continuous on any compact domain, it follows from Proposition 4 in Appendix A that the persuasion problem is regular, and thus strong duality holds.

This special case has been studied by Rayo and Segal (2010) and Nikandrova and Pancs (2017) under the assumption that the set Ω is a finite set and a strictly convex curve, respectively. They derive certain necessary properties of an optimal solution. In contrast, our approach allows Ω to be any compact set and it enables characterisation of necessary and sufficient conditions for a candidate solution to be optimal.

As an illustration, we provide conditions under which it is optimal to reveal only some linear combination of ω_1 and ω_2. A simple implication of this characterization is that it is optimal to reveal $\omega_1 + \omega_2$ if the prior is symmetric around the line $\omega_2 = \omega_1$.

Proposition 2. *Let G be the distribution of posterior means induced by revelation of $\omega_2 + a\omega_1$ where $a > 0$. If the support of G belongs to a line $\{(x_1, x_2) : x_2 = ax_1 + b\}$, with $b \in \mathbb{R}$, then*

G is an optimal solution to (P_m).

Conversely, suppose additionally that Ω is convex and has a non-empty interior; then, G is an optimal solution to (P_m) *only if the support of G belongs to a line $\{(x_1, x_2) : x_2 = ax_1 + b\}$, with $b \in \mathbb{R}$.*

Proof. Define

$$\theta = \omega_2 + a\omega_1,$$
$$\Omega(\theta) = \{\omega \in \Omega : \omega_2 + a\omega_1 = \theta\},$$
$$x(\theta) = (x_1(\theta), x_2(\theta)) = \mathbb{E}[(\omega_1, \omega_2)|(\omega_1, \omega_2) \in \Omega(\theta)] = \mathbb{E}[\omega|\theta].$$

That is, revealing that $\omega_2 + a\omega_1 = \theta$ induces the posterior mean $x(\theta)$.

If. Suppose that the support of G induced by revelation of $\omega_2 + a\omega_1$ belongs to a line $\{(x_1, x_2) : x_2 = ax_1 + b\}$, with $b \in \mathbb{R}$. Thus, a state $(\omega_1, \omega_2) \in \Omega$ induces a posterior mean $(x_1(\theta), x_2(\theta))$, with $\theta = \omega_2 + a\omega_1$, that satisfies

$$x_2(\theta) = ax_1(\theta) + b \quad \text{and} \quad x_2(\theta) + ax_1(\theta) = \theta;$$

so

$$x_1(\theta) = \frac{\theta - b}{2a} \quad \text{and} \quad x_2(\theta) = \frac{\theta + b}{2}.$$

Thus, the value of (P_m) at G is

$$\int_X x_1x_2 \mathrm{d}G(x_1, x_2) = \int_\Omega \frac{(\omega_2 + a\omega_1)^2 - b^2}{4a} \mathrm{d}F_0(\omega_1, \omega_2). \tag{5.1}$$

Consider $p \in C(X)$ given by

$$p(x_1, x_2) = x_1x_2 + \frac{(x_2 - ax_1 - b)^2}{4a}.$$

The Hessian matrix, which consists of second-order partial derivatives of p, is

$$\begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix} = \begin{bmatrix} \frac{a}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2a} \end{bmatrix}.$$

Since the Hessian matrix is positive semidefinite, p is convex on X. Moreover, we have that $p(x_1, x_2) \geq x_1x_2 = v(x_1, x_2)$ for all $(x_1, x_2) \in X$. Thus, $p \in C(X)$ is feasible for (D_m).

The value of (D_m) at p is

$$\int_\Omega p(\omega_1, \omega_2)\mathrm{d}F_0(\omega_1,\omega_2) = \int_\Omega \left(\omega_1\omega_2 + \frac{(\omega_2 - a\omega_1 - b)^2}{4a}\right)\mathrm{d}F_0(\omega_1, \omega_2)$$
$$= \int_\Omega \frac{(\omega_2 + a\omega_1)^2 - b^2}{4a}\mathrm{d}F_0(\omega_1, \omega_2)$$
$$- \frac{b}{2a}\int_\Omega (\omega_2 - (a\omega_1 + b))\mathrm{d}F_0(\omega_1,\omega_2)$$
$$= \int_\Omega \frac{(\omega_2 + a\omega_1)^2 - b^2}{4a}\mathrm{d}F_0(\omega_1, \omega_2), \tag{5.2}$$

where the last equality holds because

$$\mathbb{E}[\omega_2 - (a\omega_1 + b)] = \mathbb{E}[\mathbb{E}[\omega_2 - (a\omega_1 + b)|\theta]] = \mathbb{E}[x_2(\theta) - (ax_1(\theta) + b)] = 0.$$

Equations (5.1) and (5.2) imply that (O_m) holds, and thus, by Theorem 3, G and p are optimal solutions to (P_m) and (D_m).

Only if. The strategy for the proof is to observe that optimality of G imposes tight restrictions on a candidate price functions to achieve the lower bound identified by condition (O_m). After constructing a parametric class of candidate price functions, we show that convexity of the optimal p pins down a unique candidate for the support of G, delivering the desired result. We sketch the main argument below and relegate some technical details to Appendix B.5.

Suppose that the additional assumption holds and it is optimal to reveal the value of $\omega_2 + a\omega_1$. For any $p \in C(X)$ that is feasible for (D_m), we have

$$\mathbb{E}[p(\omega)] = \mathbb{E}[\mathbb{E}[p(\omega)|\theta]] \overset{(1)}{\geq} \mathbb{E}[p(\mathbb{E}[\omega|\theta])] = \mathbb{E}[p(x(\theta))] \overset{(2)}{\geq} \mathbb{E}[v(x(\theta))],$$

where (1) holds because p is convex, and (2) holds because $p \geq v$. By Theorem 3, there exists an optimal $p \in C(X)$ that satisfies

$$\mathbb{E}[p(\omega)] = \mathbb{E}[v(x(\theta))];$$

so the inequalities (1) and (2) above hold with equality. Since the support of F_0 is a compact convex set Ω, the support of the distribution of θ induced by F_0 is an interval $\Theta = [\underline{\theta}, \overline{\theta}]$. Taking into account that p and v are continuous gives

(1) $p(\omega)$ is affine in ω on $\Omega(\theta)$ for each $\theta \in \Theta$ (so that inequality (1) holds with equality);

(2) $p(x(\theta)) = v(x(\theta))$ for each $\theta \in \Theta$ (so that inequality (2) holds with equality).

