Approximating Nash Equilibria in Normal-Form Games via Unbiased Stochastic Optimization

Anonymous Author(s)
Affiliation
Address
email

Abstract

We propose the first, to our knowledge, loss function for approximate Nash equilibria of normal-form games that is amenable to unbiased Monte Carlo estimation. This construction allows us to deploy standard non-convex stochastic optimization techniques for approximating Nash equilibria, resulting in novel algorithms with provable guarantees. We complement our theoretical analysis with experiments demonstrating that stochastic gradient descent can outperform previous state-of-the-art approaches.

1 Introduction

Nash equilibrium famously encodes stable behavioral outcomes in multi-agent systems and is arguably the most influential solution concept in game theory. Formally speaking, if n players independently choose n, possibly mixed, strategies (x_i for $i \in [n]$) and their joint strategy ($\boldsymbol{x} = \prod_i x_i$) constitutes a *Nash equilibrium*, then no player has any incentive to unilaterally deviate from their strategy. This concept has sparked extensive research in various fields, ranging from economics [30] to machine learning [16], and has even inspired behavioral theory generalizations such as quantal response equilibria which allow for more realistic models of boundedly rational agents [28].

Unfortunately, when considering Nash equilibria beyond the special case of the 2-player, zero-sum scenario, two significant challenges arise. First, it becomes unclear how a group of n independent players would collectively identify a Nash equilibrium when multiple equilibria are possible, giving rise to the *equilibrium selection* problem [18]. Secondly, even approximating a single Nash equilibrium is known to be computationally intractable and specifically PPAD-complete [11]. Combining both problems together, e.g., testing for the existence of equilibria with welfare greater than some fixed threshold is NP-hard and it is in fact even hard to approximate (i.e., finding a Nash equilibrium with welfare greater than ω for any $\omega > 0$, even when the best equilibrium has welfare $1 - \omega$) [2].

From a machine learning (ML) practitioner's perspective, however, such computational complexity results hardly give pause for thought as collectively we have become all too familiar with the unreasonable effectiveness of ML heuristics in circumventing such obstacles. Famously, non-convex optimization is NP-hard, even if the goal is to compute a local minimizer [31], however, stochastic gradient descent (and variants thereof) succeed in training models with billions of parameters [7].

Unfortunately, computational techniques for Nash equilibrium have so far not achieved anywhere near the same level of success. In contrast, most modern Nash equilibrium solvers for n-player, m-action, general-sum, normal-form games (NFGs) are practically restricted to a handful of players and/or actions per player except in special cases (e.g., symmetric [38] or mean-field games [34]). This is partially due to the fact that an NFG is represented by a tensor with an exponential nm^n entries; even *reading* this description into memory can be computationally prohibitive. More to the point, any computational technique that presumes *exact* computation of the *expectation* of any function sampled according to $\boldsymbol{x}$ similarly does not have any hope of scaling beyond small instances.

This inefficiency arguably lies at the core of the differential success between ML optimization and equilibrium computation. For example, numerous techniques exist that reduce the problem of Nash equilibrium computation to finding the minimum of the expectation of a random variable (see related work section). Unfortunately, unlike the source of randomness in ML applications where batch learning suffices to easily produce unbiased estimators, these techniques do not extend easily to game theory which incorporates non-linear functions such as maximum, best-response amongst others. This raises our motivating goal:

Can we solve for Nash equilibria via unbiased stochastic optimization?

Our results. Following in the successful steps of the interplay between ML and stochastic optimization, we reformulate the approximation of Nash equilibria in an NFG as a stochastic non-convex optimization problem admitting unbiased Monte-Carlo estimation. This enables the use of powerful solvers and advances in parallel computing to efficiently enumerate Nash equilibria for n-player, general-sum games. Furthermore, this re-casting allows practitioners to incorporate other desirable objectives into the problem such as "find an approximate Nash equilibrium with welfare above ω" or "find an approximate Nash equilibrium nearest the current observed joint strategy" resolving the equilibrium selection problem in effectively ad-hoc and application tailored manner. Concretely, we make the following contributions by producing:

- A loss function $\mathcal{L}(\boldsymbol{x})$ 1) whose global minima coincide with interior Nash equilibria in normal form games, 2) admits unbiased Monte-Carlo estimation, and 3) is Lipschitz and bounded.
- A loss function $\mathcal{L}^{\tau}(\boldsymbol{x})$ 1) whose global minima coincide with logit equilibria (QREs) in normal form games, 2) admits unbiased Monte-Carlo estimation, and 3) is Lipschitz and bounded.
- An efficient randomized algorithm for approximating Nash equilibria in a novel class of games. The algorithm emerges by employing a recent $\mathcal{X}$-armed bandit approach to $\mathcal{L}^{\tau}(\boldsymbol{x})$ and connecting its stochastic optimization guarantees to approximate Nash guarantees. For large games, this enables approximating equilibria *faster* than the game can even be read into memory.
- An empirical comparison of stochastic gradient descent against state-of-the-art baselines for approximating NEs in large games. In some games, vanilla SGD actually improves upon previous state-of-the-art; in others, SGD is slowed by saddle points, a familiar challenge in deep learning [12].

Overall, this perspective showcases a promising new route to approximating equilibria at scale in practice. We conclude the paper with discussion for future work.

2 Preliminaries

In an n-player, normal-form game, each player $i \in \{1, \ldots, n\}$ has a strategy set $\mathcal{A}_i = \{a_{i1}, \ldots, a_{im_i}\}$ consisting of m_i pure strategies. These strategies can be naturally indexed, so we redefine $\mathcal{A}_i = \{1, \ldots, m_i\}$ as an abuse of notation. Each player i also has a utility function, $u_i : \mathcal{A} = \prod_i \mathcal{A}_i \rightarrow [0, 1]$, (equiv. "payoff tensor") that maps joint actions to payoffs in the unit-interval . Note that equilibria are invariant to payoff shift and scale [27] so we are effectively assuming we know bounds on possible payoffs. We denote the average cardinality of the players' action sets by $\bar{m} = \frac{1}{n} \sum_k m_k$ and maximum by $m^* = \max_k m_k$. Player i may play a mixed strategy by sampling from a distribution over their pure strategies. Let player i's mixed strategy be represented by a vector $x_i \in \Delta^{m_i - 1}$ where Δ^{m_i-1} is the $(m_i - 1)$-dimensional probability simplex embedded in $\mathbb{R}^{m_i}$. Each function u_i is then extended to this domain so that $u_i(\boldsymbol{x}) = \sum_{\boldsymbol{a} \in \mathcal{A}} u_i(\boldsymbol{a}) \prod_j x_{j a_j}$ where $\boldsymbol{x} = (x_1, \ldots, x_n)$ and $a_j \in \mathcal{A}_j$ denotes player j's component of the joint action $\boldsymbol{a} \in \mathcal{A}$. For convenience, let x_{-i} denote all components of $\boldsymbol{x}$ belonging to players other than player i.

The joint strategy $\boldsymbol{x} \in \prod_i \Delta^{m_i-1}$ is a Nash equilibrium if and only if, for all $i \in \{1, \ldots, n\}$, $u_i(z_i, x_{-i}) \leq u_i(\boldsymbol{x})$ for all $z_i \in \Delta^{m_i-1}$, i.e., no player has any incentive to unilaterally deviate from $\boldsymbol{x}$. Nash is typically relaxed with ϵ-Nash, our focus: $u_i(z_i, x_{-i}) \leq u_i(\boldsymbol{x}) + \epsilon$ for all $z_i \in \Delta^{m_i-1}$.

As an abuse of notation, let the atomic action $a_i = e_i$ also denote the m_i-dimensional "one-hot" vector with all zeros aside from a 1 at index a_i; its use should be clear from the context. We also introduce

Loss	Function	Obstacle
Exploitabilty	$\max_k \epsilon_k(\boldsymbol{x})$	$\max$ of r.v.
Nikaido-Isoda (NI)	$\sum_k \epsilon_k(\boldsymbol{x})$	$\max$ of r.v.
Fully-Diff. Exp	$\sum_k \sum_{a_k \in \mathcal{A}_k} [\max(0, u_k(a_k, x_{-i}) - u_k(\boldsymbol{x}))]^2$	$\max$ of r.v.
Gradient-based NI	NI w/ $\texttt{BR}_k \leftarrow \texttt{aBR}_k = \Pi_\Delta\Big(x_k + \eta \nabla_{x_k} u_k(\boldsymbol{x})\Big)$	Π_Δ of r.v.
Unconstrained	Loss + Simplex Deviation Penalty	sampling from $x_i \in \mathbb{R}^{m_k}$

Table 1: Previous loss functions for NFGs and their obstacles to unbiased estimation.

$\nabla^i_{x_i}$ as player i's utility gradient. And for convenience, denote by $H^i_{il} = \mathbb{E}_{x_{-il}}[u_i(a_i, a_l, x_{-il})]$ the bimatrix game approximation [20] between players i and l with all other players marginalized out; x_{-il} denotes all strategies belonging to players other than i and l and $u_i(a_i, a_l, x_{-il})$ separates out l's strategy x_l from the rest of the players x_{-i}. Similarly, denote by $T^i_{ilq} = \mathbb{E}_{x_{-ilq}}[u_i(a_i, a_l, a_q, x_{-ilq})]$ the 3-player tensor approximation to the game. Note player i's utility can now be written succinctly as $u_i(x_i, x_{-i}) = x_i^\top \nabla^i_{x_i} = x_i^\top H^i_{il} x_l = x_i T^i_{ilq} x_l x_q$ for any l, q where we use Einstein notation for tensor arithmetic. For convenience, define $\texttt{diag}(z)$ as the function that places a vector z on the diagonal of a square matrix, and $\texttt{diag3} : z \in \mathbb{R}^d \rightarrow \mathbb{R}^{d \times d \times d}$ as a 3-tensor of shape (d, d, d) where $\texttt{diag3}(z)_{iii} = z_i$. Following convention from differential geometry, let $T_v\mathcal{M}$ be the tangent space of a manifold $\mathcal{M}$ at v. For the interior of the d-action simplex Δ^{d-1}, the tangent space is the same at every point, so we drop the v subscript, i.e., $T\Delta^{d-1}$. We denote the projection of a vector $z \in \mathbb{R}^d$ onto this tangent space as $\Pi_{T\Delta^{d-1}}(z) = z - \frac{1}{d}\mathbf{1}^\top z$. We drop d when the dimensionality is clear from the context. Finally, let $\mathcal{U}(S)$ denote a discrete uniform distribution over elements from set S.

3 Related Work

Representing the problem of computing a Nash equilibrium as an optimization problem is not new. A variety of loss functions and pseudo-distance functions have been proposed. Most of them measure some function of how much each player can exploit the joint strategy by unilaterally deviating:

$$\epsilon_k(\boldsymbol{x}) \stackrel{\text{def}}{=} u_k(\texttt{BR}_k, x_{-k}) - u_k(\boldsymbol{x}) \text{ where } \texttt{BR}_k \in \arg\max_z u_k(z, x_{-k}). \tag{1}$$

As argued in the introduction, we believe it is important to be able to subsample payoff tensors of normal-form games in order to scale to large instances. As Nash equilibria can consist of mixed strategies, it is advantageous to be able to sample from an equilibrium to estimate its exploitability ϵ. However none of these losses is amenable to unbiased estimation under sampled play. Each of the functions currently explored in the literature is biased under sampled play either because 1) a random variable appears as the argument of a complex, nonlinear (non-polynomial) function or because 2) how to sample play is unclear. Exploitability, Nikaido-Isoda (NI) [32] (also known by `NashConv` [21] and ADI [15]), as well as fully-differentiable options ([36], p. 106, Eqn 4.31) introduce bias when a $\max$ over payoffs is estimated using samples from $\boldsymbol{x}$. Gradient-based NI [35] requires projecting the result of a gradient-ascent step onto the simplex; for the same reason as the $\max$, this is prohibitive because it is a nonlinear operation which introduces bias. Lastly, unconstrained optimization approaches ([36], p. 106) that instead penalize deviation from the simplex lose the ability to sample from strategies when iterates are no longer proper distributions. Table 1 summarizes these complications.

4 Nash Equilibrium as Stochastic Optimization

We will now develop our proposed loss function which is amenable to unbiased estimation. Our key technical insight is to pay special attention to the geometry of the simplex. To our knowledge, prior works have failed to recognize the role of the tangent space $T\Delta$. Proofs are in the appendix.

4.1 Stationarity on the Simplex Interior

Lemma 1. *Assuming player i's utility, $u_i(x_i, x_{-i})$, is concave in its own strategy x_i, a strategy in the interior of the simplex is a best response $\texttt{BR}_i$ if and only if it has zero projected-gradient[1] norm:*

[1]Not to be confused with the nonlinear (i.e., introduces bias) projected gradient operator introduced in [19].

$$\texttt{BR}_i \in \big(int\Delta \cap \arg\max_z u_i(z, x_{-i}) - u_i(x_i, x_{-i})\big) \iff (\texttt{BR}_i \in int\Delta) \wedge (||\Pi_{T\Delta}[\nabla^i_{\texttt{BR}_i}]|| = 0). \tag{2}$$

In NFGs, each player's utility is linear in x_i, thereby satisfying the concavity condition of Lemma 1.

4.2 Projected Gradient Norm as Loss

An equivalent description of a Nash equilibrium is a joint strategy $\boldsymbol{x}$ where every player's strategy is a best response to the equilibrium (i.e., $x_i = \texttt{BR}_i$ so that $\epsilon_i(\boldsymbol{x}) = 0$). Lemma 1 states that any interior best response has zero projected-gradient norm, which inspires the following loss function

$$\mathcal{L}(\boldsymbol{x}) = \sum_k \eta_k ||\Pi_{T\Delta}(\nabla^k_{x_k})||^2 \tag{3}$$

where $\eta_k > 0$ represent scalar weights, or equivalently, step sizes to be explained next.

Proposition 1. *The loss $\mathcal{L}$ is equivalent to* `NashConv`*, but where player k's best response is approximated by a single step of projected-gradient ascent with step size η_k: $\texttt{aBR}_k = x_k + \eta_k \Pi_{T\Delta}(\nabla^k_{x_k})$.*

This connection was already pointed out in prior work for unconstrained problems [15, 35], but this result is the first for strategies constrained to the simplex.

4.3 Connection to True Exploitability

In general, we can bound exploitability in terms of the projected-gradient norm as long as each player's utility is concave (this result extends beyond gradients to subgradients of non-smooth functions).

Lemma 2. *The amount a player can gain by exploiting a joint strategy $\boldsymbol{x}$ is upper bounded by a quantity proportional to the norm of the projected-gradient:*

$$\epsilon_k(\boldsymbol{x}) \leq \sqrt{2}||\Pi_{T\Delta}(\nabla^k_{x_k})||. \tag{4}$$

This bound is not tight on the boundary of the simplex, which can be seen clearly by considering x_k to be part of a pure strategy equilibrium. In that case, this analysis assumes x_k can be improved upon by a projected-gradient ascent step (via the equivalence pointed out in Proposition 1). However, that is false because the probability of a pure strategy cannot be increased beyond 1. We mention this to provide further intuition for why $\mathcal{L}(\boldsymbol{x})$ is only valid for interior equilibria.

Note that $||\Pi_{T\Delta}(\nabla^k_{x_k})|| \leq ||\nabla^k_{x_k}||$ because $\Pi_{T\Delta}$ is a projection. Therefore, this improves the naive bounds on exploitability and distance to best responses given using the "raw" gradient $\nabla^k_{x_k}$.

Lemma 3. *The exploitability of a joint strategy $\boldsymbol{x}$, is upper bounded by a function of $\mathcal{L}(\boldsymbol{x})$:*

$$\epsilon \leq \sqrt{\frac{2n}{\min_k \eta_k}}\sqrt{\mathcal{L}(\boldsymbol{x})} \stackrel{\text{def}}{=} f(\mathcal{L}). \tag{5}$$

4.4 Unbiased Estimation

As discussed in Section 3, a primary obstacle to unbiased estimation of $\mathcal{L}(\boldsymbol{x})$ is the presence of complex, nonlinear functions of random variables, with the projection of a point onto the simplex being one such example (see Π_Δ in Table 1). However, $\Pi_{T\Delta}$, *the projection onto the tangent space of the simplex, is linear*! This is the key that allows us to design an unbiased estimator (Lemma 5).

Our proposed loss requires computing the squared norm of the *expected value* of the gradient under the players' mixed strategies, i.e., the l-th entry of player k's gradient equals $\nabla^k_{x_{kl}} = \mathbb{E}_{a_{-k} \sim x_{-k}} u_k(a_{kl}, a_{-k})$. By analogy, consider a random variable Y. In general, $\mathbb{E}[Y]^2 \neq \mathbb{E}[Y^2]$. This means that we cannot just sample projected-gradients and then compute their average norm to estimate our loss. However, consider taking two independent samples from two corresponding identically distributed, independent random variables $Y^{(1)}$ and $Y^{(2)}$. Then $\mathbb{E}[Y^{(1)}]^2 = \mathbb{E}[Y^{(1)}]\mathbb{E}[Y^{(2)}] =$

	Exact	Sample Others	Sample All
Estimator of $\nabla^{k(p)}_{x_k}$	$u_k(a_{kl}, x_{-k})$	$u_k(a_{kl}, a_{-k} \sim x_{-k})$	$m_k u_k(a_{kl} \sim \mathcal{U}(\mathcal{A}_k), a_{-k} \sim x_{-k}) e_l$
$\hat{\nabla}^{k(p)}_{x_k}$ Bounds	$[0, 1]$	$[0, 1]$	$[0, m_k]$
$\hat{\nabla}^{k(p)}_{x_k}$ Query Cost	$\prod_{i=1}^n m_i$	m_k	1
$\mathcal{L}$ Bounds	$\pm\frac{1}{4}\sum_k \eta_k m_k$	$\pm\frac{1}{4}\sum_k \eta_k m_k$	$\pm\frac{1}{4}\sum_k \eta_k m_k^3$
$\mathcal{L}$ Query Cost	$n\prod_{i=1}^n m_i$	$2n\bar{m}$	$2n$

Table 2: Examples and Properties of Unbiased Estimators of Loss and Player Gradients ($\hat{\nabla}^{k(p)}_{x_k}$).

$\mathbb{E}[Y^{(1)}Y^{(2)}]$ by properties of expected value over products of independent random variables. This is a common technique to construct unbiased estimates of expectations over polynomial functions of random variables. Proceeding in this way, define $\nabla^{k(1)}_{x_k}$ as a random variable distributed according to the distribution induced by all other players' mixed strategies ($j \neq k$). Let $\nabla^{k(2)}_{x_k}$ be independent and distributed identically to $\nabla^{k(1)}_{x_k}$. Then

$$\mathcal{L}(\boldsymbol{x}) = \mathbb{E}[\sum_k \eta_k \underbrace{(\hat{\nabla}^{k(1)}_{x_k} - \frac{\mathbf{1}}{m_k}(\mathbf{1}^\top \hat{\nabla}^{k(1)}_{x_k})\mathbf{1})^\top}_{\text{projected-gradient 1}} \underbrace{(\hat{\nabla}^{k(2)}_{x_k} - \frac{\mathbf{1}}{m_k}(\mathbf{1}^\top \hat{\nabla}^{k(2)}_{x_k})\mathbf{1})}_{\text{projected-gradient 2}}] \tag{6}$$

where $\hat{\nabla}^{k(p)}_{x_k}$ is an unbiased estimator of player k's gradient. This unbiased estimator can be constructed in several ways. The most expensive, an exact estimator, is constructed by marginalizing player k's payoff tensor over all other players' strategies. However, a cheaper estimate can be obtained at the expense of higher variance by approximating this marginalization with a Monte Carlo estimate of the expectation. Specifically, if we sample a single action for each of the remaining players, we can construct an unbiased estimate of player k's gradient by considering the payoff of each of its actions against the sampled background strategy. Lastly, we can consider constructing a Monte Carlo estimate of player k's gradient by sampling only a single action from player k to represent their entire gradient. Each of these approaches is outlined in Table 2 along with the query complexity [3] of computing the estimator and bounds on the values it can take (derived via Lemma 19).

We can extend Lemma 3 to one that holds under T samples with probability $1 - \delta$ by applying, for example, a Hoeffding bound: $\epsilon \leq f\big(\hat{\mathcal{L}}(\boldsymbol{x}) + \mathcal{O}(\sqrt{\frac{1}{T}\ln(1/\delta)})\big)$.

4.5 Interior Equilibria

We discussed earlier that $\mathcal{L}(\boldsymbol{x})$ captures interior equilibria. But some games may only have *pure* equilibria. We show how to circumvent this shortcoming by considering quantal response equilibria (QREs), specifically, logit equilibria. By adding an entropy bonus to each player's utility, we can

- guarantee **all** equilibria are interior,
- still obtain unbiased estimates of our loss,
- maintain an upper bound on the exploitability ϵ of any approximate equilibrium in the original game (i.e., the game without an entropy bonus).

Define $u_k^\tau(\boldsymbol{x}) = u_k(\boldsymbol{x}) + \tau S(x_k)$ where the Shannon entropy $S(x_k) = -\sum_l x_{kl}\ln(x_{kl})$ is a 1-strongly concave function with respect to the 1-norm [6]. Also define $\mathcal{L}^\tau(\boldsymbol{x})$ as before except where $\nabla^k_{x_k}$ is replaced with $\nabla^{k\tau}_{x_k} = \nabla_{x_k} u_k^\tau(\boldsymbol{x})$, i.e., the gradient of player k's utility *with* the entropy bonus.

It is well known that Nash equilibria of entropy-regularized games satisfy the conditions for logit equilibria [23], which are solutions to the fixed point equation $x_k = \texttt{softmax}(\frac{\nabla^k_{x_k}}{\tau})$. The appearance of the `softmax` makes clear that all probabilities have positive mass at positive temperature.

Recall that in order to construct an unbiased estimate of our loss, we simply needed to construct unbiased estimates of player gradients. The introduction of the entropy term to player k's utility is special in that it depends entirely on known quantities, i.e., the player's own mixed strategy. We can directly and deterministically compute $\tau\frac{dS}{dx_k} = -\tau(\ln(x_k) + \mathbf{1})$ and add this to our estimator of $\nabla^{k(p)}_{x_k}$: $\hat{\nabla}^{k\tau(p)}_{x_k} = \hat{\nabla}^{k(p)}_{x_k} + \tau\frac{dS}{dx_k}$. Consider our refined loss function with changes in blue:

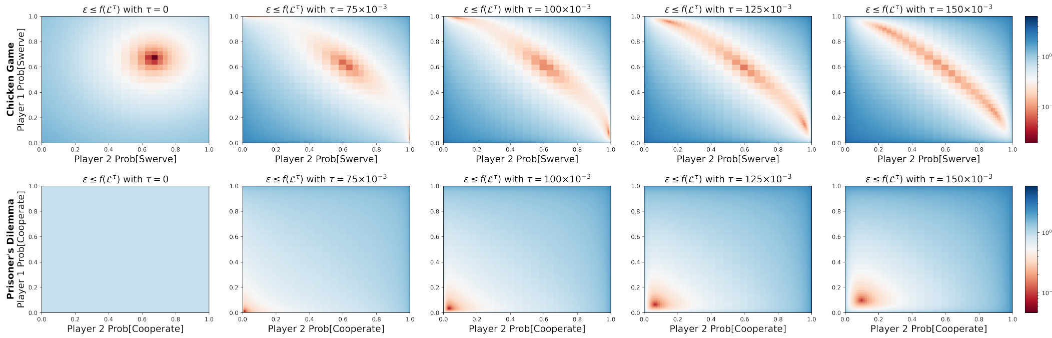

Figure 1: Upper Bound ($\epsilon \leq f(\mathcal{L}^\tau)$) Heatmap Visualization. The first row examines the loss landscape for the classic anti-coordination game of Chicken (Nash equilibria: $(0, 1), (1, 0), (2/3, 1/3)$) while the second row examines the Prisoner's dilemma (Unique Nash equilibrium: $(0, 0)$). Temperature increases for each plot moving to the right. For high temperatures, interior (fully-mixed) strategies are incentivized while for lower temperatures, nearly pure strategies can achieve minimum exploitability. For zero temperature, pure strategy equilibria (e.g., defect-defect) are not captured by the loss as illustrated by the bottom-left Prisoner's Dilemma plot with a constant loss surface.

$$\mathcal{L}^{\tau}(\boldsymbol{x}) = \sum_k \eta_k ||\Pi_{T\Delta}(\nabla^{k\tau}_{x_k})||^2. \tag{7}$$

As mentioned above, the utilities with entropy bonuses are still concave, therefore, a similar bound to Lemma 2 applies. We use this to prove the QRE counterpart to Lemma 3 where ϵ_{QRE} is the exploitability of an approximate equilibrium in a game with entropy bonuses.

