Sublinear Classical and Quantum Algorithms for General Matrix Games

Tongyang Li\textsuperscript{1,2}  Chunhao Wang\textsuperscript{3,4}  Shouvanik Chakrabarti\textsuperscript{1}  Xiaodi Wu\textsuperscript{1}

\textsuperscript{1}Joint Center for Quantum Information and Computer Science, Department of Computer Science, and Institute for Advanced Computer Studies, University of Maryland

\textsuperscript{2}Center for Theoretical Physics, Massachusetts Institute of Technology

\textsuperscript{3}Department of Computer Science and Engineering, Pennsylvania State University

\textsuperscript{4}Department of Computer Science, University of Texas at Austin
tongyang@mit.edu, cwang@psu.edu, \{shouv, xwu\}@cs.umd.edu

Abstract

We investigate sublinear classical and quantum algorithms for matrix games, a fundamental problem in optimization and machine learning, with provable guarantees. Given a matrix $A \in \mathbb{R}^{n \times d}$, sublinear algorithms for the matrix game \(\min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} y^\top Ax\) were previously known only for two special cases: (1) $\mathcal{Y}$ being the $\ell_1$-norm unit ball, and (2) $\mathcal{X}$ being either the $\ell_1$- or the $\ell_2$-norm unit ball. We give a sublinear classical algorithm that can interpolate smoothly between these two cases: for any fixed $q \in (1, 2]$, we solve the matrix game where $\mathcal{X}$ is a $\ell_q$-norm unit ball within additive error $\epsilon$ in time $O(\sqrt{n} + d)\sqrt{\log(1/\epsilon)}$. We also provide a corresponding sublinear quantum algorithm that solves the same task in time $O((\sqrt{n} + d)\sqrt{\log(1/\epsilon)})$ with a quadratic improvement in both $n$ and $d$. Both our classical and quantum algorithms are optimal in the dimension parameters $n$ and $d$ up to poly-logarithmic factors. Finally, we propose sublinear classical and quantum algorithms for the approximate Carathéodory problem and the $\ell_q$-margin support vector machines as applications.

Introduction

Motivations. Minimax games between two parties, i.e., $\min_x \max_y f(x, y)$, is a basic model in game theory and has ubiquitous connections and applications to economics, optimization and machine learning, theoretical computer science, etc. Among minimax games, one of the most fundamental cases is the bilinear minimax game, also known as the matrix game, with the following form:

$$\min_x \max_y y^\top Ax, \quad A \in \mathbb{R}^{n \times d}, \quad \mathcal{X} \subset \mathbb{R}^d, \quad \mathcal{Y} \subset \mathbb{R}^n. \quad (1)$$

Matrix games are fundamental in algorithm design due to their equivalence to linear programs (Dantzig 1998), and also in machine learning because they contain classification (Novikoff 1963; Minsky and Papert 1988) as a special case, and many other important problems.

For many common domains $\mathcal{X}$ and $\mathcal{Y}$, matrix games can be solved efficiently within approximation error $\epsilon$, i.e., to output $x' \in \mathcal{X}$ and $y' \in \mathcal{Y}$ such that $(y')^\top Ax'$ is $\epsilon$-close to the optimum in (1). For some specific choices of $\mathcal{X}$ and $\mathcal{Y}$, the matrix game can even be solved in sublinear time in the size $nd$ of $A$. When $\mathcal{X}$ and $\mathcal{Y}$ are both $\ell_1$-unit norm balls, Grigoriadis and Khachiyan (1995) can solve the matrix game in time $O((n + d)\log(n + d)/\epsilon^2)$. When $\mathcal{X}$ is the $\ell_2$-norm unit ball in $\mathbb{R}^d$ and $\mathcal{Y}$ is the $\ell_1$-norm unit ball in $\mathbb{R}^n$, Clarkson, Hazan, and Woodruff (2012) can solve the matrix game in time $O((n + d)\log n/\epsilon^2)$.

As far as we know, the $\ell_1$-$\ell_1$ and $\ell_2$-$\ell_1$ matrix games are the only two cases where sublinear algorithms are known. However, there is general interest of solving matrix games with general norms. For instance, matrix games are closely related to the Carathéodory problem for finding a sparse linear combination in the convex hull of given data points, where all the $\ell_p$-metrics with $p \geq 2$ have been well-studied (Barman 2015; Mirrokni et al. 2017; Combettes and Pokutta 2019). In addition, matrix games are common in machine learning especially support vector machines (SVMs), and general $\ell_p$-margin SVMs have also been considered by previous literature, see e.g. the book by Deng, Tian, and Zhang (2012). In all, it is a natural question to investigate sublinear algorithms for general matrix games. In addition, quantum computing has been rapidly advancing and current technology has reached “quantum supremacy” for some specific tasks (Arute et al. 2019); since previous works have given sublinear quantum algorithms for $\ell_1$-$\ell_1$ matrix games (Li, Chakrabarti, and Wu 2019; Apeldoorn and Gilyén 2019) and $\ell_2$-$\ell_1$ matrix games (Li, Chakrabarti, and Wu 2019) with running time $(\sqrt{n} + d)\sqrt{\log(1/\epsilon)}$, it is also natural to explore sublinear quantum algorithms for general matrix games.

Contributions. We conduct a systematic study of $\ell_q$-$\ell_1$ matrix games for any $q \in (1, 2]$ which corresponds to $\ell_q$-margin SVMs and the $\ell_q$-Carathéodory problem for any $p \geq 2$. We use the following entry-wise input model, the standard assumption in the sublinear algorithms in Grigoriadis and Khachiyan (1995); Clarkson, Hazan, and Woodruff (2012):

Input model: Given any $i \in [n]$ and $j \in [d]$, the $j^{th}$ entry of $A_i$ can be recovered in $O(1)$ time.

Quantumly, we consider an almost same oracle: Quantum input model: Given any $i \in [n]$ and $j \in [d]$, the $j^{th}$ entry of $A_i$ can be recovered in $O(1)$ time coherently.

The only difference is to allow coherent queries, which give quantum algorithms the ability to query different locations in superposition, and have been the standard quantiza-
tion of the classical inputs and commonly adopted in previous works (Li, Chakrabarti, and Wu [2019]; Apeldoorn and Gilyén [2019]).

**Theorem 1 (Main Theorem).** Given \( q \in (1, 2] \), Define \( p \geq 2 \) such that \( \frac{1}{p} + \frac{1}{q} = 1 \). Consider the \( \ell_q \)-\( \ell_1 \) matrix game

\[
\sigma := \max \min_{x \in B^d_q} p^\top A x,
\]

where \( B^d_q \) is the \( \ell_q \)-unit ball in \( \mathbb{R}^d \) and \( \Delta_n \) is the \( \ell_1 \)-simplex in \( \mathbb{R}^n \). Then we can find an \( \tilde{x} \in B^d_q \) s.t.

\[
\min_{i \in [n]} A_i \tilde{x} \geq \sigma - \epsilon
\]

with success probability at least 2/3, using

- \( O\left(\frac{(n+d)(p+\log n)}{\varepsilon^2}\right) \) classical queries (Theorem 2); or
- \( \tilde{O}\left(\frac{n^{2/3}d^{1/3}}{\varepsilon^2} + \frac{n^{1/3}d^{2/3}}{\epsilon} \right) \) quantum queries (Theorem 3).

When \( p = \Omega(\log d/\epsilon) \), the above bounds can be improved (by Lemma 7) to respectively

- \( O\left(\frac{(n+d)(\log d + \log n)}{\varepsilon^2}\right) \) queries to the classical input model;
- \( \tilde{O}\left(\frac{\sqrt{n}d}{\epsilon^2} + \frac{\sqrt{d}}{\varepsilon} \right) \) queries to the quantum input model.

Both results are optimal in \( n \) and \( d \) up to poly-log factors as we show \( \Omega(n+d) \) and \( \Omega(\sqrt{n} + \sqrt{d}) \) classical and quantum lower bounds respectively when \( \epsilon = \Theta(1) \) (Theorem 4).

Conceptually, our classical and quantum algorithms for general matrix games enjoy quite a few nice properties. On the one hand, they can be directly applied to

- **Convex geometry:** We give the first sublinear classical and quantum algorithms for the approximate Carathéodory problem (Corollary 1), improving the previous linear-time algorithms of Mirrokni et al. [2017]; Combettes and Pokutta [2019];

- **Supervised learning:** We provide the first sublinear algorithms for general \( \ell_q \)-margin support vector machines (SVMs) (Corollary 2).

On the other hand, our quantum algorithm is friendly for near-term applications. It uses the standard quantum input model and needs not to use any sophisticated quantum data structures. It is classical-quantum hybrid where the quantum part is isolated by pieces of state preparations connected by classical processing. Its output is completely classical.