Since $p \geq v$, conditions (1) and (2) imply that, for each $\theta \in \Theta$ and $\omega \in \Omega(\theta)$, we have

$$p(\omega) = v(x(\theta)) + \partial v(x(\theta))(\omega - x(\theta)),$$

where $\partial v(x(\theta))$ is the gradient of v at $x(\theta)$. Since $v(x_1, x_2) = x_1 x_2$, we have $\partial v(x(\theta)) = (x_2(\theta),\, x_1(\theta))$, and

$$\begin{aligned} p(\omega_1,\, \omega_2) &= x_1(\theta)x_2(\theta) + x_2(\theta)(\omega_1 - x_1(\theta)) + x_1(\theta)(\omega_2 - x_2(\theta)) \\ &= \omega_1 x_2(\theta) + \omega_2 x_1(\theta) - x_1(\theta)x_2(\theta). \end{aligned}$$

Substituting $\theta = \omega_2 + a\omega_1$ and $x_2(\theta) = \theta - a x_1(\theta)$ yields, for each $\omega \in \Omega$,

$$p(\omega_1,\, \omega_2) = \omega_1\omega_2 + a(x_1(\omega_2 + a\omega_1) - \omega_1)^2. \tag{5.3}$$

The function $p(\omega)$ is convex on a convex set Ω with a non-empty interior: We show in Lemma 7 in Appendix B.5 that this can only be true when $x(\theta)$ belongs to a line $\{(x_1,\, x_2) : x_2 = a x_1 + b\}$, with $b \in \mathbb{R}$. □

In the converse of Proposition 2, it is possible to replace the assumption that Ω is convex and has a non-empty interior with a weaker assumption that the support of G is a connected set that lies almost everywhere in the interior of X. This weaker assumption allows Ω to be a curve, as in Nikandrova and Pancs (2017). In Appendix B.5, we provide examples illustrating that the assumption on Ω cannot be entirely relaxed: For example, the converse of Proposition 2 may fail if Ω is a finite set or when its convex hull has an empty interior.

6 Extension to Constrained Persuasion

A number of papers analyze a Bayesian persuasion problem with constraints (see Doval and Skreta (2018) and references therein). Our analysis extends easily to this case.

Suppose that for each $k = 1, \ldots, K$, $g_k : \Delta(\Omega) \to \mathbb{R}$ is a continuous function, and c_k is a scalar. The primal problem is now to find $\tau \in \Delta(\Delta(\Omega))$ to

$$\begin{aligned} \text{maximize} &\int_{\Delta(\Omega)} V(\mu)\mathrm{d}\tau(\mu) \\ \text{subject to} &\int_{\Delta(\Omega)} \mu \mathrm{d}\tau(\mu) = \mu_0, \\ \int_{\Delta(\Omega)} g_k(\mu)\mathrm{d}\tau(\mu) &\leq c_k \text{ for all } k = 1, \ldots, K. \end{aligned} \tag{P$_\mathrm{c}$}$$

We assume throughout that there exists a distribution τ satisfying the constraints of ($\mathrm{P_c}$).[10]

The dual problem is to find a function $P \in C(\Omega)$ and non-negative scalars P_k, for $k = 1, \ldots, K$, to

$$\begin{aligned} &\text{minimize} \int_{\Omega} P(\omega)\mathrm{d}\mu_0(\omega) + \sum_{k=1}^{K} P_k c_k \\ &\text{subject to} \int_{\Omega} P(\omega)\mathrm{d}\mu(\omega) + \sum_{k=1}^{K} P_k g_k(\mu) \geq V(\mu) \text{ for all } \mu \in \Delta(\Omega). \end{aligned} \tag{$\mathrm{D_c}$}$$

The interpretation of the primal problem is as follows. The producer now faces additional technological constraints. There are K factors of production, and the producer is endowed with c_k units of each factor k. A process $\mu \in \Delta(\Omega)$ operated at unit level consumes the measure μ of resources and $g_k(\mu)$ units of each factor k. Since the additional technological constraints take the form of inequalities, the producer can freely dispose of each factor k.

In the dual problem, the wholesaler, who wants to buy up all the resources and factors, sets a unit price $P(\omega)$ for each resource ω and a unit price P_k for each factor k. Since the producer can freely dispose of each factor k, each price P_k is non-negative.

To establish conditions for strong duality, we follow Doval and Skreta (2018), and define a function $V^{g_1,\ldots,g_k} : \Delta(\Omega) \times \mathbb{R}^K \to \mathbb{R}$ by

$$V^{g_1,\ldots,g_k}(\mu, x) = \begin{cases} V(\mu), & \text{if } g_k(\mu) \leq x_k \text{ for } k = 1, \ldots, K, \\ -\infty, & \text{otherwise.} \end{cases}$$

Then, as Doval and Skreta show,[11] the value of the problem ($\mathrm{P_c}$) – when the prior is μ_0 and the right hand side of the constraints in ($\mathrm{P_c}$) is given by a vector $c \in \mathbb{R}^K$ – is equal to the concave closure of the function $V^{g_1,\ldots,g_k}$ at (μ_0, c), denoted $\widehat{V}^{g_1,\ldots,g_k}(\mu_0, c)$. The function $\widehat{V}^{g_1,\ldots,g_k}$ is concave and upper semi-continuous. We present the following result without a proof as it follows from the same reasoning as in Theorems 1 and 2.

Theorem 4. *If τ is feasible for ($\mathrm{P_c}$) and $(P, P_1, \ldots, P_K)$ is feasible for ($\mathrm{D_c}$), and*

$$\int_{\Omega} P(\omega)\mathrm{d}\mu_0(\omega) + \sum_{k=1}^{K} P_k c_k = \int_{\Delta(\Omega)} V(\mu)\mathrm{d}\tau(\mu), \tag{$\mathrm{O_c}$}$$

[10] While problem ($\mathrm{P_c}$) only contains inequality constraints, an equality constraint of the form $\int_{\Delta(\Omega)} g(\mu)\mathrm{d}\tau(\mu) = c$ can be restated as the two inequality constraints $\int_{\Delta(\Omega)} g(\mu)\mathrm{d}\tau(\mu) \leq c$ and $\int_{\Delta(\Omega)} (-g(\mu))\mathrm{d}\tau(\mu) \leq -c$.

[11] Formally, they work with a finite state space but their proof technique extends to the case of a compact metric state space.

then τ and P are optimal solutions to ($\mathrm{P_c}$) *and* ($\mathrm{D_c}$)*, respectively.*

The problems ($\mathrm{P_c}$) *and* ($\mathrm{D_c}$) *have optimal solutions that satisfy* ($\mathrm{O_c}$) *if and only if* $\widehat{V}^{g_1,\ldots,g_k}$ *is superdifferentiable at* (μ_0, c)*.*

Concavity of $\widehat{V}^{g_1,\ldots,g_k}$ implies that strong duality holds when Ω is finite.

7 Relationship to Existing Duality Results

Our approach builds on and extends the geometric duality of Gale (1967). Gale shows that the Kuhn-Tucker Theorem holds for a convex optimization problem in a normed space if and only if the value function is superdifferentiable at the primal constraint. The persuasion problem is a linear-programming problem on the space of probability measures, $\Delta(\Omega)$. Importantly, we cannot endow $\Delta(\Omega)$ with a norm because the weak* topology (which we need to obtain prices by applying the representation theorem for linear functionals) is not normable when Ω is infinite. Thus, Gale's result cannot be applied directly in our setting. Instead, we show that his approach can be extended to the metric space $\Delta(\Omega)$. Gale also shows that a concave function defined on a normed space is superdifferentiable at a point if and only if it has bounded steepness at this point. In Appendix A, we generalize the "if" part of this result to the metric space $\Delta(\Omega)$ and show (by means of a counterexample) that the "only if" part of the result does not extend.