Lemma 4. *The entropy regularized exploitability, ϵ_{QRE}, of a joint strategy $\boldsymbol{x}$, is upper bounded as:*

$$\epsilon_{QRE} \leq \sqrt{\frac{2n}{\min_k \eta_k}} \sqrt{\mathcal{L}^\tau(\boldsymbol{x})} \overset{\text{def}}{=} f(\mathcal{L}^\tau). \tag{8}$$

Lastly, we establish a connection between quantal response equilibria and Nash equilibria that allows us to approximate Nash equilibria in the original game via minimizing our modified loss $\mathcal{L}^\tau(\boldsymbol{x})$.

Lemma 14 ($\mathcal{L}^\tau$ Scores Nash Equilibria)**.** *Let $\mathcal{L}^\tau(\boldsymbol{x})$ be our proposed entropy regularized loss function with payoffs bounded in $[0, 1]$ and $\boldsymbol{x}$ be an approximate QRE. Then it holds that*

$$\epsilon \leq n\tau(W(1/e) + \frac{\bar{m} - 2}{e}) + 2\sqrt{\frac{n \max_k m_k}{\min_k \eta_k}} \sqrt{\mathcal{L}^\tau(\boldsymbol{x})} \tag{9}$$

where W is the Lambert function: $W(1/e) = W(\exp(-1)) \approx 0.278$.

This upper bound is plotted as a heatmap for familiar games in Figure 1. Notice how pure equilibria are not visible as minima for zero temperature, but appear for slightly warmer temperatures.

5 Analysis

In the preceding section we established a loss function that upper bounds the exploitability of an approximate equilibrium. In addition, the zeros of this loss function have a one-to-one correspondence with quantal response equilibria (which approximate Nash equilibria at low temperature).

Here, we derive properties that suggest it is "easy" to optimize. While this function is generally non-convex and may suffer from a proliferation of saddle points and local maxima (Figure 2) , it is Lipschitz continuous (over a subset of the interior) and bounded. These are two commonly made assumptions in the literature on non-convex optimization, which we leverage in Section 6. In addition, we can derive its gradient, its Hessian, and characterize its behavior around global minima.

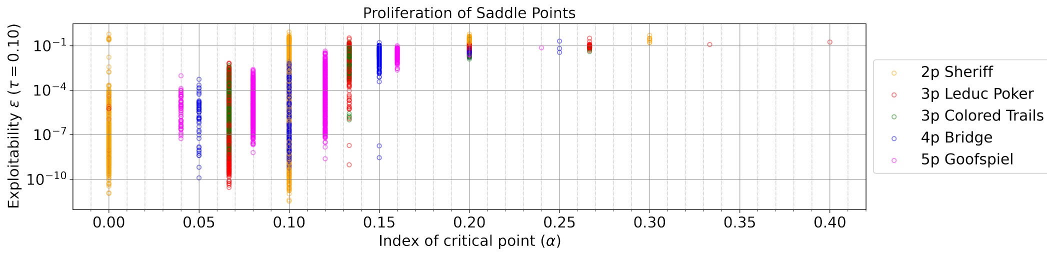

Figure 2: We reapply the analysis of [12], originally designed to understand the success of SGD in deep learning, to "slices" of several popular extensive form games. To construct a slice (or *meta-game*), we randomly sample 6 deterministic policies and then consider the corresponding n-player, 6-action normal-form game at $\tau = 0.1$ (with payoffs normalized to $[0, 1]$). The index of a critical point $\boldsymbol{x}_c$ ($\nabla_{\boldsymbol{x}} \mathcal{L}^\tau(\boldsymbol{x}_c) = \mathbf{0}$) indicates the fraction of negative eigenvalues in the Hessian of $\mathcal{L}^\tau$ at $\boldsymbol{x}_c$; $\alpha = 0$ indicates a local minimum, 1 a maximum, else a saddle point. We see a positive correlation between exploitability and α indicating a lower prevalence of local minima at high exploitability.

Lemma 15. *The gradient of $\mathcal{L}^\tau(\boldsymbol{x})$ with respect to player l's strategy x_l is*

$$\nabla_{x_l} \mathcal{L}^\tau(\boldsymbol{x}) = 2 \sum_k \eta_k B_{kl}^\top \Pi_{T\Delta}(\nabla_{x_k}^{k\tau}) \tag{10}$$

where $B_{ll} = -\tau[I - \frac{1}{m_l}\mathbf{1}\mathbf{1}^\top]\texttt{diag}(\frac{1}{x_l})$ and $B_{kl} = [I - \frac{1}{m_k}\mathbf{1}\mathbf{1}^\top] H_{kl}^k$ for $k \neq l$.

Lemma 17. *The Hessian of $\mathcal{L}^\tau(\boldsymbol{x})$ can be written*

$$\texttt{Hess}(\mathcal{L}^\tau) = 2\big[\tilde{B}^\top \tilde{B} + T\Pi_{T\Delta}(\tilde{\nabla}^\tau)\big] \tag{11}$$

where $\tilde{B}_{kl} = \sqrt{\eta_k} B_{kl}$, $\Pi_{T\Delta}(\tilde{\nabla}^\tau) = [\eta_1 \Pi_{T\Delta}(\nabla_{x_1}^{1\tau}), \ldots, \eta_n \Pi_{T\Delta}(\nabla_{x_n}^{n\tau})]$, and we augment T (the 3-player approximation to the game, T_{lqk}^k) so that $T_{lll}^l = \tau\,\texttt{diag3}(\frac{1}{x_l^2})$.

At an equilibrium, the latter term disappears because $\Pi_{T\Delta}(\nabla_{x_k}^{k\tau}) = \mathbf{0}$ for all k (Lemma 1). If $\mathcal{X}$ was $\mathbb{R}^{n\bar{m}}$, then we could simply check if $\tilde{B}$ is full-rank to determine if $Hess \succ 0$. However, $\mathcal{X}$ is a simplex product, and we only care about curvature in directions toward which we can update our equilibrium. Toward that end, define M to be the $n(\bar{m}+1) \times n\bar{m}$ matrix that stacks $\tilde{B}$ on top of a repeated identity matrix that encodes orthogonality to the simplex:

$$M(\boldsymbol{x}) = \begin{bmatrix} -\tau\sqrt{\eta_1}\Pi_{T\Delta}(\frac{1}{x_1}) & \sqrt{\eta_1}\Pi_{T\Delta}(H_{12}^1) & \cdots & \sqrt{\eta_1}\Pi_{T\Delta}(H_{1n}^1) \\ \vdots & \vdots & \vdots & \vdots \\ \sqrt{\eta_n}\Pi_{T\Delta}(H_{n1}^n) & \cdots & \sqrt{\eta_n}\Pi_{T\Delta}(H_{n,n-1}^n) & -\tau\sqrt{\eta_n}\Pi_{T\Delta}(\frac{1}{x_n}) \\ \mathbf{1}_1^\top & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & \mathbf{1}_n^\top \end{bmatrix} \tag{12}$$

where $\Pi_{T\Delta}(z \in \mathbb{R}^{a \times b}) = [I_a - \frac{1}{a}\mathbf{1}_a\mathbf{1}_a^\top]z$ subtracts the mean from each column of z and $\frac{1}{x_i}$ is shorthand for $\texttt{diag}(\frac{1}{x_i})$. If $M(x)z = \mathbf{0}$ for a nonzero vector $z \in \mathbb{R}^{n\bar{m}}$, this implies there exists a z that 1) is orthogonal to the ones vectors of each simplex (i.e., is a valid equilibrium update direction) and 2) achieves zero curvature in the direction z, i.e., $z^\top(\tilde{B}^\top\tilde{B})z = z^\top(Hess)z = 0$, and so $Hess$ is not positive definite. Conversely, if $M(\boldsymbol{x})$ is of rank $n\bar{m}$ for a quantal response equilibrium $\boldsymbol{x}$, then the Hessian of $\mathcal{L}^\tau$ at $\boldsymbol{x}$ in the tangent space of the simplex product ($\mathcal{X} = \prod_i \mathcal{X}_i$) is positive definite. In this case, we call $\boldsymbol{x}$ *well*-isolated because it implies it is not connected to any other equilibria.

By analyzing the rank of M, we can confirm that many classical matrix games including Rock-Paper-Scissors, Chicken, Matching Pennies, and Shapley's game all induce strongly convex $\mathcal{L}^\tau$'s at zero temperature (i.e., they have unique mixed Nash equilibria). In contrast, a game like Prisoner's Dilemma has a unique pure strategy that will not be captured by our loss at zero temperature.

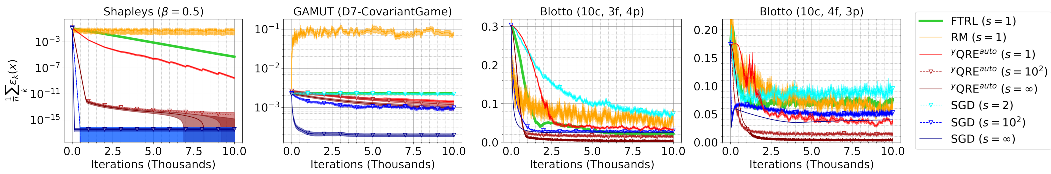

Figure 3: Comparison of SGD on $\mathcal{L}^{\tau=0}$ against baselines on four games evaluated in [15]. From left to right: 2-player, 3-action, nonsymmetric; 6-player, 5-action, nonsymmetric; 4-player, 66-action, symmetric; 3-player, 286-action, symmetric. SGD struggles at saddle points in Blotto.

6 Algorithms

We have formally transformed the approximation of Nash equilibria in NFGs into a **stochastic** optimization problem. To our knowledge, this is the first such formulation that allows one-shot unbiased Monte-Carlo estimation which is critical to introduce the use of powerful algorithms capable of solving high dimensional optimization problems. We explore two off-the-shelf approaches.

Stochastic gradient descent is the workhorse of high-dimensional stochastic optimization. It comes with guaranteed convergence to stationary points [10], however, it may converge to local, rather than global minima. It also enjoys implicit gradient regularization [4], seeking "flat" minima and performs approximate Bayesian inference [26]. Despite the lack of global convergence guarantee, in the next section, we find it performs well empirically in games previously examined by the literature.

We explore one other algorithmic approach to non-convex optimization based on minimizing regret, which enjoys finite time convergence rates. $\mathcal{X}$-armed bandits [8] systematically explore the space of solutions by refining a mesh over the joint strategy space, trading off exploration versus exploitation of promising regions.[2] Several approaches exist [5, 37] with open source implementations (e.g., [24]).

6.1 High Probability, Polynomial Convergence Rates

We use a recent $\mathcal{X}$-armed bandit approach called BLiN [14] to establish a high probability $\tilde{\mathcal{O}}(T^{-1/4})$ convergence rate to Nash equilibria in n-player, general-sum games under mild assumptions. The quality of this approximation improves as $\tau \to 0$, at the same time increasing the constant on the convergence rate via the Lipschitz constant $\sqrt{\hat{L}}$ defined below. For clarity, we assume users provide a temperature in the form $\tau = \frac{1}{\ln(1/p)}$ with $p \in (0, 1)$ which ensures all equilibria have probability mass greater than $\frac{p}{m^*}$ for all actions (Lemma 9). Lower p corresponds with lower temperature.

The following convergence rate depends on bounds on the exploitability in terms of the loss (Lemma 14), bounds on the magnitude of estimates of the loss (Lemma 8), Lipschitz bounds on the infinity norm of the gradient (Corollary 2), and the number of distinct strategies ($n\bar{m} = \sum_k m_k$).

Theorem 1 (BLiN PAC Rate)**.** *Assume $\eta_k = \eta = 2/\hat{L}$, $\tau = \frac{1}{\ln(1/p)}$, and a previously pulled arm is returned uniformly at random (i.e., $t \sim U([T])$). Then for any $w > 0$*

$$\epsilon_t \leq w\Big[\frac{n}{\ln(1/p)}\big(W(1/e) + \frac{\bar{m}-2}{e}\big) + 4(1 + (4c^2)^{1/3})\sqrt{nm^*\hat{L}}\Big(\frac{\ln T}{T}\Big)^{\frac{1}{2(d_z+2)}}\Big] \tag{13}$$

with probability $(1 - w^{-1})(1 - 2T^{-2})$ where W is the Lambert function ($W(1/e) \approx 0.278$), $m^* = \max_k m_k$, $c \leq \frac{1}{4}\frac{n\bar{m}}{\hat{L}}\Big(\frac{\ln(m^*)}{\ln(1/p)} + 2\Big)^2 \leq \frac{1}{4}\Big(\frac{\ln(m^*)}{\ln(1/p)} + 2\Big)$ upper bounds the range of stochastic estimates of $\mathcal{L}^\tau$ (see Lemma 8), and $\hat{L} = \Big(\frac{\ln(m^*)}{\ln(1/p)} + 2\Big)\Big(\frac{m^{*2}}{p\ln(1/p)} + n\bar{m}\Big)$ (see Corollary 2).

This result depends on the *near-optimality* [37] or *zooming*-dimension $d_z = n\bar{m}(\frac{\alpha_{hi}-\alpha_{lo}}{\alpha_{lo}\alpha_{hi}}) \in [0, \infty)$ (Theorem 2) where α_{lo} and α_{hi} denote the degree of the polynomials that lower and upper bound the function $\mathcal{L}^\tau \circ s$ locally around an equilibrium. For example, in the case where the Hessian is positive definite, $\alpha_{lo} = \alpha_{hi} = 2$ and $d_z = 0$. Here, $s : [0, 1]^{n(\bar{m}-1)} \to \prod_i \Delta^{m_i-1}$ is any function that maps from the unit hypercube to a product of simplices; we analyze two such maps in the appendix.

[2] Zhou et al. [39] developed a similar approach but only for pure Nash equilibria.

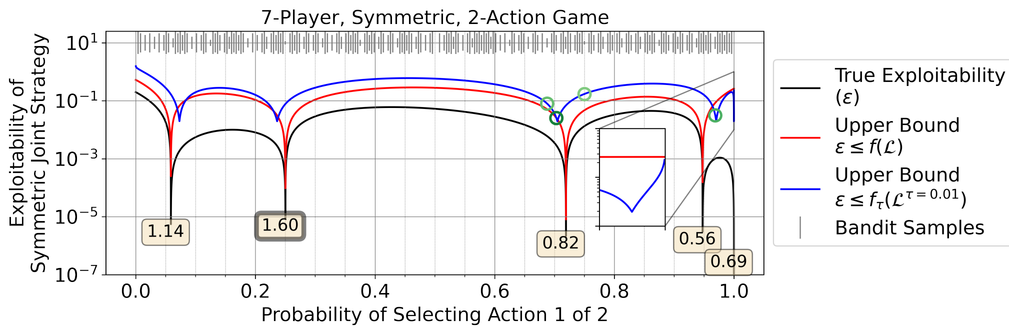

Figure 4: Bandit-based (BLiN) Nash solver applied to an artificial 7-player, symmetric, 2-action game. We search for a symmetric equilibrium, which is represented succinctly as the probability of selecting action 1. The plot shows the true exploitability ϵ of all symmetric strategies in black and indicates there exist potentially 5 NEs (the dips in the curve). Upper bounds on our unregularized loss $\mathcal{L}$ capture 4 of these equilibria, missing only the pure NE on the right. By considering our regularized loss, $\mathcal{L}^\tau$, we are able to capture this pure NE (see zoomed inset). The bandit algorithm selects strategies to evaluate, using 10 Monte-Carlo samples for each evaluation (arm pull) of $\mathcal{L}^\tau$. These samples are displayed as vertical bars above with the height of the vertical bar representing additional arm pulls. The best arms throughout search are denoted by green circles (darker indicates later in the search). The boxed numbers near equilibria display the welfare of the strategy.

Note that Theorem 1 implies that for games whose corresponding $\mathcal{L}^\tau$ has zooming dimension $d_z = 0$, NEs can be approximated with high probability in polynomial time. This general property is difficult to translate concisely into game theory parlance. For this reason, we present the following more interpretable corollary which applies to a more restricted class of games.

Corollary 1. *Consider the class of NFGs with at least one QRE(τ) whose local polymatrix approximation indicates it is isolated (i.e., M from equation (12) is rank-$n\bar{m}$ implies Hess $\succ 0$ implies $d_z = n\bar{m}(\frac{2-2}{4}) = 0$). Then by Theorem 1, BLiN is a fully polynomial-time randomized approximation scheme (FPRAS) for QREs and is a PRAS for NEs of games in this class.*

To convey the impact of stochastic optimization guarantees more concretely, assume we are given that an interior well-isolated NE exists. Then for a 20-player, 50-action game, it is $1000\times$ cheaper to compute a $1/100$-NE with probability 95% than it is to just list the nm^n payoffs that define the game.

6.2 Empirical Evaluation

Figure 3 shows SGD is competitive with scalable techniques to approximating NEs. Shapley's game induces a strongly convex $\mathcal{L}$ (see Section 5) leading to SGD's strong performance. Blotto shows signs of convergence to low, but nonzero ϵ, demonstrating the challenges of local minima.

We demonstrate BLiN (applied to $\mathcal{L}^\tau$) on a 7-player, symmetric, 2-action game. Figure 4 shows the bandit algorithm discovers two equilibria, settling on one near $\boldsymbol{x} = [0.7, 0.3] \times 7$ with a wider basin of attraction (and higher welfare). In theory, BLiN can enumerate all NEs as $T \to \infty$.

7 Conclusion

In this work, we proposed a stochastic loss for approximate Nash equilibria in normal-form games. An unbiased loss estimator of Nash equilibria is the "key" to the stochastic optimization "door" which holds a wealth of research innovations uncovered over several decades. Thus, it allows the development of new algorithmic techniques for computing equilibria. We consider bandit and vanilla SGD methods in this work, but theses are only two of the many options now at our disposal (e.g, adaptive methods [1], Gaussian processes [9], evolutionary algorithms [17], etc.). Such approaches as well as generalizations of these techniques to imperfect-information games are promising directions for future work. Similarly to how deep learning research first balked at and then marched on to train neural networks via NP-hard non-convex optimization, we hope computational game theory can march ahead to make useful equilibrium predictions of large multiplayer systems.

References

[1] K. Antonakopoulos, P. Mertikopoulos, G. Piliouras, and X. Wang. Adagrad avoids saddle points. In *International Conference on Machine Learning*, pages 731–771. PMLR, 2022.

[2] P. Austrin, M. Braverman, and E. Chlamtáč. Inapproximability of NP-complete variants of Nash equilibrium. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques: 14th International Workshop, APPROX 2011, and 15th International Workshop, RANDOM 2011, Princeton, NJ, USA, August 17-19, 2011. Proceedings*, pages 13–25. Springer, 2011.

[3] Y. Babichenko. Query complexity of approximate Nash equilibria. *Journal of the ACM (JACM)*, 63(4):36:1–36:24, 2016.

[4] D. Barrett and B. Dherin. Implicit gradient regularization. In *International Conference on Learning Representations*, 2020.

[5] P. L. Bartlett, V. Gabillon, and M. Valko. A simple parameter-free and adaptive approach to optimization under a minimal local smoothness assumption. In *Algorithmic Learning Theory*, pages 184–206. PMLR, 2019.

[6] A. Beck and M. Teboulle. Mirror descent and nonlinear projected subgradient methods for convex optimization. *Operations Research Letters*, 31(3):167–175, 2003.

[7] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, et al. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901, 2020.

[8] S. Bubeck, R. Munos, G. Stoltz, and C. Szepesvári. $\mathcal{X}$-armed bandits. *Journal of Machine Learning Research*, 12(5), 2011.

[9] D. Calandriello, L. Carratino, A. Lazaric, M. Valko, and L. Rosasco. Scaling gaussian process optimization by evaluating a few unique candidates multiple times. In *International Conference on Machine Learning*, pages 2523–2541. PMLR, 2022.

[10] A. Cutkosky, H. Mehta, and F. Orabona. Optimal stochastic non-smooth non-convex optimization through online-to-non-convex conversion. *arXiv preprint arXiv:2302.03775*, 2023.

[11] C. Daskalakis, P. W. Goldberg, and C. H. Papadimitriou. The complexity of computing a Nash equilibrium. *Communications of the ACM*, 52(2):89–97, 2009.

[12] Y. N. Dauphin, R. Pascanu, C. Gulcehre, K. Cho, S. Ganguli, and Y. Bengio. Identifying and attacking the saddle point problem in high-dimensional non-convex optimization. *Advances in neural information processing systems*, 27, 2014.

[13] A. Deligkas, J. Fearnley, A. Hollender, and T. Melissourgos. Pure-circuit: Strong inapproximability for PPAD. In *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 159–170. IEEE, 2022.

[14] Y. Feng, T. Wang, et al. Lipschitz bandits with batched feedback. *Advances in Neural Information Processing Systems*, 35:19836–19848, 2022.

[15] I. Gemp, R. Savani, M. Lanctot, Y. Bachrach, T. Anthony, R. Everett, A. Tacchetti, T. Eccles, and J. Kramár. Sample-based approximation of Nash in large many-player games via gradient descent. In *Proceedings of the 21st International Conference on Autonomous Agents and Multiagent Systems*, pages 507–515, 2022.

[16] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. *Advances in Neural Information Processing Systems*, 27, 2014.

[17] N. Hansen, S. D. Müller, and P. Koumoutsakos. Reducing the time complexity of the derandomized evolution strategy with covariance matrix adaptation (CMA-ES). *Evolutionary computation*, 11(1):1–18, 2003.

[18] J. C. Harsanyi, R. Selten, et al. A general theory of equilibrium selection in games. *MIT Press Books*, 1, 1988.

[19] E. Hazan, K. Singh, and C. Zhang. Efficient regret minimization in non-convex games. In *International Conference on Machine Learning*, pages 1433–1441. PMLR, 2017.

[20] E. Janovskaja. Equilibrium points in polymatrix games. *Lithuanian Mathematical Journal*, 8 (2):381–384, 1968.

[21] M. Lanctot, V. Zambaldi, A. Gruslys, A. Lazaridou, K. Tuyls, J. Pérolat, D. Silver, and T. Graepel. A unified game-theoretic approach to multiagent reinforcement learning. In *Advances in Neural Information Processing Systems*, pages 4190–4203, 2017.

[22] M. Lanctot, E. Lockhart, J.-B. Lespiau, V. Zambaldi, S. Upadhyay, J. Pérolat, S. Srinivasan, F. Timbers, K. Tuyls, S. Omidshafiei, D. Hennes, D. Morrill, P. Muller, T. Ewalds, R. Faulkner, J. Kramár, B. D. Vylder, B. Saeta, J. Bradbury, D. Ding, S. Borgeaud, M. Lai, J. Schrittwieser, T. Anthony, E. Hughes, I. Danihelka, and J. Ryan-Davis. OpenSpiel: A framework for reinforcement learning in games. *CoRR*, abs/1908.09453, 2019. URL `http://arxiv.org/abs/1908.09453`.