Technique-wise, we are deeply inspired by Clarkson, Hazan, and Woodruff [2012], which serves as the starting point of our algorithm design. At a high level, Clarkson et al.’s algorithm follows a primal-dual framework where the primal part applies \( \ell_2 \)-norm (online gradient descent) by Zinkevich [2003], and the dual part applies multiplicative weight updates (MWU) by \( \ell_2 \)-sampling. The choice of the \( \ell_2 \)-norm metric greatly facilitates the design and analysis of the algorithms for both parts. However, it is conceivable that more sophisticated design and analysis will be required to handle general \( \ell_q \)-\( \ell_1 \) matrix games.

Classically, our main technical contribution is to expand the primal-dual approach of Clarkson, Hazan, and Woodruff [2012] to work for more general metrics for the \( \ell_q \)-\( \ell_1 \) matrix game. Specifically, in the primal we replace OGD by a generalized \( p \)-norm OGD due to Shalev-Shwartz [2012], and in the dual we replace the \( \ell_2 \)-sampling by \( \ell_q \)-sampling. We conduct a careful algorithm design and analysis to ensure that this strategy only incurs an \( O(p/\epsilon^2) \) overhead in the number of iterations, and the error of the \( \ell_q \)-\( \ell_1 \) matrix game is still bounded by \( \epsilon \) as in (5). In a nutshell, our algorithm can be viewed as an interpolation between the \( \ell_2 \)-\( \ell_1 \) matrix game (Clarkson, Hazan, and Woodruff [2012]) and the \( \ell_1 \)-\( \ell_1 \) matrix game (Grigoriadis and Khachiyan [1995]); when \( q \) is close to 2 the algorithm is more similar to Clarkson, Hazan, and Woodruff [2012], whereas when \( q \) is close to 1, \( p \) is large and the \( p \)-norm GD becomes closer to the normalized expontential gradient (Shalev-Shwartz 2012), which is exactly the update rule in (Grigoriadis and Khachiyan [1995]).

Quantumly, our main contribution is the systematic improvement of the previous quantum algorithm for \( \ell_q \)-\( \ell_1 \) matrix games by Li, Chakrabarti, and Wu [2019]. They achieved a quantum speedup of \( O(\sqrt{n} + \sqrt{d}) \) for solving \( \ell_q \)-\( \ell_1 \) matrix games by leveraging quantum amplitude amplification and observing that \( \ell_2 \)-sampling can be readily accomplished by quantum state preparation as quantum states refer to \( \ell_1 \) unit vectors. For general \( \ell_q \)-\( \ell_1 \) matrix game \( (q \in (1, 2]) \), we likewise upgrade both primal and dual parts as in our classical algorithm: specifically, in the primal, we apply the \( p \)-norm OGD in \( O(\sqrt{d}) \) time, whereas in the dual, we apply the multiplicative weight update via an \( \ell_q \)-sampling in \( O(\sqrt{n}) \) time. To that end, we contribute to the following technical improvements, which may be of independent interest:

- In our algorithm, we cannot directly leverage quantum state preparation in the \( \ell_q \) metric because it corresponds to \( \ell_2 \)-normalized vectors. Instead, we propose Algorithm 2 for quantum \( \ell_q \)-sampling with \( O(\sqrt{n}) \) oracle calls which works with states whose amplitudes follow \( \ell_q \)-norm proportion. Measuring such states is equivalent to performing \( \ell_q \)-sampling.

- When \( p = q = 2 \), we improved the \( \epsilon \)-dependence from the \( 1/\epsilon^3 \) in the prior art by Li, Chakrabarti, and Wu [2019] to \( 1/\epsilon^7 \). This is achieved by deriving a better upper bound on the entries of the vectors in the \( p \)-norm OGD (i.e., \( y_{t,j} \) as in Eq. (38); see the supplementary material (Eqs. (38–40)) for details.

- In our lower bounds, although the hard cases are motivated by Li, Chakrabarti, and Wu [2019], the matrix game values are much more complicated in the \( \ell_q \)-\( \ell_1 \) case. In the supplementary material, we figure out two functions \( f_1 \) and \( f_2 \) that not only separate the game values of two specifically-constructed \( \ell_q \)-\( \ell_1 \) matrix games but also have monotone and nonnegative properties, which are crucial factors in our proof.
These improvements together result in Theorem 1.

Related work. Matrix games were probably first studied as zero-sum games by Neumann [1928]. The seminal work Nemirovski and Yudin [1983] proposed the mirror descent method and gave an algorithm for solving matrix games in time O(nd/ε²). This was later improved to O(nd/ε) by the prox-method due to Nemirovskii [2004] and the dual extrapolation method due to Nesterov [2007]. To further improve the cost, there have been two main focuses:

• Sampling-based methods: They focus on achieving sublinear cost in nd, the size of the matrix A. [Grigoriadis and Khachiyan (1995); Clarkson, Hazan, and Woodruff (2012)] mentioned above are seminal examples; these sublinear algorithms can also be used to solve semidefinite programs (Garber and Hazan 2011), SVMs (Hazan, Koren, and Srebro 2011), etc.

• Variance-reduced methods: They focus on the cost in 1/ε, in particular its decoupling with nd. Palaniappan and Bach (2016) showed how to apply the standard SVRG (John- son and Zhang 2013) technique for solving ℓ₂-ℓ₂ matrix games; this idea can also be extended to smooth functions using general Bregman divergences (Shi, Zhang, and Yu 2017). Variance-reduced methods for solving matrix games culminate in Carmon et al. (2019), where they show how to solve ℓ₁-ℓ₁ and ℓ₂-ℓ₂ matrix games in time O(ndz(A) + √ndz(A) · (n + d)/ε), where nz(A) is the number of nonzero elements in A.

There have been relatively few quantum results for solving matrix games. Kapoor, Wiebe, and Svore (2016) solved the ℓ₂-ℓ₂ matrix game with cost O(√nd/ε²) using an unusual input model where the representation of a data point in ℜᵈ is the concatenation of ℓ² floating point numbers. More recently, Apeldoorn and Gilyén (2019) was able to solve the ℓ₁-ℓ₁ matrix game with cost O(√nd/ε³ + √d/ε) using the standard input model above, and Li, Chakrabarti, and Wu (2019) solved the ℓ₂-ℓ₂ matrix game with cost O(√nd/ε² + √d/ε) also using the standard input model.

Preliminaries and Notations

To facilitate the reading of this paper, we introduce necessary definitions and notations here.

Preliminaries for quantum computing. Quantum mechanics can be formulated in terms of linear algebra. For the space ℂᵈ, we denote {e₀, . . . , eᵈ−¹} as its computational basis, where eᵢ = (0, . . . , 1, . . . , 0)ᵀ where 1 only appears in the (i + 1)th coordinate. These basic vectors can be written by the Dirac notation: eᵢ := |i⟩ (called a “ket”), and eᵢᵀ := ⟨i| (called a “bra”). A d-dimensional quantum state is a unit vector in ℂᵈ, i.e., |ψ⟩ = (ψ₀, . . . , ψᵈ−¹)ᵀ such that ∑ᵢ=0ᵈ−¹ |ψᵢ|² = 1.

Tensor product of quantum states is their Kronecker product: if |ψ⟩ ∈ ℂᵈ, and |φ⟩ ∈ ℂᵈ', then

|ψ⟩ ⊗ |φ⟩ := (ψ₀ |φ₀⟩, ψ₁ |φ₁⟩, . . . , ψᵈ−¹ |φᵈ−¹⟩)ᵀ.

which is a vector in ℂᵈ₁ ⊗ ℂᵈ'.

Quantum access to an input matrix, also known as a quantum oracle, is reversible and allows access to coordinates of the matrix in superposition, this is the essence of quantum speedups. In particular, to access entries of a matrix A ∈ ℜⁿ×ᵈ, we exploit a quantum oracle Oₐ, which is a unitary transformation on ℂⁿ ⊗ ℂᵈ ⊗ ℂᵈₑ (dₑbeing the dimension of a floating-point register) such that

Oₐ(|i⟩ ⊗ |j⟩ ⊗ |z⟩) = |i⟩ ⊗ |j⟩ ⊗ (z + Aᵢⱼ) (5)

for any i ∈ [n], j ∈ [d], and z ∈ ℂᵈₑ. Intuitively, Oₐ reads the entry Aᵢⱼ and stores it in the third register as a floating-point number. However, to promise that Oₐ a unitary transformation, Oₐ applies the XOR operation (⊕) on the third register. This is a natural generalization of classical reversible computation, when each entry of A can be recovered in O(1) time. Subsequently, a common assumption is that a single query to Oₐ takes O(1) cost.

Interpolation for large p. If p is large, we prove the following lemma showing that we can restrict without loss of generality to cases where p such that ½+ 1/q = 1 is O(log d/ε), since in this case the ℓₚ-ℓ₂ matrix game is ε-close to the ℓ₁-ℓ₁ matrix game in the following sense:

Lemma 1. An ℓₚ-ℓ₂ matrix game where p such that ½+ 1/q = 1 is greater than log d/ε can be solved using an algorithm for solving ℓ₁-ℓ₁ games. This introduces an error O(ε) in the objective value.