The results of this paper generalize some existing results in the persuasion literature. The linear persuasion problem in which the objective function depends only on the expected one-dimensional state has received special attention (see, for example, Gentzkow and Kamenica, 2016, Kolotilin et al., 2017, Kolotilin, 2018, Dworczak and Martini, 2019, and Dizdar and Kováč, 2019). Linear persuasion is the special case of moment persuasion when $\omega \in \mathbb{R}$ and $m(\omega) = \omega$. It is important to emphasize, however, that while our main duality result is mathematically more general than the existing ones (in the sense that it applies on a larger domain of problems), previous papers identify useful (easier to verify) regularity conditions on the primitives under which strong duality holds. Our regularity condition is necessary for strong duality so it implies all existing sufficient conditions but verifying it may be difficult in practice.

When the persuasion problem is regular, Theorem 3 generalizes Theorems 1 and 2 in Dworczak and Martini (2019): By a simple transformation, Theorem 3 establishes existence of a convex continuous function $p^\star$ and a distribution $G^\star$ such that

$$\operatorname{supp}(G^\star) \subseteq \{x \in X : p^\star(x) = v(x)\},$$
$$\int_\Omega p^\star(x) \mathrm{d}F_0(x) = \int_\Omega p^\star(x) \mathrm{d}G^\star(x),$$
$$G^\star \leq_{cx} F_0, \text{ and } p^\star \geq v.$$

Moreover, the theorem resolves (positively) the conjecture of Dworczak and Martini that if the objective function V is measurable with respect to a moment $m(\omega)$, then so is the corresponding price function.

To date, Dizdar and Kováč (2019) identify the most permissive condition on v for the linear persuasion problem to be regular (and hence for strong duality to hold). Our proof of no-duality gap (Proposition 1) uses similar techniques to theirs but the proof strategy for strong duality is different: Dizdar and Kováč show that the problem (D) has an optimal solution by demonstrating that feasible solutions can be restricted to a compact set; in contrast, we construct the optimal solution (a price function on the space of moments) from the supergradient of the concave closure of V.

Our approach also allows us to simplify the duality result for the general persuasion problem in Dworczak and Martini (2019): They define a price function on the space of beliefs $\Delta(\Omega)$ and require it to be "outer-convex" (a relaxation of convexity). Our results demonstrate that such a price function exists (strong duality holds) when the problem is regular (for example, when V is Lipschitz continuous), and that the price can in fact be taken to be *linear* by extending (linearly) the price function P defined on the state space Ω.

Kolotilin (2018) and Galperti and Perego (2018) use an alternative approach to the persuasion problem. Instead of working with an objective function $V : \Delta(\Omega) \to \mathbb{R}$, they consider a sender and a receiver whose utility functions are $w : \Omega \times A \to \mathbb{R}$ and $u : \Omega \times A \to \mathbb{R}$ where A is the space of the receiver's actions. Moreover, instead of choosing a distribution of posterior beliefs $\tau \in \Delta(\Delta(\Omega))$, the sender chooses a joint distribution $\pi \in \Delta(\Omega \times A)$ of the state ω and the recommended action a. On top of the Bayes plausibility constraint, π must satisfy the obedience constraint, which requires each recommended action to be incentive-compatible for the receiver given the beliefs it induces. Our approach is simpler because it sidesteps the obedience constraint, but the alternative approach can be more easily extended to allow for a privately informed receiver and multiple receivers.

Generally, it is possible to reformulate the alternative problem as our problem, and vice versa. To illustrate this point suppose that Ω is a finite set. Consider an alternative problem in which the sender's and receiver's utility functions are $w(\omega, a)$ and $u(\omega, a)$. Kamenica and Gentzkow (2011) show that this problem is equivalent to our problem in which the objective function is $V(\mu) = \mathbb{E}_\mu[w(\omega, a^\star(\mu))]$ with $a^\star(\mu) \in \arg\max_a \mathbb{E}_\mu[u(\omega, a)]$. Conversely, consider our problem in which the objective function is V. This problem is equivalent to the alternative problem in which the action space is $A = \Delta(\Omega)$, and the sender's and receiver's utility functions are $w(\omega, a) = V(a)$ and $u(\omega, a) = 2a(\omega) - \sum_{\omega' \in \Omega} a^2(\omega')$. Indeed, given a posterior μ, the receiver takes an action $a^\star(\mu) = \mu$, which maximizes his expected utility $\sum_{\omega \in \Omega}(2a(\omega)\mu(\omega) - a^2(\omega))$, and thus the objective function is $V(\mu)$.

To compare our and alternative approaches, consider the linear persuasion problem. Let $A = \Omega$. The alternative dual problem in Kolotilin (2018) is to find a function $q \in C(\Omega)$ and a bounded measurable function $r : A \to \mathbb{R}$ to

$$\begin{aligned} &\text{minimize} \int_\Omega q(\omega) \mathrm{d}\mu_0(\omega) \\ &\text{subject to } q(\omega) + r(a)(a - \omega) \geq v(a) \text{ for all } (\omega, a) \in \Omega \times A, \end{aligned} \tag{D$_\mathrm{a}$}$$

where q and r are multipliers for the Bayes plausibility and obedience constraints. Observe that if $p^\star$ is an optimal solution to (D$_\mathrm{m}$), then $(q^\star, r^\star) = (p^\star, \partial p^\star)$ is an optimal solution to (D$_\mathrm{a}$). This is so because $(q^\star, r^\star)$ yields the same value and is feasible, as follows from

$$p^\star(\omega) + \partial p^\star(a)(a - \omega) \geq p^\star(a) \geq v(a),$$

where the first inequality is by convexity of $p^\star$ and the second is by $p^\star \geq v$. Thus, q has the same interpretation as p, but an economic interpretation of r remains elusive. Finally, using Theorem 3, it is straightforward to reproduce the conditions in Kolotilin's Proposition 3 for "interval revelation" to be optimal.

Appendix

A Regularity of the Persuasion Problem

In this appendix, we present various sufficient conditions under which the persuasion problem is regular. We will denote by $M(\Omega)$ the set of signed Borel measures of bounded variation on Ω, and by $M^+(\Omega)$ the subset of non-negative measures. Weak convergence of (signed) measures μ_n to μ will be denoted by $\mu_n \xrightarrow{w} \mu$.

Let $d : M(\Omega) \times M(\Omega) \to \mathbb{R}$ denote the Kantorovich-Rubinstein distance between measures: For any $\mu,\, \eta \in M(\Omega)$,

$$d(\mu,\, \eta) = \sup\left\{\int_\Omega f(\omega)\mathrm{d}(\mu-\eta)(\omega) :\ f \in Lip_1(\Omega),\, \sup_{\omega\in\Omega}|f(\omega)| \le 1\right\},$$

where $Lip_1(\Omega)$ is the set of 1-Lipschitz functions on Ω. Because Ω is assumed to be a compact metric space (hence separable), d is a metric that metrizes weak* topology on the space of non-negative measures $M^+(\Omega)$ (by Theorem 8.3.2. in Bogachev, 2007).