[23] S. Leonardos, G. Piliouras, and K. Spendlove. Exploration-exploitation in multi-agent competition: convergence with bounded rationality. *Advances in Neural Information Processing Systems*, 34:26318–26331, 2021.

[24] W. Li, H. Li, J. Honorio, and Q. Song. Pyxab – a python library for $\mathcal{X}$-armed bandit and online blackbox optimization algorithms, 2023. URL `https://arxiv.org/abs/2303.04030`.

[25] C. K. Ling, F. Fang, and J. Z. Kolter. What game are we playing? end-to-end learning in normal and extensive form games. *arXiv preprint arXiv:1805.02777*, 2018.

[26] S. Mandt, M. D. Hoffman, and D. M. Blei. Stochastic gradient descent as approximate bayesian inference. *Journal of Machine Learning Research*, 18:1–35, 2017.

[27] L. Marris, I. Gemp, and G. Piliouras. Equilibrium-invariant embedding, metric space, and fundamental set of 2x2 normal-form games. *arXiv preprint arXiv:2304.09978*, 2023.

[28] R. D. McKelvey and T. R. Palfrey. Quantal response equilibria for normal form games. *Games and Economic Behavior*, 10(1):6–38, 1995.

[29] D. Milec, J. Černỳ, V. Lisỳ, and B. An. Complexity and algorithms for exploiting quantal opponents in large two-player games. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(6):5575–5583, 2021.

[30] P. R. Milgrom and R. J. Weber. A theory of auctions and competitive bidding. *Econometrica: Journal of the Econometric Society*, pages 1089–1122, 1982.

[31] K. G. Murty and S. N. Kabadi. Some NP-complete problems in quadratic and nonlinear programming. Technical report, 1985.

[32] H. Nikaidô and K. Isoda. Note on non-cooperative convex games. *Pacific Journal of Mathematics*, 5(1):807815, 1955.

[33] E. Nudelman, J. Wortman, Y. Shoham, and K. Leyton-Brown. Run the GAMUT: A comprehensive approach to evaluating game-theoretic algorithms. In *AAMAS*, volume 4, pages 880–887, 2004.

[34] J. Pérolat, S. Perrin, R. Elie, M. Laurière, G. Piliouras, M. Geist, K. Tuyls, and O. Pietquin. Scaling mean field games by online mirror descent. In *Proceedings of the 21st International Conference on Autonomous Agents and Multiagent Systems*, 2022.

[35] A. Raghunathan, A. Cherian, and D. Jha. Game theoretic optimization via gradient-based Nikaido-Isoda function. In *International Conference on Machine Learning*, pages 5291–5300. PMLR, 2019.

[36] Y. Shoham and K. Leyton-Brown. *Multiagent systems: Algorithmic, game-theoretic, and logical foundations*. Cambridge University Press, 2008.

[37] M. Valko, A. Carpentier, and R. Munos. Stochastic simultaneous optimistic optimization. In *International Conference on Machine Learning*, pages 19–27. PMLR, 2013.

[38] B. Wiedenbeck and E. Brinkman. Data structures for deviation payoffs. In *Proceedings of the 22nd International Conference on Autonomous Agents and Multiagent Systems*, 2023.

[39] Y. Zhou, J. Li, and J. Zhu. Identify the Nash equilibrium in static games with random payoffs. In *International Conference on Machine Learning*, pages 4160–4169. PMLR, 2017.

Appendix: Approximating Nash Equilibria in Normal-Form Games via Unbiased Stochastic Optimization

A Loss: Connection to Exploitability, Unbiased Estimation, and Upper Bounds **14**

A.1 KKT Conditions Imply Fixed Point Sufficiency 14

A.2 Norm of Projected-Gradient and Equivalence to NFG Exploitability with Approximate Best Responses 15

A.3 Connection to True Exploitability 15

A.4 Unbiased Estimation 17

A.5 Bound on Loss 18

B QREs Approximate NEs at Low Temperature **19**

C Gradient of Loss **22**

C.1 Unbiased Estimation 23

C.2 Bound on Gradient / Lipschitz Property 23

D Hessian of Loss **24**

E Regret Bounds **26**

F Complexity **27**

F.1 Polymatrix Games 27

F.2 Normal-Form Games 28

G Helpful Lemmas and Propositions **28**

H Maps from Hypercube to Simplex Product **30**

H.1 Hessian of Bandit Reward Function 30

I A2: Bounded Diameters and Well-shaped Cells **33**

I.1 L_2-Norm 33

I.2 L_∞-Norm 34

I.3 Near Optimality Dimension 35

J D-BLiN **39**

K Experimental Setup and Details **43**

K.1 Loss Visualization and Rank Test 43

K.2 Saddle Point Analysis 44

K.3 SGD on Classical Games 45

K.4 BLiN on Artificial Game 45

A Loss: Connection to Exploitability, Unbiased Estimation, and Upper Bounds

A.1 KKT Conditions Imply Fixed Point Sufficiency

Consider the following constrained optimization problem:

$$\max_{\boldsymbol{x} \in \mathbb{R}^d} f(\boldsymbol{x}) \tag{14}$$

$$s.t. g_i(\boldsymbol{x}) \leq 0 \ \forall i \tag{15}$$

$$h_j(\boldsymbol{x}) = 0 \ \forall j \tag{16}$$

where f is concave and g_i and h_j represent inequality and equality constraints respectively. If g_i and h_i are affine functions, then any maximizer $\boldsymbol{x}^*$ of f must satisfy the following KKT conditions (necessary and sufficient):

- Stationarity: $0 \in \partial f(\boldsymbol{x}^*) - \sum_j \lambda_j \partial h_j(\boldsymbol{x}^*) - \sum_i \mu_i \partial g_i(\boldsymbol{x}^*)$
- Primal feasibility: $h_j(\boldsymbol{x}^*) = 0$ for all j and $g_i(\boldsymbol{x}^*) \leq 0$ for all i
- Dual feasibility: $\mu_i \geq 0$ for all i
- Complementary slackness: $\mu_i g_i(\boldsymbol{x}^*) = 0$ for all i.

Lemma 1. *Assuming player i's utility, $u_i(x_i, x_{-i})$, is concave in its own strategy x_i, any best response in the interior of the simplex has zero projected-gradient norm:*

$$z^* \in \left(int\Delta \cup \arg\max_z u_i(z, x_{-i}) - u_i(x_i, x_{-i}\right) \iff (z^* \in int\Delta) \wedge (||\Pi_\Delta[\nabla^i_{z^*}]|| = 0). \tag{17}$$

Proof. Consider the problem of formally computing $exp_i(\boldsymbol{x}) = \max_{z \in int\Delta} u_i(z, x_{-i}) - u_i(x_i, x_{-i})$:

$$\max_{z \in \mathbb{R}^d} u_i(z, x_{-i}) - u_i(x_i, x_{-i}) \tag{18}$$

$$s.t. - z_i + x_{\min} \leq 0 \ \forall i \tag{19}$$

$$1 - \sum_i z_i = 0. \tag{20}$$

where $x_{\min} > 0$ is some constant that captures our given assumption that the solution z^* lies in the interior of the simplex. Note that the objective is linear (concave) in z and the constraints are affine, therefore the KKT conditions are necessary and sufficient for optimality. Mapping the KKT conditions onto this problem yields the following:

- Stationarity: $0 \in \partial u_i(z^*, x_{-i}) + \lambda \mathbf{1} + \sum_i \mu_i e_i$
- Primal feasibility: $\sum_i z_i^* = 1$ and $z_i^* \geq x_{\min}$ for all i
- Dual feasibility: $\mu_i \geq 0$ for all i
- Complementary slackness: $\mu_i z_i^* = 0$ for all i.

For any point $z \in int\Delta$, primal feasibility will be satisfied for some $x_{\min} > 0$. This implies each z_j is strictly positive. By complementary slackness and dual feasibility, each μ_i must be identically zero. This implies the stationarity condition can be simplified to $0 \in \partial u_i(z^*, x_{-i}) + \lambda \mathbf{1}$. Rearranging terms we find that for any z^*, there exists a λ such that

$$\partial u_i(z^*, x_{-i}) \in \lambda \mathbf{1}. \tag{21}$$

Equivalently, $\partial u_i(z^*, x_{-i}) \propto \mathbf{1}$ at $z^* \in int\Delta$. Any vector proportional to the ones vector has zero projected-gradient norm, completing the claim. $\square$

A.2 Norm of Projected-Gradient and Equivalence to NFG Exploitability with Approximate Best Responses

Proposition 1. *The loss $\mathcal{L}$ is equivalent to* `NashConv`*, but where player k's best response is approximated by a single step of projected-gradient ascent with step size η_k:* $\texttt{aBR}_k = x_k + \eta_k \Pi_\Delta[\nabla^k_{x_k}]$.

Proof. Define an approximate best response as the result of a player adjusting their strategy via a projected-gradient ascent step, i.e., $\texttt{aBR}_k = x_k + \eta_k \Pi_\Delta[\nabla^k_{x_k}]$ for player k.

In a normal form game, player k's utility at this new strategy is $u_k(\texttt{aBR}_k, x_{-k}) = (\nabla^k_{x_k})^\top (x_k + \eta_k \Pi_\Delta[\nabla^k_{x_k}]) = u_k(\boldsymbol{x}) + \eta_k (\nabla^k_{x_k})^\top \Pi_\Delta[\nabla^k_{x_k}]$.

Therefore, the amount player k gains by playing `aBR` is

$$
\begin{aligned}
\hat{\epsilon}_k(\boldsymbol{x}) &= u_k(\texttt{aBR}_k, x_{-k}) - u_k(\boldsymbol{x}) && (22)\\
&= \eta_k (\nabla^k_{x_k})^\top \Pi_\Delta[\nabla^k_{x_k}] && (23)\\
&= \eta_k (\nabla^k_{x_k} - \frac{1}{m_k}(\mathbf{1}^\top \nabla^k_{x_k})\mathbf{1})^\top \Pi_\Delta[\nabla^k_{x_k}] && (24)\\
&= \eta_k ||\Pi_\Delta[\nabla^k_{x_k}]||^2 && (25)
\end{aligned}
$$

where the third equality follows from the fact that the projected-gradient, $\Pi_\Delta[\nabla^k_{x_k}]$, is orthogonal to the ones vector. □

A.3 Connection to True Exploitability

Lemma 2. *The amount a player can gain by deviating is upper bounded by a quantity proportional to the norm of the projected-gradient:*

$$
\epsilon_k(\boldsymbol{x}) \leq \sqrt{2}||\Pi_\Delta(\nabla^k_{x_k})||. \tag{26}
$$

Proof. Let z be any point on the simplex. Then

$$
\begin{aligned}
u_k(z, x_{-k}) - u_k(\boldsymbol{x}) &\leq (\nabla^k_{x_k})^\top (z - x_k) && (27)\\
&= (\nabla^k_{x_k})^\top (z - x_k) - \frac{1}{m_k}(\mathbf{1}^\top \nabla^k_{x_k}) \overbrace{\mathbf{1}^\top (z - x_k)}^{1-1=0} && (28)\\
&= (\Pi_\Delta[\nabla^k_{x_k}])^\top (z - x_k) && (29)\\
&\leq \sqrt{2}||\Pi_\Delta(\nabla^k_{x_k})||. && (30)
\end{aligned}
$$

□

Continuing, we can prove a bound on NashConv in terms of projected-gradient loss:

Lemma 3. *The exploitability, ϵ, of a joint strategy $\boldsymbol{x}$, is upper bounded as a function of our proposed loss:*

$$
\epsilon \leq \sqrt{\frac{2n}{\min_k \eta_k}} \sqrt{\mathcal{L}(\boldsymbol{x})}. \tag{31}
$$

Proof.

$$\begin{aligned}
\epsilon &= \max_k \max_z u_k(z, x_{-k}) - u_k(\boldsymbol{x}) &(32)\\
&\leq \sum_k \max_z u_k(z, x_{-k}) - u_k(\boldsymbol{x}) &(33)\\
&\leq \sum_k \sqrt{2}||\Pi_\Delta(\nabla^k_{x_k})||_2 &(34)\\
&= \sqrt{2}\Big|\Big| ||\Pi_\Delta(\nabla^1_{x_1})||_2, \ldots \sqrt{2}||\Pi_\Delta(\nabla^n_{x_n})||_2 \Big|\Big|_1 &(35)\\
&\leq \sqrt{2n}\Big|\Big| ||\Pi_\Delta(\nabla^1_{x_1})||_2, \ldots ||\Pi_\Delta(\nabla^n_{x_n})||_2 \Big|\Big|_2 &(36)\\
&= \sqrt{2n}\sqrt{\sum_k ||\Pi_\Delta(\nabla^k_{x_k})||^2_2} &(37)\\
&\leq \sqrt{2n}\sqrt{\sum_k \Big(\frac{1}{\eta_k}\Big)\eta_k||\Pi_\Delta(\nabla^k_{x_k})||^2_2} &(38)\\
&\leq \sqrt{\frac{2n}{\min_k \eta_k}}\sqrt{\sum_k \eta_k||\Pi_\Delta(\nabla^k_{x_k})||^2_2} &(39)\\
&= \sqrt{\frac{2n}{\min_k \eta_k}}\sqrt{\mathcal{L}(\boldsymbol{x})} &(40)
\end{aligned}$$

□

Lemma 4. *The entropy regularized exploitability, ϵ_{QRE}, of a joint strategy $\boldsymbol{x}$, is upper bounded as a function of our proposed loss:*

$$\epsilon_{QRE} \leq \sqrt{\frac{2n}{\min_k \eta_k}}\sqrt{\mathcal{L}^\tau(\boldsymbol{x})}. \tag{41}$$

Proof. Recall that $u^\tau_k(x_k, x_{-k})$ is also concave with respect to x_k. Then

$$\begin{aligned}
\epsilon_{QRE} &= \max_k \max_z u^\tau_k(z, x_{-k}) - u^\tau_k(\boldsymbol{x}) &(42)\\
&\leq \sum_k \max_z u^\tau_k(z, x_{-k}) - u^\tau_k(\boldsymbol{x}) &(43)\\
&\leq \sum_k \sqrt{2}||\Pi_\Delta(\nabla^{k\tau}_{x_k})||_2 &(44)\\
&= \sqrt{2}\Big|\Big| ||\Pi_\Delta(\nabla^{1\tau}_{x_1})||_2, \ldots \sqrt{2}||\Pi_\Delta(\nabla^{n\tau}_{x_n})||_2 \Big|\Big|_1 &(45)\\
&\leq \sqrt{2n}\Big|\Big| ||\Pi_\Delta(\nabla^{1\tau}_{x_1})||_2, \ldots ||\Pi_\Delta(\nabla^{n\tau}_{x_n})||_2 \Big|\Big|_2 &(46)\\
&= \sqrt{2n}\sqrt{\sum_k ||\Pi_\Delta(\nabla^{k\tau}_{x_k})||^2_2} &(47)\\
&\leq \sqrt{2n}\sqrt{\sum_k \Big(\frac{1}{\eta_k}\Big)\eta_k||\Pi_\Delta(\nabla^{k\tau}_{x_k})||^2_2} &(48)\\
&\leq \sqrt{\frac{2n}{\min_k \eta_k}}\sqrt{\sum_k \eta_k||\Pi_\Delta(\nabla^{k\tau}_{x_k})||^2_2} &(49)\\
&= \sqrt{\frac{2n}{\min_k \eta_k}}\sqrt{\mathcal{L}^\tau(\boldsymbol{x})} &(50)
\end{aligned}$$

□

A.4 Unbiased Estimation

Lemma 5. *An unbiased estimate of $\mathcal{L}(\boldsymbol{x})$ can be obtained by drawing two samples (pure strategies) from each players' mixed strategy and observing payoffs.*

Proof. Define $\nabla^k_{x_k}$ as the random variable distributed according to the distribution induced by all players' mixed strategies. Let $\nabla^{k(1)}_{x_k}$ and $\nabla^{k(2)}_{x_k}$ represent two other independent random variables, distributed identically to $\nabla^k_{x_k}$. Then

$$\mathbb{E}_{a_k \sim x_k \forall k}[\mathcal{L}(\boldsymbol{x})] = \mathbb{E}_{a_k \sim x_k \forall k}[\sum_k \eta_k (||\nabla^k_{x_k}||^2 - \frac{1}{m_k}(\mathbf{1}^\top \nabla^k_{x_k})^2)] \tag{51}$$

$$= \sum_k \eta_k (\mathbb{E}_{a_k \sim x_k \forall k}[||\nabla^k_{x_k}||^2] - \frac{1}{m_k}\mathbb{E}_{a_k \sim x_k \forall k}[(\mathbf{1}^\top \nabla^k_{x_k})^2]) \tag{52}$$

$$= \sum_k \eta_k (\mathbb{E}_{a_k \sim x_k \forall k}[\sum_l (\nabla^k_{x_{kl}})^2] - \frac{1}{m_k}\mathbb{E}_{a_k \sim x_k \forall k}[(\sum_l \nabla^k_{x_{kl}})^2]) \tag{53}$$

$$= \sum_k \eta_k (\sum_l \mathbb{E}_{a_k \sim x_k \forall k}[(\nabla^k_{x_{kl}})^2] - \frac{1}{m_k}\mathbb{E}_{a_k \sim x_k \forall k}[(\sum_l \nabla^k_{x_{kl}})^2]) \tag{54}$$

$$= \sum_k \eta_k \Big(\sum_l \mathbb{E}_{a_k \sim x_k \forall k}[\nabla^{k(1)}_{x_{kl}}]\mathbb{E}_{a_k \sim x_k \forall k}[\nabla^{k(2)}_{x_{kl}}] \tag{55}$$

$$- \frac{1}{m_k}\mathbb{E}_{a_k \sim x_k \forall k}[\sum_l \nabla^{k(1)}_{x_{kl}}]\mathbb{E}_{a_k \sim x_k \forall k}[\sum_l \nabla^{k(2)}_{x_{kl}}]\Big) \tag{56}$$

$$= \sum_k \eta_k \Big(\sum_l \mathbb{E}_{a_j \sim x_j \forall j \neq k}[\nabla^{k(1)}_{x_{kl}}]\mathbb{E}_{a_j \sim x_j \forall j \neq k}[\nabla^{k(2)}_{x_{kl}}] \tag{57}$$

$$- \frac{1}{m_k}\mathbb{E}_{a_j \sim x_j \forall j \neq k}[\sum_l \nabla^{k(1)}_{x_{kl}}]\mathbb{E}_{a_j \sim x_j \forall j \neq k}[\sum_l \nabla^{k(2)}_{x_{kl}}]\Big) \tag{58}$$

$$= \sum_k \eta_k \Big(\big[\hat{\nabla}^{k(1)}_{x_k}\big]^\top \hat{\nabla}^{k(2)}_{x_k} - \frac{1}{m_k}(\mathbf{1}^\top \hat{\nabla}^{k(1)}_{x_k})(\mathbf{1}^\top \hat{\nabla}^{k(2)}_{x_k})\Big) \tag{59}$$

$$= \sum_k \eta_k \underbrace{(\hat{\nabla}^{k(1)}_{x_k} - \frac{\mathbf{1}}{m_k}(\mathbf{1}^\top \hat{\nabla}^{k(1)}_{x_k})\mathbf{1})^\top}_{\text{appx. br gradient}} \underbrace{\hat{\nabla}^{k(2)}_{x_k}}_{\text{exp. payoffs}} \tag{60}$$

where $\hat{\nabla}^{k(p)}_{x_k}$ is an unbiased estimator of player k's gradient.

$\square$

Lemma 6. *The loss formed as the sum of the squared norms of the projected-gradients, $\mathcal{L}^\tau$, can be decomposed into three terms as follows:*

$$\mathcal{L}^\tau(\boldsymbol{x}) = \underbrace{\sum_k \eta_k x_q^\top B_{kq}^\top B_{kq} x_q}_{(A)} + \underbrace{2\sum_k \eta_k E_k^\top B_{kq} x_q}_{(B)} + \underbrace{\sum_k \eta_k E_k^\top E_k}_{(C)} \tag{61}$$

where q is any player other than k.

Proof. Let $S^\tau = -\tau \sum_l x_{kl} \log(x_{kl})$ so that $\frac{\partial S^\tau}{\partial x_k} = -\tau(\ln(x_k) + \mathbf{1})$. Note that $\Pi_{T\Delta}[\frac{\partial S^\tau}{\partial x_k}] = -\tau \ln(x_k)$.

$$\mathcal{L}^\tau(\boldsymbol{x}) = \sum_k \eta_k (\Pi_{T\Delta}[\nabla^k_{x_k}])^\top \Pi_{T\Delta}[\nabla^k_{x_k}] \tag{62}$$
$$= \sum_k \eta_k [H^k_{kq} x_q + \frac{\partial S^\tau}{\partial x_k}]^\top [I - \frac{1}{m_k}\mathbf{1}\mathbf{1}^\top][I - \frac{1}{m_k}\mathbf{1}\mathbf{1}^\top][H^k_{kq} x_q + \frac{\partial S^\tau}{\partial x_k}] \tag{63}$$
$$= \sum_k \eta_k \Big(x_q^\top [H^k_{kq}]^\top [I - \frac{1}{m_k}\mathbf{1}\mathbf{1}^\top]^2 [H^k_{kq}] x_q + 2[\frac{\partial S^\tau}{\partial x_k}]^\top [I - \frac{1}{m_k}\mathbf{1}\mathbf{1}^\top]^2 [H^k_{kq} x_q] \tag{64}$$
$$+ [\frac{\partial S^\tau}{\partial x_k}]^\top [I - \frac{1}{m_k}\mathbf{1}\mathbf{1}^\top]^2 [\frac{\partial S^\tau}{\partial x_k}] \Big) \tag{65}$$
$$= \underbrace{\sum_k \eta_k x_q^\top B_{kq}^\top B_{kq} x_q}_{(A)} + \underbrace{2\sum_k \eta_k E_k^\top B_{kq} x_q}_{(B)} + \underbrace{\sum_k \eta_k E_k^\top E_k}_{(C)} \tag{66}$$

where $B_{kq} = [I - \frac{1}{m_k}\mathbf{1}\mathbf{1}^\top] H^k_{kq}$ and $E_k = [I - \frac{1}{m_k}\mathbf{1}\mathbf{1}^\top][\frac{\partial S^\tau}{\partial x_k}] = -\tau \ln(x_k)$.

□

A.5 Bound on Loss

By equation (51), we can also rewrite this loss as a weighted sum of 2-norms, $\mathcal{L}(\boldsymbol{x}) = \sum_k \eta_k ||\nabla^k_{x_k} - \mu_k||^2_2$ where $\mu_k = \frac{1}{m_k}(\mathbf{1}^\top \nabla^k_{x_k}) \in [0, 1]$ for brevity. This will allow us to more easily analyze our loss.

Lemma 7. *Assume payoffs are bounded by* 1*, then setting* $\eta_k \leq \frac{4}{nm_k}$ *or* $\eta_k \leq \frac{4}{n\bar{m}}$ *or* $\sum_k \eta_k \leq \frac{4}{\bar{m}}$ *ensures* $0 \leq \mathcal{L}(x) \leq 1$ *for all* $x \in \mathcal{X}$.

