

Mixture Selection, Mechanism Design, and Signaling

Yu Cheng, Ho Yee Cheung, Shaddin Dughmi, Ehsan Emamjomeh-Zadeh, Li Han, Shang-Hua Teng*

Department of Computer Science

University of Southern California

{yu.cheng.1, hoyeeche, shaddin, emamjome, li.han, shanghai}@usc.edu.

Abstract

We pose and study a fundamental algorithmic problem which we term *mixture selection*, arising as a building block in a number of game-theoretic applications: Given a function g from the n -dimensional hypercube to the bounded interval $[-1, 1]$, and an $n \times m$ matrix A with bounded entries, maximize $g(Ax)$ over x in the m -dimensional simplex. This problem arises naturally when one seeks to design a lottery over items for sale in an auction, or craft the posterior beliefs for agents in a Bayesian game through the provision of information (a.k.a. signaling).

We present an approximation algorithm for this problem when g simultaneously satisfies two “smoothness” properties: Lipschitz continuity with respect to the L^∞ norm, and *noise stability*. The latter notion, which we define and cater to our setting, controls the degree to which low-probability — and possibly correlated — errors in the inputs of g can impact its output. The approximation guarantee of our algorithm degrades gracefully as a function of the Lipschitz continuity and noise stability of g . In particular, when g is both $O(1)$ -Lipschitz continuous and $O(1)$ -stable, we obtain an (additive) polynomial-time approximation scheme (PTAS) for mixture selection. We also show that neither assumption suffices by itself for an additive PTAS, and both assumptions together do not suffice for an additive fully polynomial-time approximation scheme (FPTAS).

We apply our algorithm for mixture selection to a number of different game-theoretic applications, focusing on problems from mechanism design and optimal signaling. In particular, we make progress on a number of open problems suggested in prior work by easily reducing them to mixture selection: we resolve an important special case of the *small-menu lottery design* problem posed by Dughmi, Han, and Nisan [10]; we resolve the problem of *revenue-maximizing signaling in Bayesian second-price auctions* posed by Emek et al. [12] and Miltersen and Sheffet [5]; we design a quasipolynomial-time approximation scheme for the *optimal signaling problem in normal form games* suggested by Dughmi [9]; and we design an approximation algorithm for the optimal signaling problem in the voting model of Alonso and Câmara [3].

Keywords

Mechanism design, signaling, game theory, noise stability.

1 INTRODUCTION

Lotteries, beliefs, mixed strategies — all are distributions arising as important objects in game theory. It is unsurprising, therefore, that algorithmic game theory is rife with algorithmic problems which — implicitly or explicitly — optimize over the space of distributions, or equivalently the simplex. In this paper, we identify a family of algorithmic problems over the simplex which arise over and over in game theory. We term problems in this class *mixture selection*, and examine their computational complexity.

Given a function g from the solid n -dimensional hypercube to the bounded interval $[-1, 1]$, and an integer m , we define the *m -dimensional mixture selection problem* for g as follows. The input to this problem is an $n \times m$ matrix A with bounded entries, and the objective is to compute $x \in \Delta_m$ maximizing $g(Ax)$. It is natural to expect that the computational complexity of mixture selection depends crucially on the “complexity” of the function g . We therefore identify two “smoothness” parameters of the function g which control the extent to which mixture selection is tractable, and derive a simple approximation algorithm with guarantees degrading gracefully in those parameters. Moreover, we present evidence — in the form of hardness results — that smoothness in both senses is necessary for the kind of general results we obtain.

The first smoothness quantity is a familiar one, namely Lipschitz continuity in the L^∞ metric. The second quantity, which we define and term *noise stability*, borrows ideas from related definitions of stability in other contexts (e.g. [21], [27]), though is importantly different. Informally, a function g from the solid n -dimensional hypercube to the real numbers is *β -noise stable* (or *β -stable* for short) if the random corruption of an α -fraction of the n inputs to g , with no individual input disproportionately likely to be corrupted, does not decrease the output of g by more than $\alpha\beta$. We note that a Fourier-analytic notion of stability is closely-related to ours — we elaborate on this connection in Section 9.

*This work was supported by NSF grants CCF-1350900, CCF-1423618, CCF-0964481, CCF-1111270, and Shang-Hua Teng’s Simons Investigator Award from the Simons Foundation.

This paper lays out a framework for tackling mixture selection problems, and presents a number of applications in mechanism design and optimal signaling in games. Notably, we find that we resolve or make progress on a number of known open problems, and some new ones, using our framework.

Our Results

Our results for mixture selection can be viewed as generalizing the main insights of Lipton et al. [25]. First, we show that when g is noise stable and Lipschitz continuous, and $x \in \Delta_n$ is arbitrary, there is a sparse vector \tilde{x} for which $g(A\tilde{x})$ is not much smaller than $g(Ax)$. The proof of this fact proceeds by sampling from x and letting \tilde{x} be the empirical distribution, as in [25]. However, when g is sufficiently noise stable and Lipschitz continuous, we obtain a better tradeoff between the number of samples required and the error introduced into the objective than does [25], and this is crucial for our applications. Our analysis bounds the expected difference between $g(Ax)$ and $g(A\tilde{x})$ as the sum of two terms: The first term represents the error in the output of g caused by the low-probability “large errors” in its n inputs, and the second term represents the error in the output of g introduced by the higher-probability “small errors” in its n inputs. The first term is bounded using noise stability, and the second is bounded using Lipschitz continuity.

Second, we instantiate the above insight algorithmically, as does [25]. Specifically, our algorithm enumerates vectors \tilde{x} of the desired sparsity in order to find an approximately optimal solution to our mixture selection problem. We note that our guarantees are all parametrized by the Lipschitz continuity c and the noise stability β of the function g . Most notably, we obtain an additive polynomial-time approximation scheme (PTAS) whenever both β and c are constants.

Third, we rule out certain natural extensions of our results assuming well-believed complexity-theoretic conjectures. We show that neither Lipschitz continuity nor noise stability alone suffices for an additive PTAS for mixture selection, and both together do not suffice for an additive fully polynomial-time approximation scheme (FPTAS). For a function which is $O(1)$ -stable yet $O(1)$ -Lipschitz continuous only in the L^1 metric, we show approximation hardness by a reduction from the NP-hard maximum independent set problem. For a function which is $O(1)$ -Lipschitz in L^∞ yet not $O(1)$ -stable, we show approximation hardness by a reduction from the planted clique problem. Finally, for a function which is both $O(1)$ -Lipschitz in L^∞ and $O(1)$ -stable, we rule out an additive FPTAS via a reduction from the maximum independent set problem.

Despite the simplicity of our framework, we find that it has powerful implications for problems in mechanism design and optimal signaling in games. We feature four natural applications in this paper, three of which resolve or partially resolve outstanding open problems from prior work:

- 1) **Lottery design:** Dughmi, Han, and Nisan [10] examined one of the most basic problems in mechanism design: that of designing the revenue-maximizing multi-item auction for a single unit-demand buyer with valuation represented implicitly via a sampling oracle. They reduced this problem to a regularized variant of itself, namely optimally designing a small number of lottery-price pairs — a *small menu* — from which the buyer is allowed to choose. We apply our framework to resolve the special case of this problem with a single lottery — i.e., a menu of size 1. This follows from the Lipschitz continuity and noise stability of the function $g_w^{(\text{lottery})}(t) := \max_p \{p \cdot \sum_{i=1}^n w_i \cdot I[t_i \geq p]\}$ for an arbitrary weight vector $w \in \Delta_n$, where $I[\mathcal{E}]$ is the indicator function for the event \mathcal{E} .
- 2) **Revenue-maximizing signaling in probabilistic second-price auctions:** Emek et al. [12] and Miltersen and Sheffet [5] considered signaling in the context of a probabilistic second-price auction. In particular, the attributes of the item for sale are unknown, and the auctioneer must decide what information to reveal in order to maximize his revenue in this auction. This is particularly relevant in advertising auctions, where items are impressions associated with demographics that are a-priori unknown to the advertisers bidding in the auction. Whereas both papers presented a polynomial-time algorithm for this problem when bidder types are fixed, the general problem was shown to be NP-hard and its approximability was left largely open. Using our framework, a PTAS for the general problem follows easily. We use the fact that the function max2 , which simply returns the second largest entry of a vector, is Lipschitz continuous and noise stable.
- 3) **Persuasion in voting:** Alonso and Câmara [3] examine a simple election for selecting a binary outcome — say whether a ballot measure is passed — when voters are not fully informed of the consequences of the measure, and hence of their utilities. Each voter casts a Yes/No vote, and the measure passes if the fraction of Yes votes exceeds a certain pre-specified threshold. A principal — say a moderator of a political debate — can determine the protocol — or signaling scheme — through which information regarding the measure is gathered and shared with voters. We consider a principal concerned with maximizing the probability of the measure passing. [3] characterize the optimal signaling scheme and a number of its properties, though stop short of deriving an algorithm for optimal signaling. We design a multi-criteria PTAS for this problem using our framework. Along the way, we also design a bi-criteria PTAS for the related problem of maximizing the expected number of Yes votes in the election. For both results, we use the fact that the function $g^{(\text{vote-sum})}(t) = \frac{1}{n} |\{i : t_i \geq 0\}|$ is noise stable and Lipschitz continuous in a bi-criteria sense.

- 4) **Optimal signaling in normal form games:** Dughmi [9] examined the problem of optimal signaling in abstract normal form games, and ruled out an FPTAS even for two-player zero-sum games. The possibility for a PTAS for two-player zero-sum games, and a QPTAS for general games with a constant number of players, were left open. We show that a bi-criteria QPTAS for normal-form games with a constant number of players follows from our framework, and applies to a large and natural class of objective functions. We use the fact that every function is $O(n)$ -stable, and the fact that the function measuring the quality of equilibria satisfies a bi-criteria notion of Lipschitz continuity which we define.

Additional Discussion of Related Work

As previously described, our framework generalizes and refines the main insight of [25]. The recent work of Barman [4] is also similar in spirit; in particular, the approximate variant of Caratheodory’s theorem employed in that paper can be viewed as a mixture selection problem with $g(Ax) = -\|Ax - Ax^*\|_p$ for a fixed vector x^* and norm $p \geq 2$. Even though this function g is neither Lipschitz continuous in L^∞ nor noise stable, Barman exhibits a PTAS under the assumption that the columns of A have small (i.e. constant) p -norm.

2 FRAMEWORK

For a function $g : [-1, 1]^n \rightarrow [-1, 1]$ and a positive integer m , we define the following optimization problem which we term m -dimensional *mixture selection* for g : given an $n \times m$ matrix A with entries in $[-1, 1]$, find x in the m -dimensional simplex Δ_m maximizing $f(x) := g(Ax)$. In this section, we present our notion of *noise stability*, and derive approximation algorithms for this problem when the function g is simultaneously noise stable and Lipschitz continuous with respect to the L^∞ metric. Moreover, we show that neither requirement alone suffices for our results.

Our approximation guarantees will be additive — i.e., an ϵ -approximation algorithm for mixture selection outputs $x \in \Delta_m$ with $f(x) \geq \max_{y \in \Delta_m} f(y) - \epsilon$. To illustrate our techniques, we use the following function $g^{(\text{mid})} : [-1, 1]^n \rightarrow [-1, 1]$, which averages all but the top and bottom quartiles of its inputs, as a running example.

$$g^{(\text{mid})}(t) = \frac{1}{\lceil 3n/4 \rceil - \lfloor n/4 \rfloor} \sum_{i=\lfloor n/4 \rfloor + 1}^{\lceil 3n/4 \rceil} t_{[i]},$$

where $t_{[i]}$ denote the i^{th} largest entry of t . Throughout the paper, we use t_i to denote the i^{th} entry of t , and use $t_{[i]}$ to denote the i^{th} largest entry of t .

Though we present our framework for functions $g : [-1, 1]^n \rightarrow [-1, 1]$, we define mixture selection similarly for functions $g : [0, 1]^n \rightarrow [0, 1]$. The two definitions are equivalent up to normalization, and it is easy to verify that all our results and bounds for mixture selection carry through unchanged to either definition.

Our main result applies to functions g which are both noise stable and Lipschitz continuous with respect to the L^∞ metric. We now formalize these two conditions.

Lipschitz Continuity

A function $g : [-1, 1]^n \rightarrow [-1, 1]$ is c -Lipschitz continuous in L^∞ — or c -Lipschitz for short — if and only if for all t, t' in the domain of g , $|g(t) - g(t')| \leq c\|t - t'\|_\infty$. To illustrate, our example function $g^{(\text{mid})}$ is 1-Lipschitz. We note that Lipschitz continuity in L^∞ is a stronger assumption than in any other L^p norm.

Noise Stability

Our notion of noise stability captures the following desirable property of a function $g : [-1, 1]^n \rightarrow [-1, 1]$: if a random process corrupts (i.e., modifies arbitrarily) some of the inputs to g , with no individual input disproportionately likely to be corrupted, then the output of g does not decrease by much in expectation. Such random corruption patterns are captured by our notion of a *light distribution* over subsets of $[n]$, defined below.

Definition 2.1 (Light Distribution). *Let \mathcal{D} be a distribution supported on subsets of $[n]$. For $\alpha \in (0, 1]$, we say \mathcal{D} is α -Light if and only if the following holds for all $i \in [n]$: $\Pr_{R \sim \mathcal{D}}[i \in R] \leq \alpha$.*

In other words, a light distribution bounds the *marginal probability* of any individual element of $[n]$. When corrupted inputs follow a light distribution, no individual input is too likely to be corrupted. However, we note that our notion of light distribution allows arbitrary correlations between the corruption events of various inputs. We define a noise stable function as one which is robust, in an average sense, to corrupting a subset R of its n inputs when R follows a light distribution \mathcal{D} . Our notion of robustness is one-sided: we only require that our function’s output not decrease substantially in expectation. This one-sided guarantee suffices for all our applications, and is necessitated by some. We note that the light distribution \mathcal{D} , as well as the (corrupted) inputs, are chosen adversarially. We make use of the following notation in our definition: Given

vectors $t, t' \in [-1, 1]^n$ and a set $R \subseteq [n]$, we say $t' \underset{R}{\approx} t$ if $t_i = t'_i$ for all $i \notin R$. In other words, if $t' \underset{R}{\approx} t$, then t' is a result of corrupting the entries of t corresponding to R .

Definition 2.2 (Noise Stability). *Given a function $g : [-1, 1]^n \rightarrow [-1, 1]$ and a real number $\beta \geq 0$, we say g is β -stable if and only if the following holds for all $t \in [-1, 1]^n$, $\alpha \in (0, 1]$, and α -Light distributions \mathcal{D} over subsets of $[n]$:*

$$\mathbb{E}_{R \sim \mathcal{D}} \left[\min\{g(t') : t' \underset{R}{\approx} t\} \right] \geq g(t) - \alpha\beta.$$

To illustrate this definition, we show that our example function $g^{(\text{mid})}$ is 4-stable. To see this, observe that changing k entries of the input to $g^{(\text{mid})}$ can decrease its output by at most $\frac{4k}{n}$. When R is drawn from an α -Light distribution and t is an arbitrary input, 4-stability therefore follows from the linearity of expectations:

$$\mathbb{E}_{R \sim \mathcal{D}} \left[\min\{g^{(\text{mid})}(t') : t' \underset{R}{\approx} t\} \right] \geq \mathbb{E}_{R \sim \mathcal{D}} \left[g^{(\text{mid})}(t) - \frac{4|R|}{n} \right] \geq g^{(\text{mid})}(t) - 4\alpha.$$

We note that every function $g : [-1, 1]^n \rightarrow [-1, 1]$ is $2n$ -stable, which follows from the union bound.