We say that $\widehat{V} :\ \Delta(\Omega) \to \mathbb{R}$ has *bounded steepness* at μ_0 if there exists a positive constant L such that, for all $\mu \in \Delta(\Omega)$,

$$\frac{\widehat{V}(\mu) - \widehat{V}(\mu_0)}{d(\mu,\, \mu_0)} \le L.$$

Intuitively, bounded steepness rules out the possibility that the function $\widehat{V}$ has an unbounded slope from μ_0 to μ. We show that this condition is sufficient to obtain a regular persuasion problem. The proof is complicated by the fact there exist "vertical" hyperplanes separating the graph of $\widehat{V}$ from the point $(\mu_0,\, \widehat{V}(\mu_0))$: For example, a linear continuous functional $H :\ M(\Omega)\times\mathbb{R} \to \mathbb{R}$ given by $H(\mu, \alpha) = \mu(\Omega)$ is a separating hyperplane but not a proper one. We circumvent this difficulty by using the following "trick:" we first extend the domain of the function $\widehat{V}$ to non-probability measures (this is accomplished in Lemma 4 below, with the aid of Lemma 3), and only then apply the relevant version of the hyperplane separation theorem (specifically, we use a result about separating convex cones from their vertices).

Theorem 5. *If $\widehat{V}$ has bounded steepness at μ_0, then the persuasion problem is regular.*

Proof. Define the space $\Delta_2(\Omega) := \{\kappa\mu :\ \mu \in \Delta(\Omega),\, \kappa \in [0,\, 2]\}$ of non-negative measures that assign a weight of at most 2 to the space Ω. We prove three lemmas first.

Lemma 3. *Suppose that μ and η are probability measures. Then, for any $\lambda \in \mathbb{R}$, $d(\mu, \lambda\eta) \ge \frac{1}{2}d(\mu,\, \eta)$.*

Proof of Lemma 3. Suppose that $\delta = d(\mu,\, \eta)$. Since Ω is compact, the Arzelá-Ascoli theorem implies that the supremum in the definition of d is attained at some function $f \in Lip_1(\Omega)$, $\sup_{\omega\in\Omega}|f(\omega)| \le 1$ such that $\delta = \int_\Omega f(\omega)\mathrm{d}(\mu-\eta)(\omega)$. It is enough to construct a function $g \in Lip_1(\Omega)$, $\sup_{\omega\in\Omega}|g(\omega)| \le 1$ such that $\int_\Omega g(\omega)\mathrm{d}(\mu-\lambda\eta)(\omega) \ge \delta/2$. Let

$$g(\omega) = \frac{1}{2}\left(f(\omega) - \int_\Omega f(\omega')\mathrm{d}\eta(\omega')\right).$$

Because η is a probability measure and $|f(\omega)| \le 1$ for all ω, the function g is indeed feasible.

Moreover, we have

$$\int_\Omega g(\omega)\mathrm{d}(\mu-\lambda\eta)(\omega) = \int_\Omega g(\omega)\mathrm{d}(\mu-\eta)(\omega)-(\lambda-1)\underbrace{\int_\Omega g(\omega)\mathrm{d}\eta(\omega)}_{=0} = \int_\Omega \frac{f(\omega)}{2}\mathrm{d}(\mu-\eta)(\omega) = \frac{\delta}{2},$$

where the second to last inequality holds because μ and η are probability measures. □

Lemma 4. *There exists an extension of $\widehat{V}$ to $\Delta_2(\Omega)$ that is concave and has bounded steepness at μ_0.*

Proof of Lemma 4. Extend $\widehat{V}$ to $\Delta_2(\Omega)$ by

$$\widehat{V}(\eta) = \eta(\Omega)\widehat{V}\left(\frac{\eta}{\eta(\Omega)}\right),$$

and $\widehat{V}(0) = 0$. It follows (by directly checking the definition) that $\widehat{V}$ remains concave. We will show that $\widehat{V}$ has bounded steepness at μ_0 with a constant $2L+\bar{V}$, where $\bar{V}$ is the upper bound on $\widehat{V}$ on $\Delta(\Omega)$ (which exists because $\widehat{V}$ is upper-semicontinuous).

Let $\eta_n \xrightarrow{w} \mu_0$. We can represent $\eta_n = \lambda_n\mu_n$, where $\lambda_n = \eta_n(\Omega)$, and $\mu_n = \eta_n/\lambda_n \in \Delta(\Omega)$. When looking at the sequence $\widehat{V}(\mu_n)-\widehat{V}(\mu_0)$, it is without loss of generality to assume that $\widehat{V}(\mu_n) \geq \widehat{V}(\mu_0)$ for all n (by passing to a subsequence if necessary, as otherwise the bound we are trying to prove is trivially satisfied). By Lemma 3, we have

$$\frac{\widehat{V}(\eta_n)-\widehat{V}(\mu_0)}{d(\eta_n,\,\mu_0)} = \frac{\widehat{V}(\mu_n)-\widehat{V}(\mu_0)}{d(\eta_n,\,\mu_0)} + \frac{\lambda_n-1}{d(\eta_n,\,\mu_0)}\widehat{V}(\mu_n) \leq \frac{\widehat{V}(\mu_n)-\widehat{V}(\mu_0)}{\frac{1}{2}d(\mu_n,\,\mu_0)} + \frac{|\lambda_n-1|}{d(\eta_n,\,\mu_0)}\bar{V}.$$

Note that $d(\eta_n,\,\mu_0) \geq |\lambda_n-1|$ because we can always set $f=1$ and $f=-1$ in the definition of the metric d, and hence the right hand side converges to a number $L' \leq 2L+\bar{V}$. □

Lemma 5. *If μ_n is a sequence in $M(\Omega)$, and $\mu_n \xrightarrow{w} 0$, then $d(\mu_n,\,0) \to 0$.*

Proof of Lemma 5. Suppose not; then there exists $\epsilon > 0$ and a sequence of functions $f_n \in Lip_1(\Omega)$, $\sup_{\omega\in\Omega}|f_n(\omega)| \leq 1$ such that $\int_\Omega f_n(\omega)\mathrm{d}\mu_n(\omega) > \epsilon$. By the Arzelá-Ascoli theorem, the space of such functions on a compact domain is compact, so we can choose a subsequence of f_n that converges uniformly to some f (that is in particular continuous). We have

$$\int_\Omega f(\omega)\mathrm{d}\mu_n(\omega) = \int_\Omega f_n(\omega)\mathrm{d}\mu_n(\omega) + \int_\Omega (f(\omega)-f_n(\omega))\mathrm{d}\mu_n(\omega) \geq \epsilon/2,$$

for large enough n since the second term converges to 0. This directly contradicts the definition of μ_n converging to 0 in the weak* topology. □

We note that Lemma 5 has a converse only on the space of non-negative measures (because d metrizes weak* topology on that space but not on $M(\Omega)$).

