Extended Real-Valued Information Design

Nathan Yoder*

October 10, 2021

Abstract

This note shows that a finite-state Bayesian persuasion problem still has a solution when the cost of inducing some posteriors is infinite. Consequently, the concavification approach of Kamenica and Gentzkow (2011) can be applied to these settings.

In a *Bayesian persuasion* or *information design* problem (Kamenica and Gentzkow, 2011; Kamenica, 2019), a sender chooses a Blackwell experiment on a set of states Ω to maximize the expectation of some function v of the induced posterior belief. Kamenica and Gentzkow (2011) show that this is equivalent to choosing a distribution of posterior beliefs τ whose mean is the prior $p_0 \in \Delta(\Omega)$:

$$\max_{\tau \in \Delta(\Delta(\Omega))} \{E_\tau v(p) \text{ s.t. } E_\tau p = p_0\}$$

Typically, the value function v is the sender's interim payoff from a receiver's use of her experiment's result to choose a risky action. Sometimes, this payoff is modified by a function representing her cost of experimentation, as in Gentzkow and Kamenica (2014), or other payoffs from generating a certain posterior, as in Yoder (2021). Kamenica and Gentzkow (2011) show that when v is bounded, the problem has a solution whenever v is upper semicontinuous. I show that this is also true when v is unbounded, or extended real-valued. This extends their results to settings where the sender faces an infinite cost of inducing certain posteriors.

These infinite costs are natural in many settings. As Pomatto et al. (2020) point out, they are present in any application where the sender faces a constant marginal cost of experimentation. And as Doval and Skreta (2021) show, they also arise in persuasion problems where the sender faces constraints on the distributions of posterior beliefs they can induce.

*University of Georgia, Terry College of Business, Department of Economics; E-mail: nathan.yoder@uga.edu. This paper contains material from earlier drafts of Yoder (2021). I gratefully acknowledge the Terry College of Business' summer support through a Terry-Sanford research award. I also thank Alex Bloedel and Marzena Rostek for helpful suggestions.

Related literature

Among settings where some posteriors are infinitely costly to induce, the most heavily studied are those where the sender has one of the *log-likelihood ratio* (LLR) cost functions introduced by Pomatto et al. (2020). Specific cost functions in this class arise from foundations like Wald (1945) sequential sampling (Morris and Strack, 2019) and indifference to sequential learning (Bloedel and Zhong, 2021). In general, the LLR class is characterized by a cost of experimentation proportional to the Kullback-Leibler divergence between the distributions of signals conditional on different states. By formulating the information design problem as one of choosing those conditional signal distributions, Pomatto et al. (2020) observe that a solution exists when the sender has LLR costs, since Kullback-Leibler divergence is a lower semicontinuous function on the space of probability measures.

This note extends their observation beyond the LLR functional form. More generally, my results show that we need not depart from the posterior-based approach of Kamenica and Gentzkow (2011) to ensure that an extended real-valued information design problem has a solution. Instead, we can rely directly on the value function's upper semicontinuity in the induced posterior belief, just as Kamenica and Gentzkow (2011) show we can in the real-valued case.

An important environment where this result provides new insights is the constrained information design setting considered by Doval and Skreta (2021). They show that when there are constraints on the expected values of functions of the posterior, the supremal value of a persuasion problem is given by the concavification of an extended real-valued function. Proposition 1 strengthens their result by showing that as long as the underlying problem's value function is upper semicontinuous, this supremum is actually attained, and so the constrained problem has a solution.[1]