As a useful building block for proving some of our functions stable, we show that stable functions can be combined to yield other stable functions if composed with a convex, nondecreasing, and Lipschitz continuous function.

Proposition 2.3. *Fix $\beta, c \geq 0$, and let $g_1, g_2, \dots, g_k : [-1, 1]^n \rightarrow [-1, 1]$ be β -stable functions. For every convex function $h : [-1, 1]^k \rightarrow [-1, 1]$ which is nondecreasing in each of its arguments and c -Lipschitz continuous in L^∞ , the function $g(t) := h(g_1(t), \dots, g_k(t))$ is (βc) -stable.*

Proof: For all $t \in [-1, 1]^n$ and all α -Light distributions \mathcal{D} ,

$$\begin{aligned} \mathbb{E}_{R \sim \mathcal{D}} \left[\min_{t' \underset{R}{\approx} t} g(t') \right] &= \mathbb{E}_{R \sim \mathcal{D}} \left[\min_{t' \underset{R}{\approx} t} h(g_1(t'), \dots, g_k(t')) \right] \\ &\geq \mathbb{E}_{R \sim \mathcal{D}} \left[h\left(\min_{t' \underset{R}{\approx} t} g_1(t'), \dots, \min_{t' \underset{R}{\approx} t} g_k(t')\right) \right] && \text{(Since } h \text{ is nondecreasing)} \\ &\geq h\left(\mathbb{E}_{R \sim \mathcal{D}} \left[\min_{t' \underset{R}{\approx} t} g_1(t') \right], \dots, \mathbb{E}_{R \sim \mathcal{D}} \left[\min_{t' \underset{R}{\approx} t} g_k(t') \right]\right) && \text{(Jensen's inequality)} \\ &\geq h(g_1(t) - \alpha\beta, \dots, g_k(t) - \alpha\beta) && \text{(Stability of each } g_i) \\ &\geq h(g_1(t), \dots, g_k(t)) - \alpha\beta c && \text{(Lipschitz continuity of } h) \\ &= g(t) - \alpha\beta c. \end{aligned}$$

As a consequence of the above proposition, a convex combination of β -stable functions is β -stable, and the point-wise maximum of β -stable functions is β -stable. ■

Consequences of Noise Stability and Lipschitz Continuity

We now state the two main results of our framework. Both results apply to functions $g : [-1, 1]^n \rightarrow [-1, 1]$ which are simultaneously Lipschitz continuous and noise stable, and $n \times m$ matrices A with entries in $[-1, 1]$. Given a vector $x \in \Delta_m$ and integer $s > 0$, we view x as a probability distribution over $[m]$, and use the random variable $\tilde{x} \in \Delta_m$ to denote the empirical distribution of s i.i.d. samples from x . Formally, $\tilde{x} = \frac{1}{s} \sum_{i=1}^s e_{k_i}$, where $k_1, \dots, k_s \in [m]$ are drawn i.i.d. according to x , and $e_j \in \Delta_m$ denotes the j^{th} standard basis vector. Since \tilde{x} is the average of s standard basis vectors, we say it is s -uniform.

Definition 2.4. *We refer to a distribution $y \in \Delta_m$ as s -uniform if and only if it is the average of a multiset of s standard basis vectors in m -dimensional space.*

Our first result shows that when the number of samples s is chosen as a suitable function of the Lipschitz continuity and noise stability parameters, $g(A\tilde{x})$ is not much smaller than $g(Ax)$ in expectation over \tilde{x} . At a high level, we bound this difference as a sum of two error terms: one accounts for the effect of low-probability large errors in the inputs $\tilde{t} = A\tilde{x}$ to g , and the other accounts for effect of higher-probability small errors in the inputs \tilde{t} . The former error term is bounded using noise stability, and the latter error term is bounded using Lipschitz continuity.

Theorem 2.5. *Let $g : [-1, 1]^n \rightarrow [-1, 1]$ be β -stable and c -Lipschitz in L^∞ , let A be an $n \times m$ matrix with entries in $[-1, 1]$, let $\alpha, \delta > 0$, and let $s \geq 2 \ln(\frac{2}{\alpha}) / \delta^2$ be an integer. Fix a vector $x \in \Delta_m$, and let the random variable \tilde{x} denote the empirical distribution of s i.i.d. samples from probability distribution x . The following then holds: $\mathbb{E}[g(A\tilde{x})] \geq g(Ax) - \alpha\beta - c\delta$.*

Proof: Denote $t = Ax$ and $\tilde{t} = A\tilde{x}$. Note that \tilde{t} is a random variable. Also note that t_i and \tilde{t}_i can be viewed as the mean and empirical mean, respectively, of a distribution supported on $A_{i,1}, \dots, A_{i,m} \in [-1, 1]$. We say the i^{th} entry of t is *approximately preserved* if $|t_i - \tilde{t}_i| \leq \delta$, and we say it is *corrupted* otherwise. Let $R \subseteq [n]$ denote the set of corrupted entries. Hoeffding's inequality, and our choice of the number of samples s , imply that R follows an α -Light distribution.

Let t' be such that (1) $t'_i = \tilde{t}_i$ for $i \in R$, and (2) $t'_i = t_i$ otherwise. Observe that $t' \underset{R}{\approx} t$, and $\|t' - \tilde{t}\|_\infty \leq \delta$. We can now bound the expected difference between $g(t)$ and $g(\tilde{t})$ as a sum of the error introduced by corrupted entries and the error introduced by the approximately preserved entries of t :

$$g(t) - \mathbb{E}[g(\tilde{t})] = \mathbb{E}[g(t) - g(t')] + \mathbb{E}[g(t') - g(\tilde{t})] \leq \alpha\beta + c\delta. \quad \blacksquare$$

Notice that if we fix the desired approximation error ϵ , the minimum required number of samples s in Theorem 2.5 to guarantee that $\mathbb{E}[g(A\tilde{x})] \geq g(Ax) - \epsilon$ is obtained by minimizing $\lceil 2 \ln(\frac{2}{\alpha})/\delta^2 \rceil$ over $\alpha, \delta > 0$ satisfying $\alpha\beta + \delta c \leq \epsilon$. Therefore, the required number of samples depends only on the error term ϵ , the noise stability parameter β , and the Lipschitz continuity parameter c ; in particular, it is independent of n and m .

As a corollary of Theorem 2.5, we derive the following algorithmic result.

Theorem 2.6. *Let $g : [-1, 1]^n \rightarrow [-1, 1]$ be β -stable and c -Lipschitz, and let $m > 0$ be an integer. For every $\delta, \alpha > 0$, the m -dimensional mixture selection problem for g admits an $(\alpha\beta + c\delta)$ -approximation algorithm in the additive sense, with runtime $n \cdot m^{O(\log(1/\alpha)/\delta^2)} \cdot T$, where T denotes the time needed to evaluate g on a single input.*

Proof: Let $s \geq 2 \ln(2/\alpha)/\delta^2$ be an integer. Our algorithm simply enumerates all s -uniform distributions, and outputs the one maximizing $g(Ax)$. This takes time $n \cdot m^{O(s)} \cdot T$. The approximation guarantee follows from Theorem 2.5 and the probabilistic method. \blacksquare

As a consequence of Theorem 2.6, the mixture selection problem for $g^{(\text{mid})}$ admits a polynomial-time approximation scheme (PTAS) in the additive sense. The same holds for every function g which is $O(1)$ -stable and $O(1)$ -Lipschitz continuous. Specifically, by setting $\alpha = \frac{\epsilon}{2\beta}$ and $\delta = \frac{\epsilon}{2c}$, an ϵ -approximation algorithm runs in time $n \cdot m^{O(c^2 \log(\beta/\epsilon)/\epsilon^2)} \cdot T$. Interestingly, neither noise stability nor Lipschitz continuity alone suffices for such a PTAS, as we argue in the next subsection.

The Necessity of Both Noise Stability and Lipschitz Continuity

We now present evidence that both our assumptions — Noise stability and Lipschitz continuity — appear necessary for general positive results along the lines of those in Theorem 2.6.

- 1) *Stability alone is not sufficient.* In Section 8.1, we define a function $g^{(\text{slope})} : [0, 1]^n \rightarrow [0, 1]$ which is 1-stable. Furthermore, $g^{(\text{slope})}$ is $O(1)$ -Lipschitz with respect to the L^1 metric, which is a weaker property than Lipschitz continuity with respect to L^∞ . We show in Theorem 8.4 that there is a polynomial-time reduction from the maximum independent set problem on n -node graphs to the n -dimensional mixture selection for $g^{(\text{slope})}$. Moreover, the reduction precludes a polynomial-time ϵ -approximation algorithm in the additive sense for some constant $\epsilon > 0$.
- 2) *Lipschitz continuity alone is not sufficient.* One might hope to prove NP-hardness of mixture selection in the absence of stability. However, we are out of luck in this regard: since every function $g : [-1, 1]^n \rightarrow [-1, 1]$ is $2n$ -stable, Theorem 2.6 implies a quasipolynomial-time approximation scheme in the additive sense whenever g is $O(1)$ -Lipschitz. Nevertheless, we prove hardness of approximation assuming the *planted clique conjecture* ([19] and [24]). More specifically, in Section 8.2 we exhibit a reduction from the planted k -clique problem to mixture selection for the 3-Lipschitz function $g_k^{(\text{clique})}(t) = t_{[k]} - t_{[k+1]} + t_{[n]}$. When $k = \omega(\log^2 n)$ and A is the adjacency matrix of an n -node undirected graph G , we show that $\max_x g_k^{(\text{clique})}(Ax) \approx 1$ with high probability if G contains a k -clique, and $\max_x g_k^{(\text{clique})}(Ax) \approx \frac{1}{2}$ with high probability if G is the Erdős-Rényi random graph $G(n, \frac{1}{2})$.

A Bi-criteria Extension of the Framework

We have already showed that in the absence of Lipschitz continuity, one can not hope for a PTAS in general. Motivated by two of our applications, namely *Optimal signaling in normal form games* and *Persuasion in voting*, we extend our framework to the design of approximation algorithms for mixture selection with a *bi-criteria guarantee* when the function in question is stable but not Lipschitz continuous. We first define a (δ, ρ) -relaxation of a function.

Definition 2.7. *Given two functions $g, h : [-1, 1]^n \rightarrow [-1, 1]$ and parameters $\delta, \rho \geq 0$, we say h is a (δ, ρ) -relaxation of g if for all $t_1, t_2 \in [-1, 1]^n$ with $\|t_1 - t_2\|_\infty \leq \delta$, $h(t_2) \geq g(t_1) - \rho$.*

In lieu of the Lipschitz continuity condition, we prove our bounds for a relaxation of the function.

Theorem 2.8. Let $g : [-1, 1]^n \rightarrow [-1, 1]$ be β -stable, let A be an $n \times m$ matrix with entries in $[-1, 1]$, let $\alpha > 0$ and $\delta, \rho \geq 0$, and let $s \geq 2 \ln(\frac{2}{\alpha})/\delta^2$ be an integer. Fix a vector $x \in \Delta_m$, and let the random variable \tilde{x} denote the empirical distribution of s i.i.d. samples from probability distribution x . The following then holds for any (δ, ρ) -relaxation h of g ,

$$\mathbb{E}[h(A\tilde{x})] \geq g(Ax) - \alpha\beta - \rho.$$

Proof: Because the proof is almost identical to the proof of Theorem 2.5, we just mention the necessary modifications. Again, let $t = Ax$, let $\tilde{t} = A\tilde{x}$, let $R \subseteq [n]$ denote the set of corrupted inputs, and let t' be such that $t'_i = \tilde{t}_i$ for $i \in R$ and $t'_i = t_i$ otherwise. Then

$$\begin{aligned} g(t) - \mathbb{E}[h(\tilde{t})] &= \mathbb{E}[g(t) - g(t')] + \mathbb{E}[g(t') - h(\tilde{t})] \\ &\leq \alpha\beta + \mathbb{E}[g(t') - h(\tilde{t})] \\ &\leq \alpha\beta + \rho, \end{aligned}$$

where the first inequality follows by noise stability of g , and the last inequality follows from the fact that h is a (δ, ρ) -relaxation of g . ■

Having replaced Theorem 2.5 by Theorem 2.8, a similar computational result as Theorem 2.6 can be inferred in the bi-criteria sense.

3 LOTTERY DESIGN FOR REVENUE MAXIMIZATION

To illustrate the utility of our framework, we start with a simple but basic open problem in Bayesian mechanism design posed by Dughmi, Han, and Nisan [10]. An instance of the *lottery design* problem is given by a valuation matrix $A \in [0, 1]^{n \times m}$ and n non-negative weights w_1, \dots, w_n with $\sum_{i=1}^n w_i = 1$. Here n denotes the number of buyer types, m denotes the number of items, and w represents a probability distribution over types. Each $0 \leq A_{i,j} \leq 1$ is the value of item j to a buyer of type i . The goal is to design a single lottery-price pair (x, p) , with $x \in \Delta_m$ and $p \geq 0$, so that the expected revenue of the auction which offers the lottery x over items at price p to a buyer with type drawn according to w is maximized. We assume the buyer is risk neutral, and therefore accepts the offer precisely if his type i satisfies $A_i x \geq p$. Consequently, our goal is to choose (x, p) maximizing $p \cdot \sum_{i=1}^n (w_i \cdot I[A_i x \geq p])$, where $I[\mathcal{E}]$ denotes the indicator function for the event \mathcal{E} .

The lottery design problem is closely related to the general unit-demand single-buyer mechanism design problem considered in [10], where the buyer's type is drawn from a common knowledge prior distribution \mathcal{B} given by a sampling oracle, and the buyer is to be presented with a *menu* consisting of several lottery-price pairs from which to choose. [10] frame this mechanism design problem as a computational task of “learning” a good mechanism by sampling from \mathcal{B} , and use the size of the menu as a regularization constraint in order to prevent over-fitting the mechanism to the sampled data. The problem of maximizing the expected revenue by using a menu of at most a given size is the main algorithmic question in [10], and its computational complexity is left largely open. When constrained to a menu with a single lottery, the goal is to choose (x, p) maximizing the expected revenue $p \cdot \Pr_{a \sim \mathcal{B}}[ax \geq p]$.

Using our mixture selection framework, we first give an additive PTAS for the lottery design problem when the value distribution \mathcal{B} is given explicitly by the matrix A and weights w as described above. We then extend our result to cases in which \mathcal{B} can only be accessed through sampling, using fairly standard uniform convergence arguments. In Section 8.3 (Theorem 8.9), we rule out an additive FPTAS for this problem, and in doing so provide complexity-theoretic evidence that our PTAS — for both lottery design in particular and mixture selection for stable and Lipschitz-continuous functions more generally — is essentially the best we can hope for.