Now we can finish the proof of Theorem 5. Define

$$A = \mathrm{hyp}(\widehat{V}) = \left\{(\mu, \alpha) : \, \alpha \leq \widehat{V}(\mu), \, \mu \in \Delta_2(\Omega)\right\}$$

and

$$B = \left\{(\mu_0, \, \widehat{V}(\mu_0)\right\}.$$

According to Lemma 7.20 in Aliprantis and Border, separating A and B is possible if the closed convex cone generated by the set $A - B$ is not equal to the entire space $M(\Omega) \times \mathbb{R}$. That is, we have to show that the convex closure of the set

$$\left\{\lambda(\mu - \mu_0, \alpha - \widehat{V}(\mu_0)) : \, \lambda \geq 0, \, \alpha \leq \widehat{V}(\mu), \, \mu \in \Delta_2(\Omega)\right\}$$

is not equal to $M(\Omega) \times \mathbb{R}$. Notice that this set is already a convex cone, so separation is possible if there exists some $\eta \in M(\Omega)$ and $\beta \in \mathbb{R}$ such that there is no sequence $\lambda_n \geq 0$, $\alpha_n \leq \widehat{V}(\mu_n)$, $\mu_n \in \Delta_2(\Omega)$ such that $\lambda_n(\mu_n - \mu_0, \alpha_n - \widehat{V}(\mu_0)) \stackrel{w}{\to} (\eta, \, \beta)$. Take $(\eta, \, \beta) = (0, \, 1)$, and, towards a contradiction, suppose that such a sequence exists. In particular, we must have $\lambda_n(\mu_n - \mu_0) \stackrel{w}{\to} 0$. By Lemma 4, there exists L' such that

$$\frac{\alpha_n - \widehat{V}(\mu_0)}{d(\mu_n, \, \mu_0)} \leq \frac{\widehat{V}(\mu_n) - \widehat{V}(\mu_0)}{d(\mu_n, \, \mu_0)} \leq L',$$

and therefore,

$$\begin{aligned}\lim_{n\to\infty} \lambda_n(\alpha_n - \widehat{V}(\mu_0)) &= \lim_{n\to\infty} \frac{\alpha_n - \widehat{V}(\mu_0)}{d(\mu_n, \, \mu_0)} \lambda_n d(\mu_n, \, \mu_0) \\ &\leq L' \lim_{n\to\infty} \lambda_n d(\mu_n, \, \mu_0) = L' \lim_{n\to\infty} d(\lambda_n \mu_n, \, \lambda_n \mu_0) = 0,\end{aligned}$$

by the properties of d and Lemma 5. This is a contradiction.

Therefore, sets A and B can be separated: There exists a non-zero linear continuous functional on $M(\Omega) \times \mathbb{R}$, denoted $(\Phi, \, \phi)$, that achieves its maximum over the set A at $(\mu_0, \, \widehat{V}(\mu_0))$, in particular,

$$\Phi\mu_0 + \phi\widehat{V}(\mu_0) \geq \Phi\mu + \phi\widehat{V}(\mu), \forall \mu \in \Delta_2(\Omega). \tag{A.1}$$

Suppose that $\phi = 0$. Because a linear continuous functional in the weak* topology can be identified with a continuous function on Ω (by the Riesz-Markov-Kakutani theorem), we have some $f \in C(\Omega)$, not identically 0, such that

$$\int_\Omega f(\omega)\mathrm{d}\mu_0(\omega) \geq \int_\Omega f(\omega)\mathrm{d}\mu(\omega),\ \forall \mu \in \Delta_2(\Omega). \tag{A.2}$$

By considering $\mu = \frac{1}{2}\mu_0$ and $\mu = 2\mu_0$, we conclude that $\int_\Omega f(\omega)d\mu_0(\omega) = 0$. We must also have $f(\omega) \leq 0$, as otherwise there would exist a measure μ with $\int_\Omega f(\omega)d\mu(\omega) > 0$, contradicting (A.2). Hence, we have

$$0 = \int_\Omega f(\omega)\mathrm{d}\mu_0(\omega) = \int_{\{\omega:\, f(\omega)<0\}} f(\omega)\mathrm{d}\mu_0(\omega) \implies \mu_0\left(\{\omega : f(\omega) < 0\}\right) = 0$$

which is a contradiction when μ_0 has full support because the set $\{\omega : f(\omega) < 0\}$ is non-empty and open. Thus, we conclude that $\phi \neq 0$. This implies that $\phi > 0$ since the set A is unbounded from below in α.

Therefore, we can divide the two sides of equation (A.1) by ϕ to obtain

$$\widehat{V}(\mu_0) - \left(\frac{\Phi}{\phi}\right)\mu + \left(\frac{\Phi}{\phi}\right)\mu_0 \geq \widehat{V}(\mu), \forall \mu \in \Delta_2(\Omega).$$

Define $\overline{V}(\mu) = \widehat{V}(\mu_0) - \left(\frac{\Phi}{\phi}\right)\mu + \left(\frac{\Phi}{\phi}\right)\mu_0$ to be an affine continuous function on $\Delta(\Omega)$. We have that $\overline{V}(\mu_0) = \widehat{V}(\mu_0)$ and $\overline{V}(\mu) \geq \widehat{V}(\mu)$ for all $\mu \in \Delta(\Omega)$. Thus, $\overline{V}$ is a supporting hyperplane of $\widehat{V}$ at μ_0. □

Gale (1967) shows that the converse of Theorem 5 holds on normed spaces. However, as we demonstrate next by means of an example, the converse to Theorem 5 does not hold in our setting with a metric space.

Example 1 (Regularity of the persuasion problem does not imply bounded steepness of $\widehat{V}$)**.** Let $\Omega = [0, 1]$, $\mu_0 = \delta_0/2 + \lambda/2$ where λ is the Lebesgue measure on $[0, 1]$, and $\widehat{V}(\mu) = \int_{[0,1]} f(\omega)\mathrm{d}\mu(\omega)$ where $f(\omega) = \sqrt{\omega}$. The function $\widehat{V}$ is affine and continuous (since f is continuous), and thus it is concave and upper semi-continuous. Moreover, $\widehat{V}$ is superdifferentiable everywhere (with the supporting hyperplane $\overline{V} = \widehat{V}$); so the persuasion problem is regular.

Consider $\mu_n = \delta_{1/n}/2 + \lambda/2$. By the definitions of weak* convergence and the Kantorovich-Rubinstein metric, we have $\mu_n \xrightarrow{w} \mu_0$ and $d(\mu_n, \mu_0) = 1/(2n)$. Since $\widehat{V}(\mu_n) - \widehat{V}(\mu_0) = \sqrt{1/n}/2$, we have

$$\frac{\widehat{V}(\mu_n) - \widehat{V}(\mu_0)}{d(\mu_n, \mu_0)} = \sqrt{n} \to \infty,$$

showing that $\widehat{V}$ does not have bounded steepness at μ_0.