Proof.

$$0 \leq \mathcal{L}(\boldsymbol{x}) = \sum_k \eta_k ||\nabla^k_{x_k} - \mu_k||^2_2 \tag{67}$$
$$= \sum_k \eta_k m_k \Big[\frac{1}{m_k} \sum_l (\nabla^k_{x_{kl}} - \mu_k)^2\Big] \tag{68}$$
$$= \sum_k \eta_k m_k Var[\nabla^k_{x_k}] \tag{69}$$
$$\leq \frac{1}{4} \sum_k \eta_k m_k \tag{70}$$
$$\leq \frac{1}{4}(\max_k \eta_k)(\sum_k m_k) \tag{71}$$
$$= \frac{1}{4}(\max_k \eta_k) n\bar{m} \leq 1 \tag{72}$$
$$\implies (\max_k \eta_k) \leq \frac{4}{n\bar{m}}. \tag{73}$$

□

The kth element of the sum in the loss does not depend on agent k's strategy. We will rewrite the loss to make its dependence on all other players' strategies more obvious ($l, q \neq k$ below).

$$
\begin{aligned}
\mathcal{L}(\boldsymbol{x}) &= \sum_k \eta_k \big([H^k_{kl} x_l]^\top [H^k_{kq} x_q] - \frac{1}{m_k} (\mathbf{1}^\top [H^k_{kl} x_l])(\mathbf{1}^\top [H^k_{kq} x_q])\big) && (74)\\
&= \sum_k \eta_k \big([H^k_{kl} x_l]^\top [H^k_{kq} x_q] - \frac{1}{m_k} [H^k_{kl} x_l]^\top \mathbf{1}\mathbf{1}^\top [H^k_{kq} x_q]\big) && (75)\\
&= \sum_k \eta_k [H^k_{kl} x_l]^\top [I - \frac{1}{m_k} \mathbf{1}\mathbf{1}^\top][H^k_{kq} x_q] && (76)\\
&= \sum_k \eta_k x_l^\top [H^k_{kl}]^\top [I - \frac{1}{m_k} \mathbf{1}\mathbf{1}^\top][H^k_{kq}] x_q && (77)\\
&= \sum_k \eta_k x_q^\top [H^k_{kq}]^\top [I - \frac{1}{m_k} \mathbf{1}\mathbf{1}^\top][H^k_{kq}] x_q \quad \text{isolate dep. on } q && (78)\\
&= \sum_k \eta_k x_q^\top [H^k_{qk}][I - \frac{1}{m_k} \mathbf{1}\mathbf{1}^\top][H^k_{kq}] x_q && (79)\\
&= \sum_k \eta_k x_q^\top A_{qkq} x_q. && (80)
\end{aligned}
$$

where $A_{qkq} = [H^k_{qk}][I - \frac{1}{m_k}\mathbf{1}\mathbf{1}^\top][H^k_{kq}]$ does not depend on x_k.

Note this means we can also write $\mathcal{L}(\boldsymbol{x}) = \sum_k \eta_k x_l^\top A_{lkq} x_q$ for any $l, q \neq k$.

Lemma 8. *Assume payoffs are bounded in* $[0,1]$*, then*

$$
|\mathcal{L}^\tau(\boldsymbol{x})| \leq \frac{1}{4}(\max_k \eta_k) n\bar{m}\Big(\frac{\ln(m^*)}{\ln(1/p)} + 2\Big)^2 \qquad (81)
$$

for any $\boldsymbol{x}$ *such that* $x_{kl} \geq \frac{p}{m^*}\ \forall k, l$.

Proof. Starting from the definition of $\mathcal{L}^\tau$ and applying Lemma 19 along with intermediate results from Lemma 16, we find

$$
\begin{aligned}
|\mathcal{L}^\tau(\boldsymbol{x})| &= |\sum_k \eta_k ||\Pi_{T\Delta}(\nabla^k_{x_k})||^2| && (82)\\
&\leq \frac{1}{4}\sum_k \eta_k m_k (\tau \ln(\frac{1}{x_{\min}}) + 1)^2 && (83)\\
&= \frac{1}{4}\sum_k \eta_k m_k \Big(\frac{1}{\ln(1/p)} \ln\big(\frac{m^*}{p}\big) + 1\Big)^2 && (84)\\
&= \frac{1}{4}\sum_k \eta_k m_k \Big(\frac{\ln(m^*)}{\ln(1/p)} + 2\Big)^2 && (85)\\
&= \frac{1}{4}(\max_k \eta_k) n\bar{m}\Big(\frac{\ln(m^*)}{\ln(1/p)} + 2\Big)^2. && (86)
\end{aligned}
$$

□

B QREs Approximate NEs at Low Temperature

Lemma 9. *Setting* $\tau = \ln(1/p)^{-1}$ *with* $p \in [0,1)$ *ensures that all QREs contain probabilities greater than* $\frac{p}{\max_k m_k}$.

Proof. What must τ be to ensure $x_{lp} \geq x_{\min}$ for any l, p? We can check the case where $\nabla = e_i$. Let $m^* = \max_k m_k$. Then

$$x_{\min} = \min_k \min_{\nabla^k_{x_k}} \min_l \left[\texttt{softmax}(\frac{\nabla^k_{x_k}}{\tau})\right]_l \tag{87}$$
$$= \frac{e^0}{(m^*-1)e^{\frac{1}{\tau}} + e^0} \tag{88}$$
$$= \frac{1}{(m^*-1)e^{\frac{1}{\tau}} + 1} \tag{89}$$
$$\implies e^{\frac{1}{\tau}} = \frac{1}{m^*-1}(\frac{1}{x_{\min}} - 1) \tag{90}$$
$$\implies \tau = \frac{1}{\ln\left(\frac{1}{m^*-1}(\frac{1}{x_{\min}} - 1)\right)}. \tag{91}$$

If $x_{\min} = \frac{p}{m^*}$ with $p \in [0, 1]$, then

$$\tau^* = \frac{1}{\ln\left(\frac{1}{m^*-1}(\frac{1}{x_{\min}} - 1)\right)} \tag{92}$$
$$= \frac{1}{\ln\left(\frac{1}{m^*-1}(\frac{m^*}{p} - 1)\right)} \tag{93}$$
$$= \frac{1}{\ln\left(\frac{m^*-p}{m^*-1}\frac{1}{p}\right)} \tag{94}$$
$$\leq \frac{1}{\ln(\frac{1}{p})}. \tag{95}$$

This implies if we set $\tau = \ln(1/p)^{-1}$, then we are guaranteed that all QREs contain probabilities greater than $x_{\min} = \frac{p}{\max_k m_k}$. □

Lemma 10 (Repeated from Lemma 1 of [29]). *Let $\nabla^k_{x_k}$ be player k's gradient ($m_k \geq 2$) with payoffs bounded in $[0, 1]$ and $\boldsymbol{x}$ be a QRE at temperature τ. Then it holds that*

$$u_k(BR_k, x_{-k}) - u_k(\boldsymbol{x}) = \max(\nabla^k_{x_k}) - (\nabla^k_{x_k})^\top \texttt{softmax}\Big(\frac{\nabla^k_{x_k}}{\tau}\Big) \leq \tau(W(1/e) + \frac{m_k - 2}{e}) \tag{96}$$

where W is the Lambert function ($W(1/e) \approx 0.278$).

Lemma 11 (Slightly modified from Proposition 5.1a of [6]). *Let $\psi_e(x_k) = \sum_l x_{kl}\ln(x_{kl})$ if $x_k \in \Delta^{m_k-1}$ else $+\infty$. Then $\psi_e(x_k)$ is 1-strongly convex over $int\Delta^{m_k-1}$ w.r.t. the $||\cdot||_1$ and $||\cdot||_2$ norms, i.e.,*

$$\langle \nabla\psi_e(x) - \nabla\psi_e(y), x - y\rangle \geq ||x-y||_1^2 \geq ||x-y||_2^2 \tag{97}$$
$$\implies \psi_e(y) \geq \psi_e(x) + \nabla\psi_e(x)^\top(y-x) + \frac{1}{2}||y-x||_2^2. \tag{98}$$

for all $x, y \in int\Delta^{m_k-1}$.

Lemma 12. *Let $l(x|x_k) = \langle \nabla f_k(x_k), x\rangle + \frac{1}{t_k}B_{\psi_e}(x, x_k)$ where $t_k > 0$, $f_k(z) = -\epsilon_k(z) = -[u_k(z, x_{-i}) + S^\tau(z) - u_k(\boldsymbol{x}) - S^\tau(x_k)]$, and $B_{\psi_e}(x, x_k) = \psi_e(x) - \psi_e(y) - \langle x - y, \nabla\psi_e(y)\rangle$ with ψ_e defined in Lemma 11. Finally, let $x_{k+1} = \arg\min_{x\in int\Delta} l(x|x_k)$. Then*

$$||x_k - x_{k+1}|| \leq 2||\Pi_\Delta(\nabla^{k\tau}_{x_k})||. \tag{99}$$

Proof. Plugging $\psi_e(x_k) = \sum_l x_{kl}\ln(x_{kl}) = -S(x_k)$ on $int\Delta$ into $B_{\psi_e}(x, x_k)$, we find

$$B_{\psi_e}(x, x_k) = S(x_k) - S(x) - \langle \ln(x_k) + 1, x - x_k\rangle \tag{100}$$
$$= S(x_k) - S(x) - \langle \ln(x_k), x - x_k\rangle \tag{101}$$

for all $x, x_k \in int\Delta$. Note that $-S(x)$ is 1-strongly convex on $int\Delta$, therefore, $B_{\psi_e}(x, x_k)$ is also 1-strongly convex in x. Continuing, this also implies $l(x|x_k)$ is 1-strongly convex.

Let $x_{k+1} = \arg\min_{x \in int\Delta} l(x|x_k)$ and note that $\nabla f_k(x_k) = -\nabla \epsilon_k = -\nabla^{k\tau}_{x_k}$. Strong convexity of l implies

$$l(x_{k+1}) \geq l(x_k) + \nabla_x l(x_k)^\top (x_{k+1} - x_k) + \frac{1}{2}||x_{k+1} - x_k||_2^2 \tag{102}$$

$$\implies ||x_k - x_{k+1}||_2^2 \leq 2\Big[\underbrace{l(x_{k+1}) - l(x_k)}_{\leq 0} + \nabla_x l(x_k)^\top (x_k - x_{k+1})\Big] \tag{103}$$

$$\leq 2\nabla_x l(x_k)^\top (x_k - x_{k+1}) = 2(\nabla f_k(x_k) + \frac{1}{t_k}[\ln(x_k) + \mathbf{1} - \ln(x_k)])^\top (x_k - x_{k+1}) \tag{104}$$

$$= 2\nabla f_k(x_k)^\top (x_k - x_{k+1}) = 2(\nabla^{k\tau}_{x_k})^\top (x_{k+1} - x_k) = 2(\Pi_\Delta(\nabla^{k\tau}_{x_k}))^\top (x_{k+1} - x_k) \tag{105}$$

$$\leq 2||\Pi_\Delta(\nabla^{k\tau}_{x_k})||||x_k - x_{k+1}||. \tag{106}$$

Rearranging the inequality achieves the desired result. □

Lemma 13. *[Low Temperature Approximate QREs are Approximate Nash Equilibria] Let $\nabla^{k\tau}_{x_k}$ be player k's entropy regularized gradient with payoffs bounded in $[0, 1]$ and $\boldsymbol{x}$ be an approximate QRE. Then it holds that*

$$u_k(\mathit{BR}_k, x_{-k}) - u_k(\boldsymbol{x}) \leq \tau(W(1/e) + \frac{m_k - 2}{e}) + 2\sqrt{m_k}||\Pi_\Delta(\nabla^{k\tau}_{x_k})|| \tag{107}$$

where W is the Lambert function ($W(1/e) \approx 0.278$).

Proof. First note that $x_k = \texttt{softmax}(\ln(x_k))$ for $x_k \in int\Delta$. Recall that the $\texttt{softmax}$ is invariant to constant offsets to its argument, i.e., $\texttt{softmax}(z + c\mathbf{1}) = \texttt{softmax}(z)$ for any $c \in \mathbb{R}$. Then

$$\texttt{softmax}(\frac{\nabla^k_{x_k}}{\tau}) = \texttt{softmax}(\ln(x_k) - \frac{1}{\tau}[-\nabla^k_{x_k} + \tau\ln(x_k)]) \tag{108}$$

$$= \texttt{softmax}(\ln(x_k) - \frac{1}{\tau}[-\nabla^k_{x_k} + \tau\ln(x_k) + \tau\mathbf{1}]) \tag{109}$$

$$= \texttt{softmax}(\ln(x_k) - \frac{1}{\tau}\nabla f_k(x_k)) \tag{110}$$

$$= \arg\min_{x \in int\Delta} l(x|x_k) \text{ with } t_k = \frac{1}{\tau} \tag{111}$$

$$= x_k^* \tag{112}$$

where the closed-form solution to the minimization problem as a $\texttt{softmax}$ formula comes from inspecting the Entropic Descent Algorithm (EDA) of [6].

Then, beginning with the definition of exploitability, we find

$$u_k(\texttt{BR}_k, x_{-k}) - u_k(\boldsymbol{x}) = u_k(\texttt{BR}_k, x_{-k}) - (\nabla^k_{x_k})^\top x_k \tag{113}$$

$$= u_k(\texttt{BR}_k, x_{-k}) - (\nabla^k_{x_k})^\top \texttt{softmax}(\frac{\nabla^k_{x_k}}{\tau}) - (\nabla^k_{x_k})^\top (x_k - \texttt{softmax}(\frac{\nabla^k_{x_k}}{\tau})) \tag{114}$$

$$\leq \tau(W(1/e) + \frac{m_k - 2}{e}) + ||\nabla^k_{x_k}|| \cdot ||x_k - \texttt{softmax}(\frac{\nabla^k_{x_k}}{\tau})|| \tag{115}$$

$$= \tau(W(1/e) + \frac{m_k - 2}{e}) + ||\nabla^k_{x_k}|| \cdot ||x_k - x_k^*|| \tag{116}$$

$$\leq \tau(W(1/e) + \frac{m_k - 2}{e}) + 2\sqrt{m_k}||\Pi_\Delta(\nabla^{k\tau}_{x_k})||. \tag{117}$$

□

Lemma 14. *[$\mathcal{L}^\tau$ Scores Nash Equilibria] Let $\mathcal{L}^\tau(\boldsymbol{x})$ be our proposed entropy regularized loss function with payoffs bounded in $[0,1]$ and $\boldsymbol{x}$ be an approximate QRE. Then it holds that*

$$\epsilon \leq n\tau(W(1/e) + \frac{\bar{m}-2}{e}) + 2\sqrt{\frac{n \max_k m_k}{\min_k \eta_k}}\sqrt{\mathcal{L}^\tau(\boldsymbol{x})} \tag{118}$$

where W is the Lambert function ($W(1/e) \approx 0.278$).

Proof. Beginning with the definition of exploitability and applying Lemma 13, we find

$$\begin{aligned}
\epsilon &= \max_k u_k(\texttt{BR}_k, x_{-k}) - u_k(\boldsymbol{x}) && (119)\\
&\leq \sum_k u_k(\texttt{BR}_k, x_{-k}) - u_k(\boldsymbol{x}) && (120)\\
&\leq \sum_k \Big[\tau(W(1/e) + \frac{m_k-2}{e}) + 2\sqrt{m_k}||\Pi_\Delta(\nabla^{k\tau}_{x_k})||\Big] && (121)\\
&= n\tau(W(1/e) + \frac{\bar{m}-2}{e}) + 2\sum_k \sqrt{m_k}||\Pi_\Delta(\nabla^{k\tau}_{x_k})|| && (122)\\
&\leq n\tau(W(1/e) + \frac{\bar{m}-2}{e}) + 2\sqrt{\max_k m_k}\sum_k ||\Pi_\Delta(\nabla^{k\tau}_{x_k})|| && (123)\\
&\leq n\tau(W(1/e) + \frac{\bar{m}-2}{e}) + 2\sqrt{\frac{n \max_k m_k}{\min_k \eta_k}}\sqrt{\mathcal{L}^\tau(\boldsymbol{x})}. && (124)
\end{aligned}$$

where the last inequality follows from the same steps outlined in Lemma 3, which established the relationship between $\mathcal{L}(\boldsymbol{x})$ and ϵ.

□

C Gradient of Loss

Lemma 15. *The gradient of $\mathcal{L}^\tau(\boldsymbol{x})$ with respect to player l's strategy x_l is*

$$\nabla_{x_l}\mathcal{L}(\boldsymbol{x}) = 2\sum_k \eta_k B_{kl}^\top \Pi_{T\Delta}(\nabla^{k\tau}_{x_k}) \tag{125}$$

where $B_{ll} = -\tau[I - \frac{1}{m_l}\mathbf{1}\mathbf{1}^\top]\texttt{diag}(\frac{1}{x_l})$ and $B_{kl} = [I - \frac{1}{m_k}\mathbf{1}\mathbf{1}^\top]H^k_{kl}$ for $k \neq l$.

Proof. Recall from Lemma 6 that the loss can be decomposed as $\mathcal{L}^\tau(\boldsymbol{x}) = (A) + (B) + (C)$.

Then

$$D_{x_l}[(A)] = D_{x_l}[\sum_k \eta_k x_q^\top B_{kq}^\top B_{kq} x_q] = 2\sum_{k\neq l}\eta_k B_{kl}^\top B_{kl} x_l \tag{126}$$

where $q \neq k$ and $B_{kq} = [I - \frac{1}{m_k}\mathbf{1}\mathbf{1}^\top][H^k_{kq}]$ does not depend on x_k.

Also, letting $B_{ll} = -\tau[I - \frac{1}{m_l}\mathbf{1}\mathbf{1}^\top]\texttt{diag}(\frac{1}{x_l})$,

$$\begin{aligned}
D_{x_l}[(B)] &= D_{x_l}[-2\tau\sum_k \eta_k \ln(x_k)^\top B_{kq} x_q] && (127)\\
&= -2\tau\big[\eta_l D_{x_l}[\ln(x_l)^\top B_{lq} x_q] + \sum_{k\neq l}\eta_k D_{x_l}[\ln(x_k)^\top B_{kl} x_l]\big] && (128)\\
&= -2\tau\big[\eta_l \texttt{diag}(\frac{1}{x_l})B_{lq}x_q + \sum_{k\neq l}\eta_k B_{kl}^\top \ln(x_k)\big] && (129)\\
&= -2\tau\big[\eta_l([I - \frac{1}{m_l}\mathbf{1}\mathbf{1}^\top]\texttt{diag}(\frac{1}{x_l}))^\top \Pi_{T\Delta}(\nabla^l) + \sum_{k\neq l}\eta_k B_{kl}^\top \ln(x_k)\big] && (130)\\
&= 2\big[\eta_l B_{ll}^\top \Pi_{T\Delta}(\nabla^l) - \tau\sum_{k\neq l}\eta_k B_{kl}^\top \ln(x_k)\big]. && (131)
\end{aligned}$$

And

$$D_{x_l}[(C)] = D_{x_l}[\sum_k \eta_k \tau^2 \ln(x_k)^\top [I - \frac{1}{m_k}\mathbf{1}\mathbf{1}^\top] \ln(x_k)] \tag{132}$$

$$= 2\tau^2 \Big[\eta_l \texttt{diag}(\frac{1}{x_l})[I - \frac{1}{m_l}\mathbf{1}\mathbf{1}^\top] \ln(x_l)\Big] \tag{133}$$

$$= -2\tau\eta_l([I - \frac{1}{m_l}\mathbf{1}\mathbf{1}^\top]\texttt{diag}(\frac{1}{x_l}))^\top \Pi_{T\Delta}(-\tau \ln(x_l)) \tag{134}$$

$$= 2\eta_l B_{ll}^\top \Pi_{T\Delta}(-\tau \ln(x_l)). \tag{135}$$

Putting these together, we find

$$\nabla_{x_l}\mathcal{L}(\boldsymbol{x}) = 2\sum_{k \neq l} \eta_k B_{kl}^\top (B_{kl} x_l - \tau \ln(x_k)) + 2\eta_l B_{ll}^\top \left[\Pi_{T\Delta}(\nabla^l) + \Pi_{T\Delta}(-\tau \ln(x_l))\right] \tag{136}$$

$$= 2\eta_l B_{ll}^\top \Pi_{T\Delta}(\nabla_{x_k}^{k\tau}) + 2\sum_{k \neq l} \eta_k B_{kl}^\top \Pi_{T\Delta}(\nabla_{x_k}^{k\tau}) \tag{137}$$

$$= 2\sum_k \eta_k B_{kl}^\top \Pi_{T\Delta}(\nabla_{x_k}^{k\tau}). \tag{138}$$

□

C.1 Unbiased Estimation

In order to construct an unbiased estimate of A_{lkl}, we will need to form two independent unbiased estimates of H_{kl}^k. Recall that H_{kl}^k is simply the expected bimatrix game between players k and l when all other players sample their actions according to their current strategies.

C.2 Bound on Gradient / Lipschitz Property

Lemma 16. *Assume payoffs are upper bounded by* 1*, then the infinity norm of the gradient is bounded as*

$$||\nabla_{\boldsymbol{x}}\mathcal{L}^\tau(\boldsymbol{x})||_\infty \leq \frac{1}{2}(\max_k \eta_k)(\tau \ln\big(\frac{1}{x_{\min}}\big) + 1)\Big[\tau m^*\big(\frac{1}{x_{\min}} - 1\big) + n\bar{m}\Big]. \tag{139}$$

Proof. Recall from Lemma 15 that the gradient of $\mathcal{L}(\boldsymbol{x})$ with respect to player l's strategy x_l is

$$\nabla_{x_l}\mathcal{L}(\boldsymbol{x}) = 2\sum_k \eta_k B_{kl}^\top \Pi_{T\Delta}(\nabla_{x_k}^{k\tau}) \tag{140}$$

where $B_{ll} = -\tau[I - \frac{1}{m_l}\mathbf{1}\mathbf{1}^\top]\texttt{diag}(\frac{1}{x_l})$ and $B_{kl} = [I - \frac{1}{m_k}\mathbf{1}\mathbf{1}^\top]H_{kl}^k$ for $k \neq l$.

For payoffs in $[0,1]$, the entries in $\nabla_{x_k}^{k\tau} = \nabla_{x_k}^k - \tau\ln(x_k)$ are bounded within $[0, \tau\ln(\frac{1}{x_{\min}}) + 1]$ with a range $\tau\ln(\frac{1}{x_{\min}}) + 1$. Similarly, the entries in $-\tau\texttt{diag}(\frac{1}{x_l})$ are bounded within $[-\tau\frac{1}{x_{\min}}, -\tau]$ with a range of $\tau(\frac{1}{x_{\min}} - 1)$.

The infinity norm of the gradient can then be bounded as

$$||\nabla_{\boldsymbol{x}}\mathcal{L}^\tau(\boldsymbol{x})||_\infty = \max_l ||\nabla_{x_l}\mathcal{L}(\boldsymbol{x})||_\infty \tag{141}$$

$$= \max_l ||2\sum_k \eta_k B_{kl}^\top \Pi_{T\Delta}(\nabla^{k\tau}_{x_k})||_\infty \tag{142}$$

$$\leq 2\sum_k \eta_k \max_l ||B_{kl}^\top \Pi_{T\Delta}(\nabla^{k\tau}_{x_k})||_\infty \tag{143}$$

$$\leq \frac{1}{2}\sum_{k\neq l^*} \eta_k m_k (\tau \ln\big(\frac{1}{x_{\min}}\big) + 1) + \frac{1}{2}\eta_{l^*} m_{l^*} \tau\big(\frac{1}{x_{\min}} - 1\big)(\tau \ln\big(\frac{1}{x_{\min}}\big) + 1) \tag{144}$$

$$= \frac{1}{2}(\tau \ln\big(\frac{1}{x_{\min}}\big) + 1)\Big[\eta_{l^*} m_{l^*} \tau\big(\frac{1}{x_{\min}} - 1\big) + \sum_{k\neq l^*} \eta_k m_k\Big] \tag{145}$$

$$\leq \frac{1}{2}(\max_k \eta_k)(\tau \ln\big(\frac{1}{x_{\min}}\big) + 1)\Big[\tau m_{l^*}\big(\frac{1}{x_{\min}} - 1\big) + \sum_{k\neq l^*} m_k\Big] \tag{146}$$

$$\leq \frac{1}{2}(\max_k \eta_k)(\tau \ln\big(\frac{1}{x_{\min}}\big) + 1)\Big[\tau m^*\big(\frac{1}{x_{\min}} - 1\big) + n\bar{m}\Big] \tag{147}$$

where the second inequality follows from Lemma 19.