3.1 Lottery Design in the Explicit Input Model

Given the number of buyer types n and weights $w \in \Delta_n$, the lottery design problem is simply mixture selection for the function $g_w^{(\text{lottery})}(t) : [0, 1]^n \rightarrow [0, 1]$, defined below.

$$g_w^{(\text{lottery})}(t) = \max_{p \geq 0} \left\{ p \cdot \sum_{i=1}^n (w_i \cdot I[t_i \geq p]) \right\} \quad (1)$$

As the first step to applying our framework, we show that $g_w^{(\text{lottery})}$ is noise stable. The high level idea is the following: if a subset of the inputs to $g_w^{(\text{lottery})}$ is corrupted, then in the worst case each such input exceeded the price p before corruption but not after corruption. This reduces the output of $g_w^{(\text{lottery})}$ by at most the total weight of corrupted inputs. When corrupted inputs are chosen according to an α -Light distribution, their expected total weight is bounded by α .

Lemma 3.1. *The function $g_w^{(\text{lottery})}$ is 1-stable.*

Proof: Let $t \in [0, 1]^n$ be an arbitrary input to $g_w^{(\text{lottery})}$. When t' is obtained from t by corrupting the entries corresponding to $R \subseteq [n]$, an event we denote by $t' \underset{R}{\approx} t$, it is easy to see that $g_w^{(\text{lottery})}(t') \geq g_w^{(\text{lottery})}(t) - w(R)$ where $w(R) = \sum_{i \in R} w_i$ denotes the total weight of corrupted entries. Moreover, when R is a random variable drawn from an α -Light distribution \mathcal{D} , we can bound the expected loss: $\mathbb{E}[w(R)] = \sum_{i=1}^n \Pr[i \in R] \cdot w_i \leq \sum_i \alpha w_i = \alpha$. It follows that $g_w^{(\text{lottery})}$ is 1-stable.

$$\begin{aligned} \mathbb{E}_{R \sim \mathcal{D}} \left[\min\{g_w^{(\text{lottery})}(t') : t' \underset{R}{\approx} t\} \right] &\geq g_w^{(\text{lottery})}(t) - \mathbb{E}_{R \sim \mathcal{D}} [w(R)] \\ &\geq g_w^{(\text{lottery})}(t) - \alpha. \end{aligned}$$

Next, we prove Lipschitz continuity. The high-level idea is the following: if all inputs to $g_w^{(\text{lottery})}$ decrease by δ , then we need only decrease the price p from expression (1) by δ . The (weighted) fraction of inputs exceeding the price is at least the same as before. ■

Lemma 3.2. *The function $g_w^{(\text{lottery})}$ is 1-Lipschitz continuous.*

Proof: Consider $t, t' \in [0, 1]^n$ with $\|t' - t\|_\infty \leq \delta$. We prove that $g_w^{(\text{lottery})}(t') \geq g_w^{(\text{lottery})}(t) - \delta$, and by symmetry it follows that $g_w^{(\text{lottery})}(t) \geq g_w^{(\text{lottery})}(t') - \delta$. Let p be the optimal price for t — i.e. the maximizer of expression (1) — and define $p' = \max\{0, p - \delta\}$. Whenever $t_i \geq p$ we have $t'_i \geq p'$, and therefore $\sum_{i=1}^n (w_i \cdot I[t'_i \geq p']) \geq \sum_{i=1}^n (w_i \cdot I[t_i \geq p])$. It follows that $g_w^{(\text{lottery})}(t') \geq g_w^{(\text{lottery})}(t) - \delta$. ■

Combining Lemmas 3.1 and 3.2 with Theorem 2.6 yields an additive PTAS for lottery design.

Theorem 3.3. *There is an additive PTAS for the lottery design problem when the valuation distribution is given explicitly by a matrix $A \in [0, 1]^{n \times m}$ and a weight vector $w \in \Delta_n$.*

3.2 Lottery Design in the Sample Oracle Model

So far, we have assumed that the buyer's type distribution is given explicitly. We now show how to extend our results to the *sample oracle model*. Specifically, we assume the buyer's type $a \in [0, 1]^m$ is drawn from a distribution \mathcal{B} given by a sampling oracle, and seek a randomized approximation scheme with runtime (and number of samples) polynomial in m for each desired approximation guarantee ϵ . To simplify exposition we assume \mathcal{B} has finite support, though our results hold more generally. As usual, our goal is to choose a lottery $x \in \Delta_m$ and a price $p \geq 0$ to maximize $\text{Rev}_{\mathcal{B}}(x, p) = p \cdot \Pr_{a \sim \mathcal{B}}[ax \geq p]$.

Theorem 3.4. *There is an additive polynomial-time randomized approximation scheme (PRAS) for the lottery design problem in the sample oracle model.*

Proof: Recall that the PTAS in Theorem 3.3 optimizes over all s -uniform m -dimensional lotteries, where s depends only on the desired approximation guarantee $\epsilon > 0$, and is bounded by a polynomial in $\frac{1}{\epsilon}$. In particular, s is independent of the number of types n , implying that the same approach of enumerating all s -uniform lotteries \tilde{x} would succeed in the sample oracle model were it not for our inability to evaluate $\max_p \text{Rev}_{\mathcal{B}}(\tilde{x}, p)$ exactly.

We overcome this difficulty by Monte Carlo sampling from \mathcal{B} . Given n types $a_1, \dots, a_n \in [0, 1]^m$ sampled from \mathcal{B} and presented as the rows of a matrix $A \in [0, 1]^{n \times m}$, we run the PTAS from Theorem 3.3 with approximation parameter ϵ on the empirical distribution given by A and uniform weights $w_i = \frac{1}{n}$ for all i . Taking n to be a suitable polynomial in $m, \frac{1}{\epsilon}$, and $\log(\frac{1}{\gamma})$ — where $\gamma > 0$ is a parameter — guarantees that $|\text{Rev}_{\mathcal{B}}(\tilde{x}, p) - p \cdot \frac{1}{n} \sum_{i=1}^n I[A_i \tilde{x} \geq p]| \leq \epsilon$ simultaneously for all s -uniform lotteries \tilde{x} and prices p with probability at least $1 - \gamma$. This follows from standard tail bounds and the union bound, coupled with a uniform convergence argument over prices $p \in [0, 1]$. Therefore, with probability $1 - \gamma$ our algorithm outputs a lottery-price pair whose expected revenue for a buyer drawn from \mathcal{B} is within $O(\epsilon)$ from the optimal. ■

4 SIGNALING

In the next few sections, we consider a number of Bayesian games in which a key parameter θ , the *state of nature*, in part determines the payoff structure of the game. We use Θ to denote the set of all states of nature, and assume $\theta \in \Theta$ is drawn from a common-knowledge prior distribution which we denote by λ . In all our applications, we assume players a-priori know nothing about θ other than its prior distribution λ , and examine policies whereby a principal with access to the realized value of θ may commit to a policy of revealing information to the players regarding θ . This is often referred to as *signaling* (see e.g. [12], [5], [11], [9]). We restrict our attention to symmetric signaling schemes, in which the principal must reveal the same information to all players in the game. Thus, a *symmetric signaling scheme* is given by a set Σ of *signals*, and a (possibly randomized) map φ from states of nature Θ to signals Σ . The goal of the principal, who is privy to

confidential state-of-nature information, is to boost her own objective by using a signaling scheme φ that optimally affects the outcome of the game.

In this section, in addition to providing the technical background for signaling schemes, we use our framework to define an abstract signaling problem and characterize its approximation complexity. This abstract problem captures the essence of all signaling problems considered in this paper.

4.1 Background: Signaling Schemes

Let $m = |\Theta|$. Abusing notation, we use $\varphi(\theta, \sigma)$ to denote the probability of announcing signal $\sigma \in \Sigma$ conditioned on the state of nature is $\theta \in \Theta$. It is well known ([22], [9]) that signaling schemes are in one-to-one correspondence with *convex decompositions* of the prior distribution $\lambda \in \Delta_m$: Formally, a signaling scheme $\varphi : \Theta \rightarrow \Sigma$ corresponds to the convex decomposition $\lambda = \sum_{\sigma \in \Sigma} \nu_\sigma \cdot \mu_\sigma$, where (1) $\nu_\sigma = \Pr_{\theta \sim \Theta}[\varphi(\theta) = \sigma] = \sum_{\theta \in \Theta} \lambda(\theta) \varphi(\theta, \sigma)$ is the probability of announcing signal σ , and (2) $\mu_\sigma(\theta) = \Pr_{\theta \sim \Theta}[\theta | \varphi(\theta) = \sigma] = \frac{\lambda(\theta) \varphi(\theta, \sigma)}{\nu_\sigma}$ is the *posterior belief distribution* of θ conditioned on signal σ . The converse is also true: every convex decomposition of $\lambda \in \Delta_m$ corresponds to a signaling scheme. Alternatively, the reader can view a signaling scheme φ as the $m \times |\Sigma|$ matrix of pairwise probabilities $\varphi(\theta, \sigma)$ satisfying conditions (1) and (2) with respect to $\lambda \in \Delta_m$.

Note that each posterior distribution $\mu \in \Delta_m$ defines a Bayesian game, and the principal's utility depends on the outcome of the game. Given a suitable equilibrium concept and selection rule, we let $f : \Delta_m \rightarrow \mathbb{R}$ denote the principal's utility as a function of the posterior distribution μ . For example, in an auction game $f(\mu)$ may be the social welfare or principal's revenue at the induced equilibrium, or any weighted combination of players' utilities, or something else entirely. The principal's objective as a function of the signaling scheme φ can be mathematically expressed by $F(\varphi, \lambda) = \sum_{\sigma \in \Sigma} \nu_\sigma \cdot f(\mu_\sigma)$.

In this setup, the optimal choice of a signaling scheme is related to the *concave envelope* f^+ of the function f ([22], [9]).¹ Specifically, such a signaling scheme achieves $\sum_{\sigma \in \Sigma} \nu_\sigma \cdot f(\mu_\sigma) = f^+(\lambda)$. Thus, there exists a signaling scheme with $m + 1$ signals that maximizes the principal's objective, by applying Caratheodory's theorem to the hypograph of f .

4.2 An Abstract Signaling Problem and its Polynomial-Time Approximation

To connect to our mixture selection framework, we consider signaling problems in which the principal's utility $f(\mu)$ from a posterior distribution $\mu \in \Delta_m$ can be written as $g(A\mu)$ for a function $g : [-1, 1]^n \rightarrow [-1, 1]$ and a matrix $A \in [-1, 1]^{n \times m}$. As described in Section 4.1, a signaling scheme φ with signals Σ corresponds to a family of probability-posterior pairs $\{(\nu_\sigma, \mu_\sigma)\}_{\sigma \in \Sigma}$ decomposing the prior $\lambda \in \Delta_m$ into a convex combination of posterior distributions (one per signal): $\lambda = \sum_{\sigma \in \Sigma} \nu_\sigma \mu_\sigma$. The objective of our signaling problem is then

$$F(\varphi) = \sum_{\sigma \in \Sigma} \nu_\sigma f(\mu_\sigma) = \sum_{\sigma \in \Sigma} \nu_\sigma g(A\mu_\sigma).$$

We note that this signaling problem can alternatively be written as an (infinite-dimensional) linear program which searches over probability measures supported on Δ_m with expectation λ . The separation oracle for the dual of this linear program is a mixture selection problem. Whereas we do not use this infinite-dimensional formulation nor its dual directly, we nevertheless show that the same conditions — noise stability and Lipschitz continuity — on the function g which lead to an approximation scheme for mixture selection also lead to a similar approximation scheme for our signaling problem with $f(\mu) = g(A\mu)$.

Lemma 4.1. *If g is β -stable and c -Lipschitz, then for any constants $\alpha, \delta > 0$, and for any integer $s \geq 2\delta^{-2} \ln(2/\alpha)$, there exists a signaling scheme $\tilde{\varphi}$ for which every posterior distribution is s -uniform, and $F(\tilde{\varphi}) \geq OPT - (\alpha\beta + c\delta)$ where OPT denotes the value of the optimal signaling scheme.*

Proof: Let $s \geq 2\delta^{-2} \ln(2/\alpha)$, and let $\tau \in [m^s]$ index all s -uniform posteriors, with $\tilde{\mu}_\tau$ denoting the τ 'th such posterior. For an arbitrary signaling scheme $\varphi = (\Sigma, \{(\nu_\sigma, \mu_\sigma)\}_{\sigma \in \Sigma})$, we show that each posterior μ_σ can be decomposed into s -uniform posteriors without degrading the objective by more than $\alpha\beta + c\delta$; more formally:

- 1) μ_σ can be expressed as a convex combination of s -uniform posteriors as follows.

$$\mu_\sigma = \sum_{\tau \in [m^s]} \tilde{\nu}_{\sigma, \tau} \tilde{\mu}_\tau \quad \text{with} \quad \tilde{\nu}_{\sigma, \tau} \in \Delta_{m^s}. \quad (2)$$

¹ f^+ is the point-wise lowest concave function h for which $h(x) \geq f(x)$ for all x in the domain. Equivalently, the hypograph of f^+ is the convex hull of the hypograph of f .

2) The value of objective function, i.e., $g(A\mu_\sigma)$, is decreased by no more than $\alpha\beta + c\delta$ through this decomposition,

$$\sum_{\tau \in [m^s]} \tilde{\nu}_{\sigma,\tau} \cdot g(A\tilde{\mu}_\tau) \geq g(A\mu_\sigma) - (\alpha\beta + c\delta). \quad (3)$$

The existence of such a decomposition follows from Theorem 2.5: Fix σ , and let $\tilde{\mu} \in \Delta_m$ be the empirical distribution of s i.i.d. samples from distribution $\mu_\sigma \in \Delta_m$. The vector $\tilde{\mu}$ is itself a random variable supported on s -uniform posteriors, its expectation is μ_σ , and by Theorem 2.5 we have $\mathbb{E}[g(A\tilde{\mu})] \geq g(A\mu_\sigma) - (\alpha\beta + c\delta)$. Therefore, by taking $\tilde{\nu}_{\sigma,\tau} = \Pr[\tilde{\mu} = \tilde{\mu}_\tau]$ for each $\tau \in [m^s]$ we get the desired decomposition of μ_σ .

The lemma follows by composing the decomposition φ with the decompositions of the posterior beliefs μ_σ to yield a signaling scheme $\tilde{\varphi}$ with only s -uniform posteriors and $F(\tilde{\varphi}) \geq F(\varphi) - (\alpha\beta + c\delta)$. Specifically, the signals of $\tilde{\varphi}$ are $\Sigma \times [m^s]$, where signal (σ, τ) has probability $\nu_\sigma \cdot \tilde{\nu}_{\sigma,\tau}$ and induces the posterior $\tilde{\mu}_\tau$.² Using Equation (2) and (3), it is easy to verify that this describes a valid signaling scheme with $F(\tilde{\varphi}) \geq F(\varphi) - (\alpha\beta + c\delta)$. ■

Lemma 4.1 permits us to restrict attention to s -uniform posteriors without much loss in our objective. Since there are only m^s such posteriors, a simple linear program with m^s variables computes an approximately optimal signaling scheme.