Notation

Let $\overline{\mathbb{R}}$ denote the affinely extended real numbers $\mathbb{R} \cup \{\pm\infty\}$. For a set S, let $\mathrm{conv}(S)$ denote its convex hull; $\mathrm{aff}(S)$, its affine hull; $\dim(S)$, its dimension; $\mathrm{ri}(S)$, its relative interior; and $\Delta(S)$, the set of Borel probability distributions on S. For a function $v : S \to \overline{\mathbb{R}}$, let $\mathrm{dom}(v) \equiv \{s \in S | v(s) \in \mathbb{R}\}$ denote its effective domain; $\mathrm{Gr}(v) \equiv \{(s, v(s)) | s \in \mathrm{dom}(v)\}$, its graph; $\mathrm{hypo}(v) \equiv \{(s, y) | s \in \mathrm{dom}(v), y \leq v(s)\}$, its hypograph; and $\mathrm{epi}(v)$, its epigraph. For a probability distribution τ, let $\mathrm{supp}\, \tau$ denote its support.

[1] Note that the function that Doval and Skreta (2021) concavify takes the right-hand sides of the constraints $g_i(p) \leq \gamma_i$ as arguments, and so has domain $\Delta(\Omega) \times \mathbb{R}^K$. While this is not compact (a requirement for Proposition 1), if the g_i are continuous, the proof of their main result shows that the function to be concavified can be restricted to $\Delta(\Omega) \times \prod_{i=1}^{K} [-3 \max_p |g_i(p)|, 3 \max_p |g_i(p)|]$ without loss.

Main results

The approach taken by Kamenica and Gentzkow (2011) makes use of the value function's concavification V: the smallest concave function that majorizes v. Formally, V is defined as the concave envelope of v's graph: $V(p) \equiv \sup\{z|(p,z) \in \mathrm{conv}(\mathrm{Gr}(v))\}$.[2] It is straightforward to show that the value of the Bayesian persuasion problem cannot exceed $V(p_0)$, the concavification at the prior. Kamenica and Gentzkow (2011) show that this value is actually attained.

In a finite-state problem, their argument proceeds roughly as follows.[3] If v is upper semicontinuous, its hypograph $\mathrm{hypo}(v)$ is closed. If v is also bounded, then the interesting part of its hypograph, $H = \{(p,z) \in \mathrm{hypo}(v)|z \geq \inf v(p)\}$ is compact, and so its convex hull is too. Clearly, V is also the upper envelope of *that* convex hull, which is compact, and therefore must contain $(p_0, V(p_0))$. Consequently, $(p_0, V(p_0))$ can be represented as a convex combination of points on the the graph of v; the weights in that convex combination give the distribution of posteriors that solves the sender's problem.

This argument must be modified when v is unbounded below: in this case, H will not be compact, and so we cannot guarantee that its convex hull will inherit its closedness. But we are only interested in whether the convex hull's "top" is closed. Proposition 1 shows that it is.

Proposition 1 ($V(p_0)$ is Attainable)**.** *Suppose $S \subseteq \mathbb{R}^n$ is compact and convex and $v : S \to \mathbb{R} \cup \{-\infty\}$ is upper semicontinuous; let $V : S \to \overline{\mathbb{R}}$ be the concavification of v. If $p_0 \in \mathrm{dom}\, v$, then $(p_0, V(p_0)) \in \mathrm{conv}(\mathrm{Gr}(v))$.*

Intuitively, suppose that we construct a sequence $\{z_k\}$ converging to $V(p_0)$ from below, and write each (p_0, z_k) as a convex combination of points on the graph of v. If some of those points descend vertically without bound, then their weight in the convex combination must approach zero. Then in the limit, $(p_0, V(p_0))$ can be written as a convex combination of the others.

Given Proposition 1, we can extend several of the key results from Kamenica and Gentzkow (2011) to the unbounded case. In particular, Proposition 2 shows that a solution to the persuasion problem exists, the value of the problem is given by the concavification at the prior, and the support of the solution need not have more elements than the set of states.

[2]Note that V is defined on all of S, not just $\mathrm{dom}\, v$: on the extended real number line, $\sup \emptyset = -\infty$.

[3]See the proof of Lemma 4 in Kamenica and Gentzkow (2009); in their online appendix, Kamenica and Gentzkow (2011) give a more general argument which also applies to settings with bounded value functions and infinitely many states.