Proof. Assume without loss of generality that ε ≤ 1/2. Let p ≥ log d/ε ≥ log d/((- log(1 − ε))). It can be easily verified that Bᵈ₁ ⊆ Bᵈ₂ ⊆ Bᵈ₁ + (1 − d⁻¹/p) Bᵈ₁. Thus Bᵈ₂ ⊆ Bᵈ₁ + ε Bᵈ₁.

Consider applying an algorithm to solve an ℓ₁-ℓ₁ matrix game instead of the ℓₚ-ℓ₂ matrix game as required in (2). Let the optimal solution to (2) be x∗ ∈ Bᵈ₂,p∗ ∈ Δⁿ. By the previous analysis, there is a point x ∈ Bᵈ₂ such that ||x − x∗||₁ ≤ ε. Thus the solution x, p∗ has an error at most O(ε) from the true objective, and the algorithm for solving ℓ₁-ℓ₁ games finds a solution at least as good as this.

Notations. Throughout the paper, we denote p, q > 1 to be two real numbers such that ½+ 1/q = 1; p ∈ [2, +∞) and q ∈ (1, 2]. For any s > 1, we use Bᵈᵱ to denote the d-dimensional unit ball in ℓₛ-norm, i.e., Bᵈᵱ := {x : ∑ᵢ=0ᵈ−¹ |xᵢ|ˡ ≤ 1}; we use Δⁿ to denote the n-dimensional unit simplex {p ∈ ℜⁿ : pᵱ ≥ 0, ∑ᵢ=₁ⁿ pᵱ = 1}, and use Δ₁ to denote the n-dimensional all-one vector. We denote A ∈ ℜⁿ×ᵈ to be the matrix whose ℓᵱᵱ row is Aᵱ for all i ∈ [n]. We define sgn : ℜ → {−1, 0, 1} such that sgn(x) = −1 if x < 0, sgn(x) = 1 if x > 0, and sgn(0) = 0.

A Sublinear Classical Algorithm for General Matrix Games

For any q ∈ (1, 2], we consider the ℓₚ-ℓ₁ matrix game:

σ := max |x|∈Bᵈ₁,p∈Δₙ Aₓ. (6)
The goal is to find a $\tilde{x}$ that approximates the equilibrium of the matrix game within additive error $\epsilon$:

$$\min_{\bar{x} \in [n]} A_i \bar{x} \geq \sigma - \epsilon.$$  \hfill (7)

Throughout the paper, we assume $A_1, \ldots, A_n \in \mathbb{R}^d$, i.e., all the $n$ data points are normalized to have $\ell_2$-norm at most 1.

**Algorithm 1:** A sublinear algorithm for $\ell_q$-$\ell_2$ games.

**Input:** $\epsilon > 0$; $p \in [2, +\infty)$, $q \in (1, 2]$ such that $\frac{1}{p} + \frac{1}{q} = 1$; $A \in \mathbb{R}^{n \times d}$ with $A_i \in \mathbb{R}^d$ for $i \in [n]$.

**Output:** $\tilde{x}$ that satisfies (7), $w_1 = 1_n$;

1. Let $T = \left\lfloor \frac{995 \log n + 4p}{\epsilon^2} \right\rfloor$, $y_1 = 0_d$, $\eta = \sqrt{\frac{11 \log n}{12T}}$, $w_1 = 1_n$;

2. for $t = 1$ to $T$ do

   3. $p_t \leftarrow \frac{w_{t+1,i}}{\|w_{t+1,i}\|}$, $x_t \leftarrow \max_{y_t, \|y_t\| \leq 1} \eta y$;

   4. Choose $i_t \in [n]$ by $i_t \leftarrow \sim$ with probability $p_t(i)$;

   5. Define $y_{t+1,j} \leftarrow y_t + \frac{\eta \text{sgn}(A_{i_t,j}) A_{i_t,j}}{\sqrt{2T} \|A_{i_t,j}\|^{p-1}}$

6. Choose $j_t \in [d]$ by $j_t \leftarrow \sim$ with probability $\frac{x_t(j_t)^p}{\|x_t\|_p^p}$;

7. for $i = 1$ to $n$ do

   8. $\tilde{v}_t(i) \leftarrow A_t(j_t) \|x_t\|_q / x_t(j_t)^{q-1}$;

   9. $v_t(i) \leftarrow \text{clip}(\tilde{v}_t(i), \frac{1}{2})$

10. $\forall v, M \leftarrow \min\{M, \max\{-M, v\}\}$

11. $w_{t+1} \leftarrow w_t(i)(1 - \eta v_t(i) + \eta^2 \tilde{v}_t(i)^2)$;

12. Return $\tilde{x} = \frac{1}{T} \sum_{t=1}^T x_t$.

**Theorem 2.** The output of Algorithm 1 satisfies (7) with probability at least $2/3$, and its total running time is $O\left(\frac{(n+d)(p+\log n)}{\epsilon^2} \right)$ where $p \geq 2$ such that $\frac{1}{p} + \frac{1}{q} = 1$.

Our sublinear algorithm follows the primal-dual approach of Algorithm 1 of Clarkson, Hazan, and Woodruff (2012), which solves $\ell_1$-$\ell_2$ matrix games. Here for $\ell_q$-$\ell_2$ matrix games, the solution vector $x$ now lies in $\mathbb{R}^d$. Hence, the most natural adaptations are to use $\ell_q$-sampling instead of $\ell_2$-sampling in the primal updates, and to use a $p$-norm OGD by Shalev-Shwartz (2012) which generalizes the online gradient descent by Zinkevich (2003) in $\ell_2$-norm. In the following, we use various technical tools to show these natural adaptations actually work.

**Proposition 1** (Shalev-Shwartz [2012] Corollary 2.18). Consider a set of vectors $u_1, \ldots, u_T \in \mathbb{R}^d$ such that $\|u_i\|_q \leq 1$. Set $i = \sqrt{\frac{T}{2q^2}}$. Let $x_0 \leftarrow 0_d$, $x_{t+1} \leftarrow \frac{1}{\|u_t\|_q} u_t x_t + \frac{\text{sgn}(u_t) u_t}{\|u_t\|_q} \eta^{q-1}$ for all $i \in [d]$, and $x_{t+1} \leftarrow \max_i \{1, \|x_{t+1}\|_q\}$.

Then

$$\max_{x \in \mathbb{R}^d} \sum_{t=1}^T u_t^\top x - \sum_{t=1}^T u_t^\top x_t \leq \sqrt{\frac{2T}{q-1}}.$$  \hfill (8)

The analysis of Algorithm 1 uses the following lemma, adapted from the variance multiplicative weight lemma and martingale tail bounds in Clarkson, Hazan, and Woodruff (2012).

**Lemma 2** (Section 2 of Clarkson, Hazan, and Woodruff 2012). In Algorithm 1 the parameters $p_t$ in Line 3 and $v_t$ in Line 9 satisfy

$$\sum_{t \in [T]} p_t^2 v_t^2 \leq \min_{\bar{v} \in [n]} \sum_{t \in [T]} v_t(i) + \eta \sum_{t \in [T]} p_t v_t^2 + \frac{\log n}{\eta}$$  \hfill (9)

where $v_t^2$ is defined as $v_t^2 := (v_t)^2$ for all $i \in [n]$, as long as the update rule of $w_t$ is as in Line 10 and $\text{Var}[v_t(i)^2] \leq 1$ for all $i \in [T]$ and $i \in [n]$. Furthermore, with probability at least $1 - O(1/n)$,

$$\max_{i \in [n]} \sum_{t \in [T]} v_t(i) \leq 4\eta T;$$  \hfill (10)

$$\left| \sum_{t \in [T]} A_{i,t} x_t - \sum_{t \in [T]} p_t v_t \right| \leq 10\eta T,$$  \hfill (11)

with probability at least $5/7$, $\sum_{t \in [T]} p_t^2 v_t^2 \leq 7T$.

We also need to prove the following inequality on different moments of random variables.

**Lemma 3.** Suppose that $X$ is a random variable on $\mathbb{R}$, and $p \geq 2$. If $\mathbb{E}[|X|^p] \leq 1$, then $\mathbb{E}[X^2] \leq 1$.

**Proof.** Denote the probability density of $X$ as $\mu$. Then

$$\int_{-\infty}^{+\infty} |x|^p d \mu_x = \mathbb{E}[|X|^p] \leq 1.$$  \hfill (13)

By Hölder’s inequality, we have

$$\int_{-\infty}^{+\infty} |x|^2 \cdot 1^{2-p} d \mu_x = \int_{-\infty}^{+\infty} x^2 d \mu_x \leq \left(\int_{-\infty}^{+\infty} |x|^p d \mu_x\right)^{2/p} \left(\int_{-\infty}^{+\infty} 1 d \mu_x\right)^{1-2/p}$$

$$\geq \left(\int_{-\infty}^{+\infty} 1 d \mu_x\right)^{1-2/p} \int_{-\infty}^{+\infty} x^2 d \mu_x,$$  \hfill (12)

hence the result follows.