Theorem 4.2 (Polynomial-Time Signaling). *If g is β -stable and c -Lipschitz, then for any constant $\alpha, \delta > 0$, there exists a deterministic algorithm that constructs a signaling scheme with objective value at least $OPT - (\alpha\beta + c\delta)$, where OPT is the value of the optimal signaling scheme. Moreover, the algorithm runs in time $\text{poly}(m^{\delta^{-2} \ln(1/\alpha)}) \cdot n \cdot T$, where T denotes the time needed to evaluate g on a single input.*

Proof: Let s be an integer with $s \geq (2\delta^{-2} \ln(2/\alpha))$, denote $M = m^s$, and let $\{\mu_1, \dots, \mu_M\}$ be the set of all s -uniform posteriors. Lemma 4.1 shows that restricting to s -uniform posteriors only introduces an $\alpha\beta + c\delta$ additive loss in the objective. Thus it suffices to compute the optimal signaling scheme supported only on s -uniform posteriors. This can be done using the following linear program:

$$\begin{aligned} & \text{maximize} && \sum_{j \in [M]} \nu_j \cdot g(A\mu_j) \\ & \text{subject to} && \sum_{j \in [M]} \nu_j \mu_j = \lambda \\ & && \nu \in \Delta_M \end{aligned} \quad (4)$$

Note μ_j is the j 'th s -uniform posterior — the only variables in this LP are ν_1, \dots, ν_M . ■

Our proofs can be adapted to obtain a bi-criteria guarantee in the absence of Lipschitz continuity, as in Section 2. The following Theorem follows easily, and we omit the details.

Theorem 4.3 (Polynomial-Time Signaling (Bi-criteria)). *Let $g, h : [-1, 1]^n \rightarrow [-1, 1]$ be such that g is β -stable and h is a (δ, ρ) -relaxation of g , and let $\alpha > 0$ be a parameter. There exists a deterministic algorithm which, when given as input a matrix $A \in [-1, 1]^{n \times m}$ and a prior distribution $\lambda \in \Delta_m$, constructs a signaling scheme $\varphi = \{(\nu_\sigma, \mu_\sigma)\}_{\sigma \in \Sigma}$ such that*

$$\sum_{\sigma \in \Sigma} \nu_\sigma h(A\mu_\sigma) \geq OPT - \alpha\beta - \rho,$$

where OPT is the maximizer of $F(\varphi^*) = \sum_{\sigma \in \Sigma^*} \nu_\sigma^* g(A\mu_\sigma^*)$ over signaling schemes $\varphi^* = \{(\nu_\sigma^*, \mu_\sigma^*)\}_{\sigma \in \Sigma^*}$. Moreover, the algorithm runs in time $\text{poly}(m^{\delta^{-2} \ln(1/\alpha)}) \cdot n \cdot T$, where T denotes the time needed to evaluate h on a single input.

Remarks: We note that our proof suggests an extension of the result in Theorem 4.2 to cases in which f is given by a “black box” oracle, so long as we are promised that it is of the form $f(\mu) = g(A\mu)$. In this model the runtime of our algorithm does not depend on n , but instead depends on the cost of querying f . We also point out that that even though we precompute the quality of all m^s posteriors, we can guarantee that our output signaling scheme uses at most $m + 1$ signals; this is because LP (4) has only $m + 1$ constraints, and therefore admits an optimal solution where at most $m + 1$ variables are non-zero.

5 SIGNALING IN PROBABILISTIC SECOND-PRICE AUCTIONS

We examine signaling in probabilistic second-price auctions, as considered by Emek et al. [12] and Miltersen and Sheffet [5]. In this setting, the item being auctioned is probabilistic, and the instantiation of the item is known to the auctioneer but not to the bidders. The auctioneer commits to a signaling scheme for (partially) revealing information about the item for sale before subsequently running a second-price auction. We consider a probabilistic second-price auction described by the following parameters:

- An integer n denoting the number of bidders. We index the players by the set $[n] = \{1, \dots, n\}$.

²Note, however, that we can also “merge” all signals with the same posterior $\tilde{\mu}_\tau$ without loss.

- An integer m denoting the number of *states of nature*. We index states of nature by the set $\Theta = \{1, \dots, m\}$. Each $\theta \in \Theta$ represents a possible instantiation of the item being sold.
- A common-knowledge prior distribution $\lambda \in \Delta_m$ on the states of nature.
- A common-knowledge prior distribution \mathcal{D} on *valuation matrices* $\mathcal{V} \in [0, 1]^{n \times m}$, given either explicitly or as a “black-box” sampling oracle. For a valuation matrix \mathcal{V} , entry $\mathcal{V}_{i,j}$ denotes the value of player i for the item corresponding to state of nature j .

The game being played is the following: (a) The auctioneer first commits to a signaling scheme $\varphi : \Theta \rightarrow \Sigma$; (b) A state of nature $\theta \in \Theta$ is drawn according to λ and revealed to the auctioneer but not the bidders; (c) The auctioneer reveals a public signal $\sigma \sim \varphi(\theta)$ to all the bidders; (d) A valuation matrix $\mathcal{V} \in [0, 1]^{n \times m}$ is drawn according to \mathcal{D} , and each player i learns his value $\mathcal{V}_{i,j}$ for each potential item j ; (e) Finally, a second-price auction for the item is run.

As an example, consider an auction for an umbrella: the state of nature θ can be the weather tomorrow, which determines the utility $\mathcal{V}_{i,\theta}$ of an umbrella to player i . We assume that λ and \mathcal{D} are independent. We also emphasize that a bidder knows nothing about θ other than its distribution λ and the public signal σ , and the auctioneer knows nothing about \mathcal{V} besides its distribution \mathcal{D} prior to running the auction.

We adopt the (unique) dominant-strategy truth-telling equilibrium as our solution concept. Specifically, given a signaling scheme $\varphi : \Theta \rightarrow \Sigma$ and a signal $\sigma \in \Sigma$, in the subgame corresponding to σ it is a dominant strategy for player i to bid $\mathbb{E}_{\theta \sim \lambda}[\mathcal{V}_{i\theta} | \varphi(\theta) = \sigma]$ — his posterior expected value for the item conditioned on the received signal σ . Therefore the item goes to the player with maximum posterior expected value, at a price equal to the second-highest posterior expected value.

The algorithmic problem we consider is the one faced by the auctioneer in step (a) — namely computing an optimal signaling scheme — assuming the auctioneer looks to maximize expected revenue. It was shown in [12], [5] that polynomial-time algorithms exist for several special cases of this problem. However, the general problem was shown to be NP-hard even with 3 bidders — specifically, no additive FPTAS exists unless $P = NP$. In this section, we resolve the approximation complexity of this basic signaling problem by giving an additive PTAS. We note that variations of this problem were considered in [16], [17], with different constraints on the signaling scheme — the results in these works are not directly relevant to our model.

5.1 Revenue is Stable

Given a signaling scheme φ expressed as a decomposition $\{\nu_\sigma, \mu_\sigma\}_{\sigma \in \Sigma}$ of the prior distribution λ , we can express the auctioneer’s expected revenue as

$$\sum_{\sigma \in \Sigma} \nu_\sigma \mathbb{E}_{\mathcal{V} \sim \mathcal{D}} \max_2(\mathcal{V}\mu_\sigma),$$

where the function \max_2 returns the second largest entry of a given vector, i.e. $\max_2(t) = t_{[2]}$. To apply our main theorem, we need to show that the revenue in a subgame with posterior distribution $\mu \in \Delta_m$ — namely $\mathbb{E}_{\mathcal{V} \sim \mathcal{D}} \max_2(\mathcal{V}\mu)$ — can be written in the form $g(W\mu)$ for a matrix W . To facilitate our discussion we assume that the valuation distribution \mathcal{D} has finite support size C , though this is without loss of generality. Imagine we form a large matrix W by stacking matrices in the support of \mathcal{D} on top of each other. Formally, $W = [\mathcal{V}_1^T, \mathcal{V}_2^T, \dots, \mathcal{V}_C^T]^T$ where \mathcal{V}_i is the i th matrix in the support of \mathcal{D} . When matrix \mathcal{V}_i is drawn from \mathcal{D} , we take the second-highest bid from the rows of W corresponding to \mathcal{V}_i (rows $(i-1) \cdot n + 1$ to $i \cdot n$, where n is the number players). For $S \subseteq [nC]$ and $t \in [0, 1]^{nC}$, let $\max_{2_S}(t)$ denote the second-highest value among entries of t indexed by S . Then we can write the auctioneer’s expected revenue as

$$g^{(\text{rev})}(W\mu) = \mathbb{E}_{\mathcal{V} \sim \mathcal{D}} \max_{2_{S(\mathcal{V})}}(W\mu)$$

where $S(\mathcal{V})$ is the set of rows in W corresponding to \mathcal{V} .

Lemma 5.1 (Smooth and Stable Revenue). *The function $g^{(\text{rev})}(t) = \mathbb{E}_{\mathcal{V} \sim \mathcal{D}} \max_{2_{S(\mathcal{V})}}(t)$ is 1-Lipschitz and 2-stable.*

Proof: Because \max_{2_S} is 1-Lipschitz for a fixed set of indices S , it follows that $g^{(\text{rev})}$, which is a convex combination of these 1-Lipschitz functions, is also 1-Lipschitz.

To show that $g^{(\text{rev})}$ is stable, we first show that the function $\max_2 : [0, 1]^n \rightarrow [0, 1]$ is stable. Given $t \in [0, 1]^n$ and a random set $R \subseteq [n]$ drawn from an α -light distribution \mathcal{D} , the union bound implies that R includes neither of the two largest entries of t with probability at least $1 - 2\alpha$. In this case, the value of \max_2 is not affected by corruption of the entries indexed by R . Hence

$$\mathbb{E}_{R \sim \mathcal{D}} \left[\min\{\max_2(t') : t' \underset{R}{\approx} t\} \right] \geq (1 - 2\alpha) \cdot \max_2(t) + 2\alpha \cdot 0 \geq \max_2(t) - 2\alpha.$$

Therefore \max_2 is 2-stable, which implies that $\max_{2_S} : [0, 1]^{n_C} \rightarrow [0, 1]$ is also 2-stable for any fixed set of indices S . The function $g^{(\text{rev})}$ is a convex combination of functions of the form \max_{2_S} , and is therefore also 2-stable by Proposition 2.3. ■

Theorem 5.2. *The revenue-maximizing signaling problem in probabilistic second-price auctions admits an additive PTAS when the valuation distribution is given explicitly, and an additive PRAS when the valuation distribution is given by a sampling oracle.*

Proof: Lemma 5.1 shows that the function $g^{(\text{rev})}$ is 2-stable and 1-Lipschitz. If the valuation distribution \mathcal{D} is explicitly given with support size C , the function $g^{(\text{rev})}$ can be evaluated in $\text{poly}(n, m, C)$ time. Then for any $\epsilon > 0$, it follows from Theorem 4.2 by setting $\alpha = \epsilon/4$ and $\delta = \epsilon/2$ that there is a deterministic algorithm that computes a signaling scheme with expected revenue $\text{OPT} - \epsilon$, in time $\text{poly}(n, m^{\epsilon^{-2} \ln(1/\epsilon)}, C)$.

If \mathcal{D} is given via a sampling oracle, standard tail bounds and the union bound imply that $C = \Theta((s \log m + \log(\gamma^{-1}))/\epsilon^2)$ samples from \mathcal{D} suffice to estimate to within $O(\epsilon)$ the revenue associated with every s -uniform posterior in Δ_m , with success probability $1 - \gamma$. Since revenue is $O(1)$ -stable and $O(1)$ -Lipschitz, Lemma 4.1 implies that we can restrict attention to signaling schemes with s -uniform posteriors for $s = \text{poly}(\frac{1}{\epsilon})$. Proceeding as in Theorem 4.2, using the revenue estimates from Monte-Carlo sampling in lieu of exact values, we can construct a signaling scheme with revenue $\text{OPT} - \epsilon$ in time $\text{poly}(n, m^{\epsilon^{-2} \ln(1/\epsilon)}, \log(\frac{1}{\gamma}))$, with success probability $1 - \gamma$. ■

6 PERSUADING VOTERS

In this section, we apply our mixture selection framework to natural signaling problems encountered in the context of social choice, as introduced by Alonso and Câmara [3]. Consider an election with two possible outcomes, ‘Yes’ and ‘No’. For example, voters may need to choose whether to adopt a new law or social policy; board members of a company may need to decide whether to invest in a new project; and members of a jury must decide whether a defendant is declared guilty or not guilty. As in [3], we focus on the scenario in which voters have uncertainty regarding their utilities for the two possible outcomes (e.g., the risks and rewards of the new project). Specifically, voters’ utilities are parameterized by an a-priori unknown state of nature θ drawn from a common-knowledge prior distribution. We adopt the perspective of a principal with access to the realization of θ , and looking to influence the outcome of the election by signaling.

Formally, we consider a voting setting with n voters and m states of nature. We index the voters by the set $[n] = \{1, \dots, n\}$, and states of nature by the set $\Theta = \{1, \dots, m\}$. We assume voters’ preferences are given by a matrix $U \in [-1, 1]^{n \times m}$, where $U_{i,j}$ denotes voter i ’s utility in the event of a ‘Yes’ outcome in state of nature j . Without loss of generality, we assume utilities are normalized so that each voter’s utility for a ‘No’ outcome is 0 in each state of nature. A voter i who believes that the state of nature follows a distribution $\mu \in \Delta_m$ has expected utility $u(i, \mu) = \sum_{j \in \Theta} U_{i,j} \mu_j$ for a ‘Yes’ outcome. In most voting systems with a binary outcome, including for example threshold voting rules, it is a dominant strategy to vote ‘Yes’ if the utility $u(i, \mu)$ is at least 0 and ‘No’ otherwise. For our approximation algorithms, we also allow implementation in approximate dominant strategies — i.e., we sometimes assume a voter votes ‘Yes’ if his utility $u(i, \mu)$ is at least $-\delta$ for a small parameter δ .³ We assume that the state of nature $\theta \in \Theta$ is drawn from a common prior $\lambda \in \Delta_m$, and a principal with access to θ reveals a public signal σ prior to voters casting their votes. As usual, we adopt the perspective of a principal looking to commit to a signaling scheme $\varphi : \Theta \rightarrow \Sigma$, for some set of signals Σ .

Alonso and Câmara [3] consider a principal interested in maximizing the probability that at least 50% (or some given threshold) of the voters vote ‘Yes’, in expectation over states of nature. They characterize optimal signaling schemes analytically, though stop short of prescribing an algorithm for signaling. Theirs is the natural objective when the election employs a majority (or threshold) voting rule, and the principal is interested in influencing the outcome of the vote. Approximating this objective requires nontrivial modifications to our framework, and therefore we begin this section by examining a different, yet also natural, objective: the expected number of ‘Yes’ votes. We design a bi-criteria approximation scheme for this objective, then describe the necessary modifications for the threshold function objective of [3].