Proposition 2 (Solutions to Extended-Real Valued Persuasion Problems)**.** *Suppose that for compact, convex* $S \subseteq \mathbb{R}^n$*,* $v : S \to \mathbb{R} \cup \{-\infty\}$ *is upper semicontinuous; let* $V : S \to \overline{\mathbb{R}}$ *be the concavification of* v*. Then for all* $p_0 \in \operatorname{ri}(\operatorname{dom} v)$*,*

i. $V(p_0) = \max_{\tau \in \Delta(S)}\{E_\tau v(p) \text{ s.t. } E_\tau p = p_0\}$.

ii. There exists $\tau^* \in \arg\max_\tau\{E_\tau v(p) \text{ s.t. } E_\tau p = p_0\}$ *such that* $\operatorname{supp}\tau^*$ *has no more than* $\dim(S) + 1$ *elements and* $\operatorname{supp}\tau^* \subseteq \operatorname{dom} v$*.*

Note that in a persuasion problem where the set of states Ω is finite, part (ii) bounds the number of posteriors in the support of the problem's solution by $|\Omega|$, since the simplex $\Delta(\Omega)$ has dimension $|\Omega| - 1$.

The existence of a solution to the persuasion problem follows from Proposition 1 in a relatively straightforward manner, as does the fact that the problem's value is given by the concavification at the prior. However, showing that the solution's support need not have more elements than the set of states (Part (ii)) requires a modification of the argument from Kamenica and Gentzkow (2011): The Fenchel-Bunt Theorem can't always be applied to the value function's hypograph when v takes extended real values, because the hypograph might not be connected.

Instead, Proposition 2 (ii) relies on the following alternative argument. Suppose S has dimension $|\Omega| - 1$. If the concavification lies above the value function at the prior, it coincides with a hyperplane supported by points on the value function's graph. Because that hyperplane has dimension no greater than $|\Omega| - 1$, Carathéodory's theorem says that we can write $(p_0, V(p_0))$ as a convex combination of at most $|\Omega|$ of those points. Consequently, an experiment with value $V(p_0)$ need not induce more than $|\Omega|$ posteriors.

References

BLOEDEL, A. W. AND W. ZHONG (2021): "The Cost of Optimally-Acquired Information," Working paper.

DOVAL, L. AND V. SKRETA (2021): "Constrained Information Design: Toolkit," *arXiv preprint arXiv:1811.03588*.

GENTZKOW, M. AND E. KAMENICA (2014): "Costly Persuasion," *American Economic Review: Papers and Proceedings*, 104, 457–462.

HIRIART-URRUTY, J.-B. AND C. LEMARÉCHAL (2001): *Fundamentals of Convex Analysis*, Springer-Verlag.

KAMENICA, E. (2019): "Bayesian Persuasion and Information Design," *Annual Review of Economics*, 11, 249–272.

KAMENICA, E. AND M. GENTZKOW (2009): "Bayesian Persuasion," Working Paper 15540, National Bureau of Economic Research.

——— (2011): "Bayesian Persuasion," *American Economic Review*, 101.

MORRIS, S. AND P. STRACK (2019): "The Wald Problem and the Equivalence of Sequential Sampling and Ex-Ante Information Costs," *Available at SSRN 2991567*.

POMATTO, L., P. STRACK, AND O. TAMUZ (2020): "The Cost of Information," *arXiv preprint arXiv:1812.04211*.

WALD, A. (1945): "Sequential Tests of Statistical Hypotheses," *The Annals of Mathematical Statistics*, 16, 117–186.

YODER, N. (2021): "Designing Incentives for Heterogeneous Researchers," Working paper.