Now we are ready to prove our main theorem.

**Proof of Theorem 2** First, $\tilde{v}_t(i)$ is an unbiased estimator of $A_i x_t$ as

$$\mathbb{E}[	ilde{v}_t(i)] = \sum_{j=1}^d \frac{x_t(j)^q}{\|x_t\|_q^q} \cdot A_i(j) \|x_t\|_q^q = A_i x_t.$$  \hfill (13)

Furthermore,

$$\mathbb{E}[|\tilde{v}_t(i)|^p] = \sum_{j=1}^d \frac{x_t(j)^q}{\|x_t\|_q^q} \cdot |A_i(j)|^p \|x_t\|_q^q = |A_i|^p \|x_t\|_q^p \leq 1,$$  \hfill (14)

The proof follows from the proofs of Lemmas 2.3, 2.4, 2.5, and 2.6 in Section 2 and Appendix B of Clarkson, Hazan, and Woodruff (2012), with only small modifications to fit our new parameter choices. For instance, the original statement requires that $\eta \geq \sqrt{\frac{\log n}{12T}}$, but the proofs actually work for $\eta \geq \sqrt{\frac{11 \log n}{12T}}$. 

---

4The proof follows from the proofs of Lemmas 2.3, 2.4, 2.5, and 2.6 in Section 2 and Appendix B of Clarkson, Hazan, and Woodruff (2012), with only small modifications to fit our new parameter choices. For instance, the original statement requires that $\eta \geq \sqrt{\frac{\log n}{12T}}$, but the proofs actually work for $\eta \geq \sqrt{\frac{11 \log n}{12T}}$. 

---

---

---
where the second equality follows from the identities $q = p(q - 1)$ and $p = q(p - 1)$, and the last inequality follows from the assumption that $A_i \in \mathbb{B}_d^n \forall i \in [n]$. By Lemma 3, $\mathbb{E}[v_t(i)^2] \leq 1$. Because the clip function in Line 9 only makes the variance smaller, this means that the conditions of Lemma 2 are satisfied and hence we have (9), rewritten below:

$$\sum_{t \in [T]} p_t^i v_t \leq \min_{i \in [n]} \sum_{t \in [T]} v_t(i) + \eta \sum_{t \in [T]} p_t^i v_t^2 + \frac{\log n}{\eta}. \quad (15)$$

Furthermore, Lemma 2 implies that with probability $5/7 - O(1/n)$ we have

$$\sum_{t \in [T]} A_i x_t \leq \min_{i \in [n]} \sum_{t \in [T]} v_t(i) + 17\eta T + \frac{\log n}{\eta}. \quad (16)$$

Moreover, (10) gives $\sum_{t \in [T]} [v_t(i) - A_i x_t] \leq 4\eta T$, and hence $\min_{i \in [n]} \sum_{t \in [T]} A_i x_t$. Plugging this into (16), we have

$$\sum_{t \in [T]} A_i x_t \leq \min_{i \in [n]} \sum_{t \in [T]} p_t^i v_t + 10\eta T \leq \min_{i \in [n]} \sum_{t \in [T]} A_i x_t + 21\eta T + \frac{\log n}{\eta}, \quad (17)$$

with probability $(5/7 - O(1/n)) \cdot (1 - O(1/n)) \geq 2/3$. On the other hand, by taking $u_t = A_i$ in Proposition 1, we have

$$T\sigma \leq \max_{x \in \mathbb{B}_d^n} \sum_{t=1}^{T} A_i x \leq \sum_{t=1}^{T} A_i x_t + \sqrt{2T} p, \quad (18)$$

since $\frac{1}{q_T} = \frac{\sqrt{d}}{d} \leq p$. Combining (17) and (18), we have

$$\min_{i \in [n]} \sum_{t \in [T]} A_i x_t \geq T\sigma - \sqrt{2T} p - 21\eta T - \frac{\log n}{\eta}. \quad (19)$$

Consequently, the return $\bar{x} = \frac{1}{T} \sum_{t=1}^{T} x_t$ of Algorithm 1 in Line 11 satisfies

$$\min_{i \in [n]} A_i \bar{x} \geq \sigma - \sqrt{\frac{2p}{T}} - 21\eta - \frac{\log n}{\eta}. \quad (20)$$

To prove (7), it remains to show that $\sqrt{\frac{2p}{T}} + \frac{\log n}{\eta T} \leq \epsilon$, which is equivalent to $\sqrt{\frac{2p}{T}} + \sqrt{\frac{11 \log n}{12}} + \sqrt{\frac{12 \log n}{11}} \leq \sqrt{T} \epsilon$ by the definition of $\eta$. This is true because the AM-GM inequality implies that LHS is at most $2(\sqrt{\frac{2p}{T}})^2 + 2(2\sqrt{\frac{11 \log n}{12}} + \sqrt{\frac{12 \log n}{11}})^2 \leq 4p + 895 \log n \leq T\epsilon^2$. \hfill \Box

Lemma 1 combined with Theorem 2 yields the classical result in Theorem 1.

A Sublinear Quantum Algorithm for General Matrix Games

In this section, we give a quantum algorithm for solving the general $\ell_q$-$\ell_1$ matrix games. It closely follows our classical algorithm because they both use a primal-dual approach, where

the primal part is composed of $p$-norm online gradient descent and the dual part is composed of multiplicative weight updates. However, we adopt quantum techniques to achieve speedup on both.

The intuition behind the quantum algorithm and the quantum speedup is that we measure quantum states to obtain random samples. These quantum states can be efficiently prepared (with cost $O(\sqrt{n})$ and $O(\sqrt{d})$). Mathematically, a quantum state can be represented by an $\ell_2$-normalized complex vector $\psi$ in the sense that measuring this quantum states yields outcome $i$ with probability $|\psi(i)|^2$ (thus for every probability distribution there is a quantum state corresponding to it). Let us denote the quantum state for sampling from $w$ by $|w\rangle$ and the quantum state for sampling from $x$ by $|x\rangle$ (different from the notation in Algorithm 3). If we can maintain $|w\rangle$ and $|x\rangle$ in each iteration, then there is no need for classical updates, and preparing $|w\rangle$ and $|x\rangle$ becomes the bottleneck of the quantum algorithm.

The source of our quantum speedup comes from an important subroutine, Algorithm 2, which is designed to prepare states for $\ell_q$-sampling. It uses standard Grover-based techniques to prepare states but we carefully keep track of the normalizing factor to facilitate $\ell_q$-sampling. We showed in Proposition 2 in the supplementary material that preparing $|w\rangle$ costs $O(\sqrt{n})$ and preparing $|x\rangle$ costs $O(\sqrt{d})$. In the following, we give the high-level ideas of Algorithm 2.

1. We first create a quantum state corresponding to the uniform distribution, which is easy using Hadamard gates.

2. For each entry, we create a state with the desired amplitude associated with 0, and an undesired amplitude associated with 1 (the unitarity of quantum operations necessitates the existence of this undesired term).

3. Finally we use a technique called amplitude amplification to amplify the portion of the state corresponding to 0 for each entry, to get a state with only the desired amplitudes.

The details of our quantum algorithm for solving the general $\ell_q$-$\ell_1$ matrix games, Algorithm 3, is rather technical. To simplify the presentation, we postpone its pseudocode (Algorithm 3) to the supplementary material and highlight how it is different from Algorithm 1 in the following.

- For the primal part, we prepare a quantum state $|y_t\rangle$ for the $q$-norm OGD and measure it in Line 7 to obtain a sample $j_t \in [d]$. The subtlety here is that we need to perform the $\ell_q$-sampling to the vector $y_t$: this is different from the $\ell_2$-sampling in Li, Chakrabarti, and Wu (2019) which uses the fact that pure quantum states are $\ell_2$-normalized. To this end, we design Algorithm 2 for $\ell_q$-quantum state sampling, which may be of independent interest; this algorithm is built upon a clever use of quantum amplitude amplification, the technique behind the Grover search (Grover (1996)). Note that sampling according to $y_t$ is equivalent to sampling according to $x_t$ in Algorithm 1 because $x_t(j) y_t^q = y_t(j)^q / \|y_t\|_q$. Moreover, it suffices to replace $|x_t||\|x_t||_q^{-1}$ with $||y_t||_q^{-1} / \|y_t(j_t)||^{-1}$ max{1, $||y_t||_q$} in Line 8 of Algorithm 3. Similar to preparing $|p_t\rangle$, we use $O(\sqrt{d})$ queries
Algorithm 2: Prepare an $\ell_p$-pure state given an oracle to its coefficients.