6.1 Maximizing Expected Number of Votes

We now examine bi-criteria approximation algorithms for maximizing the expected number of ‘Yes’ votes. For our benchmark, we use the function $g^{(\text{vote-sum})}(t) := \sum_{i \in [n]} \frac{1}{n} I[t_i \geq 0]$, where $I[\mathcal{E}]$ denotes the indicator function for event \mathcal{E} . Assuming voters vote ‘Yes’ precisely when their posterior expected utility for a ‘Yes’ outcome is nonnegative, the number of ‘Yes’ votes when voters have preferences $U \in [-1, 1]^{n \times m}$ and posterior belief $\mu \in \Delta_m$ equals $g^{(\text{vote-sum})}(U\mu)$. When the state of nature is distributed according to a common prior λ , and voters are informed according to signaling φ inducing a

³Such relaxations seem necessary for our results. Moreover, depending on the context, modes of intervention for shifting the votes of voters who are close to being indifferent may be realistic.

decomposition $\{\mu_\sigma, \nu_\sigma\}_{\sigma \in \Sigma}$ of λ , the expected number of ‘Yes’ votes equals $F^{(\text{vote-sum})}(\varphi, U, \lambda) := \sum_{\sigma \in \Sigma} \nu_\sigma g^{(\text{vote-sum})}(U\mu_\sigma)$. We use $\text{OPT}^{(\text{vote-sum})}(U, \lambda)$ to denote the maximum value of $F^{(\text{vote-sum})}(\varphi, U, \lambda)$ over signaling schemes φ .

As the first step to apply our framework, we prove that $g^{(\text{vote-sum})}$ is stable.

Lemma 6.1. *The function $g^{(\text{vote-sum})}$ is 1-stable.*

Proof: For each voter $i \in [n]$, let $g_i : [-1, 1]^n \rightarrow \{0, 1\}$ be the function indicating whether voter i prefers the ‘Yes’ outcome, i.e., $g_i(t) = I[t_i \geq 0]$. Each individual g_i is 1-stable, because as long as the i ’th input t_i is not corrupted the output of g_i does not change. Therefore $g^{(\text{vote-sum})}(t) = \frac{1}{n} \sum_{i=1}^n g_i(t)$, being a convex combination of 1-stable functions, is 1-stable by Proposition 2.3. \blacksquare

Unfortunately, $g^{(\text{vote-sum})}$ is not $O(1)$ -Lipschitz. We therefore employ the bi-criteria extension to our framework from Definition 2.7. Specifically, for a parameter $\delta > 0$, we assume a voter votes ‘Yes’ as long as his expected utility from a ‘Yes’ outcome is at least $-\delta$. Correspondingly, we define the relaxed function $g_\delta^{(\text{vote-sum})}(t) := \sum_{i \in [n]} \frac{1}{n} I[t_i \geq -\delta]$; the expected number of ‘Yes’ votes from a signaling scheme $\varphi = \{\mu_\sigma, \nu_\sigma\}_{\sigma \in \Sigma}$ can analogously be written as $F_\delta^{(\text{vote-sum})}(\varphi, U, \lambda) := \sum_{\sigma \in \Sigma} \nu_\sigma g_\delta^{(\text{vote-sum})}(U\mu_\sigma)$.

It is easy to verify that $g_\delta^{(\text{vote-sum})}$ is an $(\delta, 0)$ -relaxation of $g^{(\text{vote-sum})}$; combining this fact with Theorem 4.3 yields a bi-criteria approximation scheme for the problem of maximizing the expected number of ‘Yes’ votes.

Theorem 6.2. *Let $\epsilon, \delta > 0$ be parameters, let $U \in [-1, 1]^{n \times m}$ describe the preferences of n voters in m states of nature, and let $\lambda \in \Delta_m$ be the prior of states of nature. There is an algorithm with runtime $\text{poly}(m^{\frac{\ln(1/\epsilon)}{\delta^2}}, n)$ for computing a signaling scheme φ such that $F_\delta^{(\text{vote-sum})}(\varphi, U, \lambda) \geq \text{OPT}^{(\text{vote-sum})}(U, \lambda) - \epsilon$.*

Using the same techniques as in Section 3.2, we can extend this result to the case where the valuations of voters are drawn from a distribution give either explicitly or by a sampling oracle. We omit the details.

6.2 Maximizing Probability of a Majority Vote

We now sketch the necessary modifications when the principal is interested in maximizing the probability of a ‘Yes’ outcome, assuming a majority voting rule. We make two relaxations, which appear necessary for our framework: we assume a voter votes ‘Yes’ as long as his expected utility from a ‘Yes’ outcome is at least $-\delta$, and assume that the ‘Yes’ outcome is attained when at least a $(0.5 - \delta)$ fraction of voters vote ‘Yes’. Our benchmark will be the maximum probability of a ‘Yes’ outcome in the absence of these two relaxations.

We define our benchmark using the function $g^{(\text{vote-thresh})}(t) = I[g^{(\text{vote-sum})}(t) \geq 0.5]$ which evaluates to 1 if at least half of its n inputs are nonnegative, and to 0 otherwise. This function is not $O(1)$ -stable, so we work with a more stringent benchmark which is. Specifically, for a parameter $\delta > 0$, we use the function $g_\delta^{(\text{vote-smooth-thresh})}$ which is pointwise greater than or equal to $g^{(\text{vote-thresh})}$, defined as follows:

$$g_\delta^{(\text{vote-smooth-thresh})}(t) = \begin{cases} \frac{1}{\delta} (g^{(\text{vote-sum})}(t) - 0.5 + \delta) & \text{if } g^{(\text{vote-sum})}(t) \in [0.5 - \delta, 0.5] \\ g^{(\text{vote-thresh})}(t) & \text{otherwise.} \end{cases}$$

Observe that $g_\delta^{(\text{vote-smooth-thresh})}$ applies a continuous piecewise-linear function to the output of $g^{(\text{vote-sum})}$. It is easy to verify that $g_\delta^{(\text{vote-smooth-thresh})}$ is $\frac{1}{\delta}$ -stable, and upperbounds $g^{(\text{vote-thresh})}$.

Finally, to measure the quality of our output we define the relaxed function $g_\delta^{(\text{vote-thresh})} : [-1, 1]^n \rightarrow \{0, 1\}$, which outputs 1 if at least a $(0.5 - \delta)$ fraction of its inputs exceed $-\delta$, and outputs 0 otherwise. By Definition 2.7, $g_\delta^{(\text{vote-thresh})}$ is a $(\delta, 0)$ -relaxation of $g_\delta^{(\text{vote-smooth-thresh})}$ (and, consequently, also of $g^{(\text{vote-thresh})}$).

As usual, let $F^{(\text{vote-thresh})}(\varphi, U, \lambda)$ and $F_\delta^{(\text{vote-thresh})}(\varphi, U, \lambda)$ denote the functions which evaluate the quality of a signaling φ scheme using $g^{(\text{vote-thresh})}$ and $g_\delta^{(\text{vote-thresh})}$, respectively. Moreover, let $\text{OPT}^{(\text{vote-thresh})}(U, \lambda)$ be the maximum value of $F^{(\text{vote-thresh})}(\varphi, U, \lambda)$ over signaling schemes φ . We apply Theorem 4.3 to $g_\delta^{(\text{vote-thresh})}$ and $g_\delta^{(\text{vote-smooth-thresh})}$, setting $\alpha = \epsilon\delta$, and use the fact that $g_\delta^{(\text{vote-smooth-thresh})}$ upperbounds our true benchmark $g^{(\text{vote-thresh})}$, to conclude the following.

Theorem 6.3. *Let $\epsilon, \delta > 0$ be parameters, let $U \in [-1, 1]^{n \times m}$ describe the preferences of n voters in m states of nature, and let $\lambda \in \Delta_m$ be the prior of states of nature. There is an algorithm with runtime $\text{poly}(m^{\frac{\ln(1/\epsilon\delta)}{\delta^2}}, n)$ for computing a signaling scheme φ such that $F_\delta^{(\text{vote-thresh})}(\varphi, U, \lambda) \geq \text{OPT}^{(\text{vote-thresh})}(U, \lambda) - \epsilon$.*

6.3 Connection to Maximum Feasible Subsystem of Linear Inequalities

Turning our attention away from signaling, we note that $g^{(\text{vote-sum})}(Ax)$ simply counts the number of satisfied inequalities in the system $Ax \geq 0$. Mixture selection for $g^{(\text{vote-sum})}$ is therefore the problem of maximizing the number of satisfied inequalities over the simplex. Using our framework from Section 2, we obtain a bi-criteria PTAS for this problem. Moreover, using Monte-Carlo sampling, our bi-criteria PTAS extends to the model in which A is given implicitly; specifically, the rows of A correspond to the sample space of a distribution \mathcal{D} over $[-1, 1]^m$, and are weighted accordingly. In this implicit model, we can think of mixture selection for $g^{(\text{vote-sum})}$ as the problem of finding $x \in \Delta_m$ which maximizes the probability that $a \cdot x \geq 0$ for $a \sim \mathcal{D}$.

Motivated by systems applications, Daskalakis et al. [7] consider a special case of this problem termed *Fault-Tolerant Distributed Storage*. Their problem is equivalent to mixture selection for $g^{(\text{vote-sum})}$ in the implicit model, with the additional restriction that \mathcal{D} is a product distribution over binary vectors with marginal probabilities given explicitly. They present an additive EPTAS for this problem in a uni-criteria sense. Our framework relaxes their restrictions on \mathcal{D} , at the cost of a bi-criteria guarantee and exponential dependence on the error parameters.

7 SIGNALING IN BAYESIAN NORMAL FORM GAMES

We consider normal form games of incomplete information, in which payoffs are parameterized by a state of nature θ . A principal has access to the exact realization of θ , whereas the players initially share a prior belief on θ and form a posterior belief based on the information revealed by the principal. The goal of the principal is then to commit to revealing certain information about θ — i.e., a signaling scheme — to induce a favorable equilibrium over the resulting Bayesian subgames.

Signaling in normal form games has recently been examined from a complexity-theoretic perspective. Dughmi [9] considered the special case of two-player zero-sum games, and examined the design of symmetric signaling schemes with the goal of maximizing the expected utility of one of the players. It was shown that no FPTAS is possible for the signaling problem for zero sum games, assuming the planted clique conjecture. In this section, we complement the impossibility result of [9] with a bi-criteria quasi-polynomial time approximation scheme (QPTAS) which applies to normal form games with a constant number of players, slightly relaxing both the equilibrium definition and the polynomial-time restriction. It remains open if signaling for Bayesian zero-sum games admits a PTAS.

7.1 Background and Notation

We make heavy use of *tensors* in describing multi-player games. Specifically, we focus on n -dimensional tensors of order k with entries in $[-1, 1]$, where n is typically the number of strategies per player and k is the number of players. We think of these tensors as functions $T : [n]^k \rightarrow [-1, 1]$ mapping a strategy profile $s \in [n]^k$ to a number in $[-1, 1]$. Such a tensor T also naturally describes a multilinear map; overloading notation, given k vectors $x_1, \dots, x_k \in \mathbb{R}^n$ we write $T(x_1, \dots, x_k) = \sum_{s_1, \dots, s_k \in [n]} \left(T(s_1, \dots, s_k) \cdot \prod_{i=1}^k x_i(s_i) \right)$. This is most natural when $x_1, \dots, x_k \in \Delta_n$ form a mixed strategy profile, in which case $T(x_1, \dots, x_k)$ evaluates the expected value of T over pure strategy profiles drawn from (x_1, \dots, x_k) .

A Bayesian normal form game is defined by the following parameters:

- An integer k denoting the number of players, indexed by the set $[k] = \{1, \dots, k\}$.
- An integer n bounding the number of pure strategies of each player. Without loss of generality, we assume each player has exactly n pure strategies, and index them by the set $[n] = \{1, \dots, n\}$.
- An integer m denoting the number of *states of nature*. We index states of nature by the set $\Theta = \{1, \dots, m\}$, and use the variable θ to represent a state of nature.
- A *common prior distribution* $\lambda \in \Delta_m$ on states of nature.
- A family of payoff tensors $\mathcal{A}_i^\theta : [n]^k \rightarrow [-1, 1]$, one per player i and state of nature θ , where $\mathcal{A}_i^\theta(s_1, \dots, s_k)$ is the payoff to player i when the state of nature is θ and each player j plays strategy s_j .

Note that a game of complete information is the special case with $m = 1$ — i.e., the state of nature is fixed and known to all. In a general Bayesian normal form game, absent any information about the state of nature beyond the prior λ , risk neutral players will behave as in the complete information game $E_{\theta \sim \lambda}[\mathcal{A}^\theta]$. We consider signaling schemes which partially and symmetrically inform players by publicly announcing a signal σ , correlated with θ ; this induces a common posterior belief on the state of nature for each value of σ . When players' posterior belief over θ is given by $\mu \in \Delta_m$, we use \mathcal{A}^μ to denote the equivalent complete information game $E_{\theta \sim \mu}[\mathcal{A}^\theta]$. As shorthand, we use $\mathcal{A}_i^\mu(x_1, \dots, x_k)$ to denote $E[\mathcal{A}_i^\theta(s_1, \dots, s_k)]$ when $\theta \sim \mu \in \Delta_m$ and $s_i \sim x_i \in \Delta_n$. In the event that the state of nature is θ and players play the pure strategy profile s_1, \dots, s_k , we refer to the tuple $(\theta, s_1, \dots, s_k)$ as the *state of play*. For our result, we assume that a Bayesian game (\mathcal{A}, λ) is represented explicitly as a vector $\lambda \in \Delta_m$ and a list of tensors $\{\mathcal{A}_i^\theta \in [-1, 1]^{n^k} : i \in [k], \theta \in [m]\}$.

We adopt the approximate Nash equilibrium as our equilibrium concept. There are two variants.

Definition 7.1. Let $\epsilon \geq 0$. In a k -player n -action normal form game with expected payoffs in $[-1, 1]$ given by tensors $\mathcal{A}_1, \dots, \mathcal{A}_k$, a mixed strategy profile $x_1, \dots, x_k \in \Delta_n$ is an ϵ -Nash Equilibrium (ϵ -NE) if

$$\mathcal{A}_i(x_1, \dots, x_k) \geq \mathcal{A}_i(t_i, x_{-i}) - \epsilon$$

for every player i and alternative pure strategy $t_i \in [n]$.

Definition 7.2. Let $\epsilon \geq 0$. In a k -player n -action normal form game with expected payoffs in $[-1, 1]$ given by tensors $\mathcal{A}_1, \dots, \mathcal{A}_k$, a mixed strategy profile $x_1, \dots, x_k \in \Delta_n$ is an ϵ -well-supported Nash equilibrium (ϵ -WSNE) if

$$\mathcal{A}_i(s_i, x_{-i}) \geq \mathcal{A}_i(t_i, x_{-i}) - \epsilon$$

for every player i , strategy s_i in the support of x_i , and alternative pure strategy $t_i \in [n]$.