□

Corollary 2. *If τ is set according to Lemma 9, then the infinity norm of the gradient is bounded as*

$$||\nabla_{\boldsymbol{x}}\mathcal{L}^\tau(\boldsymbol{x})||_\infty \leq \frac{1}{2}(\max_k \eta_k)\Big[\frac{\ln(m^*)}{\ln(1/p)} + 2\Big]\Big[\frac{m^{*2}}{p\ln(1/p)} + n\bar{m}\Big] = \frac{1}{2}(\max_k \eta_k)\hat{L} \tag{148}$$

where $m^* = \max_k m_k$ and $\hat{L}$ is defined implicitly for convenience in other derivations.

Proof. Starting with Lemma 16 and applying Lemma 9 (i.e., $\tau = \ln(1/p)^{-1}$ and $x_{\min} = \frac{p}{m^*}$ where $m^* = \max_k m_k$), we find

$$||\nabla_{\boldsymbol{x}}\mathcal{L}^\tau(\boldsymbol{x})||_\infty \leq \frac{1}{2}(\max_k \eta_k)(\tau \ln\big(\frac{1}{x_{\min}}\big) + 1)\Big[\tau m^*\big(\frac{1}{x_{\min}} - 1\big) + n\bar{m}\Big] \tag{149}$$

$$= \frac{1}{2}(\max_k \eta_k)\Big[\frac{\ln(m^*/p)}{\ln(1/p)} + 1\Big]\Big[\frac{m^*}{\ln(1/p)}\big(\frac{m^*}{p} - 1\big) + n\bar{m}\Big] \tag{150}$$

$$\leq \frac{1}{2}(\max_k \eta_k)\Big[\frac{\ln(m^*)}{\ln(1/p)} + 2\Big]\Big[\frac{m^{*2}}{p\ln(1/p)} + n\bar{m}\Big]. \tag{151}$$

As $p \rightarrow 0^+$, the norm of the gradient blows up because the gradient of Shannon entropy blows up for small probabilities. As $p \rightarrow 1$, the norm of the gradient blows up because we require infinite temperature τ to guarantee all QREs are nearly uniform; recall τ is the regularization coefficient on the entropy bonus terms which means our modified utilities blow up for large τ. In practice, setting p to $\mathcal{O}(1)$, e.g., $p = \frac{1}{10}$ is sufficient. □

D Hessian of Loss

We will now derive the Hessian of our loss. This will be useful in establishing properties about global minima that enable the application of tailored minimization algorithms. Let $D_z[f(z)]$ denote the differential operator applied to (possibly multivalued) function f with respect to z. For example, $D_{x_q}[H^k_{lk}] = D_{x_q}[x_q T^k_{qlk}] = T^k_{qlk}$ where T^k_{qlk} is player k's payoff tensor according to the three-way approximation between players k, l, and q to the game at $\boldsymbol{x}$.

Lemma 17. *The Hessian of $\mathcal{L}^\tau(\boldsymbol{x})$ can be written*

$$\texttt{Hess}(\mathcal{L}^\tau) = 2\tilde{B}^\top \tilde{B} + T\Pi_{T\Delta}(\tilde{\nabla}^\tau) \tag{152}$$

where $\tilde{B}_{kl} = \sqrt{\eta_k} B_{kl}$, $\Pi_{T\Delta}(\tilde{\nabla}^\tau) = [\eta_1 \Pi_{T\Delta}(\nabla^{1\tau}_{x_1}), \ldots, \eta_n \Pi_{T\Delta}(\nabla^{n\tau}_{x_n})]$, and we augment T (the 3-player tensor approximation to the game, T^k_{lqk}) so that $T^l_{lll} = \tau\, \texttt{diag3}(\frac{1}{x_l^2})$ and otherwise 0.

Proof. Recall the gradient of our proposed loss:

$$\nabla_{x_l}\mathcal{L}(\boldsymbol{x}) = 2\sum_k \eta_k B_{kl}^\top \Pi_{T\Delta}(\nabla_{x_k}^{k\tau}) \tag{153}$$

where $B_{ll} = -\tau[I - \frac{1}{m_l}\mathbf{1}\mathbf{1}^\top]\texttt{diag}(\frac{1}{x_l})$ and $B_{kl} = [I - \frac{1}{m_k}\mathbf{1}\mathbf{1}^\top]H_{kl}^k$ for $k \neq l$.
Consider the following Jacobians, which will play an auxiliary role in our derivation of the Hessian:

$$\begin{aligned}
D_l[B_{ll}] &= \tau[I - \frac{1}{m_l}\mathbf{1}\mathbf{1}^\top]\texttt{diag3}(\frac{1}{x_l^2}) && (154)\\
D_q[B_{ll}] &= \mathbf{0} && (155)\\
D_l[B_{kl}] &= \mathbf{0} && (156)\\
D_q[B_{kl}] &= [I - \frac{1}{m_k}\mathbf{1}\mathbf{1}^\top]T_{klq}^k && (157)\\
D_k[\Pi_{T\Delta}(\nabla_{x_k}^{k\tau})] &= [I - \frac{1}{m_k}\mathbf{1}\mathbf{1}^\top]D_k[\nabla_{x_k}^{k\tau}] && (158)\\
&= [I - \frac{1}{m_k}\mathbf{1}\mathbf{1}^\top]D_k[\nabla_{x_k}^{k} - \tau\ln(x_k)] && (159)\\
&= [I - \frac{1}{m_k}\mathbf{1}\mathbf{1}^\top][-\tau\texttt{diag}(\frac{1}{x_k})] && (160)\\
&= B_{kk} && (161)\\
D_l[\Pi_{T\Delta}(\nabla_{x_k}^{k\tau})] &= [I - \frac{1}{m_k}\mathbf{1}\mathbf{1}^\top]D_l[\nabla_{x_k}^{k\tau}] && (162)\\
&= [I - \frac{1}{m_k}\mathbf{1}\mathbf{1}^\top]D_l[\nabla_{x_k}^{k} - \tau\ln(x_k)] && (163)\\
&= [I - \frac{1}{m_k}\mathbf{1}\mathbf{1}^\top][H_{kl}^k] && (164)\\
&= B_{kl}. && (165)
\end{aligned}$$

We can derive the diagonal blocks of the Hessian as

$$\begin{aligned}
D_{ll}[\mathcal{L}(\boldsymbol{x})] &= D_l[\nabla_{x_l}\mathcal{L}(\boldsymbol{x})] && (166)\\
&= 2D_l[\sum_k \eta_k B_{kl}^\top \Pi_{T\Delta}(\nabla_{x_k}^{k\tau})] && (167)\\
&= 2\Big[\eta_l D_l\big[B_{ll}^\top \Pi_{T\Delta}(\nabla_{x_l}^{l\tau})\big] + \sum_{k\neq l}\eta_k D_l\big[B_{kl}^\top \Pi_{T\Delta}(\nabla_{x_k}^{k\tau})\big]\Big] && (168)\\
&= 2\Big[\eta_l\big[D_l[B_{ll}]^\top \Pi_{T\Delta}(\nabla_{x_l}^{l\tau}) + B_{ll}^\top D_l[\Pi_{T\Delta}(\nabla_{x_l}^{l\tau})]\big] && (169)\\
&+ \sum_{k\neq l}\eta_k\big[\cancel{D_l[B_{kl}]}^\top \Pi_{T\Delta}(\nabla_{x_k}^{k\tau}) + B_{kl}^\top D_l[\Pi_{T\Delta}(\nabla_{x_k}^{k\tau})]\big]\Big] && (170)\\
&= 2\Big[\eta_l\big[\tau\texttt{diag3}(\frac{1}{x_l^2})[I - \frac{1}{m_l}\mathbf{1}\mathbf{1}^\top]\Pi_{T\Delta}(\nabla_{x_l}^{l\tau}) + B_{ll}^\top B_{ll}\big] + \sum_{k\neq l}\eta_k B_{kl}^\top B_{kl}\Big] && (171)\\
&= 2\Big[\tau\eta_l\texttt{diag}\big([\frac{1}{x_l^2}] \odot \Pi_{T\Delta}(\nabla_{x_l}^{l\tau})\big) + \sum_k \eta_k B_{kl}^\top B_{kl}\Big] && (172)
\end{aligned}$$

and the off-diagonal blocks as

$$\begin{aligned}
D_{lq}[\mathcal{L}(\boldsymbol{x})] &= D_q[\nabla_{x_l}\mathcal{L}(\boldsymbol{x})] && (173)\\
&= 2D_q[\sum_k \eta_k B_{kl}^\top \Pi_{T\Delta}(\nabla_{x_k}^{k\tau})] && (174)\\
&= 2\Big[\eta_l D_q\big[B_{ll}^\top \Pi_{T\Delta}(\nabla_{x_l}^{l\tau})\big] + \sum_{k\neq l}\eta_k D_q\big[B_{kl}^\top \Pi_{T\Delta}(\nabla_{x_k}^{k\tau})\big]\Big] && (175)\\
&= 2\Big[\eta_l\big[\cancel{D_q[B_{ll}]^\top}\Pi_{T\Delta}(\nabla_{x_l}^{l\tau}) + B_{ll}^\top D_q[\Pi_{T\Delta}(\nabla_{x_l}^{l\tau})]\big] && (176)\\
&+ \sum_{k\neq l}\eta_k\big[D_q[B_{kl}]^\top \Pi_{T\Delta}(\nabla_{x_k}^{k\tau}) + B_{kl}^\top D_q[\Pi_{T\Delta}(\nabla_{x_k}^{k\tau})]\big]\Big] && (177)\\
&= 2\Big[\eta_l B_{ll}^\top B_{lq} + \sum_{k\neq l}\eta_k\big[T_{lqk}^k[I - \frac{1}{m_k}\mathbf{1}\mathbf{1}^\top]\Pi_{T\Delta}(\nabla_{x_k}^{k\tau}) + B_{kl}^\top B_{kq}\big]\Big] && (178)\\
&= 2\Big[\sum_k \eta_k B_{kl}^\top B_{kq} + \sum_{k\neq l}\eta_k T_{lqk}^k \Pi_{T\Delta}(\nabla_{x_k}^{k\tau})\Big]. && (179)
\end{aligned}$$

Therefore, the Hessian can be written concisely as

$$2\big[\tilde{B}^\top \tilde{B} + T\Pi_{T\Delta}(\tilde{\nabla}^\tau)\big] \tag{180}$$

where $\tilde{B}_{kl} = \sqrt{\eta_k}B_{kl}$, $\Pi_{T\Delta}(\tilde{\nabla}^\tau) = [\eta_1\Pi_{T\Delta}(\nabla_{x_1}^{1\tau}), \ldots, \eta_n\Pi_{T\Delta}(\nabla_{x_n}^{n\tau})]$, and we augment T (the 3-player tensor approximation to the game, T_{lqk}^k) so that $T_{lll}^l = \tau\texttt{diag3}(\frac{1}{x_l^2})$ and otherwise 0.

□

E Regret Bounds

Lemma 18. *[Loss Regret to Exploitability Regret] Assume exploitability of a joint strategy $\boldsymbol{x}$ is upper bounded by $f(\mathcal{L}^\tau(\boldsymbol{x}))$ where f is a concave function and $\mathcal{L}^\tau$ is a loss function. Let $\boldsymbol{x}_t$ be a joint strategy randomly drawn from the set of predictions made by an online learning algorithm $\mathcal{A}$ over T steps. Then the expected exploitability of $\boldsymbol{x}_t$ is bounded by the average regret of $\mathcal{A}$:*

$$\mathbb{E}[\epsilon_t] \leq f(\frac{1}{T}\sum_t \mathcal{L}_t). \tag{181}$$

Proof.

$$\begin{aligned}
\mathbb{E}[\epsilon_t] &= \mathbb{E}[f(\mathcal{L}(\boldsymbol{x}_t))] && (182)\\
&\leq f(\mathbb{E}[\mathcal{L}(\boldsymbol{x}_t)]) && (183)\\
&= f(\frac{1}{T}\sum_t \mathcal{L}(\boldsymbol{x}_t)) && (184)
\end{aligned}$$

where the second inequality follows from Jensen's inequality. □

Theorem 1. *[BLiN PAC Rate] Assume $\eta_k = \eta = 2/\hat{L}$ as defined in Lemma 2, $\tau = \frac{1}{\ln(1/p)}$ so that all equilibria place at least $\frac{p}{m^*}$ mass on each strategy, and a previously pulled arm is returned uniformly at random (i.e., $t \sim U(T)$). Then for any $w > 0$,*

$$\epsilon_t \leq w\Big[n\tau(W(1/e) + \frac{\bar{m}-2}{e}) + 4(1+(4c^2)^{1/3})\sqrt{nm^*\hat{L}}\Big(\frac{\ln T}{T}\Big)^{\frac{1}{2(d_z+2)}}\Big] \tag{185}$$

with probability $(1-w^{-1})(1-2T^{-2})$ where W is the Lambert function ($W(1/e) \approx 0.278$), $m^* = \max_k m_k$, and $c \leq \frac{1}{4}\frac{n\bar{m}}{\hat{L}}\Big(\frac{\ln(m^*)}{\ln(1/p)} + 2\Big)^2$ is an upper bound on the maximum sampled value from $\mathcal{L}^\tau$ (see Lemma 8).

Proof. Assume $\eta_k = \eta = \frac{2}{\hat{L}}$ as defined in Lemma 2 so that $\mathcal{L}^\tau$ is 1-Lipschitz with respect to $||\cdot||_\infty$. Also assume a previously pulled arm is returned uniformly at random. Starting with Lemma 14 and applying Corollary 9, we find

$$\mathbb{E}[\epsilon_t] \leq n\tau(W(1/e) + \frac{\bar{m}-2}{e}) + 2\sqrt{\frac{n \max_k m_k}{\min_k \eta_k}}\sqrt{\mathcal{L}^\tau(\boldsymbol{x})} \tag{186}$$

$$= \frac{n}{\ln(1/p)}(W(1/e) + \frac{\bar{m}-2}{e}) + \sqrt{2nm^*\hat{L}}\sqrt{8(1+(4c^2)^{1/3})^2 T^{\frac{-1}{(d_z+2)}} \ln T^{\frac{1}{(d_z+2)}}} \tag{187}$$

$$= \frac{n}{\ln(1/p)}(W(1/e) + \frac{\bar{m}-2}{e}) + 4(1+(4c^2)^{1/3})\sqrt{nm^*\hat{L}}\Big(\frac{\ln T}{T}\Big)^{\frac{1}{2(d_z+2)}} \tag{188}$$

with probability $1 - 2T^{-2}$ where W is the Lambert function ($W(1/e) \approx 0.278$), $m^* = \max_k m_k$, and $c \leq \frac{1}{4}\frac{n\bar{m}}{\hat{L}}\Big(\frac{\ln(m^*)}{\ln(1/p)} + 2\Big)^2$ is an upper bound on the range of sampled values from $\mathcal{L}^\tau$ (see Lemma 8).

Recall $\hat{L} = \Big[\frac{\ln(m^*)}{\ln(1/p)} + 2\Big]\Big[\frac{m^{*2}}{p\ln(1/p)} + n\bar{m}\Big]$. Therefore,

$$c \leq \frac{1}{4}\frac{n\bar{m}}{\hat{L}}\Big(\frac{\ln(m^*)}{\ln(1/p)} + 2\Big)^2 \tag{189}$$

$$= \frac{1}{4}n\bar{m}\Big(\frac{\frac{\ln(m^*)}{\ln(1/p)} + 2}{\frac{m^{*2}}{p\ln(1/p)} + n\bar{m}}\Big). \tag{190}$$

Markov's inequality then allows us to bound the pointwise exploitability of any arm returned by the algorithm as

$$\epsilon_t \leq w\Big[\frac{n}{\ln(1/p)}(W(1/e) + \frac{\bar{m}-2}{e}) + 4(1+(4c^2)^{1/3})\sqrt{nm^*\hat{L}}\Big(\frac{\ln T}{T}\Big)^{\frac{1}{2(d_z+2)}}\Big] \tag{191}$$

with probability $(1-w^{-1})(1-2T^{-2})$ for any $w > 0$. $\square$

F Complexity

F.1 Polymatrix Games

Interestingly, at zero temperature (where QRE = Nash), M is constant for a polymatrix game, so the rank of this matrix can be computed just once to extract information about all possible interior equilibria in the game. Furthermore, the Hessian is positive semi-definite over the entire joint strategy space, implying the loss function is convex (see Figure 1 (left) for empirical support). This indicates, by convex optimization theory, 1) all mixed Nash equilibria in polymatrix games form a convex set (i.e., they are connected) and 2) assuming mixed equilibria exist, they can be computed simply by stochastic gradient descent on $\mathcal{L}$. If M is rank-$n\bar{m}$, then this interior equilibrium is unique.

Complexity Approximation of Nash equilibria in polymatrix games is known to be PPAD-hard [13]. In contrast, if we restrict our class of polymatrix games to those with at least one interior Nash equilibrium, our analysis proves we can find an approximate Nash equilibrium in deterministic, polynomial time (Corollary 3). This follows directly from the fact that $\mathcal{L}$ is convex, our decision set $\mathcal{X} = \prod_i \mathcal{X}_i$ is convex, and convex optimization theory admits polynomial time approximation algorithms (e.g., gradient descent). We consider the assumption of the existence of an interior Nash equilibrium to be relatively mild[3], so this positive complexity result is surprising.

Also, note that the Hessian of the loss at Nash equilibria is encoded entirely by the polymatrix approximation at the equilibrium. Therefore, approximating the Hessian of $\mathcal{L}$ about the equilibrium (which amounts to observing near-equilibrium behavior [25]) allows one to recover this polymatrix approximation (up to constant offsets of the columns which equilibria are invariant to [27]).

[3]Marris et al. [27] shows 2-player, 2-action polymatrix games with interior Nash equilibria make up a non-trivial $1/4$ of the space of possible 2×2 games.

Corollary 3 (Approximating Nash Equilibria of Polymatrix Games with Interior Equilibria)**.** *Consider the class of polymatrix games with interior Nash equilibria. This class of games admits a fully polynomial time deterministic approximation scheme (FPTAS).*

Proof. Lemma 3 relates the approximation of Nash equilibria to the minimization of the loss function $\mathcal{L}(\boldsymbol{x})$. By Lemma 1, this loss function attains its minimum value of zero if and only if $\boldsymbol{x}$ is a Nash equilibrium. For polymatrix games, Hessian of this loss function is everywhere finite and positive definite (Lemma 17), therefore, this loss function is convex. The decision set for this minimization problem is the product space of simplices, therefore it is also convex. Given that we only consider polymatrix games with interior Nash equilibria, we know that our loss function attains a global minimum within this set. By convex optimization theory, this function can be approximately minimized in a polynomial number of steps by, for example, (projected) gradient descent. Gradient descent requires computing the gradient of the loss function at each step. From Lemma 15, we see that computing the gradient (at zero temperature) simply requires reading the polymatrix description of the game (i.e., each bi-matrix game H^k_{kl} between players), which is clearly polynomial in the size of the input (the polymatrix description). The remaining computational steps of gradient descent (e.g., projection onto simplices) are polynomial as well. In conclusion, gradient descent approximates a Nash equilibrium in polynomial number of steps (logarithmic if strongly-convex), each of which costs polynomial time, therefore the entire scheme is polynomial. □

F.2 Normal-Form Games

Corollary 1. *Consider the class of NFGs with at least one QRE(τ) whose local polymatrix approximation indicates it is isolated (i.e., M from equation (12) is rank-$n\bar{m}$ implies Hess $\succ 0$ implies $d_z = n\bar{m}(\frac{2-2}{4}) = 0$). Then by Theorem 1, BLiN is a fully polynomial-time randomized approximation scheme (FPRAS) for QREs and is a PRAS for NEs of games in this class.*

Proof. If $\alpha = 0$, an ϵ-QRE can be obtained with BLiN in a number of iterations that is polynomial in the game description length (nm^n). The same holds for an ϵ-NE, however, the temperature must be exponentially small to achieve a given ϵ; hence, we lose the *fully* qualifier. Specifically,

$$p \leq e^{-\frac{8n}{\epsilon}\left(W(1/e)+\frac{\bar{m}-2}{e}\right)}. \tag{192}$$

This, in turn, causes the Lipschitz constant $\hat{L}$ to grow exponentially large, leading to an exponential blow up in the number of iterations required for convergence. □

F.2.1 Concrete Example

The end of Section 6 stated a concrete result for a 20-player, 50-action game *assuming* we are given that the game as an interior Nash equilibrium. This result requires re-deriving a rate similar to Theorem 1, but for the unregularized game.

For example, revisiting Corollary 2 but for zero temperature, we find $\hat{L} = n\bar{m}$. Let $\eta = \frac{2}{\hat{L}}$ as before. Now, consulting Table 2, we find that samples from $\mathcal{L}$ are constrained to a range of size $c = \frac{1}{2}n\bar{m}\eta = 1$. Applying Corollary 9 to Lemma 3, we find:

$$\epsilon_t \leq w\Big[2\sqrt{2}(1+4^{1/3})n\sqrt{\bar{m}}\Big(\frac{\ln T}{T}\Big)^{\frac{1}{4}}\Big] \tag{193}$$

with probability $(1-w^{-1})(1-2T^{-2}) = 0.95(1-2T^{-2})$. Plugging in $w = 20$, $n = 20$, and $m = 50$ and solving for T numerically, we find that $T \leq 10^{28.7}$. For such large T, $0.95(1-2T^{-2}) \approx 0.95$. Again consulting Table 2, each call (arm pull) of BLiN costs $2nm$, implying a total query cost of $10^{32.0}$. In contrast, there exist $10^{35.2}$ scalar entries in the nm^n payoff tensor, which is a factor larger by 1000.