1. Apply the minimum finding algorithm (Dür and Høyer 1996) to find $a_q[n] := \max_{i \in [n]} |a_i|^q$ in $O(\sqrt{n})$ time;
2. Prepare the uniform superposition $\frac{1}{\sqrt{n}} \sum_{i \in [n]} |i\rangle$;
3. Perform the following unitary transformations:
   
   $$\frac{1}{\sqrt{n}} \sum_{i \in [n]} |i\rangle \xrightarrow{O_A} \frac{1}{\sqrt{n}} \sum_{i \in [n]} |i\rangle |a_i\rangle$$
   
   $$\xrightarrow{\sum_{i \in [n]} |i\rangle |a_i\rangle \left( \frac{a_i^{q/2}}{\|a\|_q^2} |0\rangle + \sqrt{1 - \frac{|a_i|^q}{\|a\|_q^2}} |1\rangle \right) \xrightarrow{OA} \frac{1}{\sqrt{n}} \sum_{i \in [n]} |i\rangle |0\rangle \left( \frac{a_i^{q/2}}{\|a\|_q^2} |0\rangle + \sqrt{1 - \frac{|a_i|^q}{\|a\|_q^2}} |1\rangle \right);$$
4. Discard the second register above and rewrite the state as
   
   $$\left\lVert a\right\rVert_q^{q/2} \left( \frac{1}{\|a\|_q^{q/2}} \sum_{i \in [n]} a_i^{q/2} |i\rangle \right) |0\rangle + |a^-\rangle |1\rangle,$$
   
   where $|a^-\rangle := \frac{1}{\sqrt{n}} \sum_{i \in [n]} \sqrt{1 - \frac{|a_i|^q}{\|a\|_q^2}} |i\rangle$;
5. Apply amplitude amplification (Brassard et al. 2002) for the state in (21) conditioned on the second register being 0. Return the output.

To $O_A$ to prepare $y_t$, while classically we need to compute all the entries of $y_t$, which takes $O(d)$ queries.

- For the dual part, we prepare the multiplicative weight vector as a quantum state $|p_t\rangle$ and measure it (in Line 3) to obtain a sample $i_t \in [n]$. This adaption enables us to achieve the $O(\sqrt{n})$ dependence by using quantum amplitude amplification in the quantum state preparation: in Line 8 we implement the oracle $O_t$ and in Line 9 we use $O(\sqrt{n})$ queries to $O_t$ to prepare the state $|p_{t+1}\rangle$ for the next iteration. In contrast, classically we need to compute all the entries of $w_{t+1}$ to obtain the probability distribution $p_{t+1}$ for the next iteration, which takes $O(n)$ queries.

In general, Algorithm 3 can be viewed as a template for achieving quantum speedups for online mirror descent methods: in this work, we focus on the general matrix games where the primal and dual are in the special relationship of $\ell_p$ and $\ell_q$ norms, but in principle it may be applicable to study other dualities in online learning.

We summarize the main quantum result as the following theorem, which states the correctness and time complexity of Algorithm 3. The relevant technical proofs are deferred to the supplementary material.

Theorem 3. Algorithm 3 returns a succinct classical repre-
classical algorithm with running time $O\left((n+d)(p+\log n)^2\right)$ or a quantum algorithm with running time $\tilde{O}\left(\sqrt{\frac{n}{\epsilon}} + \frac{\sqrt{d}}{\epsilon}\right)$.

We can also assume $p = O\left((\log d/\epsilon)^2\right)$ (Lemma 1) and result in running time $O\left((n+d)(\log d/\epsilon+\log n)\right)$ and $\tilde{O}\left(\sqrt{\frac{n}{\epsilon}} + \frac{\sqrt{d}}{\epsilon}\right)$, respectively.

Proof. We denote the matrix $V := (v_1; v_2; \ldots; v_n)$ where $v_i$ is the $i$th element in $S$. Note that the approximate Carathéodory problem can be formed as

$$\min_{p\in\Delta_n} \max_{x\in\mathbb{R}^d} (p^\top V - u^\top)x.$$  

We denote $U = (u; u; \ldots; u) \in \mathbb{R}^{n \times d}$, i.e., all the $n$ rows of $U$ are $u$. Then we have $(p^\top V - u^\top)x = 2p^\top x - 2Ux$. Furthermore, since $u, v_i \in \mathbb{R}^d$ for all $i \in [n]$, each row of $2U$ is also in $u, v_i \in \mathbb{R}^d$. Finally, by the Sion’s Theorem (Sion 1958) we can switch the order of the min and max in (23).

In all, to solve the approximate Carathéodory problem with precision $\epsilon$, it suffices to solve the maximin game

$$\max_{p\in\Delta_n} \min_{x\in\mathbb{R}^d} p^\top V - U/2.$$  

with precision $\frac{\epsilon}{2}$. This is exactly (6), thus the result follows from Theorem 2 and Theorem 3.

Compared to Mirrokni et al. (2017), we pay a $\log n$ overhead in the cardinality of the convex combination, but in time complexity the dominating term $nd$ is significantly improved to $n + d$. We also give the first sublinear quantum algorithm. Note that as Mirrokni et al. (2017) pointed out, the approximate Carathéodory problem has wide applications in machine learning and optimization, including support vector machines (SVMs), rounding in polytopes, submodular function minimization, etc. We elaborate the details of SVMs below, and leave out the details of other applications as the reductions are direct.

### $\ell_q$-margin support vector machine (SVM)

When we solve the $\ell_q$-$\ell_1$ matrix game in Algorithm 1, we apply $\ell_q$-sampling where $j_i = j$ with probability $x(j_i)/\|x\|_q^q$ for any $j \in [d]$. The key reason of the success of Algorithm 1 is because the expectation of the random variable $A_j(j_i)\|x_i\|_q^q x(j_i)_{j_i}^{q-1}$ in Line 8 is $A_jx$, which is unbiased.

If we consider some alternate random variables, we can potentially solve a maximin game in $\ell_q$-$\ell_1$ norm with respect to some nonlinear functions of the matrix. A specific problem of significant interest is the $\ell_q$-margin support vector machine (SVM), where we are given data points $X_1, \ldots, X_n$ in $\mathbb{R}^d$ and a label vector $y \in \{1, -1\}^n$. The goal is to find a separating hyperplane $w \in \mathbb{R}^d$ of these data points with the largest margin under the $\ell_q$-norm loss, i.e.,

$$\sigma_{\text{SVM}} := \max_{w \in \mathbb{R}^d} \min_{x \in [n]} 2y_i \cdot X_i^\top w - \|w\|_q^q.$$  

Without loss of generality, we assume $y_i = 1$ for all $i \in [n]$, otherwise we take $X_i \leftarrow (-1)^{y_i} \cdot X_i$. In this case, the random variable $2X_i(j)\|w\|_q^q (w(j))^{q-1} - \|w\|_q^q$ is unbiased under $\ell_q$-sampling on $j$:

$$\mathbb{E} \left[ 2X_i(j)\|w\|_q^q (w(j))^{q-1} - \|w\|_q^q \right] = 2X_i^\top w - \|w\|_q^q.$$  

(26)

Note that $\sigma_{\text{SVM}} \geq 0$ since $2X_i^\top w - \|w\|_q^q \geq 0$ for all $i \in [n]$ when $w = 0$. For the case $\sigma_{\text{SVM}} > 0$ and taking $0 < \epsilon < \sigma_{\text{SVM}}$, similar to Theorem 2 and Theorem 3 we have:

**Corollary 2.** To return a vector $\bar{w} \in \mathbb{R}^d$ such that with probability at least $2/3$,

$$\min_{i\in[n]} 2X_i \bar{w} - \|\bar{w}\|_q^q \geq \sigma_{\text{SVM}} - \epsilon > 0,$$  

there is a classical algorithm that achieves this with $O\left(\frac{(n+d)(p+\log n)}{\epsilon^2}\right)$ time and a quantum algorithm that achieves this with $\tilde{O}\left(\frac{p^2}{\epsilon^2} + \frac{d}{\epsilon\sigma_{\text{SVM}}}\right)$ time.

We also give the first sublinear quantum algorithm for $\ell_q$-SVMs in Li, Chakrabarti, and Wu (2019). We essentially generalize their results to the $\ell_q$-norm cases based on our new general matrix game solvers in Theorem 2 and Theorem 3.

### Classical and Quantum Lower Bounds

For both our classical and quantum algorithms for general matrix games, we can prove matching classical and quantum lower bounds in $n$ and $d$ for constant $\epsilon$:

**Theorem 4.** Assume $0 < \epsilon < 0.04$. Then to return an $\bar{x} \in \mathbb{R}^d$ satisfying

$$A_j\bar{x} \geq \max_{x \in \mathbb{R}^d} \min_{x \in [n]} A_jx - \epsilon \quad \forall j \in [n],$$  

with probability at least $2/3$, we need $\Omega(n + d)$ classical queries or $\Omega(\sqrt{n} + \sqrt{d})$ quantum queries.

Due to the space limitation, we postpone the proof details of Theorem 4 to the supplementary material.