Clearly, every ϵ -WSNE is also an ϵ -NE. When $\epsilon = 0$, both correspond to the exact Nash Equilibrium. Note that we omitted reference to the state of nature in the above definitions — in a subgame corresponding to posterior beliefs $\mu \in \Delta_m$, we naturally use tensors $\mathcal{A}_1^\mu, \dots, \mathcal{A}_k^\mu$ instead.

Fixing an equilibrium concept (NE, ϵ -NE, or ϵ -WSNE), a Bayesian game (\mathcal{A}, λ) , and a signaling scheme $\varphi : \Theta \rightarrow \Sigma$, an *equilibrium selection rule* distinguishes an equilibrium strategy profile $(x_1^\sigma, \dots, x_k^\sigma)$ to be played in each subgame σ — we call the tuple $X = \{x_i^\sigma : \sigma \in \Sigma, i \in [k]\}$ a *Bayesian equilibrium* of the game (\mathcal{A}, λ) with signaling scheme φ . Together with the prior λ , the Bayesian equilibrium X induces a distribution $\Gamma \in \Delta_{\Theta \times [n]^k}$ over states of play — we refer to Γ as a *distribution of play*. This is analogous to implementation of allocation rules in traditional mechanism design.

Our results concern objectives which depend only on the state of play, and we seek to maximize the objective in expectation over the distribution of play. These include, but are not restricted to, the social welfare of the players, as well as weighted combinations of player utilities. Formally, our objective is described by a family of tensors $\mathcal{F}^\theta : [n]^k \rightarrow [-1, 1]$, one for each state of nature $\theta \in \Theta$. Equivalently, we may think of the objective as describing the payoffs of an additional player in the game — namely the principal. For a distribution μ over states of nature, we use $\mathcal{F}^\mu = \mathbb{E}_{\theta \sim \mu} \mathcal{F}^\theta$ to denote the principal's expected utility in a subgame with posterior beliefs μ , as a function of players' strategies.

For a signaling scheme φ and associated (approximate) equilibria $X = \{x_i^\sigma : \sigma \in \Sigma, i \in [k]\}$, our objective function can be written as $F(\varphi, X) = \mathbb{E}_{\theta \sim \lambda} \mathbb{E}_{\sigma \sim \varphi(\theta)} \mathbb{E}_{\vec{s} \sim x^\sigma} [\mathcal{F}(\theta, \vec{s})]$. When φ corresponds to a convex decomposition $\{(\mu_\sigma, \nu_\sigma)\}_{\sigma \in \Sigma}$ of the prior distribution, this can be equivalently written as $F(\varphi, X) = \sum_{\sigma \in \Sigma} \nu_\sigma \mathcal{F}^{\mu_\sigma}(x^\sigma)$. Let $OPT = OPT(\mathcal{A}, \lambda, \mathcal{F})$ denote the maximizer of $F(\varphi^*, X^*)$ over signaling schemes φ^* and (exact) Nash equilibria X^* . We seek a signaling scheme $\varphi : \Theta \rightarrow \Sigma$, as well as a Bayesian ϵ -NE (or ϵ -WSNE) X such that $F(\varphi, X) \geq OPT - \epsilon$.

We will use the following Lemma, which follows easily from the results of Lipton et al. [25], to restrict attention to equilibria with small support.

Lemma 7.3. Let tensors $\mathcal{A}_1, \dots, \mathcal{A}_k : [n]^k \rightarrow [-1, 1]$ describe a k -player game of complete information with n pure strategies per player, and let $\mathcal{F} : [n]^k \rightarrow [-1, 1]$ be a tensor describing an objective function on mixed strategies. Define the function $r(\epsilon) = \frac{3(k+1)^2 \ln((k+1)^2 n)}{\epsilon^2}$. For each $\epsilon > 0$, integer $s \geq r(\epsilon)$, and mixed strategy profile $x = (x_1, \dots, x_k)$, there is a profile $\tilde{x} = (\tilde{x}_1, \dots, \tilde{x}_k)$ of s -uniform mixed strategies such that $|\mathcal{A}_i(x) - \mathcal{A}_i(\tilde{x})| \leq \epsilon$ for all players i , $|\mathcal{F}(x) - \mathcal{F}(\tilde{x})| \leq \epsilon$, and if x is a Nash equilibrium of \mathcal{A} then \tilde{x} is an ϵ -equilibrium of \mathcal{A} . This holds for both NE and WSNE.

Proof: We can think of the tensor $\mathcal{F}^\mu : [n]^k \rightarrow [-1, 1]$ as describing the utility of an additional player in the game with a trivial strategy set. The rest follows from [25, Theorem 2]. ■

7.2 QPTAS for Signaling in Normal Form Games

We prove the following bi-criteria result.

Theorem 7.4. Let $\epsilon > 0$ denote an approximation parameter, let (\mathcal{A}, λ) be a Bayesian normal form game with $k = O(1)$ players, n actions, and m states of nature, and let $\mathcal{F} : [m] \times [n]^k \rightarrow [-1, 1]$ be an objective function given as a tensor. There is an algorithm with runtime $\text{poly}(m^{\frac{\ln(n/\epsilon)}{\epsilon^2}}, n^{\frac{\ln n}{\epsilon^2}})$ which outputs a signaling scheme φ and corresponding Bayesian ϵ -equilibria X satisfying $F(\varphi, X) \geq OPT(\mathcal{A}, \lambda, \mathcal{F}) - \epsilon$. This holds for both approximate NE and approximate WSNE.

In other words, when the number of players is a constant we can in quasi-polynomial time approximate the optimal reward from signaling while losing an additive ϵ in the objective as well as in the incentive constraints, as compared to the optimal signaling scheme / Nash equilibrium combination.

Fix $\epsilon > 0$. To prove this theorem, we define functions g and g_ϵ which each take as input a k -player n -action game of complete information \mathcal{B} , given as payoff tensors $\mathcal{B}_1, \dots, \mathcal{B}_k : [n]^k \rightarrow [-1, 1]$, and an objective tensor $\mathcal{G} : [n]^k \rightarrow [-1, 1]$, and output a number in $[-1, 1]$. Specifically, $g(\mathcal{B}, \mathcal{G}) = \max\{g(x) : x \in EQ(\mathcal{B})\}$ and $g_\epsilon(\mathcal{B}, \mathcal{G}) = \max\{g(x) : x \in EQ_\epsilon(\mathcal{B})\}$, where $EQ(\mathcal{B})$ denotes the set of Nash equilibria of the game \mathcal{B} , and $EQ_\epsilon(\mathcal{B})$ denotes the (non-empty) set of $\lceil r(\epsilon/4) \rceil$ -uniform ϵ -Nash equilibria (or ϵ -WSNE) for r as given in Lemma 7.3. Recall that $\mathcal{G}(x)$ denotes evaluating the multilinear map described by tensor \mathcal{G} at the mixed strategy profile $x \in \Delta_n^k$.

Now suppose we fix a Bayesian game (\mathcal{A}, λ) and objective tensor \mathcal{F} as in the statement of Theorem 7.4. For a subgame with a posterior distribution $\mu \in \Delta_m$ over states of nature, the principal's expected utility at the "best" Nash equilibrium of this subgame can be written as $g(\mathcal{A}^\mu, \mathcal{F}^\mu)$. Similarly, the principal's expected utility at the "best" $\lceil r(\epsilon/4) \rceil$ -uniform ϵ -NE (or ϵ -WSNE) can be written as $g_\epsilon(\mathcal{A}^\mu, \mathcal{F}^\mu)$. Observe that the input to both g and g_ϵ is a linear function of μ , as need to apply the results in Section 4. For a signaling scheme φ corresponding to a decomposition $\lambda = \sum_{\sigma \in \Sigma} \nu_\sigma \cdot \mu_\sigma$ of the prior distribution λ into posterior distributions (see Section 4.1), we can write the principal's expected utility assuming the "best" Nash equilibrium as $F(\varphi) = \sum_{\sigma \in \Sigma} \nu_\sigma g(F^{\mu_\sigma}, \mathcal{A}^{\mu_\sigma})$, and assuming the "best" $\lceil r(\epsilon/4) \rceil$ -uniform ϵ -equilibrium as $F_\epsilon(\varphi) = \sum_{\sigma \in \Sigma} \nu_\sigma g_\epsilon(F^{\mu_\sigma}, \mathcal{A}^{\mu_\sigma})$. We use OPT to denote the maximum value of F over all signaling schemes.

We prove Theorem 7.4 by exhibiting an algorithm for computing a signaling scheme φ such that $F_\epsilon(\varphi) \geq OPT - \epsilon$. The proof hinges on two main lemmas.

Lemma 7.5. *The function g is $2(k+1)n^k$ -stable.*

Proof: As noted in Section 2, any function mapping a hypercube $[-1, 1]^N$ to the interval $[-1, 1]$ is $2N$ stable. The function g is such a function with $N = (k+1)n^k$. ■

Lemma 7.6. *The function g_ϵ is an $(\epsilon/4, \epsilon/2)$ -relaxation of g .*

Proof: Consider tensors $\mathcal{G}, \tilde{\mathcal{G}} : [n]^k \rightarrow [-1, 1]$ with $|\mathcal{G}(s) - \tilde{\mathcal{G}}(s)| \leq \epsilon/4$ for all $s \in [n]^k$, and two k -player n -action games $\mathcal{B} = (\mathcal{B}_1, \dots, \mathcal{B}_k)$ and $\tilde{\mathcal{B}} = (\tilde{\mathcal{B}}_1, \dots, \tilde{\mathcal{B}}_k)$ with $|\mathcal{B}_i(s) - \tilde{\mathcal{B}}_i(s)| \leq \epsilon/4$ for all $s \in [n]^k$. It suffices to show that $g_\epsilon(\tilde{\mathcal{B}}, \tilde{\mathcal{G}}) \geq g(\mathcal{B}, \mathcal{G}) - \epsilon/2$. Let $x \in \Delta_n^k$ be the Bayesian equilibrium of \mathcal{B} for which $\mathcal{G}(x) = g(\mathcal{B}, \mathcal{G})$. By Lemma 7.3, there is a profile \tilde{x} of $\lceil r(\epsilon/4) \rceil$ -uniform mixed strategies such that \tilde{x} is an $\epsilon/4$ -equilibrium of \mathcal{B} , and $\mathcal{G}(\tilde{x}) \geq \mathcal{G}(x) - \epsilon/4$. Since $\tilde{\mathcal{B}}$ differs from \mathcal{B} by at most $\epsilon/4$ everywhere, it follows that \tilde{x} is an ϵ -equilibrium of $\tilde{\mathcal{B}}$, i.e. $\tilde{x} \in EQ_\epsilon(\tilde{\mathcal{B}})$. Similarly, since $\tilde{\mathcal{G}}$ differs from \mathcal{G} by at most $\epsilon/4$ everywhere, it follows that $\tilde{\mathcal{G}}(\tilde{x}) \geq \mathcal{G}(\tilde{x}) - \epsilon/4 \geq \mathcal{G}(x) - \epsilon/2$. We conclude that $g_\epsilon(\tilde{\mathcal{B}}, \tilde{\mathcal{G}}) \geq \tilde{\mathcal{G}}(\tilde{x}) \geq g(\mathcal{B}, \mathcal{G}) - \epsilon/2$. ■

We now complete the proof of Theorem 7.4 by instantiating Theorem 4.3 with $g, h = g_\epsilon$, and $\alpha = \frac{\epsilon}{4(k+1)n^k}$. The runtime is $\text{poly}(m^{\frac{\ln(1/\alpha)}{\epsilon^2}}, (k+1)n^k, T)$, where T is the time needed to evaluate g_ϵ (and compute the corresponding $\lceil r(\epsilon/4) \rceil$ -uniform ϵ -equilibrium) on a given input. Recall that $k = O(1)$ and $\alpha = \frac{\epsilon}{\text{poly}(n)}$. Moreover, using brute-force enumeration of all $\lceil r(\epsilon/4) \rceil$ -uniform mixed strategy profiles we conclude that T is bounded by a polynomial in $n^{\frac{\ln n}{\epsilon^2}}$. Therefore our total runtime is $\text{poly}(m^{\frac{\ln(1/\alpha)}{\epsilon^2}}, n^{\frac{\ln n}{\epsilon^2}})$, as needed.

Remarks: In the special case of two-player zero-sum games and a principal interested in maximizing one player's utility, as studied in [9], our techniques lead to a more efficient approximation scheme and a uni-criteria guarantee. This is because the principal's payoff tensor \mathcal{G} equals the payoff tensor \mathcal{B} of one of the players (say, player 1), and consequently the function $g(\mathcal{B}, \mathcal{G}) = g(\mathcal{B}, \mathcal{B}) = \max_x \min_y x^\top \mathcal{B} y$ is n^2 -stable and 2-Lipschitz. Its Lipschitz continuity follows from the fact that an ϵ -equilibrium of a zero-sum game leads to utilities within ϵ of the equilibrium utilities. Moreover, evaluating g now takes time $T = \text{poly}(m, n)$. Theorem 4.2 instantiated with $\alpha = \frac{\epsilon}{4n^2}$ and $\delta = \epsilon/4$, leads to an algorithm with runtime $\text{poly}(m^{\frac{\ln(n/\epsilon)}{\epsilon^2}}, n)$, which outputs a signaling scheme φ and corresponding Bayesian (exact) Nash-equilibria X satisfying $F(\varphi, X) \geq OPT(\mathcal{A}, \lambda, \mathcal{F}) - \epsilon$.

8 HARDNESS RESULTS

In this section, we present hardness results which justify our assumptions, and exhibit the limitations of our techniques. Specifically, we show in Sections 8.1 and 8.2 that neither stability nor Lipschitz continuity alone suffices for an additive PTAS. In Section 8.3, we show that even in the presence of Lipschitz continuity and noise stability, obtaining an additive FPTAS would imply $P = NP$.

8.1 NP-Hardness in the Absence of Lipschitz Continuity

We now show that stability alone does not suffice for an additive PTAS for mixture selection, in general. First, we show that mixture selection for the 1-stable function $g^{(\text{vote-sum})}$, presented in Section 6, does not admit a (uni-criteria) additive PTAS unless $P = NP$. Being that $g^{(\text{vote-sum})}$ is not continuous in any metric, we drive the point home by exhibiting a "smoothed"

function $g^{(\text{slope})}$ which is 1-stable and $O(1)$ -Lipschitz with respect to L^1 , but not $O(1)$ -Lipschitz with respect to L^∞ , and show that mixture selection for $g^{(\text{slope})}$ still does not admit an additive PTAS unless $P = NP$.

Both NP-hardness results share a similar reduction from the *maximum independent set* problem. We use a consequence of the result by [23], namely that there exists a constant ϵ such that it is NP-hard to approximate maximum independent set to within an additive error of ϵn , where n denotes the number of vertices.

Given an n -node undirected graph G , let $\text{OPT}_{\text{IS}} = \text{OPT}_{\text{IS}}(G)$ be the size of its largest independent set. We define the $n \times n$ matrix $A = A(G)$ as follows:

- Diagonal entries of A are all $\frac{1}{2}$ ($A_{i,i} = \frac{1}{2}$ for all $1 \leq i \leq n$).
- When vertices i and j share an edge in G , both $A_{i,j}$ and $A_{j,i}$ are -1 .
- All other entries of A , namely $A_{i,j}$ for non-adjacent distinct vertices i and j , are $-\frac{1}{4n}$.