G Helpful Lemmas and Propositions

Proposition 2. *The matrix $I - \frac{1}{m_k}\mathbf{1}\mathbf{1}^\top$ is a projection matrix and therefore idempotent. It is also symmetric, which implies it is its own square root.*

Proof.

$$[I - \frac{1}{m_k}\mathbf{1}\mathbf{1}^\top]^\top [I - \frac{1}{m_k}\mathbf{1}\mathbf{1}^\top] = I - \frac{2}{m_k}\mathbf{1}\mathbf{1}^\top + \frac{1}{m_k^2}\mathbf{1}(\mathbf{1}^\top\mathbf{1})\mathbf{1}^\top \tag{194}$$

$$= I - \frac{2}{m_k}\mathbf{1}\mathbf{1}^\top + \frac{1}{m_k}\mathbf{1}\mathbf{1}^\top \tag{195}$$

$$= [I - \frac{1}{m_k}\mathbf{1}\mathbf{1}^\top]. \tag{196}$$

□

Proposition 3. *The matrix $I - \frac{1}{m_k}\mathbf{1}\mathbf{1}^\top$ is positive semi-definite.*

Proof. Let $z \in \mathbb{R}^{m_k}$. Then

$$z^\top [I - \frac{1}{m_k}\mathbf{1}\mathbf{1}^\top] z = ||z||_2^2 - \frac{1}{m_k}\langle z, \mathbf{1}\rangle^2 \tag{197}$$

$$\geq ||z||_2^2 - \frac{1}{m_k}\langle |z|, \mathbf{1}\rangle^2 \tag{198}$$

$$= ||z||_2^2 - \frac{1}{m_k}||z||_1^2 \tag{199}$$

$$\geq ||z||_2^2 - ||z||_2^2 = 0 \ \forall z \tag{200}$$

$$\implies [I - \frac{1}{m_k}\mathbf{1}\mathbf{1}^\top] \succeq 0. \tag{201}$$

□

Proposition 4. *The matrix $I - \frac{1}{m_k}\mathbf{1}\mathbf{1}^\top$ has rank $m_k - 1$ and its 1-d nullspace lies along $\mathbf{1}_k$.*

Proof. Note that $rank(A+B) \leq rank(A) + rank(B)$ for matrices A and B of the same dimension. Let $A = I - \frac{1}{m_k}\mathbf{1}\mathbf{1}^\top$ and $B = \frac{1}{m_k}\mathbf{1}\mathbf{1}^\top$ and apply $rank(A) \geq rank(A+B) - rank(B)$:

$$rank(I - \frac{1}{m_k}\mathbf{1}\mathbf{1}^\top) \geq rank(I) - rank(\frac{1}{m_k}\mathbf{1}\mathbf{1}^\top) = m_k - 1. \tag{202}$$

We can confirm the nullspace by inspection:

$$[I - \frac{1}{m_k}\mathbf{1}\mathbf{1}^\top]\mathbf{1} = \mathbf{1} - \frac{m_k}{m_k}\mathbf{1} = 0. \tag{203}$$

□

Lemma 19. *The product $A[I_m - \frac{1}{m}\mathbf{1}_m\mathbf{1}_m^\top]^p B$ for any $p > 0$ has entries whose absolute value is bounded by $\frac{m}{4}(A_{\max} - A_{\min})(B_{\max} - B_{\min})$ where $A_{\min}, A_{\max}, B_{\min}, B_{\max}$ represent the minima and maxima of the matrices respectively.*

Proof. The matrix $[I - \frac{1}{m}\mathbf{1}\mathbf{1}^\top]$ is idempotent so we can rewrite the product for any p as

$$A[I - \frac{1}{m}\mathbf{1}\mathbf{1}^\top][I - \frac{1}{m}\mathbf{1}\mathbf{1}^\top]B. \tag{204}$$

The matrix $[I - \frac{1}{m}\mathbf{1}\mathbf{1}^\top]$ has the property that it removes the mean from every row of a matrix when right multiplied against it, i.e., $A[I - \frac{1}{m}\mathbf{1}\mathbf{1}^\top]$ removes the means from the rows of A. Similarly, left multiplying it removes the means from the column. Let $\tilde{A}$ and $\tilde{B}$ represent these mean-centered results respectively. The absolute value of the ijth entry in the resulting product can then be recognized as

$$|\sum_k \tilde{A}_{ik}\tilde{B}_{kj}| = |\sum_k (A_{ik} - \frac{1}{m}\sum_{k'} A_{ik'})(B_{kj} - \frac{1}{m}\sum_{k'} B_{k'j})| \tag{205}$$

$$= |m \cdot Corr(A_{i,\cdot}, B_{\cdot,j}) \cdot \sigma_{A_{i,\cdot}}\sigma_{B_{\cdot,j}}| \tag{206}$$

$$\leq m\sigma_{A_{i,\cdot}}\sigma_{B_{\cdot,j}}. \tag{207}$$

The variance of a bounded random variable X is upper bounded by $Var[X] \leq \frac{1}{4}(\max_X - \min_X)^2$. Hence its standard deviation is bounded by $Std[X] \leq \frac{1}{2}(\max_X - \min_X)$. Plugging these bounds for A and B into equation (207) completes the claim. □

H Maps from Hypercube to Simplex Product

In this section, we derive properties of a map s from the unit-hypercube to the simplex product. This map is necessary to to adapt our proposed loss $\mathcal{L}^\tau$ to the commonly assumed setting in the $\mathcal{X}$-armed bandit literature [8]. We derive relevant properties of two such maps: the `softmax` and a mapping that interprets dimensions of the hypercube as angles on a unit-sphere that is then ℓ_1-normalized.

Lemma 20. *Let $f(x) = -\mathcal{L}(s(x))$. Then $||\nabla f(x)||_\infty \leq ||J(s(x))^\top||_\infty ||\nabla \mathcal{L}(s(x))||_\infty$.*

Proof.

$$||\nabla f(x)||_\infty = ||J(s(x))^\top \nabla \mathcal{L}(s(x))||_\infty \leq ||J(s(x))^\top||_\infty ||\nabla \mathcal{L}(s(x))||_\infty. \tag{208}$$

□

Lemma 21. *The ∞-norm of the Jacobian-transpose of a transformation $s(x)$ applied elementwise to a product space is bounded by the ∞-norm of the Jacobian-transpose of a single transformation from that product space, i.e., $||J(s(\boldsymbol{x}))^\top||_\infty \leq \max_{x_i \in \mathcal{X}_i} ||J(s(x_i))^\top||_\infty$ for any i.*

Proof. Let $\boldsymbol{x} \in \mathcal{X} = \prod_{i=1}^n \mathcal{X}_i$, $\mathcal{Z} = \prod_{i=1}^n \mathcal{Z}_i$ and $S : \mathcal{X} \to \mathcal{Z} = [s(x_1); \cdots ; s(x_n)]^\top$ where ; denotes column-wise stacking, $x_i \in \mathcal{X}_i$. Also, $\mathcal{X}_i = \mathcal{X}_j$ and $\mathcal{Z}_i = \mathcal{Z}_j$for all i and j. Then the Jacobian of $S(\boldsymbol{x})$ is

$$J(S(\boldsymbol{x}))^\top = \begin{bmatrix} J(s(x_1))^\top & 0 \ldots & 0 \\ 0 & J(s(x_2))^\top \ldots & 0 \\ 0 & 0 \ddots & 0 \\ 0 & 0 \ldots & J(s(x_n))^\top \end{bmatrix}. \tag{209}$$

The ∞-norm of this matrix is the max 1-norm of any row. This matrix is diagonal, therefore, the ∞-norm of each elementwise Jacobian-transpose represents the max 1-norm of the rows spanned by its block. Given that the domains, ranges, and transformations s for all blocks are the same, their ∞-norms are also the same. The max ∞ over the blocks is then equal to the ∞-norm of any individual $J(s(x_i))^\top$. □

H.1 Hessian of Bandit Reward Function

Lemma 22. *Let $s(x)$ be a function that maps the unit hypercube to the simplex product (mixed strategy space). Then the objective function $f(x) = -\mathcal{L}(s(x))$. The Hessian of $-f(x)$ at an optimum x^* in direction Δ is $\Delta x^\top [Ds(x)^\top H_\mathcal{L}(x) Ds(x)]\Big|_{x^*} \Delta x$ where $H_\mathcal{L}$ is the Hessian of $\mathcal{L}$ and $Ds(x)$ is the Jacobian of $s(x)$.*

Proof.

$$(D^2(\mathcal{L} \circ s)(x^*))(\Delta x, \Delta x) = \Delta x^\top \Big[\sum_i \overbrace{\partial_i \mathcal{L}(s(x))}^{=0 \text{ at } x=x^*} D^2 h_i(x) \Big]\Big|_{x^*} \Delta x + \Delta x^\top [Ds(x)^\top H_\mathcal{L}(x) Ds(x)]\Big|_{x^*} \Delta x \tag{210}$$

$$= \Delta x^\top [Ds(x)^\top H_\mathcal{L}(x) Ds(x)]\Big|_{x^*} \Delta x. \tag{211}$$

□

Lemma 23. *Let $s(x) : \mathcal{X} \to \prod_k \Delta^{m_k - 1}$ be an injective function, i.e., $x \neq y \implies s(x) \neq s(y)$. Also let $J = J(s(x))$ be the Jacobian of s with respect to x and Δx be a nonzero vector in the tangent space of $\mathcal{X}$. Then*

$$J \Delta x \neq \mathbf{0}. \tag{212}$$

Proof. Recall that the ijth entry of the Jacobian represents $\frac{\partial s_i}{\partial x_j}$ so that the ith entry of $J\Delta x$ is

$$[J\Delta x]_i = \sum_j \frac{\partial s_i}{\partial x_j} \Delta x_j = ds_i. \tag{213}$$

Assume $J\Delta x = \mathbf{0}$. This would imply a change in $x \in \mathcal{X}$ results in no change in s ($ds = \mathbf{0}$), contradicting the fact that s is injective. Therefore, we must conclude the claim that $J\Delta x \neq \mathbf{0}$. $\square$

Lemma 24. *Let J be the Jacobian of the softmax operator. Then $||J||_\infty \leq 2$ and $||J^\top||_\infty \leq 2$.*

Proof. Let S_i represent the ith entry of $S = \texttt{softmax}(z)$ for any $z \in \mathbb{R}^m$. Then the 1-norm of row i is upper bounded as

$$\begin{aligned}
D_j S_i &= S_i(\delta_{ij} - S_j) && (214)\\
\implies \sum_j |D_j S_i| &= \sum_j |S_i(\delta_{ij} - S_j)| && (215)\\
&\leq \sum_j |\delta_{ij} S_i| + |S_i S_j| && (216)\\
&= S_i + \sum_j S_i S_j && (217)\\
&= S_i + S_i \sum_j S_j && (218)\\
&= 2S_i && (219)\\
&\leq 2 \ \forall i. && (220)
\end{aligned}$$

Also, the 1-norm of row j is upper bounded similarly as

$$\tag{221}$$

$$\begin{aligned}
\sum_i |D_j S_i| &= \sum_i |S_i(\delta_{ij} - S_j)| && (222)\\
&\leq \sum_i |\delta_{ij} S_i| + |S_i S_j| && (223)\\
&= S_j + \sum_i S_i S_j && (224)\\
&= S_i + S_j \sum_i S_i && (225)\\
&= 2S_j && (226)\\
&\leq 2 \ \forall j. && (227)
\end{aligned}$$

The ∞-norm of a matrix is the maximum 1-norm of any row. Therefore, $||J||_\infty$ and $||J^\top||_\infty$ are both upper bounded by 2. $\square$

Lemma 25. *Let $J = J(s(x))$ be the Jacobian of any composition of transformations $s = s_t \circ \ldots s_1$ where $s_t(z) = [z_i / \sum_j z_j]_i$. Then $J\Delta x$ lies in the tangent space of the simplex.*

Proof. We aim to show $\mathbf{1}^\top J \Delta x = \mathbf{0}$ for any Δx and x. By chain rule, the Jacobian of s is $J = J(s) = \prod_{t'=t}^{t'=1} J(s'_t)$. Therefore, $\mathbf{1}^\top J\Delta x = \mathbf{1}^\top (\prod_{t'=t}^{t'=1} J(s'_t))\Delta x$. Consider the first product:

$$\mathbf{1}^\top J(s_t) = \mathbf{0} \tag{228}$$

by Lemma 27. Therefore $\mathbf{1}^\top J\Delta x = \mathbf{1}^\top J(s_t)(\prod_{t'=t-1}^{t'=1} J(s'_t))\Delta x = \mathbf{0}^\top (\prod_{t'=t-1}^{t'=1} J(s'_t))\Delta x = 0$. This implies $J\Delta x$ is orthogonal to $\mathbf{1}$ for any $x \in \mathcal{X}$ and Δx, therefore $J\Delta x$ lies in the tangent space of the simplex for any $x \in \mathcal{X}$ and Δx. $\square$

For spherical coordinates, $s(x) = n(l(c(x)))$ where $c(x) = \pi/2x$, $l(\psi)$ maps angles to the unit sphere, and $n(z) = [z_i / \sum_j z_j]_i$.

Definition 1. *Define $l(\psi)$ as the transformation to the unit-sphere using spherical coordinates:*

$$
\begin{aligned}
l_1(\psi) &= \cos(\psi_1) && (229)\\
l_2(\psi) &= \sin(\psi_1)\cos(\psi_2) && (230)\\
l_3(\psi) &= \sin(\psi_1)\sin(\psi_2)\cos(\psi_3) && (231)\\
\vdots &= \vdots && (232)\\
l_{m-1}(\psi) &= \sin(\psi_1)\sin(\psi_2)\ldots\cos(\psi_{m-1}) && (233)\\
l_m(\psi) &= \sin(\psi_1)\sin(\psi_2)\ldots\sin(\psi_{m-1}). && (234)
\end{aligned}
$$

Lemma 26. *Let J be the Jacobian of the transformation to the unit-sphere using spherical coordinates, i.e. $z = l(\psi)$ where $||l||^2 = 1$ and $\psi_i \in [0, \frac{\pi}{2}]$ represents an angle for each i. Then $||J||_F \leq \sqrt{m}$.*

Proof. The Jacobian of the transformation is

$$
J(l) = \begin{bmatrix}
-\sin(\psi_1) & 0 & \cdots & 0\\
\cos(\psi_1)\cos(\psi_2) & -\sin(\psi_1)\sin(\psi_2) & \cdots & 0\\
\vdots & \vdots & \ddots & \vdots\\
\cos(\psi_1)\sin(\psi_2)\ldots\cos(\psi_{m-1}) & \cdots & \cdots & -\sin(\psi_1)\ldots\sin(\psi_{m-2})\sin(\psi_{m-1})\\
\cos(\psi_1)\sin(\psi_2)\ldots\sin(\psi_{m-1}) & \cdots & \cdots & \sin(\psi_1)\ldots\sin(\psi_{m-2})\cos(\psi_{m-1})
\end{bmatrix} \tag{235}
$$

and it square is

$$
J(l) = \begin{bmatrix}
t_1 & 0 & \cdots & 0\\
\cos(\psi_1)^2\cos(\psi_2)^2 & \sin(\psi_1)^2 t_2 & \cdots & 0\\
\vdots & \vdots & \ddots & \vdots\\
\cos(\psi_1)^2\sin(\psi_2)^2\ldots\cos(\psi_{m-1})^2 & \cdots & \cdots & \sin(\psi_1)^2\ldots\sin(\psi_{m-2})^2 t_{m-1}\\
\cos(\psi_1)^2\sin(\psi_2)^2\ldots\sin(\psi_{m-1})^2 & \cdots & \cdots & \sin(\psi_1)^2\ldots\sin(\psi_{m-2})^2 t_m
\end{bmatrix} \tag{236}
$$

where

$$
\begin{aligned}
\delta_{im} &= 1 \text{ if } i = m, 0 \text{ else} && (237)\\
t_i &= \delta_{im}\cos^2(\psi_{i-1}) + (1-\delta_{im})\sin^2(\psi_i) \leq 1. && (238)
\end{aligned}
$$

To compute the Frobenius norm, we will need the sum of the squares of all entries. We will consider the sum of each row individually using the following auxiliary variable $R_{i,k\leq i}$ where $\sum_j J_{ij}^2 = R_{i,1}$ and apply a recursive inequality.

$$
\tag{239}
$$

$$
\begin{aligned}
R_{i,k\leq i} &= \sum_{k'=k}^{i-1}\cos^2(\psi_{k'})\Big[\prod_{l=k,l\neq k'}^{i-1}\sin^2(\psi_l)\Big]\cos^2(\psi_i) + t_i\prod_{l=k}^{i-1}\sin^2(\psi_l) && (240)\\
&= \cos^2(\psi_k)\underbrace{\Big[\prod_{l=k+1}^{i-1}\sin^2(\psi_l)\Big]\cos^2(\psi_i)}_{\leq 1} && (241)\\
&+ \sin^2(\psi_k)\sum_{k'=k+1}^{i-1}\cos^2(\psi_{k'})\Big[\prod_{l=k+1,l\neq k'}^{i-1}\sin^2(\psi_l)\Big]\cos^2(\psi_i) && (242)\\
&+ \sin^2(\psi_k)t_i\prod_{l=k+1}^{i-1}\sin^2(\psi_l) && (243)\\
&\leq \cos^2(\psi_k) && (244)\\
&+ \sin^2(\psi_k)\Big(\sum_{k'=k+1}^{i-1}\cos^2(\psi_{k'})\Big[\prod_{l=k+1,l\neq k'}^{i-1}\sin^2(\psi_l)\Big]\cos^2(\psi_i) + t_i\prod_{l=k+1}^{i-1}\sin^2(\psi_l)\Big) && (245)\\
&= \cos^2(\psi_k) + \sin^2(\psi_k)R_{i,k+1}. && (246)
\end{aligned}
$$

Note then that $R_{i,k+1} \leq 1 \implies R_{i,k} \leq 1$. We know $R_{i,i} = t_i \leq 1$, therefore, $R_{i,1} \leq 1$ by applying the inequality recursively. Finally, $\sum_j J_{ij}^2 = R_{i,1} \leq 1$ implies the claim $||J||_F^2 = \sum_i R_{i,1} \leq m$. $\square$

Lemma 27. *Let J be the Jacobian of $n(z) = z/Z$ where $Z = \sum_k z_k$. Then $\mathbf{1}^\top J = \mathbf{0}^\top$.*

Proof. The ijth entry of the Jacobian of $n(z)$ is

$$J(n)_{ij} = \frac{1}{Z^2}(-z_i + \delta_{ij} Z). \tag{247}$$

Therefore $[\mathbf{1}^\top J]_j = \sum_i J(n)_{ij} = \frac{1}{Z^2}(-Z + Z) = 0$ where z is a point on the unit-sphere in the positive orthant. $\square$

I A2: Bounded Diameters and Well-shaped Cells

We assume the feasible set is a unit-hypercube of dimensionality d where cells are evenly split along the longest edge to give b new partitions and $x_{h,i}$ represents the center of each cell.

There exists a decreasing sequence $w(h) > 0$, such that for any depth $h \geq 0$ and for any cell $\mathcal{X}_{h,i}$ of depth h, we have $\sup_{x \in \mathcal{X}_{h,i}} \ell(x_{h,i}, x) \leq w(h)$. Moreover, there exists $\nu > 0$ such that for any depth $h \geq 0$, any cell $\mathcal{X}_{h,i}$ contains an ℓ-ball of radius $\nu w(h)$ centered at $x_{h,i}$.

$\ell(x,y)$	c	γ	ν
$\ell(x,y) = \|\|x-y\|\|_2^\alpha$	$d^{\alpha/2}\left(\frac{b}{2}\right)^\alpha$	$b^{-\alpha/d}$	$d^{-\alpha/2} b^{-2\alpha}$
$\ell(x,y) = \|\|x-y\|\|_\infty^\alpha$	$\left(\frac{b}{2}\right)^\alpha$	$b^{-\alpha/d}$	$b^{-2\alpha}$

Table 3: Bounding Constants: $\sup_{x \in \mathcal{X}_{h,i}} \ell(x_{h,i}, x) \leq w(h) = c\gamma^h$.

I.1 L_2-Norm

Lemma 28 (L_2-Norm Bounding Ball)**.** *Let $\ell(x,y) = ||x-y||_2^\alpha$. Then $\sup_{x \in \mathcal{X}_{h,i}} \ell(x_{h,i}, x) \leq w_2(h) = c\gamma^h$ where $c = \left(\frac{db^2}{4}\right)^{\alpha/2}$ and $\gamma = b^{-\alpha/d}$.*

Proof.

$$w(0) = \big[\sum_{i=1}^{d}(1/2)^2\big]^{\alpha/2} = \big(\frac{d}{4}\big)^{\alpha/2} \tag{248}$$

$$w(1) = \big[(1/b\cdot 1/2)^2 + \sum_{i=2}^{d}(1/2)^2\big]^{\alpha/2} = [(1/b^2)(1/4) + (d-1)(1/4)]^{\alpha/2} \tag{249}$$

$$= \big(\frac{d-1+1/b^2}{4}\big)^{\alpha/2} \tag{250}$$

$$w(d) = \big[\sum_{i=1}^{d}(1/b\cdot 1/2)^2\big]^{\alpha/2} = \big(\frac{d}{4\cdot b^2}\big)^{\alpha/2} \tag{251}$$

$$w(h) = \big[r(1/b)^{2(q+1)}(1/2)^2 + \sum_{i=r}^{d}(1/b)^{2q}(1/2)^2\big]^{\alpha/2} \tag{252}$$

$$= \big[(1/b)^{2q}(1/2)^2\big(r(1/b)^2 + (d-r)\big]^{\alpha/2} \tag{253}$$

$$= \big[(1/b^2)^q(1/4)\big(d - r(1-\frac{1}{b^2})\big)\big]^{\alpha/2} \tag{254}$$

$$\leq \big[(1/b^2)^q(1/4)d\big]^{\alpha/2} \tag{255}$$

$$\leq \big[(1/b^2)^{h/d-1}(1/4)d\big]^{\alpha/2} \tag{256}$$

$$= \big[(1/b^2)^{h/d}(b^2/4)d\big]^{\alpha/2} \tag{257}$$

$$= \big(\frac{db^2}{4}\big)^{\alpha/2}(1/b)^{\frac{\alpha}{d}h} \tag{258}$$

$$= c\gamma^h \tag{259}$$

where $q, r = divmod(h, d) \implies q \geq h/d - 1$, $c = \big(\frac{db^2}{4}\big)^{\alpha/2}$, and $\gamma = (1/b)^{\alpha/d} = b^{-\alpha/d}$. $\square$

Lemma 29 (L_2-Norm Inner Ball)**.** *Let $\ell(x, y) = ||x-y||_2^\alpha$. Any cell $\mathcal{X}_{h,i}$ contains an ℓ-ball of radius $\nu w_2(h)$ where $\nu = (db^4)^{-\alpha/2}$.*

Proof. Any cell $\mathcal{X}_{h,i}$ contains an ℓ-ball of radius equal to its shortest axis:

$$r_{\min} = \big[(1/4)(1/b^2)^{\lceil h/d\rceil}\big]^{\alpha/2} \tag{260}$$

$$\geq \big[(1/4)(1/b^2)^{h/d+1}\big]^{\alpha/2} \tag{261}$$

$$= \big(\frac{1}{b^2\cdot 4}\big)^{\alpha/2}(1/b)^{\frac{\alpha}{d}h} \tag{262}$$

$$= w(h)\cdot\big(\frac{1}{db^4}\big)^{\alpha/2}. \tag{263}$$

$\square$

I.2 L_∞-Norm

Lemma 30 (L_∞-Norm Bounding Ball)**.** *Let $\ell(x, y) = ||x-y||_\infty^\alpha$. Then $\sup_{x\in\mathcal{X}_{h,i}} \ell(x_{h,i}, x) \leq w_\infty(h) = c\gamma^h$ where $c = \big(\frac{b}{2}\big)^\alpha$ and $\gamma = b^{-\alpha/d}$.*

Proof. Any cell $\mathcal{X}_{h,i}$ is contained by an ℓ-ball of radius equal to its longest axis:

$$r_{\max} = \big[(1/4)(1/b^2)^{\lfloor h/d\rfloor}\big]^{\alpha/2} \tag{264}$$

$$\leq \big[(1/4)(1/b^2)^{h/d-1}\big]^{\alpha/2} \tag{265}$$

$$= \big(\frac{b^2}{4}\big)^{\alpha/2}(1/b)^{\frac{\alpha}{d}h} \tag{266}$$

$$= c\gamma^h \tag{267}$$

where $c = \left(\frac{b^2}{4}\right)^{\alpha/2}$, and $\gamma = (1/b)^{\alpha/d} = b^{-\alpha/d}$. □

Lemma 31 (L_∞-Norm Inner Ball)**.** *Let $\ell(x, y) = ||x - y||_\infty^\alpha$. Any cell $\mathcal{X}_{h,i}$ contains an ℓ-ball of radius $\nu w_\infty(h)$ where $\nu = b^{-2\alpha}$.*

Proof. Any cell $\mathcal{X}_{h,i}$ contains an ℓ-ball of radius equal to its shortest axis:

$$r_{\min} = \left[(1/4)(1/b^2)^{\lceil h/d \rceil}\right]^{\alpha/2} \tag{268}$$

$$\geq \left[(1/4)(1/b^2)^{h/d+1}\right]^{\alpha/2} \tag{269}$$

$$= \left(\frac{1}{b^2 \cdot 4}\right)^{\alpha/2}(1/b)^{\frac{\alpha}{d}h} \tag{270}$$

$$= w(h) \cdot \left(\frac{1}{b^4}\right)^{\alpha/2}. \tag{271}$$

□

I.3 Near Optimality Dimension

This is written in terms of a maximizing f.