We now provide a simple sufficient condition on the primitive function V for $\widehat{V}$ to have bounded steepness.

Proposition 3. *If V is Lipschitz continuous, then $\widehat{V}$ has bounded steepness at μ_0.*

Proof. Recall that

$$\widehat{V}(\mu_0) = \inf\{H(\mu_0) |\, H : \Delta(\Omega) \to \mathbb{R},\, H \geq V,\, H \text{ is affine and continuous}\}.$$

Because the infimum is taken over affine functions that lie everywhere above V, when V is L-Lipschitz continuous, it is without loss of generality to only consider affine functions that are also L-Lipschitz continuous:

$$\widehat{V}(\mu_0) = \inf\{H(\mu_0) |\, H \geq V,\, H \text{ is affine and } L\text{-Lipschitz continuous}\}.$$

It now follows from the Arzelá-Ascoli theorem that the infimum is attained at some $H^\star$. Clearly, $H^\star(\mu_0) = \widehat{V}(\mu_0)$ and $H^\star(\mu) \geq \widehat{V}(\mu)$ for any μ. We thus have, for any μ,

$$\frac{\widehat{V}(\mu) - \widehat{V}(\mu_0)}{d(\mu,\, \mu_0)} \leq \frac{H^\star(\mu) - H^\star(\mu_0)}{d(\mu,\, \mu_0)} \leq L$$

which shows that $\widehat{V}$ has bounded steepness at μ_0. □

Finally, we use Proposition 3 to show that a sufficient condition for V to be Lipschitz continuous in the moment persuasion problem is that v is Lipschitz continuous.

Proposition 4. *In the setting of Section 4, if $v : X \to \mathbb{R}$ is Lipschitz continuous, then the persuasion problem is regular.*

Proof. Since we can redefine the state as $m(\omega)$ without violating any assumption in Sections 2 and 4, it suffices to prove the lemma for the case $m(\omega) = \omega$. Recall that X denotes the space of moments, which, under the above normalization, is equal to the convex hull of Ω.

Suppose that v is L-Lipschitz. By Proposition 3 and Theorem 5, it is enough to prove that the function $V(\mu) = v(\mathbb{E}_\mu[\omega])$ is Lipshitz continuous. We have, for any $\mu,\, \eta \in \Delta(\Omega)$, where ρ denotes a metric on X,

$$\frac{V(\mu) - V(\eta)}{d(\mu,\, \eta)} = \frac{v(\mathbb{E}_\mu[\omega]) - v(\mathbb{E}_\eta[\omega])}{d(\mu,\, \eta)} = \frac{v(\mathbb{E}_\mu[\omega]) - v(\mathbb{E}_\eta[\omega])}{\rho(\mathbb{E}_\mu[\omega],\, \mathbb{E}_\eta[\omega])} \frac{\rho(\mathbb{E}_\mu[\omega],\, \mathbb{E}_\eta[\omega])}{d(\mu,\, \eta)} \leq L \frac{\rho(\mathbb{E}_\mu[\omega],\, \mathbb{E}_\eta[\omega])}{d(\mu,\, \eta)}.$$

Because we have assumed that X is a subset of a finite-dimensional Euclidean space, we can without loss of generality take ρ to be the uniform metric:

$$\rho(\mathbb{E}_\mu[\omega],\, \mathbb{E}_\eta[\omega]) = \max_{i=1,\dots,N} \left|\mathbb{E}_\mu[\omega_i] - \mathbb{E}_\eta[\omega_i]\right|.$$

Because the function $f(\omega) = \omega_i$ is bounded and Lipschitz continuous on the compact domain Ω, there exists a constants L' such that

$$\left|\mathbb{E}_\mu[\omega_i] - \mathbb{E}_\eta[\omega_i]\right| = \left|\int_\Omega \omega_i \mathrm{d}(\mu - \eta)(\omega)\right| \leq L' d(\mu,\, \eta).$$

Therefore,

$$\rho(\mathbb{E}_\mu[\omega],\, \mathbb{E}_\eta[\omega]) \leq L' d(\mu,\, \eta),$$

and thus V is Lipschitz continuous with a constant $L \cdot L'$. □

B Omitted Proofs

B.1 Proof of Theorem 1

To prove the first part, note that the constraint in (D) implies

$$\int_{\Delta(\Omega)} \int_\Omega P(\omega)\mathrm{d}\mu(\omega)\mathrm{d}\tau(\mu) \geq \int_{\Delta(\Omega)} V(\mu)\mathrm{d}\tau(\mu).$$

The constraint in (P) and Fubini's theorem imply

$$\int_{\Delta(\Omega)} \int_\Omega P(\omega)\mathrm{d}\mu(\omega)\mathrm{d}\tau(\mu) = \int_\Omega P(\omega)\mathrm{d}\mu_0(\omega).$$

Combining these two conditions gives (G).

We now prove the second part. Consider any other feasible $\widetilde{\tau} \in \Delta(\Delta(\Omega))$. By the first part of the theorem,

$$\int_{\Delta(\Omega)} V(\mu)\mathrm{d}\widetilde{\tau}(\mu) \leq \int_\Omega P(\omega)\mathrm{d}\mu_0(\omega).$$

Combining this inequality with (O) gives

$$\int_{\Delta(\Omega)} V(\mu)\mathrm{d}\widetilde{\tau}(\mu) \leq \int_{\Delta(\Omega)} V(\mu)\mathrm{d}\tau(\mu),$$

showing that τ is an optimal solution to (P).

Similarly, consider any other feasible $\widetilde{P} \in C(\Omega)$. By the first part of the theorem,

$$\int_\Omega \widetilde{P}(\omega)\mathrm{d}\mu_0(\omega) \geq \int_{\Delta(\Omega)} V(\mu)\mathrm{d}\tau(\mu).$$

Combining this inequality with (O) gives

$$\int_\Omega \widetilde{P}(\omega)\mathrm{d}\mu_0(\omega) \geq \int_\Omega P(\omega)\mathrm{d}\mu_0(\omega),$$

showing that P is an optimal solution to (D).

B.2 Proof of Proposition 1

Suppose first that the objective function V is continuous. Then, the conclusion follows immediately from standard results by treating the problem (D) as the primal problem.[12] Indeed, if (D) is treated as a primal minimization problem in the space of continuous functions on a compact domain Ω endowed with the topology of uniform convergence, then it follows that *(i)* the problem (P) is the topological dual of (D), and *(ii)* for any μ_0, there exists a P feasible for (D) such that $\int_\Omega P(\omega)\mathrm{d}\mu(\omega)$ lies in the interior of the feasible set, that is, is bounded away from $V(\mu)$ uniformly (because V is upper semi-continuous, and thus bounded by some constant K, it is enough to take $P(\mu) = K + \epsilon$, for all μ, for some $\epsilon > 0$). The proposition follows from Theorem 3.13 in Anderson and Nash (1987).