We relate OPT_{IS} to convex combinations of the columns of A as follows.

Observation 8.1. *Let \mathcal{I} be an independent set of G with $|\mathcal{I}| = k$. There exists $x \in \Delta_n$ such that k entries of Ax are at least $\frac{1}{4n}$, and all remaining entries are strictly negative.*

Proof: Let $x \in \Delta_n$ be the normalized indicator vector of \mathcal{I} — i.e., $x_i = \frac{1}{k}$ if $i \in \mathcal{I}$ and $x_i = 0$ otherwise. By construction $(Ax)_i = \frac{1}{k}(\frac{1}{2} - (k-1)\frac{1}{4n}) \geq \frac{1}{4n}$ whenever $i \in \mathcal{I}$, and $(Ax)_i \leq -\frac{1}{4n}$ otherwise. ■

Observation 8.2. *For any $x \in \Delta_n$, nonnegative entries of Ax correspond to an independent set of G . Consequently, Ax can have at most OPT_{IS} nonnegative entries.*

Proof: Let $t = Ax$. Consider an edge $\{i, j\}$ of graph G , and without loss of generality assume that $x_i \geq x_j$. If $x_i = 0$, then $t_i \leq -\frac{1}{4n} < 0$ by construction. Otherwise, $t_j \leq \frac{x_j}{2} - x_i < 0$. Therefore, t_i and t_j cannot be both nonnegative. We conclude that the nonnegative coordinates of t correspond to an independent set of G . ■

Observations 8.1 and 8.2 imply that $\max_{x \in \Delta_n} g^{(\text{vote-sum})}(Ax) = \frac{\text{OPT}_{\text{IS}}}{n}$. Combined with the fact that obtaining an additive PTAS for the maximum independent set problem is NP-hard, we get the following theorem.

Theorem 8.3. *Mixture selection for the 1-stable function $g^{(\text{vote-sum})}$ admits no additive PTAS unless $P = NP$.*

Noting that $g^{(\text{vote-sum})}$ is a discontinuous function, for emphasis we exhibit a function $g^{(\text{slope})}$ which is Lipschitz continuous in L^1 (but not in L^∞) and 1-noise stable, but for which the same impossibility result holds by an identical reduction. Informally, $g^{(\text{slope})}$ “smooths” the threshold behavior of $g^{(\text{vote-sum})}$ as follows: each input t_i contributes 0 to $g^{(\text{slope})}(t)$ when $t_i \leq 0$, contributes $\frac{1}{n}$ when $t_i \geq \frac{1}{4n}$, and the contribution is a linear function of t_i increasing from 0 to $\frac{1}{n}$ for $t_i \in [0, \frac{1}{4n}]$. Formally, we define $g^{(\text{slope})}(t) = \sum_{i=1}^n \min\{4 \max\{0, t_i\}, \frac{1}{n}\}$. Since each entry of t contributes at most $\frac{1}{n}$ to $g^{(\text{slope})}(t)$, it is easy to verify that $g^{(\text{slope})}$ is 1-stable. Moreover, since the partial derivatives of $g^{(\text{slope})}(t)$ are upper-bounded by 4, it is 4-Lipschitz continuous with respect to the L^1 metric. Observations 8.1 and 8.2 imply that $\max_{x \in \Delta_n} g^{(\text{slope})}(Ax) = \frac{\text{OPT}_{\text{IS}}}{n}$, ruling out an additive PTAS for mixture selection for $g^{(\text{slope})}$.

Theorem 8.4. *The function $g^{(\text{slope})}$ is 1-stable and $O(1)$ -Lipschitz with respect to L^1 , and yet mixture selection for $g^{(\text{slope})}$ admits no additive PTAS unless $P = NP$.*

8.2 Planted-Clique Hardness in the Absence of Stability

We now present evidence that Lipschitz continuity alone does not suffice for a PTAS for mixture selection. Recalling that a quasipolynomial time algorithm follows from our framework whenever a function is $O(1)$ -Lipschitz, we reduce from the *planted clique problem*—for which a quasipolynomial time algorithm exists, and yet a polynomial-time algorithm is conjectured not to exist—rather than from an NP-hard problem.

In the planted clique problem, one must distinguish the n -node Erdős-Rényi random graph $\mathcal{G}(n, \frac{1}{2})$, in which each edge is included independently with probability $\frac{1}{2}$, from the graph $\mathcal{G}(n, \frac{1}{2}, k)$ formed by “planting” a clique in $\mathcal{G}(n, \frac{1}{2})$ at a randomly (or, equivalently, adversarially) chosen set of k nodes. This problem was first considered by Jerrum [19] and Kučera [24], and has been the subject of a large body of work since. A quasi-polynomial time algorithm exists when $k \geq 2 \log n$, and the best polynomial-time algorithms only succeed for $k = \Omega(\sqrt{n})$ (see e.g., [2] [8] [14] [6]). Several papers suggest that the problem is hard for $k = o(\sqrt{n})$ by ruling out natural classes of algorithmic approaches (e.g. [19], [13], [15]). The planted clique problem has therefore found use as a hardness assumption in a variety of applications (e.g. [1], [20], [18], [26], [9]). We use the following well-believed conjecture as our hardness assumption.

Assumption 8.5. *For some function $k = k(n)$ satisfying $k = \omega(\log^2 n)$ and $k = o(\sqrt{n})$, there is no probabilistic polynomial-time algorithm that can distinguish between a random graph drawn from $\mathcal{G}(n, \frac{1}{2})$ and a random graph drawn from $\mathcal{G}(n, \frac{1}{2}, k)$ with success probability $1 - o(1)$.*

We let $k = k(n)$ be as in Assumption 8.5, and consider mixture selection for the function $g_k^{(\text{clique})} : [0, 1]^n \rightarrow [0, 1]$ with $g_k^{(\text{clique})}(t) = t_{[k]} - t_{[k+1]} + t_{[n]}$, where $t_{[i]}$ denotes the i 'th largest entry of the vector t . It is easy to verify that $g_k^{(\text{clique})}$ is 3-Lipschitz with respect to the L^∞ metric, yet is not $O(1)$ -stable. We prove the following theorem.

Theorem 8.6. *Assumption 8.5 implies that there is no additive PTAS for mixture selection for $g_k^{(\text{clique})}$.*

To prove Theorem 8.6, we show that $\max_{x \in \Delta_n} g_k^{(\text{clique})}(Ax)$ is arbitrarily close to 1 with high probability when A is the adjacency matrix of $G \sim \mathcal{G}(n, \frac{1}{2}, k)$, and is bounded away from 1 with high probability when A is the adjacency matrix of $G \sim \mathcal{G}(n, \frac{1}{2})$. For convenience, and without loss of generality, we assume that both random graphs include each self-loop with probability $\frac{1}{2}$ — i.e., diagonal entries of the adjacency matrix A are independent uniform draws from $\{0, 1\}$ in both cases. Our argument is captured by the following two lemmas.

Lemma 8.7. *Fix a constant $\epsilon > 0$. Let $G \sim \mathcal{G}(n, \frac{1}{2}, k)$, and let A be its adjacency matrix. With probability $1 - o(1)$, there exists an $x \in \Delta_n$ such that $g_k^{(\text{clique})}(Ax) \geq 1 - \epsilon$.*

Proof: Let \mathcal{C} denote the vertices of the planted k -clique. We set $x_i = \frac{1}{k}$ if $i \in \mathcal{C}$ and 0 otherwise. Let $t = Ax$. For $i \in \mathcal{C}$, $t_i \geq 1 - \frac{1}{k}$. On the other hand, all other entries of t concentrate around $\frac{1}{2}$ with high probability. For $i \notin \mathcal{C}$, t_i is simply the average of k independent Bernoulli random variables by definition of $\mathcal{G}(n, \frac{1}{2}, k)$; using Hoeffding's inequality, we bound the probability that t_i deviates from its expectation by more than a constant $\delta > 0$, to be chosen later:

$$\Pr \left[\left| t_i - \frac{1}{2} \right| > \delta \right] \leq 2e^{-2\delta^2 k}.$$

By the union bound, $t_i \in [\frac{1}{2} - \delta, \frac{1}{2} + \delta]$ simultaneously for all $i \notin \mathcal{C}$ with probability at least $1 - 2\log n - \Omega(k) = 1 - o(1)$. Thus $t_{[k+1]} - t_{[n]} \leq 2\delta$ and $g_k^{(\text{clique})}(t) = t_{[k]} - (t_{[k+1]} - t_{[n]}) \geq 1 - \frac{1}{k} - 2\delta$ with probability $1 - o(1)$. Choosing $\delta = \epsilon/3$, we conclude that $g_k^{(\text{clique})}(t) \geq 1 - \epsilon$ with probability $1 - o(1)$. ■

Lemma 8.8. *Fix a constant $\epsilon > 0$. Let $G \sim \mathcal{G}(n, \frac{1}{2})$, and let A be its adjacency matrix. With probability $1 - o(1)$, $g_k^{(\text{clique})}(Ax) \leq \frac{3}{4} + \epsilon$ for all $x \in \Delta_n$.*

Proof: Recall that $g_k^{(\text{clique})}$ is $O(1)$ -Lipschitz and — like any other function from the hypercube to the bounded interval — $O(n)$ -stable. If there exists x^* such that $g_k^{(\text{clique})}(Ax^*) \geq \frac{3}{4} + \epsilon$, then Theorem 2.5 implies that there is an integer $s = O(\log n)$ and an s -uniform vector \tilde{x} such that $g_k^{(\text{clique})}(A\tilde{x}) > \frac{3}{4}$. There are n^s such vectors. We next show that for an arbitrary fixed vector $x \in \Delta_n$ the probability that $g_k^{(\text{clique})}(Ax) > \frac{3}{4}$ is at most $2^{-\Omega(k)}$. This would complete the proof by the union bound, since $1 - n^s \cdot 2^{-\Omega(k)} = 1 - o(1)$.

Fix $x \in \Delta_n$, and let $t = Ax$. Define \mathcal{D} as the distribution supported on $[0, 1]$ which is sampled as follows: draw a uniformly from $\{0, 1\}^n$, and output $a \cdot x$. Since A is the adjacency matrix of $G \sim \mathcal{G}(n, \frac{1}{2})$, each entry t_i of t can be viewed as an independent draw from \mathcal{D} . We exploit a key property of \mathcal{D} in our proof, namely the fact that \mathcal{D} is *symmetric* about $\frac{1}{2}$. Formally we mean that $\Pr_{\mathcal{D}}[r] = \Pr_{\mathcal{D}}[1 - r]$ for all $r \in [0, 1]$, and this follows easily from the definition of \mathcal{D} .

Symmetry of \mathcal{D} implies that $\Pr_{r \sim \mathcal{D}}[r \geq \frac{1}{2}] = \Pr_{r \sim \mathcal{D}}[r \leq \frac{1}{2}] \geq \frac{1}{2}$. Recalling that $k = o(n)$ and that entries of t are independent draws from \mathcal{D} , the Chernoff bound implies that the following holds with probability at least $1 - 2^{-\Omega(n)}$:

$$t_{[n]} \leq \frac{1}{2} \leq t_{[k+1]}. \quad (5)$$

If $g_k^{(\text{clique})}(t) > \frac{3}{4}$, then the following two conditions must hold:

- 1) $t_{[k]} > \frac{3}{4}$, and
- 2) $t_{[k+1]} - t_{[n]} < \frac{1}{4}$.

Condition 1 implies that the k largest entries of t are all at least $\frac{3}{4}$. Furthermore, unless Inequality (5) is violated — an event with small probability $2^{-\Omega(n)}$ — Condition 2 implies that remaining entries of t are all strictly between $\frac{1}{4}$ and $\frac{3}{4}$. Let p denote $\Pr_{r \sim \mathcal{D}}[r \leq \frac{1}{4}]$, also equal to $\Pr_{r \sim \mathcal{D}}[r \geq \frac{3}{4}]$ by symmetry of \mathcal{D} . The probability that k entries of t are at least $\frac{3}{4}$ and all remaining entries are in $(\frac{1}{4}, \frac{3}{4})$ is given by $\binom{n}{k} p^k (1 - 2p)^{n-k}$, which is maximized at $p = \frac{k}{2n}$, with maximum value $2^{-\Omega(k)}$. In summary, the probability that $g_k^{(\text{clique})}(Ax) > \frac{3}{4}$ is at most $2^{-\Omega(k)} + 2^{-\Omega(n)} = 2^{-\Omega(k)}$, as needed. ■

8.3 NP-hardness of an additive FPTAS

Our last hardness proof rules out an additive FPTAS for the lottery design problem, a.k.a. mixture selection for the function $g^{(\text{lottery})}$, as defined in Section 3. We restrict attention to the weight vector w assigning equal weight to all inputs to $g^{(\text{lottery})}$, and therefore omit references to the weight vector for the remainder of this section.

Theorem 8.9. *The lottery design problem, a.k.a. mixture selection for $g^{(\text{lottery})}$, admits no additive FPTAS unless $P = NP$.*

Proof: The proof involves a reduction from the independent set problem which is very similar to the reduction in Section 8.1, so we only detail the necessary modifications. Given an n -node undirected graph G , we define an $n \times n$ matrix $A = A(G)$ as in Section 8.1, though shifted and normalized so entries lie in $[0, 1]$. Specifically, we set diagonal entries of A to $\frac{3}{4}$, and we set an off-diagonal entry A_{ij} to 0 if i and j share an edge and to $\frac{1}{2} - \frac{1}{8n}$ otherwise. Observation 8.2 implies that for every $x \in \Delta_n$, entries of Ax which are at least $\frac{1}{2}$ correspond to an independent set of G . Observation 8.1 implies that there is a vector $x^* \in \Delta_n$ so that Ax^* has exactly $\text{OPT}_{\text{IS}}(G)$ entries no less than $p^* := \frac{1}{2} + \frac{1}{8n}$.

Our input to the lottery design problem will be a valuation matrix B with $8n^2 + n$ rows and n columns, obtained from A by adding $8n^2$ “dummy” rows, each of which is (p^*, \dots, p^*) . Setting a price of p^* for the lottery x^* results in expected revenue $r^* := p^* \cdot \frac{8n^2 + \text{OPT}_{\text{IS}}}{8n^2 + n} = (\frac{1}{2} + \frac{1}{8n}) \cdot \frac{8n^2 + \text{OPT}_{\text{IS}}}{8n^2 + n} > \frac{1}{2}$. Therefore, r^* lower-bounds $\max_{x \in \Delta_n} g^{(\text{lottery})}(Bx)$. We claim that r^* is in fact the maximum value of this mixture selection problem, and prove this by fixing an arbitrary lottery $x \in \Delta_n$ and conducting a simple case analysis on the associated price p :

- When $p^* < p \leq 1$, none of the “dummy” types purchase the lottery x , resulting in expected revenue at most $\frac{n}{8n^2 + n} < \frac{1}{2} < r^*$.
- When $\frac{1}{2} \leq p \leq p^*$, all dummy types purchase x , and the i 'th non-dummy type purchases x only if $(Ax)_i \geq \frac{1}{2}$. Since entries of Ax which are no less than $\frac{1}{2}$ correspond to an independent set of G , at most $8n^2 + \text{OPT}_{\text{IS}}$ types purchase the lottery x , resulting in expected revenue at most $p \cdot \frac{8n^2 + \text{OPT}_{\text{IS}}}{8n^2 + n} \leq r^*$.
- When $0 \leq p < \frac{1}{2}$, the best case scenario is that all buyer types purchase the lottery at price p , yielding expected revenue at most $p < \frac{1}{2} < r^*$.