Assumption 1. *Locally around each interior x^*, $-f(x)$ is lower bounded by $-f(x^*) + \sigma_-||x - x^*||^{\alpha_{hi}}$ and upper bounded by $-f(x^*) + \ell(x, x^*)$ where $\ell(x, x^*) = \sigma_+||x - x^*||^{\alpha_{lo}}$ with $\alpha_{lo} \leq \alpha_{hi}$ and $\sigma_- \leq \sigma_+$ if $\alpha_{lo} = \alpha_{hi}$. In other words, for all $f(x) \geq f(x^*) - \eta$:*

$$f(x^*) - f(x) \leq \sigma_+||x - x^*||^{\alpha_{lo}} \tag{272}$$

$$f(x^*) - f(x) \geq \sigma_-||x - x^*||^{\alpha_{hi}} \tag{273}$$

where we have left the precise norm unspecified for generality.

Definition 2. $\mathcal{X}_\epsilon \overset{\text{def}}{=} \{x \in \mathcal{X} | f(x) \geq f(x^*) - \epsilon\}$

Definition 3. $\mathcal{X}_\epsilon^{lower} \overset{\text{def}}{=} \{x \in \mathcal{X} | f(x^*) - \sigma_-||x - x^*||^{\alpha_{hi}} \geq f(x^*) - \epsilon\}$

Corollary 4. $\mathcal{X}_\epsilon \subseteq \mathcal{X}_\epsilon^{lower}$.

Proof. By Assumption 1, $f(x^*) - \sigma_-||x - x^*||^{\alpha_{hi}} \geq f(x)$. Therefore, any $x \in \mathcal{X}$ that satisfies the requirement for an element of $\mathcal{X}_\epsilon$, $f(x) \geq f(x^*) - \epsilon$, will also satisfy the requirement for an element of $\mathcal{X}_\epsilon^{lower}$. □

Definition 4. *The ψ-near optimality dimension is the smallest $d' > 0$ such that there exists $C > 0$ such that for any $\epsilon > 0$, the maximum number of disjoint ℓ-balls of radius $\psi\epsilon$ and center in $\mathcal{X}_\epsilon$ is less than $C\epsilon^{-d'}$.*

Theorem 2. *The ψ-near optimality dimension of $f : x \in [0, 1]^d \to [-1, 1]$ under ℓ is $d' = d(\frac{\alpha_{hi} - \alpha_{lo}}{\alpha_{lo}\alpha_{hi}})$ with constant*

$$C = \max\left\{1, S_d^{-1}\left(r_\eta^{\frac{\alpha_{hi}}{\alpha_{lo}}} \sigma_-^{\left(\frac{\alpha_{hi} - \alpha_{lo}}{\alpha_{lo}\alpha_{hi}}\right)}\right)^{-d}\right\}\left(\frac{\sigma_+}{\psi\sigma_-^{\alpha_{lo}/\alpha_{hi}}}\right)^{d/\alpha_{lo}}. \tag{274}$$

Proof. First, let us define $r_\eta = \left(\frac{\eta}{\sigma_-}\right)^{1/\alpha_{hi}}$ as in equation (285) which implies $\eta = \sigma_- r_\eta^{\alpha_{hi}}$. Then apply Lemmas 32 ($N_{\epsilon \leq \eta} \leq C_{\epsilon \leq \eta}\epsilon^{-d'}$) and 34 ($N_{\epsilon \geq \eta} \leq C_{\epsilon \geq \eta}$) which bound the number of ℓ-balls required to pack $\mathcal{X}_\epsilon$ when ϵ is less than and greater than η respectively:

$$C_{\epsilon \leq \eta} = \left(\frac{\sigma_+}{\psi\sigma_-^{\alpha_{lo}/\alpha_{hi}}}\right)^{d/\alpha_{lo}} \tag{275}$$

$$d' = d\left(\frac{\alpha_{hi} - \alpha_{lo}}{\alpha_{lo}\alpha_{hi}}\right) \tag{276}$$

and

$$C_{\epsilon\geq\eta} = S_d^{-1}\Big(\frac{\sigma_+}{\psi\eta}\Big)^{d/\alpha_{lo}} \tag{277}$$

$$= S_d^{-1}\eta^{-d/\alpha_{lo}}\sigma_-^{d/\alpha_{hi}}\Big(\frac{\sigma_+}{\psi\sigma_-^{\alpha_{lo}/\alpha_{hi}}}\Big)^{d/\alpha_{lo}} \tag{278}$$

$$= S_d^{-1}\eta^{-d/\alpha_{lo}}\sigma_-^{d/\alpha_{hi}}C_{\epsilon\leq\eta} \tag{279}$$

$$= S_d^{-1}r_\eta^{-d\alpha_{hi}/\alpha_{lo}}\sigma_-^{-d/\alpha_{lo}}\sigma_-^{d/\alpha_{hi}}C_{\epsilon\leq\eta} \tag{280}$$

$$= S_d^{-1}r_\eta^{-d\frac{\alpha_{hi}}{\alpha_{lo}}}\sigma_-^{-d\left(\frac{\alpha_{hi}-\alpha_{lo}}{\alpha_{lo}\alpha_{hi}}\right)}C_{\epsilon\leq\eta} \tag{281}$$

$$= S_d^{-1}\Big(r_\eta^{\frac{\alpha_{hi}}{\alpha_{lo}}}\sigma_-^{\left(\frac{\alpha_{hi}-\alpha_{lo}}{\alpha_{lo}\alpha_{hi}}\right)}\Big)^{-d}C_{\epsilon\leq\eta} \tag{282}$$

where S_d is the volume constant for a d-sphere under the given norm. S_d^{-1} has been upper bounded for the 2-norm in Lemma 33. For the ∞-norm, $S_d^{-1} = 2^{-d}$. We have written $C_{\epsilon\geq\eta}$ in terms of $C_{\epsilon\leq\eta}$ to clarify which is larger.

Therefore,

$$C = \max\Big\{1, S_d^{-1}\Big(r_\eta^{\frac{\alpha_{hi}}{\alpha_{lo}}}\sigma_-^{\left(\frac{\alpha_{hi}-\alpha_{lo}}{\alpha_{lo}\alpha_{hi}}\right)}\Big)^{-d}\Big\}C_{\epsilon\leq\eta} \tag{283}$$

$$= \max\Big\{1, S_d^{-1}\Big(r_\eta^{\frac{\alpha_{hi}}{\alpha_{lo}}}\sigma_-^{\left(\frac{\alpha_{hi}-\alpha_{lo}}{\alpha_{lo}\alpha_{hi}}\right)}\Big)^{-d}\Big\}\Big(\frac{\sigma_+}{\psi\sigma_-^{\alpha_{lo}/\alpha_{hi}}}\Big)^{d/\alpha_{lo}}. \tag{284}$$

Intuitively, if the radius for which the polynomial bounds hold (r_η) is large and the minimum curvature constant σ_- is also large, then the bound $C_{\epsilon\leq\eta}$ holds for large deviations from optimality η. The number of η-radius ℓ-balls required to cover the remaining space, $C_{\epsilon\geq\eta}$, will be comparatively small. □

Corollary 5 (Zooming Dimension)**.** *The zooming dimension of* $f : x \in [0,1]^d \to [-1,1]$ *under* $\ell(x,y) = ||x-y||_\infty$ *is* $d_z = d(\frac{\alpha_{hi}-\alpha_{lo}}{\alpha_{lo}\alpha_{hi}})$.

Proof. Mapping the definition of zooming dimension onto ψ-near optimality, we find $\psi\epsilon = r/2$ and $\epsilon = 16r$. Then we can infer $\psi = 1/32$. This result only effects the constant C_z, not the zooming dimension.

If $\epsilon = 8(1+\sqrt{c_1/c_2})r_m$, then $\psi = \frac{1}{16(1+\sqrt{c_1/c_2})}$. □

Lemma 32 ($N_{\epsilon\leq\eta} \leq C_{\epsilon\leq\eta}\epsilon^{-d'}$)**.** *The number of disjoint* ℓ*-balls that can pack into a set* $\mathcal{X}_{\epsilon\leq\eta}$, $N_{\epsilon\leq\eta}$, *is upper bounded by* $C_{\epsilon\leq\eta}\epsilon^{-d'}$ *where* $C_{\epsilon\leq\eta} = \Big(\frac{\sigma_+}{\psi\sigma_-^{\alpha_{lo}/\alpha_{hi}}}\Big)^{d/\alpha_{lo}}$ *and* $d' = d(\frac{\alpha_{hi}-\alpha_{lo}}{\alpha_{lo}\alpha_{hi}})$ *and* S_d *is the volume constant for a* d*-sphere under the given norm* $||\cdot||$.

Proof. The number of disjoint ℓ-balls of radius $\psi\epsilon$ and center in $\mathcal{X}_{\epsilon\leq\eta}$ can be upper bounded as follows.

Rewrite $\mathcal{X}_\epsilon^{lower}$ by rearranging terms as

$$\mathcal{X}_\epsilon^{lower} = \{x \in \mathcal{X} |\ ||x-x^*|| \leq \Big(\frac{\epsilon}{\sigma_-}\Big)^{1/\alpha_{hi}} \stackrel{\text{def}}{=} r_\epsilon\} \tag{285}$$

and recall that from Corollary 4 that $\mathcal{X}_\epsilon \subseteq \mathcal{X}_\epsilon^{lower}$. Furthermore, an ℓ-ball of radius $\psi\epsilon$ implies

$$\sigma_+||x-y||^{\alpha_{lo}} \leq \psi\epsilon \implies ||x-y|| \leq \Big(\frac{\psi\epsilon}{\sigma_+}\Big)^{1/\alpha_{lo}} \stackrel{\text{def}}{=} r_\ell. \tag{286}$$

The number of disjoint ℓ-balls that can pack into a set $\mathcal{X}_\epsilon$, $N_{\epsilon\leq\eta}$, is upper bounded by the ratio of the volumes of the two sets:

$$N_{\epsilon\leq\eta} \leq \frac{Vol(\mathcal{X}_\epsilon)}{Vol(\mathcal{B}_\ell)} \tag{287}$$
$$\leq \frac{Vol(\mathcal{X}_\epsilon^{lower})}{Vol(\mathcal{B}_\ell)} \tag{288}$$
$$= \frac{S_d r_\epsilon^d}{S_d r_\ell^d} \tag{289}$$
$$\leq \frac{\left(\frac{\epsilon}{\sigma_-}\right)^{d/\alpha_{hi}}}{\left(\frac{\psi\epsilon}{\sigma_+}\right)^{d/\alpha_{lo}}} \tag{290}$$
$$= \left(\frac{\sigma_+^{1/\alpha_{lo}}\psi^{-1/\alpha_{lo}}}{\sigma_-^{1/\alpha_{hi}}}\right)^d \epsilon^{d(1/\alpha_{hi}-1/\alpha_{lo})} \tag{291}$$
$$= \left(\frac{\sigma_+}{\psi\sigma_-^{\alpha_{lo}/\alpha_{hi}}}\right)^{d/\alpha_{lo}} \epsilon^{-d(\frac{\alpha_{hi}-\alpha_{lo}}{\alpha_{lo}\alpha_{hi}})} \tag{292}$$
$$= C_{\epsilon\leq\eta}\epsilon^{-d'} \tag{293}$$

where $C_{\epsilon\leq\eta} = \left(\frac{\sigma_+}{\psi\sigma_-^{\alpha_{lo}/\alpha_{hi}}}\right)^{d/\alpha_{lo}}$ and $d' = d(\frac{\alpha_{hi}-\alpha_{lo}}{\alpha_{lo}\alpha_{hi}})$ and S_d is the volume constant for a d-sphere under the given norm $||\cdot||$, e.g., $S_d = 2^d$ for $||\cdot||_\infty$. □

Corollary 6. *If $\alpha_{lo} = \alpha_{hi} = \alpha$,*

$$N_{\epsilon\leq\eta} \leq \left(\frac{\kappa}{\psi}\right)^{d/\alpha}. \tag{294}$$

In other words, $N_{\epsilon\leq\eta} \leq C_{\epsilon\leq\eta}\epsilon^{-d'}$ where $C_{\epsilon\leq\eta} = \left(\frac{\sigma_+}{\psi\sigma_-}\right)^{d/\alpha}$ and $d' = 0$.

Corollary 7. *If Assumption 1 is given in terms of the 2-norm, these can be translated to bounds in terms of the ∞-norm resulting in the same ψ-near optimality dimension but incurring an additional exponential factor in the constant $C^{(\infty)}_{\epsilon\leq\eta} \leftarrow C^{(2)}_{\epsilon\leq\eta}d^{d/2}$.*

Proof. Recall that $||\cdot||_\infty \leq ||\cdot||_2 \leq \sqrt{d}||\cdot||_\infty$, therefore

$$f(x^*) - f(x) \leq \sigma_{+2}||x - x^*||_2^{\alpha_{lo}} \leq \sigma_{+\infty}||x - x^*||_\infty^{\alpha_{lo}} \tag{295}$$
$$f(x^*) - f(x) \geq \sigma_{-2}||x - x^*||_2^{\alpha_{hi}} \geq \sigma_{-\infty}||x - x^*||_\infty^{\alpha_{hi}} \tag{296}$$

where $\sigma_{+\infty} = \sigma_{+2}d^{\alpha_{lo}/2}$ and $\sigma_{-\infty} = \sigma_{-2}$. Then

$$C^{(\infty)}_{\epsilon\leq\eta} = \left(\frac{\sigma_{+2}d^{\alpha_{lo}/2}}{\psi\sigma_{-2}^{\alpha_{lo}/\alpha_{hi}}}\right)^{d/\alpha_{lo}} = \left(\frac{\sigma_{+2}}{\psi\sigma_{-2}^{\alpha_{lo}/\alpha_{hi}}}\right)^{d/\alpha_{lo}} d^{d/2} = C^{(2)}_{\epsilon\leq\eta}d^{d/2}. \tag{297}$$

□

Recall, these results apply when $f(x) \geq f(x^*) - \eta$, i.e., when $\epsilon \leq \eta$. Otherwise, we can upper bound the number of ℓ-balls by considering the entire set $\mathcal{X}$ which has volume 1. First, we will bound the constant associated with the volume of a d-sphere.

Lemma 33. *The volume of a d-sphere with radius r and d even is given by $S_d r^d$ where $S_d^{-1} \leq \sqrt{2\pi d}\left(\frac{d}{2\pi e}\right)^{d/2}$.*

Proof. First, we recall Stirling's bounds on the factorial: $\sqrt{2\pi n}(\frac{n}{e})^n e^{\frac{1}{12n+1}} < n! < \sqrt{2\pi n}(\frac{n}{e})^n e^{\frac{1}{12n}}$. This will be useful for bounding the Gamma function: $\Gamma(d) = (d-1)!$ for even d.

Given d is even, we start with the exact formula for S_d:

$$
\begin{aligned}
S_d^{-1} &= \frac{\Gamma(d/2+1)}{\pi^{d/2}} && (298)\\
&= \frac{(d/2)!}{\pi^{d/2}} && (299)\\
&< \frac{\sqrt{2\pi(d/2)}(\frac{d/2}{e})^{d/2}e^{\frac{1}{12(d/2)}}}{\pi^{d/2}} && (300)\\
&= \frac{\pi^{1/2}d^{1/2}(\frac{d}{2e})^{d/2}e^{\frac{1}{6d}}}{\pi^{d/2}} && (301)\\
&= \frac{\pi^{1/2}d^{(d+1)/2}e^{\frac{1}{6d}}}{(2\pi e)^{d/2}} && (302)\\
&\leq \sqrt{2\pi d}\Big(\frac{d}{2\pi e}\Big)^{d/2}. && (303)
\end{aligned}
$$

□

Lemma 34 ($N_{\epsilon\geq\eta} \leq C_{\epsilon\geq\eta}$)**.** *The number of disjoint ℓ-balls that can pack into a set $\mathcal{X}_{\epsilon\geq\eta}$, $N_{\epsilon\geq\eta}$, is upper bounded by $C_{\epsilon\geq\eta}$ where $C_{\epsilon\geq\eta} = S_d^{-1}\Big(\frac{\sigma_+}{\psi\eta}\Big)^{d/\alpha_{lo}}$ and S_d is the volume constant for a d-sphere under a given norm.*

Proof. We can upper bound the number of ℓ-balls needed to pack the entire space as follows:

$$
\begin{aligned}
N_{\epsilon\geq\eta} &\leq \frac{Vol(\mathcal{X})}{Vol(\mathcal{B}_\ell)} && (304)\\
&= \frac{1}{S_d r_\ell^d} && (305)\\
&\leq \frac{1}{S_d\Big(\frac{\psi\eta}{\sigma_+}\Big)^{d/\alpha_{lo}}} && (306)\\
&= S_d^{-1}\Big(\frac{\sigma_+}{\psi\eta}\Big)^{d/\alpha_{lo}} && (307)\\
&= C_{\epsilon\geq\eta} && (308)
\end{aligned}
$$

where r_l was defined in equation (286). S_d^{-1} has been upper bounded for the 2-norm in Lemma 33. For the ∞-norm, $S_d^{-1} = 2^{-d}$. □

Corollary 8. *If Assumption 1 is given in terms of the 2-norm, these can be translated to bounds in terms of the ∞-norm resulting in the same ψ-near optimality dimension but incurring an additional exponential factor in the constant $C^{(\infty)}_{\epsilon\geq\eta} = \Big(\frac{\sigma_{+2}^{1/\alpha_{hi}}}{2\eta^{1/\alpha_{lo}}}\Big)^d C^{(\infty)}_{\epsilon\leq\eta} = \Big(\frac{\sigma_{+2}^{1/\alpha_{hi}}}{2\eta^{1/\alpha_{lo}}}\Big)^d C^{(2)}_{\epsilon\leq\eta} d^{d/2}$.*

Proof.

$$
\begin{aligned}
C^{(\infty)}_{\epsilon\geq\eta} &= 2^{-d}\Big(\frac{\sigma_{+2}}{\psi\eta}\Big)^{d/\alpha_{lo}} d^{d/2} && (309)\\
&= 2^{-d}\eta^{-d/\alpha_{lo}}\sigma_{-2}^{d/\alpha_{hi}}\Big(\frac{\sigma_{+2}}{\psi\sigma_{-2}^{\alpha_{lo}/\alpha_{hi}}}\Big)^{d/\alpha_{lo}} d^{d/2} && (310)\\
&= 2^{-d}\eta^{-d/\alpha_{lo}}\sigma_{-2}^{d/\alpha_{hi}} C^{(\infty)}_{\epsilon\leq\eta} && (311)\\
&= \Big(\frac{\sigma_{+2}^{1/\alpha_{hi}}}{2\eta^{1/\alpha_{lo}}}\Big)^d C^{(\infty)}_{\epsilon\leq\eta}. && (312)
\end{aligned}
$$

□

If we further assume $\alpha = \alpha_{lo} = \alpha_{hi} = 2$, then we can bound the number of ℓ-balls required with a constant, independent of ϵ, as

$$C = \max\{N_{\epsilon\leq\eta}, N_{\epsilon\geq\eta}\} \tag{313}$$
$$= \max\Big\{\Big(\frac{\kappa}{\psi}\Big)^{d/2}, \sqrt{2\pi d}\Big(\frac{d\sigma_{max}}{2\pi e\psi\eta}\Big)^{d/2}\Big\} \tag{314}$$
$$= \beta^{d/2}\psi^{-d/2}d^{\xi/2(d+1)} \tag{315}$$

where $\beta = \kappa$, $\xi = 0$ for $N_{\epsilon\leq\eta}$ and $\beta = \frac{\sigma_{\max}(2\pi)^{1/d}}{2\pi e\eta} < \frac{2\sigma_{\max}}{\pi e\eta} = \frac{2\kappa}{\pi e r_\eta^2} < \frac{\kappa}{(2r_\eta)^2}$ for $d \geq 2$, $\xi = 1$ for $N_{\epsilon\geq\eta}$. $N_{\epsilon\geq\eta}$ dominates for large d. The cross over occurs at

$$\Big(\frac{\kappa}{\psi}\Big)^{d/2} = \sqrt{2\pi d}\Big(\frac{d\sigma_{max}}{2\pi e\psi\eta}\Big)^{d/2} \tag{316}$$
$$\implies \frac{\kappa}{\psi} = (2\pi d)^{1/d}\Big(\frac{d\sigma_{max}}{2\pi e\psi\eta}\Big) \tag{317}$$
$$\implies r_\eta^2 = \frac{\eta}{\sigma_-} = (2\pi d)^{1/d}\Big(\frac{d}{2\pi e}\Big) = z(d). \tag{318}$$

where r_η was defined in equation (285). As d grows and $z(d)$ exceeds r_η^2, $N_{\epsilon\geq\eta}$ begins to dominate, therefore we will upper bound C as

$$C \leq \Big(\frac{\kappa}{\psi(2r_\eta)^2}\Big)^{d/2} d^{\frac{1}{2}(d+1)}. \tag{319}$$

Locality	C
$(*)\ r_\eta^2 \leq z(d)$	$N_{\epsilon\geq\eta} \leq \Big(\frac{\kappa}{\psi(2r_\eta)^2}\Big)^{d/2} d^{\frac{1}{2}(d+1)} = \Big(\frac{3\kappa b^2}{(2r_\eta)^2}\Big)^{d/2} d^{d+\frac{1}{2}}$
$r_\eta^2 > z(d)$	$N_{\epsilon\leq\eta} \leq \Big(\frac{\kappa}{\psi}\Big)^{d/2} = \big(3\kappa b^2\big)^{d/2} d^{d/2}$

Table 4: Bounding Constants for $\ell(x, y) = ||x - y||_2^2$, $\psi = \nu/3 = (3b^2 d)^{-1}$ and $z(d) = (2\pi d)^{1/d}\Big(\frac{d}{2\pi e}\Big)$ with smoothness radius r_η and $\psi = \nu/3$. $(*)$ indicates the case that is more likely for difficult problems.

For convenience, we repeat the other relevant constants in Table 5.

$\ell(x, y)$	c	γ	ν
$\ell(x, y) = \|\|x - y\|\|_2^2$	$d\big(\frac{b}{2}\big)^2$	$b^{-2/d}$	$d^{-1}b^{-2}$

Table 5: Bounding Constants

J D-BLiN

The regret bound for Doubling BLiN [14] was originally proved assuming a standard normal distribution, however, the authors state their proof can be easily adapted to any sub-Gaussian distribution, which includes bounded random variables. This matches our setting with bounded payoffs, so we repeat their analysis here for that setting.

Definition 5 (Global Arm Accuracy). $\mathcal{E} \stackrel{\text{def}}{=} \Big\{|\mu(x) - \hat{\mu}_m(C)| \leq r_m + \sqrt{c_1 \frac{\ln T}{n_m}},\ \forall 1 \leq m \leq B_{stop} - 1,\ \forall C \in \mathcal{A}_m,\ \forall x \in C\Big\}$.

Define: $n_m = c_2 \frac{\ln T}{r_m^2} \implies r_m = \sqrt{c_2 \frac{\ln T}{n_m}}$.