For the case when V is only upper semi-continuous, we can use an approximation argument developed by Villani (2009) and adapted to the linear persuasion problem by Dizdar and Kováč (2019). Because we study a general persuasion problem and because (D) need not admit an optimal solution even for a continuous V, we present the complete argument below.

By Tong (1952), using the fact that the space $\Delta(\Omega)$ is perfectly normal (because it is a metric space, see p. 45 of Aliprantis and Border, 2006), we can approximate the upper semi-continuous function V by a monotonically non-increasing sequence V_k of continuous functions (in particular $V_k \geq V$ for all k). For every k, we know from the first part of the proof that there is no duality gap. Using the fact that the problem (P) with objective V_k has an optimal solution $\tau_k^\star$, this means that there exists a sequence of function P_k^j feasible

[12] Dizdar and Kováč (2019) were first to reverse the roles of the primal and dual problems to show that there is no duality gap in the linear persuasion problem (in which the objective function depends only on the expected one-dimensional state).

for (D) such that

$$\int_{\Delta(\Omega)} V_k(\mu)\mathrm{d}\tau_k^\star(\mu) = \lim_{j\to\infty}\int_\Omega P_k^j(\omega)\mathrm{d}\mu_0(\omega).$$

Because the sequence $\{\tau_k^\star\}$ lives in a compact set, we can assume without loss of generality that $\tau_k^\star \to \tau^\star$ in the weak topology. Then, we have, for any l,

$$\begin{aligned}\int_{\Delta(\Omega)} V_l(\mu)\mathrm{d}\tau^\star(\mu) &= \lim_{k\to\infty}\int_{\Delta(\Omega)} V_l(\mu)\mathrm{d}\tau_k^\star(\mu)\\ &\geq \limsup_{k\to\infty}\int_{\Delta(\Omega)} V_k(\mu)\mathrm{d}\tau_k^\star(\mu) = \limsup_{k\to\infty}\lim_{j\to\infty}\int_\Omega P_k^j(\omega)\mathrm{d}\mu_0(\omega),\end{aligned}$$

where we have used the fact that $V_k \leq V_l$ for $k \geq l$. By the monotone convergence theorem,

$$\lim_{l\to\infty}\int_{\Delta(\Omega)} V_l(\mu)\mathrm{d}\tau^\star(\mu) = \int_{\Delta(\Omega)} V(\mu)\mathrm{d}\tau^\star(\mu),$$

which gives us

$$\int_{\Delta(\Omega)} V(\mu)\mathrm{d}\tau^\star(\mu) \geq \limsup_{k\to\infty}\lim_{j\to\infty}\int_\Omega P_k^j(\omega)\mathrm{d}\mu_0(\omega).$$

But because $\int_\Omega P_k^j(\omega)\mathrm{d}\mu(\omega) \geq V_k(\mu) \geq V(\mu)$, each function P_k^j is feasible for the problem (D). Thus, by weak duality, we also have for any j and k,

$$\int_\Omega P_k^j(\omega)\mathrm{d}\mu_0(\omega) \geq \int_{\Delta(\Omega)} V(\mu)\mathrm{d}\tau^\star(\mu).$$

Therefore, there is no duality gap.

B.3 Proof of Existence of an Optimal Solution to the Problem (P)

Because Ω is a compact metric space, by Theorem 15.11 in Aliprantis and Border (2006), the space $\Delta(\Omega)$ is also compact and metrizable in the weak* topology. By the same argument, $\Delta(\Delta(\Omega))$ is compact in the weak* topology. Because the function $\tau \to \int_{\Delta(\Omega)} \mu \mathrm{d}\tau(\mu)$ is continuous, the set of distributions that satisfy the constraint in (P) is closed. A closed subset of a compact space is compact, and thus the set of feasible measures τ is compact.

By one of (equivalent) definitions of convergence in the weak* topology, the function $U(\tau) = \int_{\Delta(\Omega)} V(\mu)\mathrm{d}\tau(\mu)$ is upper semi-continuous whenever V is upper semi-continuous and bounded from above. By assumption, $V(\mu)$ is upper semi-continuous in the weak* topology on $\Delta(\Omega)$, and it is bounded from above because $\Delta(\Omega)$ is compact.

Therefore, by Weierstrass theorem, U admits a maximum on the set of feasible measures.

B.4 Proof of Corollary 1

If (C) holds, and τ satisfies the constraint of (P), then (O) holds. Thus, τ and P are optimal solutions, by Theorem 1. Conversely, if τ and P are optimal solutions, then, by Proposition 1, (O) holds, which, together with the constraint of (P), gives

$$\int_{\Delta(\Omega)} \left(\int_{\Omega} P(\omega) \mathrm{d}\mu(\omega) - V(\mu) \right) \mathrm{d}\tau(\mu) = 0.$$

Since P is feasible for (D), the integrand is non-negative and thus (C) follows.

B.5 The Converse of Proposition 2

We maintain the additional assumption imposed in the converse of Proposition 2. We also maintain the notation used in the proof of Proposition 2.

Lemma 6. *The support of G is ordered: for all x, $x' \in supp(G)$, either $x' \geq x$ or $x \geq x'$.*

Proof. Suppose that there exist unordered x, $x' \in \operatorname{supp}(G)$ that have probabilities β, $\beta' > 0$, so that either $(x'_1, -x'_2) > (x_1, -x_2)$ or $(-x'_1, x'_2) > (-x_1, x_2)$. Then, we can strictly increase the objective by pooling x and x', because

$$(\beta + \beta') \frac{\beta x_1 + \beta' x'_1}{\beta + \beta'} \frac{\beta x_2 + \beta' x'_2}{\beta + \beta'} - \beta x_1 x_2 - \beta' x'_1 x'_2 = -\frac{\beta \beta'}{\beta + \beta'} (x'_1 - x_1)(x'_2 - x_2) > 0.$$

More generally, if the support of G is not ordered, then there exists a strictly positive measure of unordered pairs x, $x' \in \operatorname{supp}(G)$. By the above argument, we can strictly improve the objective conditional on any such pair, and hence also in expectation. □

Lemma 7. *If p defined by (5.3) is continuous and convex on Ω, then $x_2(\theta) = a x_1(\theta) + b$ for all θ and some $b \in \mathbb{R}$.*

Proof. Let ω, $\omega' \in \operatorname{int}(\Omega)$, $\theta \neq \theta'$, and $\lambda \in (0, 1)$. Denote

$$\begin{aligned}
\delta &= \theta' - \theta, \\
\varepsilon &= \omega_1 - x_1(\theta), \\
\varepsilon' &= \omega'_1 - x_1(\theta), \\
\gamma &= \frac{a(x_1(\lambda\theta + (1-\lambda)\theta') - x_1(\theta))}{(1-\lambda)\delta}, \\
\gamma' &= \frac{a(x_1(\theta') - x_1(\lambda\theta + (1-\lambda)\theta'))}{\lambda\delta}.
\end{aligned}$$

Since $\text{supp}(G) = \{(x_1(\theta), x_2(\theta)) : \theta \in \Theta\}$ and $x_2(\theta) + ax_1(\theta) = \theta$ for all $\theta \in \Theta$, Lemma 6 implies that $x_1(\theta)$, $x_2(\theta)$ are non-decreasing in θ and thus $\gamma, \gamma' \in [0, 1]$.