### Conclusions

We give sublinear algorithms for solving general $\ell_q$-$\ell_1$ matrix games for any $q \in (1, 2]$. Our classical and algorithms run in time $O\left((n+d)(p+\log n)\right)$ and $\tilde{O}\left(\frac{p^2}{\epsilon^2} + \frac{d}{\epsilon\sigma_{\text{SVM}}}\right)$, respectively; both bounds are tight up to poly-logarithmic factors in $n$ and $d$. Our results can be applied to solve the approximate Carathéodory problem and the $\ell_q$-margin SVMs.

Our paper raises a couple of natural open questions for future work. For instance:

- Can we give sublinear algorithms for $\ell_q$-$\ell_1$ matrix games where $p > 2$? Technically, this will probably require a $q^\theta$ moment multiplicative weight lemma to replace Lemma 2.
- Can we give quantum algorithms that achieve speedup of variance-reduced methods for solving matrix games, such as the state-of-the-art result in Carmon et al. (2019)?
Ethics Statement
This work is purely theoretical. Researchers working on learning theory and quantum computing may benefit from our results. In the long term, once fault-tolerant quantum computers have been built, our results may find practical applications in matrix game scenarios arising in the real world. As far as we are aware, our work does not have immediate negative ethical impact.

Acknowledgements
TL thanks Adrian Vladu for many helpful discussions, as well as Yair Carmon for the discussions about his paper (Carmon et al. 2019). TL was supported by an IBM PhD Fellowship, an QISE-NET Triplet Award (NSF grant DMR-1747426), the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, Quantum Algorithms Teams program, ARO contract W911NF-17-1-0433, and NSF grant PHY-1818914. CW was supported by Scott Aaronson’s Vannevar Bush Faculty Fellowship. SC and XW were partially supported by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, Quantum Algorithms Team program, and were also partially supported by the U.S. National Science Foundation grant CCF-1755800, CCF-1816695, and CCF-1942837 (CAREER).

References
Aaronson, S. 2015. Read the fine print. Nature Physics 11(4): 291.
Apelidoorn, J. v.; and Gilyén, A. 2019. Quantum algorithms for zero-sum games. arXiv:1904.03180
Arute et al., F. 2019. Quantum supremacy using a programmable superconducting processor. Nature 574(7779): 505–510. arXiv:1910.11833
Barman, S. 2015. Approximating Nash equilibria and dense bipartite subgraphs via an approximate version of Carathéodory theorem. In Proceedings of the 47th Annual ACM Symposium on Theory of Computing, 361–369. arXiv:1406.2296
Bennett, C. H.; Bernstein, E.; Brassard, G.; and Vazirani, U. 1997. Strengths and weaknesses of quantum computing. SIAM Journal on Computing 26(5): 1510–1523. arXiv:quant-ph/9701001
Brassard, G.; Höyer, P.; Mosca, M.; and Tapp, A. 2002. Quantum amplitude amplification and estimation. Contemporary Mathematics 305: 53–74. arXiv:quant-ph/0005055
Carmon, Y.; Jin, Y.; Sidford, A.; and Tian, K. 2019. Variance reduction for matrix games. In Advances in Neural Information Processing Systems, 11377–11388. arXiv:1907.02056
Clarkson, K. L.; Hazan, E.; and Woodruff, D. P. 2012. Sublinear optimization for machine learning. Journal of the ACM (JACM) 59(5): 23. arXiv:1010.4408
Cobetess, C. W.; and Pokutta, S. 2019. Revisiting the Approximate Carathéodory Problem via the Frank-Wolfe Algorithm. arXiv:1911.04415
Dantzig, G. B. 1998. Linear programming and extensions, volume 48. Princeton University Press.
Deng, N.; Tian, Y.; and Zhang, C. 2012. Support vector machines: optimization based theory, algorithms, and extensions. CRC press.
Dür, C.; and Höyer, P. 1996. A quantum algorithm for finding the minimum. arXiv:quant-ph/9607014
Garber, D.; and Hazan, E. 2011. Approximating semidefinite programs in sublinear time. In Advances in Neural Information Processing Systems, 1080–1088.
Grigoriadis, M. D.; and Khachiyan, L. G. 1995. A sublinear-time randomized approximation algorithm for matrix games. Operations Research Letters 18(2): 53–58.
Grover, L. K. 1996. A fast quantum mechanical algorithm for database search. In Proceedings of the Twenty-eighth Annual ACM Symposium on Theory of Computing, 212–219. ACM. arXiv:quant-ph/9605043
Hazen, E.; Koren, T.; and Srebro, N. 2011. Beating SGD: Learning SVMs in sublinear time. In Advances in Neural Information Processing Systems, 1233–1241.
Johnson, R.; and Zhang, T. 2013. Accelerating stochastic gradient descent using predictive variance reduction. In Advances in Neural Information Processing Systems, 315–323.
Kapoor, A.; Wiebe, N.; and Svore, K. 2016. Quantum perceptron models. In Proceedings of the 30th Conference on Neural Information Processing Systems, 3999–4007. arXiv:1602.04799
Li, T.; Chakrabarti, S.; and Wu, X. 2019. Sublinear quantum algorithms for training linear and kernel-based classifiers. In Proceedings of the 36th International Conference on Machine Learning, 3815–3824. arXiv:1904.02276
Minsky, M.; and Papert, S. A. 1988. Perceptrons: An introduction to computational geometry. MIT Press.
Mirroknii, V.; Leme, R. P.; Vladu, A.; and Wong, S. C.-w. 2017. Tight bounds for approximate Carathéodory and beyond. In Proceedings of the 34th International Conference on Machine Learning, 2440–2448. arXiv:1512.08602
Nemirovski, A. 2004. Prox-method with rate of convergence $O(1/t)$ for variational inequalities with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems. SIAM Journal on Optimization 15(1): 229–251.
Nemirovski, A. S.; and Yudin, D. B. 1983. Problem complexity and method efficiency in optimization.
Nesterov, Y. 2007. Dual extrapolation and its applications to solving variational inequalities and related problems. Mathematical Programming 109(2-3): 319–344.
Neumann, J. v. 1928. Zur theorie der gesellschaftsspiele. Mathematische Annalen 100(1): 295–320.
Novikoff, A. B. 1963. On convergence proofs for perceptron models. In Advances in Neural Information Processing Systems, 291–292.
Palaniappan, B.; and Bach, F. 2016. Stochastic variance reduction methods for saddle-point problems. In Advances in Neural Information Processing Systems, 1233–1241.
Shawe-Taylor, J.; and Cristianini, N. 2004. Kernel methods for pattern analysis. Cambridge University Press.
Preskill, J. 2018. Quantum Computing in the NISQ era and beyond. *Quantum* 2: 79. ISSN 2521-327X. [arXiv:1801.00862](https://arxiv.org/abs/1801.00862)

Shalev-Shwartz, S. 2012. Online learning and online convex optimization. *Foundations and Trends® in Machine Learning* 4(2): 107–194.

Shi, Z.; Zhang, X.; and Yu, Y. 2017. Bregman divergence for stochastic variance reduction: saddle-point and adversarial prediction. In *Advances in Neural Information Processing Systems*, 6031–6041.

Sion, M. 1958. On general minimax theorems. *Pacific Journal of Mathematics* 8(1): 171–176.

Zinkevich, M. 2003. Online convex programming and generalized infinitesimal gradient ascent. In *Proceedings of the 20th International Conference on Machine Learning*, 928–936.
Sublinear Quantum Algorithm for General Matrix Games: Proof Details

We first give the details of our quantum algorithm.

**Algorithm 3:** A sublinear quantum algorithm for \( \ell_q-\ell_1 \) matrix games.

**Input:** \( \epsilon > 0; \ p \in [2, +\infty), \ q \in (1, 2] \) such that \( \frac{1}{p} + \frac{1}{q} = 1; \ A \in \mathbb{R}^{n \times d} \) with \( A_i \in \mathbb{R}_p^d \) for all \( i \in [n] \).