Definition 6 (Elimination Rule)**.** *Eliminate $C \in \mathcal{A}_m$ if $\hat{\mu}_m^{\max} - \hat{\mu}_m(C) \geq 2(1+\sqrt{c_1/c_2})r_m = 2(\sqrt{c_2}+\sqrt{c_1})\sqrt{\frac{\ln T}{n_m}}$ where $\hat{\mu}_m^{\max} \overset{\text{def}}{=} \max_{C\in\mathcal{A}_m} \hat{\mu}_m(C)$.*

Lemma 35. $Pr[\mathcal{E}] \geq 1 - 2T^{-2(c_1/c^2-1)}$.

Proof. Assume $y_{C,i} \in [a,b]$ with $c = b-a$ and $\hat{\mu}(C) = \frac{1}{n_m}\sum_{i=1}^{n_m} y_{C,i}$. Applying a Hoeffding inequality gives

$$Pr\Big[|\hat{\mu}(C) - \mathbb{E}[\hat{\mu}(C)]| \geq \sqrt{c_1\frac{\ln T}{n_m}}\Big] \leq 2e^{-2c_1 \ln T/c^2} \tag{320}$$

$$= 2(e^{\ln T})^{-2c_1/c^2} \tag{321}$$

$$= 2T^{-2c_1/c^2}\ \forall C. \tag{322}$$

By Lipschitzness of μ we also have

$$|\mathbb{E}[\hat{\mu}(C)] - \mu(x)| \leq r_m,\ \forall x \in C. \tag{323}$$

Then consider

$$\sup_{x\in C}|\mu(x) - \hat{\mu}(C)| = \sup_{x\in C}|\mu(x) - \mathbb{E}[\hat{\mu}(C)] + \mathbb{E}[\hat{\mu}(C)] - \hat{\mu}(C)| \tag{324}$$

$$\leq \sup_{x\in C}\Big(|\mu(x) - \mathbb{E}[\hat{\mu}(C)]| + |\mathbb{E}[\hat{\mu}(C)] - \hat{\mu}(C)|\Big) \tag{325}$$

$$= \sup_{x\in C}|\mu(x) - \mathbb{E}[\hat{\mu}(C)]| + |\mathbb{E}[\hat{\mu}(C)] - \hat{\mu}(C)| \tag{326}$$

$$\leq \sqrt{c_1\frac{\ln T}{n_m}} + r_m \tag{327}$$

with probability $1 - 2T^{-2c_1/c^2}$. The first inequality follows by triangle inequality and the second follows from equation (323) and considering the complement of equation (322).

The complement of this result occurs with probability

$$Pr\Big[\sup_{x\in C}|\mu(x) - \hat{\mu}(C)| \geq r_m + \sqrt{c_1\frac{\ln T}{n_m}}\Big] \leq 2T^{-2c_1/c^2}. \tag{328}$$

At least 1 arm is played in each cube $C \in \mathcal{A}_m$ for $1 \leq m \leq B_{stop}-1$, therefore, $|\mathcal{A}_m| \leq T$ must be true given the exit condition of the algorithm. In addition, assume $B_{stop} \leq T$ (B_{stop} will be defined such that this is true). Then a union bound over all T^2 events gives

$$Pr\Big[\exists m \in [1, B_{stop}-1], C \in \mathcal{A}_m\ s.t.\ \sup_{x\in C}|\mu(x) - \hat{\mu}(C)| \geq r_m + \sqrt{c_1\frac{\ln T}{n_m}}\Big] \tag{329}$$

$$\leq \sum_{m=1}^{B_{stop}-1}\sum_{C\in\mathcal{A}_m} Pr\Big[\sup_{x\in C}|\mu(x) - \hat{\mu}(C)| \geq r_m + \sqrt{c_1\frac{\ln T}{n_m}}\Big] \tag{330}$$

$$\leq \sum_{m=1}^{B_{stop}-1}\sum_{C\in\mathcal{A}_m} 2T^{-2c_1/c^2} \tag{331}$$

$$\leq 2T^{-2c_1/c^2}T^2. \tag{332}$$

Taking the complement of this event and noting that $\sup_{x\in C}|\mu(x) - \hat{\mu}(C)| \leq r_m + \sqrt{c_1\frac{\ln T}{n_m}} \implies |\mu(x) - \hat{\mu}(C)| \leq r_m + \sqrt{c_1\frac{\ln T}{n_m}}\ \forall x \in C$ gives the desired result. ◻

Lemma 36 (Optimal Arm Survives)**.** *Under event $\mathcal{E}$, the optimal arm $x^* = \arg\max \mu(x)$ is not eliminated after the first $B_{stop}-1$ batches.*

Proof. Let C_m^* denote the cube containing x^* in $\mathcal{A}_m$. Under event $\mathcal{E}$, for any cube $C \in \mathcal{A}_m$ and $x \in C$, the following relation shows that C_m^* avoids the elimination rule in round m:

$$\hat{\mu}(C) - \hat{\mu}(C_m^*) \leq \Big(\mu(x) + r_m + \sqrt{c_1 \frac{\ln T}{n_m}}\Big) + \Big(-\mu(x^*) + r_m + \sqrt{c_1 \frac{\ln T}{n_m}}\Big) \tag{333}$$

$$= \underbrace{(\mu(x) - \mu(x^*))}_{\leq 0} + 2r_m + 2\sqrt{c_1 \frac{\ln T}{n_m}} \tag{334}$$

$$\leq 2\sqrt{c_2 \frac{\ln T}{n_m}} + 2\sqrt{c_1 \frac{\ln T}{n_m}} \tag{335}$$

$$= 2(\sqrt{c_1} + \sqrt{c_2})\sqrt{\frac{\ln T}{n_m}} \tag{336}$$

where the first inequality follows from applying Lemma 35 to upper bound $\hat{\mu}(C)$ and $\hat{\mu}(C_m^*)$ individually. The remaining steps use the optimality of x^*, the definition of r_m, and the elimination rule. $\square$

Lemma 37. *Under event $\mathcal{E}$, for any $1 \leq m \leq B_{stop}$, any $C \in A_m$ and any $x \in C$, Δ_x satisfies*

$$\Delta_x \leq 4(1 + \sqrt{c_1/c_2}) r_{m-1} \tag{337}$$

Proof. For $m = 1$, recall that r_m is the side length of a cube $C \in \mathcal{A}_m$, therefore, $\Delta_x \leq r_{m-1} \leq 4(1 + \sqrt{c_1/c_2}) r_{m-1}$ holds directly from the Lipschitzness of μ.

For $m > 1$, let $C_{m-1}^* \in \mathcal{A}_{m-1}$ be the cube containing x^*. From Lemma 36, this cube has not been eliminated under event $\mathcal{E}$. For any cube $C \in \mathcal{A}_m$ and $x \in C$, it is clear that x is also in the parent of C, denoted C_{par} ($x \in C \subset C_{par}$). Then for any $x \in C$, it holds that

$$\Delta_x = \mu(x^*) - \mu(x) \leq \Big(\hat{\mu}_{m-1}(C_{m-1}^*) + r_{m-1} + \sqrt{c_1 \frac{\ln T}{n_{m-1}}}\Big) + \Big(-\hat{\mu}_{m-1}(C_{par}) + r_{m-1} + \sqrt{c_1 \frac{\ln T}{n_{m-1}}}\Big) \tag{338}$$

$$= (\hat{\mu}_{m-1}(C_{m-1}^*) - \hat{\mu}_{m-1}(C_{par})) + 2(\sqrt{c_1} + \sqrt{c_2})\sqrt{\frac{\ln T}{n_{m-1}}} \tag{339}$$

$$\leq (\hat{\mu}_{m-1}^{\max} - \hat{\mu}_{m-1}(C_{par})) + 2(\sqrt{c_1} + \sqrt{c_2})\sqrt{\frac{\ln T}{n_{m-1}}} \tag{340}$$

$$\leq 2(\sqrt{c_1} + \sqrt{c_2})\sqrt{\frac{\ln T}{n_{m-1}}} + 2(\sqrt{c_1} + \sqrt{c_2})\sqrt{\frac{\ln T}{n_{m-1}}} \tag{341}$$

$$= 4(\sqrt{c_1} + \sqrt{c_2})\sqrt{\frac{\ln T}{n_{m-1}}} \tag{342}$$

$$= 4(1 + \sqrt{c_1/c_2}) r_{m-1} \tag{343}$$

where we have applied Lemma 35 similarly as in Lemma 36 and also used the definition of r_{m-1}. The last two inequalities use the fact that $\hat{\mu}_{m-1}(C_{m-1}^*) \leq \hat{\mu}_{m-1}^{\max}$ and C_{par} was not eliminated. $\square$

Theorem 3. *With probability exceeding $1 - 2T^{-2(c_1/c^2-1)}$, the T-step total regret $R(T)$ of BLiN with Doubling Edge-length Sequence (D-BLiN) [14] satisfies*

$$R(T) \leq 8(1 + \sqrt{c_1/c_2})(2c_2 + 1)\ln(T)^{\frac{1}{d_z+2}} T^{\frac{d_z+1}{d_z+2}} \tag{344}$$

where d_z is the zooming dimension of the problem instance. In addition, D-BLiN only needs no more than $B^* = \frac{\log 2(T) - \log 2(\ln(T))}{d_z+2} + 2$ rounds of communications to achieve this regret rate.

Proof. Since $r_m = \frac{r_{m-1}}{2} \implies r_{m-1} = 2r_m$ for the Doubling Edge-length Sequence, Lemma 37 implies that every cube $C \in A_m$ is a subset of $S(8(1+\sqrt{c_1/c_2})r_m)$. Thus from the definition of zooming number (Corollary 5 with appropriate condition), we have

$$|\mathcal{A}_m| \leq N_{r_m} \leq C_z r_m^{-d_z}. \tag{345}$$

Fix any positive number B. Also by Lemma 37, we know that any arm played after batch B incurs a regret bounded by $8(1+\sqrt{c_1/c_2})r_B$, since the cubes played after batch B have edge length no larger than r_B. Then the total regret that occurs after batch B is bounded by $8(1+\sqrt{c_1/c_2})r_B T$ (where T is an upper bound on the number of arms).

Thus the regret can be bounded as

$$R(T) \leq \sum_{m=1}^{B} \sum_{C \in \mathcal{A}_m} \sum_{i=1}^{n_m} \Delta_{x_{C,i}} + 8(1+\sqrt{c_1/c_2})r_B T \tag{346}$$

where the first term bounds the regret in the first B batches of D-BLiN, and the second term bounds the regret after the first B batches. If the algorithm stops at batch $\tilde{B} < B$, we define $\mathcal{A}_m =$ for any $\tilde{B} < m \leq B$ and inequality equation (346) still holds.

By Lemma 37, we have $\Delta_{x_{C,i}} \leq 8(1+\sqrt{c_1/c_2})r_m$ for all $C \in \mathcal{A}_m$. We can thus bound equation (346) by

$$\begin{aligned}
R(T) &\leq \sum_{m=1}^{B} |\mathcal{A}_m| \cdot n_m \cdot 8(1+\sqrt{c_1/c_2})r_m + 8(1+\sqrt{c_1/c_2})r_B T && (347)\\
&\leq \sum_{m=1}^{B} N_{r_m} \cdot n_m \cdot 8(1+\sqrt{c_1/c_2})r_m + 8(1+\sqrt{c_1/c_2})r_B T && (348)\\
&= \sum_{m=1}^{B} N_{r_m} \cdot c_2 \frac{\ln T}{r_m^2} \cdot 8(1+\sqrt{c_1/c_2})r_m + 8(1+\sqrt{c_1/c_2})r_B T && (349)\\
&= \sum_{m=1}^{B} N_{r_m} \cdot \frac{\ln T}{r_m} \cdot 8c_2(1+\sqrt{c_1/c_2}) + 8(1+\sqrt{c_1/c_2})r_B T && (350)
\end{aligned}$$

where equation (348) uses equation (345), and equation (349) uses equality $n_m = c_2 \frac{\ln T}{r_m^2}$. Since $r_m = 2^{-m+1}$ and $N_{r_m} \leq r_m^{-d_z} \leq 2^{(m-1)d_z}$, we have

$$\begin{aligned}
R(T) &\leq \sum_{m=1}^{B} 2^{(m-1)d_z} \cdot \frac{\ln T}{2^{-m+1}} \cdot 8c_2(1+\sqrt{c_1/c_2}) + 8(1+\sqrt{c_1/c_2})2^{-B+1}T && (351)\\
&= 8(1+\sqrt{c_1/c_2})\Big[c_2 \ln T \sum_{m=1}^{B} 2^{(m-1)(d_z+1)} + 2^{-B+1}T\Big]. && (352)
\end{aligned}$$

Continuing we find

$$R(T) \leq 8(1+\sqrt{c_1/c_2})\Big[c_2 \ln T \sum_{m=1}^{B} 2^{(m-1)(d_z+1)} + 2^{-B+1}T\Big] \tag{353}$$

$$= 8(1+\sqrt{c_1/c_2})\Big[c_2 \ln T \sum_{m=1}^{B} \big(2^{d_z+1}\big)^{m-1} + 2^{-B+1}T\Big] \tag{354}$$

$$= 8(1+\sqrt{c_1/c_2})\Big[c_2 \ln T \sum_{m=0}^{B-1} \big(2^{d_z+1}\big)^{m} + 2^{-B+1}T\Big] \tag{355}$$

$$= 8(1+\sqrt{c_1/c_2})\Big[c_2 \ln T \Big(\frac{2^{B(d_z+1)}-1}{2^{d_z+1}-1}\Big) + 2^{-B+1}T\Big] \quad \text{via geometric series} \tag{356}$$

$$\leq 8(1+\sqrt{c_1/c_2})\Big[c_2 \ln T \Big(\frac{2^{B(d_z+1)}}{2^{d_z+1}-1}\Big) + 2^{-B+1}T\Big] \tag{357}$$

$$\leq 8(1+\sqrt{c_1/c_2})\Big[c_2 \ln T \Big(2 \cdot \frac{2^{B(d_z+1)}}{2^{d_z+1}}\Big) + 2^{-B+1}T\Big] \tag{358}$$

$$= 8(1+\sqrt{c_1/c_2})\Big[2c_2 2^{(B-1)(d_z+1)} \ln T + 2^{-(B-1)}T\Big]. \tag{359}$$

This inequality holds for any positive B. By choosing $B^* = 1 + \frac{\log_2(\frac{T}{\ln T})}{d_z+2}$, we have

$$R(T) \leq 8(1+\sqrt{c_1/c_2})\Big[2c_2 \Big(\frac{T}{\ln T}\Big)^{\frac{(d_z+1)}{(d_z+2)}} \ln T + \Big(\frac{\ln T}{T}\Big)^{\frac{1}{(d_z+2)}} T\Big] \tag{360}$$

$$= 8(1+\sqrt{c_1/c_2})\Big[2c_2 T^{\frac{(d_z+1)}{(d_z+2)}} \ln T^{1-\frac{(d_z+1)}{(d_z+2)}} + T^{1-\frac{1}{(d_z+2)}} \ln T^{\frac{1}{(d_z+2)}}\Big] \tag{361}$$

$$= 8(1+\sqrt{c_1/c_2})\Big[2c_2 T^{\frac{(d_z+1)}{(d_z+2)}} \ln T^{\frac{1}{(d_z+2)}} + T^{\frac{(d_z+1)}{(d_z+2)}} \ln T^{\frac{1}{(d_z+2)}}\Big] \tag{362}$$

$$= 8(1+\sqrt{c_1/c_2})(2c_2+1) T^{\frac{(d_z+1)}{(d_z+2)}} \ln T^{\frac{1}{(d_z+2)}}. \tag{363}$$

□

Corollary 9. *Setting* $c_1 = 2c^2$ *and* $c_2 = \left(\frac{c^2}{2}\right)^{1/3}$ *simplifies Theorem 3 such that*

$$R(T) \leq 8(1+(4c^2)^{1/3})^2 T^{\frac{(d_z+1)}{(d_z+2)}} \ln T^{\frac{1}{(d_z+2)}}. \tag{364}$$

with probability $1 - 2T^{-2}$.

Proof. □

K Experimental Setup and Details

Here we provide further details on the experiments.

K.1 Loss Visualization and Rank Test

Figure 1 and claims made in Section 5 analyze several classical matrix games. We report the payoff matrices in standard row-player / column-player payoff form below. All games are then shifted and scaled so payoffs lie in $[0, 1]$ (i.e., first by subtracting the minimum and then scaling by the max).

RPS:

$$\begin{bmatrix} 0/0 & -1/1 & 1/-1 \\ 1/-1 & 0/0 & -1/1 \\ -1/1 & 1/-1 & 0/0 \end{bmatrix}. \tag{365}$$

Chicken:

$$\begin{bmatrix} 0/0 & -1/1 \\ 1/-1 & -3/-3 \end{bmatrix}. \tag{366}$$

Matching Pennies:

$$\begin{bmatrix} 1/-1 & -1/1 \\ -1/1 & 1/-1 \end{bmatrix}. \tag{367}$$

Modified-Shapleys:

$$\begin{bmatrix} 1/-0.5 & 0/1 & 0.5/0 \\ 0.5/0 & 1/-0.5 & 0/1 \\ 0/1 & 0.5/0 & 1/-0.5 \end{bmatrix}. \tag{368}$$

Prisoner's Dilemma:

$$\begin{bmatrix} -1/-1 & -3/0 \\ 0/-3 & -2/-2 \end{bmatrix}. \tag{369}$$

K.1.1 NashConv is Biased

We use Chicken to demonstrate the effect of sampled play on the bias of the popular NashConv loss. NashConv is unable to capture the interior Nash equilibrium because of its high bias. In contrast, our proposed loss $\mathcal{L}^\tau$ is guaranteed to capture all equilibria at low temperature τ.

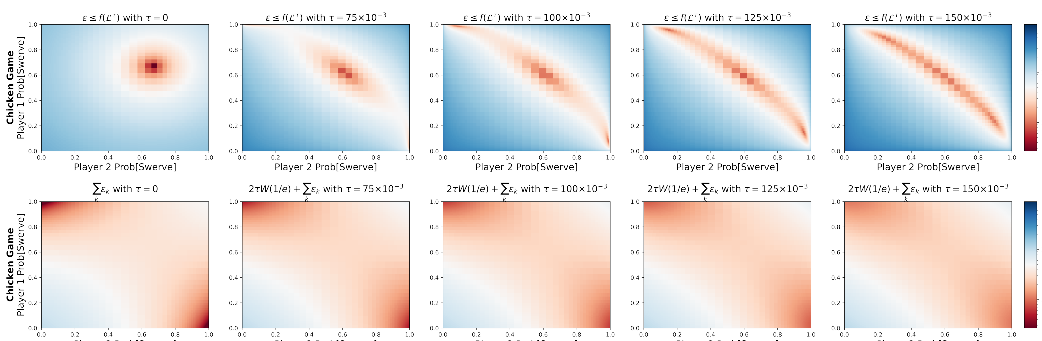

Figure 5: Effect of Sampled Play on a Biased Loss. The first row displays the expected upper bound guaranteed by our proposed loss $\mathcal{L}^\tau$ (also displayed in Figure 1). The second row displays the expectation of NashConv under sampled play, i.e., $\sum_k \epsilon_k$ where $\epsilon_k = \mathbb{E}_{a_{-k} \sim x_{-k}}[\max_{a_k} u_k^\tau(\boldsymbol{a})] - \mathbb{E}_{\boldsymbol{a} \sim \boldsymbol{x}}[u_k^\tau(\boldsymbol{a})]$. To be consistent, we add the offset $n\tau W(1/e) + \sum_k \epsilon_k$ to NashConv per Lemma 14, which relates the exploitability at positive temperature to that at zero temperature. The resulting loss surface clearly shows NashConv fails to recognize the interior Nash equilibrium due to its inherent bias. NashConv succeeds in finding pure equilibria because sampling from a pure joint equilibrium is a deterministic process (no noise means no bias).

K.2 Saddle Point Analysis

To generate Figure 2, we follow a procedure similar to the study of MNIST in [12] (Section 3 of Supp.). Their recommended procedure searches for critical points in two ways. The first repeats a randomized, iterative optimization process 20 times. They then sample one these 20 trials at random, select a random point along the descent trajectory, and search for a critical point (using Newton's method) nearby. They repeat this sampling process 100 times. The second approach randomly selects a feasible point in the decision set and searches for a critical point nearby (again using Newton's method). They also perform this 100 times.

Our protocol differs from theirs slightly in a few respects. One, we use SGD, rather than the saddle-free Newton algorithm to trace out an initial descent trajectory. Two, we do not add noise to strategies along the descent trajectory prior to looking for critical points. Lastly, we use different experimental hyperparameters. We run SGD for 1000 iterations rather than 20 epochs and rerun SGD 100 times rather than 20. We sample 1000 points for each of the two approaches for finding critical points.

K.3 SGD on Classical Games

The games examined in Figure 3 were all taken from [15]. Each is available via open source implementations in OpenSpiel [22] or GAMUT [33].

We compare against several other baselines, replicating the experiments in [15]. RM indicates regret-matching and FTRL indicates follow-the-regularized-leader. These are, arguably, the two most popular scalable stochastic algorithms for approximating Nash equilibria. ^{y}QRE^{auto} is a stochastic algorithm developed in [15].

For each of the experiments, we sweep over learning rates in log-space from 10^{-3} to 10^{2} in increments of 1. We also consider whether to run SGD with the projected-gradient and whether to constrain iterates to the simplex via Euclidean projection or entropic mirror descent [6]. We then presented the results of the best performing hyperparameters. This was the same approach taken in [15].

Saddle Points in Blotto To confirm the existence of saddle points, we computed the Hessian of $\mathcal{L}(\boldsymbol{x_{10k}})$ for SGD ($s = \infty$), deflated the matrix by removing from its eigenvectors all directions orthogonal to the simplex, and then computed its top-$(n\bar{m} - n)$ eigenvalues. We do this because there always exists a n-dimensional nullspace of the Hessian at zero temperature that lies outside the tangent space of the simplex, and we only care about curvature within the tangent space. Specificaly, at an equilibrium $\boldsymbol{x}$, if we compute $z^\top Hess(\mathcal{L})z$ where z is formed as a linear combination of the vectors $\{[x_1, 0, \ldots, 0]^\top, \ldots, [0, \ldots, x_n]^\top$, then each block $\tilde{B}_{kl}$ is identically zero at an equilibrium: $\tilde{B}_{kl}x_l = \sqrt{\eta_k}[I - \frac{1}{m_k}\mathbf{1}\mathbf{1}^\top]H^k_{kl}x_l = \sqrt{\eta_k}\Pi_{T\Delta}(\nabla^k_{x_k}) = 0$. By Lemma 17, this implies there is zero curvature of the loss in the direction z: $z^\top Hess(\mathcal{L})z = 0$.

K.4 BLiN on Artificial Game

To construct the 7-player, 2-action, symmetric, artifical game in Figure 4, we used the following coefficients (discovered by trial-and-error):

$$\begin{bmatrix} 0.09906873 & 0 & 0.23116037 & 0 & 0.62743528 & 0 & 0.19813746 \\ 0 & 0.33022909 & 0 & 0.03302291 & 0 & 0.62743528 & 0 \end{bmatrix}. \tag{370}$$

The first row indicates the payoffs received when player i plays action 0 and the background population plays any of the possible joint actions (number of combinations with replacement). For example, the first column indicates the payoff when all background players play action 0. The second column indicates all background players play action 0 except for one which plays action 1, and so on. The last column indicates all background players play action 1. These $2n$ scalars uniquely define the payoffs of a symmetric game.

Given that this game only has two actions, we represent a mixed strategy by a single scalar $p \in [0, 1]$, i.e., the probability of the first action. Furthermore, this game is symmetric and we seek a symmetric equilibrium, so we can represent a full Nash equilibrium by this single scalar p. This reduces our search space from $7 \times 2 = 14$ variables to 1 variable (and obviates any need for a map s from the unit hypercube to the simplex—see Lemma 24).