Since p is convex on Ω, we have

$$\begin{aligned}0 \leq& \lambda p(\omega) + (1-\lambda)p(\omega') - p(\lambda\omega + (1-\lambda)\omega') \\ =& \lambda(1-\lambda)\delta\left[(1-2\gamma')\varepsilon' + (2\gamma - 1)\varepsilon + \frac{1}{a}\left((\gamma')^2 - (1-\lambda)(\gamma'-\gamma)^2\right)\delta\right],\end{aligned}$$

where the equality holds by rearrangement. Since $x(\theta) \in \text{int}(\Omega)$, the above inequality holds for all ω, ω', $x(\theta')$ in a neighborhood of $x(\theta)$; so there exists $r > 0$ such that

$$\delta\left[(1-2\gamma')\varepsilon' + (2\gamma - 1)\varepsilon + \frac{1}{a}\left((\gamma')^2 - (1-\lambda)(\gamma'-\gamma)^2\right)\delta\right] \geq 0, \ \forall\, \varepsilon,\, \varepsilon',\, \delta \in [-r, r]. \quad \text{(B.1)}$$

If $\gamma \neq 1/2$, then (B.1) is violated at $\varepsilon' = 0$, $\varepsilon = -r\text{sgn}(2\gamma - 1)$, and $\delta = \min\{r, ra|2\gamma - 1|/2\}$, as follows from

$$(\gamma')^2 - (1-\lambda)(\gamma'-\gamma)^2 \leq (\gamma')^2 \leq 1.$$

If $\gamma' \neq 1/2$, then (B.1) is violated at $\varepsilon = 0$, $\varepsilon' = -r\text{sgn}(1-2\gamma')$ and $\delta = \min\{r, ra|1-2\gamma'|/2\}$.

Thus, $\gamma = \gamma' = 1/2$; so, for θ, $\theta' \in (\underline{\theta}, \overline{\theta})$, we have

$$x_1(\theta') - x_1(\theta) = \frac{1}{2a}(\theta' - \theta)$$

and

$$x_2(\theta') - x_2(\theta) = \theta' - \theta - a(x_1(\theta') - x_1(\theta)) = \frac{1}{2}(\theta' - \theta).$$

Finally, since $p(\omega)$ is continuous on Ω, we get that $x(\theta)$ is continuous on $[\underline{\theta}, \overline{\theta}]$ and the above equations hold for all θ, $\theta' \in [\underline{\theta}, \overline{\theta}]$. □

The examples below illustrate that the converse of Proposition 2 fails without an additional assumption.

Example 2. Suppose that Ω is the segment of the line $\omega_2 = \omega_1$ from (0, 0) and (1, 1). Thus, $X = \Omega$ is convex but has an empty interior. By Lemma 1 in Rayo and Segal (2010), full disclosure is optimal. Thus, revealing $\omega_2 + 2\omega_1$ is one of optimal disclosure rules (as it is informationally equivalent to full disclosure), but the induced posterior means belong to a line with slope 1 rather than 2.

Example 3. Suppose that Ω is the segment of the parabola $\omega_2 = \omega_1^2$ from (0, 0) to (1, 1). Thus, the support of G is a connected set, but it consists of the extreme points $x_2 = x_1^2$

of X. By Lemma 1 in Rayo and Segal (2010), full disclosure is optimal. That is, revealing $\omega_2 + \omega_1$ is optimal, but the induced posterior means $x_2 = x_1^2$ do not belong to a line.

Example 4. Suppose that the prior assigns probability 1/6 to each of the six states (0.1, 0.3), (0.3, 0.1), (0.4, 0.6), (0.6, 0.4), (0.8, 0.9), and (1, 0.7). Thus, the support of G consists of points (0.2, 0.2), (0.5, 0.5), and (0.9, 0.8), which are not extreme points of X, but they do not constitute a connected set. By Lemmas 1–3 in Rayo and Segal (2010), it is optimal to pool the following pairs of states: {(0.1, 0.3), (0.3, 0.1)}, {(0.4, 0.6), (0.6, 0.4)}, and {(0.8, 0.9), (1, 0.7)}. That is, revealing $\omega_2 + \omega_1$ is optimal, but the induced posterior means (0.2, 0.2), (0.5, 0.5), and (0.9, 0.8) do not belong to a line.

References

Aliprantis, C. D. and K. Border (2006): Infinite Dimensional Analysis: A Hitchhiker's Guide, Springer.

Anderson, E. J. and P. Nash (1987): Linear Programming in Infinite-Dimensional Space, Wiley.

Aumann, R. J. and M. Maschler (1995): Repeated Games with Incomplete Information, MIT press.

Bergemann, D. and S. Morris (2019): "Information Design: A Unified Perspective," Journal of Economic Literature, 57, 44–95.

Bogachev, V. (2007): Measure Theory, Springer.

Brøndsted, A. and R. T. Rockafellar (1965): "On the Subdifferentiability of Convex Functions," Proceedings of the American Mathematical Society, 16, 605–611.

Dizdar, D. and E. Kováč (2019): "A Simple Proof of Strong Duality in the Linear Persuasion Problem," .

Doval, L. and V. Skreta (2018): "Constrained Information Design: Toolkit," .

Dworczak, P. and G. Martini (2019): "The Simple Economics of Optimal Persuasion," Journal of Political Economy, 127, 1993–2048.

Gale, D. (1967): "A Geometric Duality Theorem with Economic Applications," Review of Economic Studies, 34, 19–24.

——— (1989): The Theory of Linear Economic Models, University of Chicago press.

Galperti, S. and J. Perego (2018): "A Dual Perspective on Information Design," .

Gentzkow, M. and E. Kamenica (2016): "A Rothschild-Stiglitz Approach to Bayesian Persuasion," American Economic Review, Papers & Proceedings, 106, 597–601.

Kamenica, E. (2019): "Bayesian Persuasion and Information Design," Annual Review of Economics, 11.

Kamenica, E. and M. Gentzkow (2011): "Bayesian Persuasion," American Economic Review, 101, 2590–2615.

Kolotilin, A. (2018): "Optimal Information Disclosure: A Linear Programming Approach," Theoretical Economics, 13, 607–636.

Kolotilin, A., T. Mylovanov, A. Zapechelnyuk, and M. Li (2017): "Persuasion of a Privately Informed Receiver," Econometrica, 85, 1949–1964.

Nikandrova, A. and R. Pancs (2017): "Conjugate Information Disclosure in an Auction with Learning," Journal of Economic Theory, 171, 174–212.

Rayo, L. and I. Segal (2010): "Optimal Information Disclosure," Journal of Political Economy, 118, 949–987.

Tong, H. (1952): "Some Characterizations of Normal and Perfectly Normal Spaces," Duke Mathematical Journal, 19, 289–292.

Villani, C. (2009): Optimal Transport, Old And New, Springer-Verlag, Berlin.