Proofs

Lemma 1 (Properties of the Concavification). *Let $S \subseteq \mathbb{R}^n$ be convex and compact; let $v : S \to \mathbb{R} \cup \{-\infty\}$ be upper semicontinuous with* $\operatorname{dom} v \neq \emptyset$*; and let V be the concavification of v. Then V is a proper concave function;* $\operatorname{dom} V = \operatorname{conv}(\operatorname{dom} v)$*; and* $\operatorname{ri}(\operatorname{dom} v) \subseteq \operatorname{ri}(\operatorname{dom} V)$.

Proof. Concavity of V follows immediately from the definition. Since v is upper semicontinuous and S is compact, v is bounded above: for some $\bar{z} \in \mathbb{R}$, we have $v(p) \leq \bar{z}$ for all $p \in S$. Then $\{z | (p, z) \in \operatorname{conv}(\operatorname{Gr}(v))\}$ is also bounded above: if $(p, z) \in \operatorname{conv}(\operatorname{Gr}(v))$, then for some d, $\{p_i\}_{i=1}^d \subseteq S$, and $(\lambda_1, \ldots, \lambda_d) \in \Delta(\{1, \ldots, d\})$, $z = \sum_{i=1}^d \lambda_i v(p_i) \leq \bar{z}$. Then V is bounded above: $V(p) \equiv \sup\{z | (p, z) \in \operatorname{conv}(\operatorname{Gr}(v))\} \leq \bar{z} < \infty$ for each p.

Then $p \in \operatorname{dom} V \Leftrightarrow V(p) > -\infty$. By definition, we have $V(p) > -\infty$ if and only if there exists $(p, z) \in \operatorname{conv}(\operatorname{Gr}(v))$ with $z > -\infty$. This is true if and only if there exist d, $\{p_i\}_{i=1}^d \subseteq \operatorname{dom} v$, and $\{\lambda_i\}_{i=1}^d \subset (0, 1]$ such that $\sum_{i=1}^d \lambda_i = 1$, $\sum_{i=1}^d \lambda_i v(p_i) > -\infty$, and $\sum_{i=1}^d p_i = p$. Now for any $\{\lambda_i\}_{i=1}^d \subset (0, 1]$, $\sum_{i=1}^d \lambda_i v(p_i) > -\infty$ if and only if $\{p_i\}_{i=1}^d \subseteq \operatorname{dom} v$; it follows that $V(p) > -\infty \Leftrightarrow p \in \operatorname{conv}(\operatorname{dom} v)$. Hence $\operatorname{dom} V = \operatorname{conv}(\operatorname{dom} v)$. Then the relative topologies of $\operatorname{dom} v$ and $\operatorname{dom} V$ coincide, since $\operatorname{aff}(\operatorname{dom} V) = \operatorname{aff}(\operatorname{conv}(\operatorname{dom} v)) = \operatorname{aff}(\operatorname{dom} v)$; then since $\operatorname{dom} v \subseteq \operatorname{conv}(\operatorname{dom} v)$, $\operatorname{ri}(\operatorname{dom} v) \subseteq \operatorname{ri}(\operatorname{dom} V)$. Moreover, since $\operatorname{dom} v \neq \emptyset$, it follows that $\operatorname{dom} V = \operatorname{conv}(\operatorname{dom} v) \neq \emptyset$; then since V is bounded above and concave, it is proper concave. □