Recalling that there exists a constant $\epsilon > 0$ for which the maximum independent set problem does not admit an additive ϵn -approximation algorithm in polynomial time, we conclude that mixture selection for $g^{(\text{lottery})}$ does not admit a polynomial-time additive ϵ' -approximation algorithm for any $\epsilon' = \epsilon'(n) \leq (\frac{1}{2} + \frac{1}{8n}) \frac{\epsilon}{8n+1}$. This rules out an additive FPTAS. ■

9 CONNECTION TO BOOLEAN FUNCTION ANALYSIS

In Definition 2.2, stability of a function from the solid hypercube to the bounded interval was defined as robustness to corruption patterns which follow a light distribution. In this section, we exhibit a closely-related algebraic notion of stability, based on Fourier analysis of Boolean functions. Instead of using light distributions directly, we describe the effects of corruption on a function g at input t in the solid hypercube by a Boolean function h_t . In particular, h_t takes as input a vector $z \in \{-1, 1\}^n$, interprets $z_i = 1$ as forbidding corruption of input t_i and $z_i = -1$ as permitting corruption of t_i , and considers the worst-case such corruption of t . Formally, denoting $\mathcal{I}^+(z) := \{i \in [n] : z_i = 1\}$ and $\mathcal{I}^-(z) := \{i \in [n] : z_i = -1\}$, we define Boolean extensions of g at an input t .

Definition 9.1 (Boolean extension). *Let $g : [-1, 1]^n \rightarrow [-1, 1]$, and let $t \in [-1, 1]^n$. A Boolean function $h_t : \{-1, 1\}^n \rightarrow [-1, 1]$ is called a Boolean extension of g at t if it satisfies:*

$$h_t(\mathbf{1}) = g(t) \tag{6}$$

$$h_t(z) \leq \min\{g(t') : t' \underset{\mathcal{I}^-(z)}{\approx} t\}. \tag{7}$$

Next we define Algebraic Stability for $g(t)$, using the notion of the Fourier transform of a Boolean function (e.g. [28]).

Definition 9.2 (Algebraic Stability). *A function $g(t) : [-1, 1]^n \rightarrow [-1, 1]$ is algebraically k -stable if for every $t \in [-1, 1]^n$ there exists a Boolean extension $h_t(z)$ at t such that:*

$$\widehat{h}_t(S) \geq 0 \text{ for all } S \subseteq [n] \tag{8}$$

$$\widehat{h}_t(S) = 0 \text{ for all } S \text{ such that } |S| > k, \tag{9}$$

where $\widehat{h}_t(S)$ is the Fourier coefficient of h_t at $S \subseteq [n]$.

In the parlance of Boolean function analysis, g is algebraically stable if for all t , the Fourier spectrum of h_t is both nonnegative and low-degree. We prove the following analogue of Theorem 2.5.

Theorem 9.3. *Let $g : [-1, 1]^n \rightarrow [-1, 1]$ be algebraically k -stable and c -Lipschitz in L^∞ , let A be an $n \times m$ matrix with entries in $[-1, 1]$, let $\delta, \epsilon > 0$, and let $s > \frac{1}{\delta^2} \cdot \log \frac{2k}{\epsilon}$ be an integer.*

Fix a vector $x \in \Delta_m$, and let the random variable \tilde{x} denote the empirical distribution of s i.i.d. samples from probability distribution x . The following then holds:

$$\mathbb{E}_{\tilde{x}}[g(A\tilde{x})] \geq g(Ax) - \epsilon - c\delta, \tag{10}$$

Proof sketch: Let $t = Ax$. Consider the random variable $\tilde{t} = A\tilde{x}$ where \tilde{x} is the (random) empirical distribution. Let E_i denote the event that $|\tilde{t}_i - t_i| \leq \delta$ for some $\delta \in (0, 1)$, i.e., that coordinate t_i is approximately preserved. Define a Boolean function $p_{\mathcal{D}} : \{-1, 1\}^n \rightarrow [0, 1]$,

$$p_{\mathcal{D}}(z) := \Pr_{\tilde{x}}[(\cap_{i \in \mathcal{I}^+(z)} E_i) \cap (\cap_{i \in \mathcal{I}^-(z)} \overline{E_i})]. \quad (11)$$

In particular, $p_{\mathcal{D}}(z)$ is the probability that *only* coordinates in $\mathcal{I}^+(z)$ are approximately preserved. It is easy to see that $\sum_{z \in \{-1, 1\}^n} p_{\mathcal{D}}(z) = 1$ and $\Pr[E_i] = \Pr[|\tilde{t}_i - t_i| \leq \delta] = \sum_{z: z_i=1} p_{\mathcal{D}}(z)$.

Next, we state some standard equalities for Boolean functions. Let $\mathbf{1}$ denote the all-one n -dimensional vector $(1, 1, \dots, 1)$ and h_t be a Boolean extension of g_t as stated in Definition 9.2.

$$\sum_{z \in \{-1, 1\}^n} h_t(z) \cdot p_{\mathcal{D}}(z) = 2^n \sum_{S \subseteq [n]} \widehat{p_{\mathcal{D}}}(S) \cdot \widehat{h_t}(S) \quad (12)$$

$$\sum_{S \subseteq [n]} \widehat{h_t}(S) = h_t(\mathbf{1}) \quad (13)$$

By Definition 9.1, $h_t(\mathbf{1}) = g(t)$. Since $g(t)$ is c -Lipschitz in L^∞ , we know:

$$\mathbb{E}_{\tilde{x}}[g(A\tilde{x})] = \mathbb{E}_{\tilde{t}}[g(\tilde{t})] \geq \sum_{z \in \{-1, 1\}^n} h_t(z) \cdot p_{\mathcal{D}}(z) - c\delta, \quad (14)$$

so it suffices to prove $\sum_{S \subseteq [n]} \widehat{h_t}(S) - 2^n \sum_{S \subseteq [n]} \widehat{p_{\mathcal{D}}}(S) \cdot \widehat{h_t}(S) \leq \epsilon$, which can be rewritten as:

$$\sum_{S: |S| \leq k} (1 - 2^n \widehat{p_{\mathcal{D}}}(S)) \widehat{h_t}(S) + \sum_{S: |S| > k} (1 - 2^n \widehat{p_{\mathcal{D}}}(S)) \widehat{h_t}(S) \quad (15)$$

As the latter term in (15) is zero by definition 9.2, we only need to upper bound $\sum_{S: |S| \leq k} (1 - 2^n \widehat{p_{\mathcal{D}}}(S)) \widehat{h_t}(S)$. Using a simple union-bound argument, one can prove that with sample size $s \geq \frac{1}{\delta^2} \log \frac{2k}{\epsilon}$, $\widehat{p_{\mathcal{D}}}(S) \geq \frac{1-\epsilon}{2^n}$ for all $S \subseteq [n]$ with $|S| \leq k$. As $\widehat{h_t}(S) \geq 0$ for S with $|S| \leq k$, we have $\sum_{S: |S| \leq k} (1 - 2^n \widehat{p_{\mathcal{D}}}(S)) \widehat{h_t}(S) \leq \epsilon \sum_{S \subseteq [n]} \widehat{h_t}(S) \leq \epsilon$. Combined with equation (14) we have $\mathbb{E}[g(A\tilde{x})] \geq g(t) - \epsilon - c\delta$. ■

9.1 Algebraic Stability for our Applications

We observe that algebraic stability holds for the objective functions in some our applications, and leave open whether it holds for all. Specifically, the relevant functions in lottery design, revenue maximizing signaling, and one of our two voting problems, are algebraically $O(1)$ -stable, and therefore an additive PTAS for each of those applications follows from Theorem 9.3 just as it did from Theorem 2.5. We list the Boolean extension associated with each of these applications.

Lemma 9.4 (Lottery design). *For a fixed price p , a Boolean extension of the objective function $g_{w,p}^{(lottery)}(t) := p \cdot \sum_{i=1}^n (w_i \cdot I[t_i \geq p])$ at t is $h_t^{(lottery)}(z) = p \cdot \sum_{i \in [n]} w_i \frac{z_i+1}{2} I[t_i \geq p]$. We enumerate over all prices p , up to a suitable discretization, in the algorithmic solution.*

Lemma 9.5 (Persuasion in voting). *A Boolean extension of $g^{(vote-sum)}$ at t is $h_t^{(vote-sum)}(z) = \sum_{i \in [n]} \frac{I[t_i \geq 0]}{n} \frac{z_i+1}{2}$.*

Lemma 9.6 (Revenue maximization). *A Boolean extension of the function $\max_2 : [0, 1]^n \rightarrow [0, 1]$ at t is $h_t^{(\max_2)}(z) = t_j \frac{z_i+1}{2} \frac{z_j+1}{2}$, where i and j denotes the indices of the largest and second-largest entry of t , respectively.*

REFERENCES

- [1] N. Alon, A. Andoni, T. Kaufman, K. Matulef, R. Rubinfeld, and N. Xie, “Testing k -wise and almost k -wise independence,” in *Proceedings of the 39th annual ACM Symposium on Theory of Computing (STOC)*, 2007, pp. 496–505.
- [2] N. Alon, M. Krivelevich, and B. Sudakov, “Finding a large hidden clique in a random graph,” in *Proceedings of the 9th ACM Symposium on Discrete Algorithms (SODA)*. Society for Industrial and Applied Mathematics, 1998, pp. 594–598.
- [3] R. Alonso and O. Câmara, “Persuading voters,” Manuscript, 2014.
- [4] S. Barman, “Approximating Nash equilibria and dense bipartite subgraphs via an approximate version of Caratheodory’s theorem,” in *Proceedings of the 47th Annual ACM on Symposium on Theory of Computing (STOC)*, 2015, pp. 361–369.
- [5] P. Bro Miltersen and O. Sheffet, “Send mixed signals: earn more, work less,” in *Proceedings of the 13th ACM Conference on Electronic Commerce (EC)*, 2012, pp. 234–247.

- [6] A. Coja-Oghlan, “Graph partitioning via adaptive spectral techniques,” *Combinatorics, Probability & Computing*, vol. 19, no. 2, p. 227, 2010.
- [7] C. Daskalakis, A. De, I. Diakonikolas, A. Moitra, and R. A. Servedio, “A polynomial-time approximation scheme for fault-tolerant distributed storage,” in *Proceedings of the 25th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2014, pp. 628–644.
- [8] Y. Dekel, O. Gurel-Gurevich, and Y. Peres, “Finding hidden cliques in linear time with high probability,” in *ANALCO*. SIAM, 2011, pp. 67–75.
- [9] S. Dughmi, “On the hardness of signaling,” in *Proceedings of the 55th IEEE Symposium on Foundations of Computer Science (FOCS)*, 2014.
- [10] S. Dughmi, L. Han, and N. Nisan, “Sampling and representation complexity of revenue maximization,” in *Proceedings of the 10th Conference on Web and Internet Economics (WINE)*, 2014.
- [11] S. Dughmi, N. Immorlica, and A. Roth, “Constrained signaling in auction design,” in *Proceedings of the 25th ACM Symposium on Discrete Algorithms (SODA)*, 2014.
- [12] Y. Emek, M. Feldman, I. Gamzu, R. Paes Leme, and M. Tennenholtz, “Signaling schemes for revenue maximization,” in *Proceedings of the 13th ACM Conference on Electronic Commerce (EC)*, 2012, pp. 514–531.
- [13] U. Feige and R. Krauthgamer, “The probable value of the Lovász–Schrijver relaxations for maximum independent set,” *SIAM Journal on Computing*, vol. 32, no. 2, pp. 345–370, 2003.
- [14] U. Feige and D. Ron, “Finding hidden cliques in linear time,” *DMTCS Proceedings*, no. 01, pp. 189–204, 2010.
- [15] V. Feldman, E. Grigorescu, L. Reyzin, S. Vempala, and Y. Xiao, “Statistical algorithms and a lower bound for detecting planted cliques,” in *Proceedings of the 45th annual ACM Symposium on Theory of Computing (STOC)*. ACM, 2013, pp. 655–664.
- [16] A. Ghosh, H. Nazerzadeh, and M. Sundararajan, “Computing optimal bundles for sponsored search,” in *Workshop on Internet and Network Economics (WINE)*, 2007.
- [17] M. Guo and A. Deligkas, “Revenue maximization via hiding item attributes,” in *Proceedings of the 23rd international joint conference on Artificial Intelligence*. AAAI Press, 2013, pp. 157–163.
- [18] E. Hazan and R. Krauthgamer, “How hard is it to approximate the best Nash equilibrium?” *SIAM Journal on Computing*, vol. 40, no. 1, pp. 79–91, 2011.
- [19] M. Jerrum, “Large cliques elude the Metropolis process,” *Random Structures & Algorithms*, vol. 3, no. 4, pp. 347–359, 1992.
- [20] A. Juels and M. Peinado, “Hiding cliques for cryptographic security,” *Designs, Codes and Cryptography*, vol. 20, no. 3, pp. 269–280, 2000.
- [21] J. Kahn, G. Kalai, and N. Linial, “The influence of variables on boolean functions,” in *Proceedings of the 29th Annual Symposium on Foundations of Computer Science*, ser. SFCS ’88, 1988, pp. 68–80.
- [22] E. Kamenica and M. Gentzkow, “Bayesian persuasion,” National Bureau of Economic Research, Tech. Rep., 2009.
- [23] S. Khot and R. Saket, “Hardness of finding independent sets in almost q -colorable graphs,” in *Proceedings of the 53rd IEEE Symposium on Foundations of Computer Science (FOCS)*, 2012, pp. 380–389.
- [24] L. Kučera, “Expected complexity of graph partitioning problems,” *Discrete Applied Mathematics*, vol. 57, no. 2, pp. 193–212, 1995.
- [25] R. Lipton, E. Markakis, and A. Mehta, “Playing large games using simple strategies,” in *Proceedings of the 4th ACM conference on Electronic commerce (EC)*, 2003, pp. 36–41.
- [26] L. Minder and D. Vilenchik, “Small clique detection and approximate Nash equilibria,” in *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*. Springer, 2009, pp. 673–685.
- [27] E. Mossel, R. O’Donnell, and K. Oleszkiewicz, “Noise stability of functions with low influences: Invariance and optimality,” *Annals of Mathematics*, vol. 171, no. 1, pp. 295–341, 2010.
- [28] R. O’Donnell, *Analysis of Boolean Functions*. Cambridge University Press, 2014.