**Output:** \( \bar{x} \) that satisfies (7).

1. Let \( T = \left\lceil \frac{1346 \log n + 4p}{\epsilon^2} \right\rceil \), \( y_1 = 0_d \), \( \eta = \sqrt{\frac{11 \log n}{12T}} \), \( w_1 = 1_n \), \( |p_1| = \frac{1}{\sqrt{n}} \sum_{i \in [n]} |i| \);

2. for \( t = 1 \) to \( T \) do
   1. Measure the state \( |p_t\rangle \) and denote the output basis as \( i_t \in [n] \);
   2. For each \( i \in [t] \), estimate \( \|A_i\|_p^q \) by **Lemma 4** with precision \( \delta = \eta^2 \). Output:= \( \|A_i\|_p^q \);
   3. Define \( y_{t+1} \) by \( y_{t+1,j} \leftarrow y_t + \frac{1+\frac{1}{q} \text{sgn}(A_{it,j}) |A_{it,j}|^{p-1}}{\|A_t\|_p^q} \) for all \( j \in [d] \);
   4. Apply **Lemma 4** to \( \left\lceil \log T \right\rceil \) times to estimate \( \|y_t\|_q^q \) with precision \( \eta^2 \), and take the median of the \( 2\left\lceil \log T \right\rceil \) outputs, denoted by \( \|y_t\|_q^q \);
   5. Choose \( j_t \in [d] \) by \( j_t = j \) with probability \( y_t(j)^q / \|y_t\|_q^q \), which is achieved by applying **Algorithm 2** to prepare the quantum state \( |y_t\rangle \) and measure in the computational basis;
   6. For all \( i \in [n] \), denote \( \tilde{v}_t(i) := A_i(j_t) |y_t\|_q^q / (y_t(j_t)^q - \max \{1, \|y_t\|_q^q \}) \), \( v_t(i) = \text{clip}(\tilde{v}_t(i), \eta^2, \eta) \), and \( u_{t+1}(i) = u_t(i)(1 - \eta v_t(i)) + \eta^2 v_t(i)^2 \). Prepare an oracle \( O_t \) such that \( O_t(|i\rangle|0\rangle) = |i\rangle|u_{t+1}(i)\rangle \) for all \( i \in [n] \), using \( 2t \) queries to \( O_a \) and \( O_t \) additional arithmetic computations;
   7. Prepare \( |p_{t+1}\rangle = \frac{1}{\sqrt{\|u_{t+1}\|_q^q}} \sum_{i \in [n]} u_{t+1}(i) |i\rangle \) using **Algorithm 2** (with \( q = 2 \) therein) and \( O_t \);
   8. Return \( \bar{x} = \frac{1}{T} \sum_{t=1}^T u_{t+1}(i) |i\rangle \).

We need the following lemma to estimate the norm of a vector:

**Lemma 4 (Li, Chakrabarti, and Wu 2019, Lemma 6).** Given a function \( F : [d] \rightarrow [0, 1] \) with a quantum oracle \( O_F : |i\rangle|0\rangle \rightarrow |i\rangle|F(i)\rangle \) for all \( i \in [d] \), let \( m = \frac{1}{d} \sum_{i=1}^d F(i) \). Then for any \( \delta < 0 \), there is a quantum algorithm that uses \( O(\sqrt{d}/\delta) \) queries to \( O_F \) and returns an \( \bar{m} \) such that \( |m - \bar{m}| \leq 6m \delta \) with probability at least 2/3.

We use the procedure below for preparing a quantum state given an oracle to a power of its coefficients:

**Proposition 2.** Assume that \( a \in \mathbb{C}^n \), and we are given a unitary oracle \( O_a \) such that \( O_a(|i\rangle|0\rangle) = |i\rangle|a_i\rangle \) for all \( i \in [n] \). Then **Algorithm 2** takes \( O(\sqrt{n}) \) calls to \( O_a \) for preparing the quantum state \( \frac{1}{\|a\|_q^{q/2}} \sum_{i \in [n]} a_i^{q/2} |i\rangle \) with success probability \( 1 - O(1/n) \).

**Proof.** Note that **Algorithm 2** of Li, Chakrabarti, and Wu (2019) had given a quantum algorithm for preparing an \( \ell_q \)-norm pure state given an oracle to its coefficients, and **Algorithm 2** essentially generalize this result to the \( \ell_q \)-norm case by replacing all \( a_i^{q/2} \) as in **Algorithm 2**. Note that the coefficient in (21) satisfies \( \|a\|_q^{q/2} \geq \frac{1}{\sqrt{n}} \). As a result, applying amplitude amplification for \( O(\sqrt{n}) \) times indeed promises that we obtain 0 in the second system with success probability \( 1 - O(1/n) \), i.e., the state \( \frac{1}{\|a\|_q^{q/2}} \sum_{i \in [n]} a_i^{q/2} |i\rangle \) is prepared.

We need the following lemma.

**Lemma 5.** For all \( i \in [n] \), Define

\[
\hat{v}_{t,\text{approx}}(i) := \frac{A_i(j_t) \|y_t\|_q^q}{y_t(j_t)^{q-1} \max \{1, \|y_t\|_q^q \}}; \quad \hat{v}_{t,\text{true}}(i) := \frac{A_i(j_t) \|y_t\|_q^q}{y_t(j_t)^{q-1} \max \{1, \|y_t\|_q^q \}}.
\]

where \( \|y_t\|_q^q \) and \( \|y_t\|_q^q \) satisfy

\[
\|y_t\|_q^q - \|y_t\|_q^q \leq \delta \|y_t\|_q^q
\]

\(^4\text{Here we do not write down the whole vector } y_{t+1}, \text{ but we construct any query to its entries in } O(1) \text{ time.}\)
with probability at least $1 - o(1)$. Also assume that $	ilde{v}_{t, \text{approx}}(i), \tilde{v}_{t, \text{true}}(i) \leq 1/\eta$. Then, it holds that for all $i \in [n]$,

$$|\tilde{v}_{t, \text{approx}}(i) - \tilde{v}_{t, \text{true}}(i)| \leq \frac{\delta}{\eta} \quad \forall i \in [n],$$  

(31)

with probability at least $1 - o(1)$.

**Proof.** First note that

$$|\tilde{v}_{t, \text{approx}}(i) - \tilde{v}_{t, \text{true}}(i)| = \frac{\tilde{v}_{t, \text{approx}}(i)}{\tilde{v}_{t, \text{true}}(i)}|\tilde{v}_{t, \text{approx}}(i)| - 1 \leq \frac{1}{\eta} |\tilde{v}_{t, \text{approx}}(i)| - 1.$$

(32)

When $\|y_t\|_q \geq 1$, we have $|\tilde{v}_{t, \text{approx}}(i)| = \|y_t\|_q^{-1}$, and when $\|y_t\|_q \leq 1$, we have $|\tilde{v}_{t, \text{approx}}(i)| = \|y_t\|_q^q$. By assumption, with probability at least $1 - o(1)$, it holds that $\|\tilde{v}_{t, \text{approx}}(i)\| - 1 \leq \delta$. Since $1 \leq \|y_t\|_q^{-1} \leq \|y_t\|_q^q$ when $\|y_t\|_q \geq \|y_t\|_q$, and $1 \geq \|\tilde{v}_{t, \text{approx}}(i)\| - 1 \leq \|\tilde{v}_{t, \text{approx}}(i)\| - 1 \leq \delta$. Putting this into (32), we have the desired inequality. \qed

Now, we are ready to prove the main quantum result.

**Proof of Theorem 3.** First note that in Line 4, we use an estimation $\tilde{\|A_i\|}_p$ of $\|A_i\|_p$ with relative error at most $\delta$. Then in Line 5 $\|A_i\|_p^{p-2}$ is an estimation of $\|A_i\|_p^{p-2}$ with relative error at most $\delta$ because $p \geq 2$ and $\|A_i\|_p^{p-2} = (\|A_i\|_p^p)^{(p-2)/p}$.

Hence, $y_{t+1}$ has a relative error of at most $\delta$ compared to its true value defined by

$$y_t + \sqrt{\frac{q - 1}{2T} \frac{\text{sgn}(A_{i_{t+1}})|A_{i_{t+1}}|^{p-1}}{\|A_i\|_p^{p-2}}}.$$

(33)

Consider Line 6. The estimate $\|y_t\|_q$ is the median of $2\lceil\log T\rceil$ executions of Lemma 4. It implies that, with failure probability at most $1 - (2/3)^{2\lceil\log T\rceil} = 1 - T^{-2}$, (30) holds. Since there are $T$ iterations in total, the probability that (30) holds is at least $1 - T \cdot O(1/T^2) = 1 - o(1)$. Also consider (29). It is easy to see that $\tilde{v}_{t, \text{approx}}(i), \tilde{v}_{t, \text{true}}(i) \leq 1/\eta$ because of Line 8. Therefore, the conditions of Lemma 5 hold and its result follows.

As $\delta = \eta^2$, by Lemma 5 and Lemma 2, we have that with probability at least $5/7 - O(1/n)$,

$$\sum_{t \in [T]} A_i x_t \leq \sum_{t \in [T]} p_t v_t + 11\eta T \leq \min_{i \in [n]} \sum_{t \in [T]} v_t(i) + 21\eta T + \frac{\log n}{\eta}. $$

(34)

Moreover, by Lemma 5 and Eq. (10), we have $\min_{i \in [n]} \sum_{t \in [T]} v_t(i) \leq 4\eta T + \eta T + \min_{i \in [n]} \sum_{t \in [T]} A_i x_t$. Plugging this into (34), we have

$$\sum_{t \in [T]} A_i x_t \leq \sum_{t \in [T]} p_t v_t + 11\eta T \leq \min_{i \in [n]} \sum_{t \in [T]} A_i x_t + 26\eta T + \frac{\log n}{\eta}.$$

(35)

with probability $(5/7 - O(1/n)) \cdot (1 - O(1/n)) \geq 2/3$.