Proof of Proposition 1 ($V(p_0)$ is Attainable) From Lemma 1, $\operatorname{dom} V = \operatorname{conv}(\operatorname{dom} v)$. Hence, since $p_0 \in \operatorname{dom} v$, $V(p_0) \in \mathbb{R}$; then $\{z | (p_0, z) \in \operatorname{conv}(\operatorname{Gr}(v))\}$ must be nonempty (since $V(p_0) \equiv \sup\{z | (p_0, z) \in \operatorname{conv}(\operatorname{Gr}(v))\} > -\infty$). Then there is some sequence $\{(p_0, z_k)\}_{k=1}^\infty \subset \operatorname{conv}(\operatorname{Gr}(v))$ such that $z_k \to \sup\{z | (p_0, z) \in \operatorname{conv}(\operatorname{Gr}(v))\} = V(p_0)$. By Carathéodorory's theorem, for each k there exists $(p_1^k, \ldots, p_{n+1}^k, \lambda_1^k, \ldots, \lambda_{n+1}^k) \in (\operatorname{dom} v)^{n+1} \times \Delta(\{1, \ldots, n+1\}) \subseteq S^{n+1} \times \Delta(\{1, \ldots, n+1\})$ such that $\sum_{i=1}^{n+1} \lambda_i^k p_i^k = p_0$ and $\sum_{i=1}^{n+1} \lambda_i^k v(p_i^k) = z_k$. Since S and the simplex $\Delta(\{1, \ldots, n+1\})$ are both compact, so is the Cartesian product

$S^{n+1} \times \Delta(\{1, \ldots, n+1\})$. Then $\{(p_1^k, \ldots, p_{n+1}^k, \lambda_1^k, \ldots, \lambda_{n+1}^k)\}_{k=1}^{\infty}$ has a convergent subsequence

$$\{(p_1^{k_\ell}, \ldots, p_{n+1}^{k_\ell}, \lambda_1^{k_\ell}, \ldots, \lambda_{n+1}^{k_\ell})\} \to (p_1^*, \ldots, p_{n+1}^*, \lambda_1^*, \ldots, \lambda_{n+1}^*) \in S^{n+1} \times \Delta(\{1, \ldots, n+1\}).$$

Since $\sum_{i=1}^{n+1} \lambda_i^{k_\ell} p_i^{k_\ell} = p_0$ for each ℓ, we have

$$\sum_{i=1}^{n+1} \lambda_i^* p_i^* = \sum_{i=1}^{n+1} (\lim_{\ell \to \infty} \lambda_i^{k_\ell})(\lim_{\ell \to \infty} p_i^{k_\ell}) = \lim_{\ell \to \infty} \sum_{i=1}^{n+1} \lambda_i^{k_\ell} p_i^{k_\ell} = p_0.$$

Since v is upper semicontinuous, we have $\limsup_{\ell \to \infty} v(p_i^{k_\ell}) \leq v(p_i^*)$ for each i. Since, in addition, S is compact, v is bounded above: we have $v(p) \leq \bar{z}$ for some $\bar{z} \in \mathbb{R}$. Let $J = \{i \in \{1, \ldots, n+1\} | \limsup_{\ell \to \infty} v(p_i^{k_\ell}) > -\infty\}$. Then for each $i \notin J$, we have $\lim_{\ell \to \infty} v(p_i^{k_\ell}) = -\infty$. Suppose that there is some $i \notin J$ such that $\lambda_i^* \neq 0$. Then $\lim_{\ell \to \infty} \lambda_i^{k_\ell} v(p_i^{k_\ell}) = (\lim_{\ell \to \infty} \lambda_i^{k_\ell})(\lim_{\ell \to \infty} v(p_i^{k_\ell})) = -\infty$, and so we have

$$\begin{aligned} -\infty = -\infty + \sum_{j \neq i} \lambda_j^* \bar{z} = \lim_{\ell \to \infty} \lambda_i^{k_\ell} v(p_i^{k_\ell}) + \lim_{\ell \to \infty} \sum_{j \neq i} \lambda_j^{k_\ell} \bar{z} &= \lim_{\ell \to \infty} \left(\lambda_i^{k_\ell} v(p_i^{k_\ell}) + \sum_{j \neq i} \lambda_j^{k_\ell} \bar{z} \right) \\ &\geq \lim_{\ell \to \infty} \sum_{j=1}^{n+1} \lambda_j^{k_\ell} v(p_j^{k_\ell}) = \lim_{\ell \to \infty} z_{k_\ell} = V(p_0). \end{aligned}$$

This is a contradiction, since $V(p_0) \geq v(p_0)$ by definition and $p_0 \in \operatorname{dom} v$. Then $\lambda_i^* = 0$ for each $i \notin J$, and so $\sum_{i \in J} \lambda_i^* = 1$ and $\sum_{i \in J} \lambda_i^* p_i^* = \sum_{i=1}^{n+1} \lambda_i^* p_i^* = p_0$. Moreover, by upper semicontinuity of v, $v(p_i^*) > \limsup_{\ell \to \infty} v(p_i^{k_\ell}) > -\infty$, and hence $p_i^* \in \operatorname{dom} v$, for each $i \in J$; it follows from the definition of V that $V(p_0) \geq \sum_{j \in J} \lambda_j^* v(p_j^*)$.

Now for each $i \notin J$, $\lim_{\ell \to \infty} v(p_i^{k_\ell}) = -\infty$, and so there exists L_i such that for all $\ell \geq L_i$, $v(p_i^{k_\ell}) \leq 0$. Let $L = \max_{i \notin J} L_i$. Then for all $\ell \geq L$, $z_{k_\ell} = \sum_{j=1}^{n+1} \lambda_j^{k_\ell} v(p_j^{k_\ell}) \leq \sum_{j \in J} \lambda_j^{k_\ell} v(p_j^{k_\ell})$. Then

$$\begin{aligned} V(p_0) \geq \sum_{j \in J} \lambda_j^* v(p_j^*) \geq \sum_{j \in J} \lambda_j^* \limsup_{\ell \to \infty} v(p_j^{k_\ell}) &= \sum_{j \in J} \left(\lim_{\ell \to \infty} \lambda_j^{k_\ell} \right) \left(\limsup_{\ell \to \infty} v(p_j^{k_\ell}) \right) \\ &= \sum_{j \in J} \limsup_{\ell \to \infty} \lambda_j^{k_\ell} v(p_j^{k_\ell}) \\ &\geq \limsup_{\ell \to \infty} \sum_{j \in J} \lambda_j^{k_\ell} v(p_j^{k_\ell}) \geq \limsup_{\ell \to \infty} z_{k_\ell} = V(p_0), \end{aligned}$$

and hence $V(p_0) = \sum_{i \in J} \lambda_i^* v(p_i^*)$. Then since $p_0 = \sum_{i \in J} \lambda_i^* p_i^*$ and $\{p_i^*\}_{i \in J} \subset \operatorname{dom} v$, we have $(p_0, V(p_0)) \in \operatorname{conv}(\operatorname{Gr}(v))$, as desired. $\square$

Proof of Proposition 2 (i) (Existence of Solutions) By Lemma 1, $p_0 \in \text{ri}(\text{dom}\, V)$. Then since V is proper concave by Lemma 1, it is superdifferentiable at p_0:[4] There exists $x \in \mathbb{R}^n$ such that $V(p_0) + x \cdot (p - p_0) \geq V(p)$ for each $p \in S$. Then since by definition, $V(p) \geq v(p)$ for each p, for any Bayes-plausible $\tau \in \Delta(S)$ we have

$$V(p_0) = \int_S (V(p_0) + x \cdot (p - p_0)) d\tau(p) \geq \int_S v(p) d\tau(p).$$

So $V(p_0) \geq \sup_{\tau \in \Delta(S)} \{E_\tau v(p) \text{ s.t. } E_\tau p = p_0\}$. By Proposition 1, $(p_0, V(p_0)) \in \text{conv}(\text{Gr}(v))$; hence, there exist $\{p_i\}_{i=1}^d \subseteq \text{dom}\, v$ and $\{\lambda_i\}_{i=1}^d \subset (0,1]$ with $\sum_{i=1}^d \lambda_i = 1$ such that $p_0 = \sum_{i=1}^d \lambda_i p_i$ and $V(p_0) = \sum_{i=1}^d \lambda_i v(p_i)$. It follows by letting $\tau^*(\{p_i\}) = \lambda_i$ for each $i \in \{1, \ldots, d\}$ that $V(p_0) \leq \sup_{\tau \in \Delta(S)} \{E_\tau v(p) \text{ s.t. } E_\tau p = p_0\}$ as well. Then

$$V(p_0) = \sup_{\tau \in \Delta(S)} \{E_\tau v(p) \text{ s.t. } E_\tau p = p_0\};$$

since τ^* attains this supremum, the statement follows. □

Lemma 2 (Properties of the Concavification at the Optimum). *Let $S \subseteq \mathbb{R}^n$ be convex and compact, let $v : S \to \mathbb{R} \cup \{-\infty\}$ be upper semicontinuous, let $p_0 \in \text{ri}(\text{dom}\, v)$, and let V be the concavification of v. If $\tau^* \in \arg\max_{\tau \in \Delta(S)} \{E_\tau v(p) \text{ s.t. } E_\tau p = p_0\}$, then $v(p) = V(p)$ for all $p \in \text{supp}\, \tau^*$, and V is affine and finite on $\text{conv}(\text{supp}\, \tau^*)$.*

Proof. Since $p_0 \in \text{ri}(\text{dom}\, v) \subseteq \text{ri}(\text{dom}\, V)$ and V is proper concave by Lemma 1, V is superdifferentiable at p_0: There exists $x \in \mathbb{R}^n$ such that $V(p_0) + x \cdot (p - p_0) \geq V(p) \geq v(p)$ for each $p \in S$. The rest follows identically to the proof of Lemma 3 in Yoder (2021), relying on Proposition 2 (i) instead of Kamenica and Gentzkow (2011). □

Proof of Proposition 2 (ii) (Existence of $n+1$-ary Solutions)

By Proposition 1, $(p_0, V(p_0)) \in \text{conv}(\text{Gr}(v))$, and there exist $\{p_i\}_{i=1}^d \subseteq \text{dom}(v)$ and $\{\lambda_i\}_{i=1}^d \subset (0,1]$ with $\sum_{i=1}^d \lambda_i = 1$ such that $p_0 = \sum_{i=1}^d \lambda_i p_i$ and $V(p_0) = \sum_{i=1}^d \lambda_i v(p_i)$.

Let $\hat{\tau}(\{p_i\}) = \lambda_i$ for each $i \in \{1, \ldots, d\}$. Then part (i) implies $\hat{\tau} \in \arg\max_{\tau \in \Delta(S)} \{E_\tau v(p) \text{ s.t. } E_\tau p = p_0\}$. Then by Lemma 2, V is affine on $\text{conv}(\{p_i\}_{i=1}^d)$, and coincides with v at each p_i, $i \in \{1, \ldots, d\}$.

By construction, $p_0 \in \text{conv}(\{p_i\}_{i=1}^d)$. Then by Carathéodorory's theorem, there exists $J \subseteq \{1, \ldots, d\}$ with $|J| \leq \dim(S) + 1$ and $\{\gamma_i\}_{i \in J} \subseteq (0,1]$ such that $\sum_{i \in J} \gamma_i = 1$ and $p_0 = \sum_{i \in J} \gamma_i p_i$. Since V is affine on $\text{conv}(\{p_i\}_{i=1}^d)$, and coincides with v at each p_i, we have $\sum_{i \in J} \gamma_i v(p_i) = \sum_{i \in J} \gamma_i V(p_i) = V(p_0)$. Let $\tau^*(\{p_i\}) = \gamma_i$ for each $i \in J$; it follows from part (i) that $\tau^* \in \arg\max_{\tau \in \Delta(S)} \{E_\tau v(p) \text{ s.t. } E_\tau p = p_0\}$ and $|\text{supp}\, \tau^*| = J \leq \dim(S) + 1$, as desired. □

[4] See, e.g., Hiriart-Urruty and Lemaréchal (2001) Proposition B.1.2.1.