Similar to the proof of Theorem 2, we have

$$\min_{i \in [n]} A_i x \geq \sigma - \sqrt{\frac{2p}{T} - 26\eta - \frac{\log n}{\eta T}}. $$

(36)

By the choices of $p$ and $\eta$ in Algorithm 3, the desired error bound for (22) holds because

$$\left(\frac{2p}{T} + 26\eta + \frac{\log n}{\eta T}\right)^2 \leq 2 \left(\frac{2p}{T}\right)^2 + 2 \left(26\eta + \frac{\log n}{\eta T}\right)^2 \leq \frac{4p + 1346 \log n}{T} \leq \epsilon^2,$$

(37)

where the first inequality follows from the AM-GM inequality and the last inequality follows from the choice of $T$ in Algorithm 3. Now, we analyze the time complexity. In Line 4 of Algorithm 3, the number of queries to $O_A$ for Lemma 4 is $O(\sqrt{\delta/\epsilon}) = \tilde{O}(p\sqrt{\delta/\epsilon^2})$. In Line 5, we have

$$y_{t,j} = \sqrt{\frac{q - 1}{2T} \sum_{\tau=1}^{T} \frac{\text{sgn}(A_{i_{t,j}})|A_{i_{t,j}}|^{p-1}}{\|A_i\|_p^{p-2}}}.$$

(38)
An oracle for \( y_t \) can be implemented with \( \tilde{O}(p/\epsilon^2) \) queries to \( O_A \). To estimate \( \|y_t\|_q \), we first need to normalize \( y_t \). The summand in (43) is in the range \([-1, 1]\); to see this, note that

\[
\frac{|A_{t,j}|^{p-1}}{||A_{i,\cdot}||_p^{p-2}} \leq \frac{|A_{t,j}|^{p-1}}{(|A_{i,j}|^p)^{(p-2)/p}} = |A_{t,j}| \leq 1.
\] (39)

Therefore, \( y_{t,j} = \tilde{O}(\sqrt{p}/\epsilon) = \tilde{O}(\sqrt{p}/\epsilon) \). Since the precision is \( \delta = \eta^2 = \tilde{O}(\epsilon^2/p) \), the cost for amplitude estimation is \( \tilde{O}(p\sqrt{d}/\epsilon^2) \). Finally, there are \( T = \tilde{O}(p/\epsilon^2) \) iterations in total. The total complexity in Line 5 is

\[
\tilde{O}
\left( \frac{p}{\epsilon^2} \right) \cdot \tilde{O}
\left( \frac{\sqrt{p}}{\epsilon} \right) \cdot \tilde{O}
\left( \frac{p\sqrt{d}}{\epsilon^2} \right) \cdot \tilde{O}
\left( \frac{p^3.5\sqrt{d}}{\epsilon^5} \right).
\] (40)

For Line 6 we need to prepare the state \(|y_t\rangle\). To simulate a query to an coefficient of \( y_t \), we need \( \tilde{O}(p/\epsilon^2) \) queries to \( O_A \). The query complexity for Algorithm 2 is \( O(\sqrt{d}) \), and there are \( T = \tilde{O}(p/\epsilon) \) iterations in total. The total complexity in Line 6 is

\[
\tilde{O}
\left( \frac{p}{\epsilon^2} \right) \cdot O(\sqrt{d}) \cdot \tilde{O}
\left( \frac{p}{\epsilon^2} \right) = \tilde{O}
\left( \frac{p^2\sqrt{d}}{\epsilon^4} \right),
\] (41)

which is dominated by (40).

For Line 8 we need to implement one query to \( O_A \); we need \( 2t \) queries to \( O_A \) with \( \tilde{O}(t) \) additional arithmetic computations. For Line 9 to prepare the state \(|y_{t+1}\rangle\), we need \( O(\sqrt{n}) \) queries to \( O_A \), which can be implemented by \( O(\sqrt{nt}) \) queries to \( O_A \) by Line 8 and \( O(\sqrt{nt}) \) additional arithmetic computations. Therefore, the total complexity for Line 9 is

\[
\sum_{t=1}^{T} \tilde{O}(\sqrt{nt}) = \tilde{O}(\sqrt{n}T^2) = \tilde{O}
\left( \frac{p^2\sqrt{n}}{\epsilon^4} \right).
\] (42)

The time complexity of this algorithm is established by (40) and (42).

Finally, \( \bar{x} \) has a succinct classical representation: using \( i_1, \ldots, i_r \) obtained from Line 3 and \( ||y_1||_q, \ldots, ||y_T||_q \) obtained from Line 6, a coordinate of \( \bar{x} \) can be restored in time \( T = \tilde{O}(p/\epsilon^2) \).

**Classical and Quantum Lower Bounds**

Recall that the input of the general matrix game is a matrix \( A \in \mathbb{R}^{n \times d} \) such that \( A_i \in \mathbb{B}_p^d \) for all \( i \in [n] \) (\( A_i \) being the \( i \)-th row of \( A \)), and the goal is to approximately solve

\[
\sigma := \max_{p \in \Delta_n} \min_{x \in \mathbb{B}_p^d} p^T A x,
\] (43)

where \( p \in [2, +\infty), q \in [1, 2], \) and \( \frac{1}{p} + \frac{1}{q} = 1 \). Classically, we are given an oracle that inputs \( i \in [n], j \in [d] \) and outputs \( A_{ij} \); our sublinear classical algorithm in Theorem 2 solves the general matrix game (43) in \( O(\frac{n + d}{\epsilon^2 \log n}) \) time. Quantumly, we are given the quantum oracle \( O_A \) such that \( O_A(i)[j][0] = |i\rangle[j]\langle j|A_{ij}| \forall i \in [n], j \in [d] | \); our quantum algorithm in Theorem 3 solves the general matrix game (43) in \( \tilde{O}(\frac{\epsilon^2}{\sqrt{n}} + \frac{\epsilon^{3.5}}{\sqrt{d}}) \) time. We prove matching classical and quantum lower bounds in \( n \) and \( d \) for constant \( \epsilon \) and \( p \):

**Theorem 5.** Assume \( 0 < \epsilon < 0.04 \). Then to return an \( \bar{x} \in \mathbb{B}_p^d \) satisfying

\[
A_j \bar{x} \geq \max_{x \in \mathbb{B}_p^d} A_i x - \epsilon \quad \forall j \in [n]
\] (44)

with probability at least \( 2/3 \), we need \( O(n + d) \) classical queries or \( \Omega(\sqrt{n} + \sqrt{d}) \) quantum queries.

The proof of Theorem 5 is inspired by Li, Chakrabarti, and Wu (2019), but for the \( \ell_q-\ell_1 \) matrix game the construction is different and the analysis is more intricate as seen below.

**Proof.** Assume we are given the promise that \( A \) is from one of the two cases below:

1. There exists an \( l \in \{2, \ldots, d\} \) such that \( A_{11} = -\frac{1}{21^{1/p}}, A_{1l} = \frac{1}{21^{1/p}}; A_{21} = A_{2l} = \frac{1}{21^{1/p}} \); there exists a unique \( k \in \{3, \ldots, n\} \) such that \( A_{k1} = 1, A_{kj} = 0; A_{ij} = \frac{1}{21^{1/p}} \) for all \( i \in \{3, \ldots, n\}/\{k\}, j \in \{1, l\}, \) and \( A_{ij} = 0 \) for all \( i \in [n], j \notin \{1, l\} \).
2. There exists an \( l \in \{2, \ldots, d\} \) such that \( A_{11} = -\frac{1}{21^{1/p}}, A_{1l} = \frac{1}{21^{1/p}}; A_{21} = A_{2l} = \frac{1}{21^{1/p}} \); \( A_{ij} = \frac{1}{21^{1/p}} \) for all \( i \in \{3, \ldots, n\}, \) \( j \in \{1, l\}, \) and \( A_{ij} = 0 \) for all \( i \in [n], j \notin \{1, l\} \).
Notice that the only difference between these two cases is a row where the first entry is 1 and the $l$th entry is 0; they have the following pictures, respectively.

$$\begin{align*}
\text{Case 1: } A &= \begin{pmatrix}
-\frac{1}{2^{1/p}} & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\
\frac{1}{2^{1/p}} & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \ddots & \vdots \\
\frac{1}{2^{1/p}} & 0 & \cdots & 0 & \frac{1}{2^{1/p}} & 0 & \cdots & 0 \\
0 & 1 & \cdots & 0 & 0 & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \ddots & \vdots \\
\frac{1}{2^{1/p}} & 0 & \cdots & 0 & \frac{1}{2^{1/p}} & 0 & \cdots & 0
\end{pmatrix}.
\end{align*}$$  \hfill (45)

$$\begin{align*}
\text{Case 2: } A &= \begin{pmatrix}
-\frac{1}{2^{1/p}} & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\
\frac{1}{2^{1/p}} & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \ddots & \vdots \\
\frac{1}{2^{1/p}} & 0 & \cdots & 0 & \frac{1}{2^{1/p}} & 0 & \cdots & 0 \\
0 & 1 & \cdots & 0 & 0 & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \ddots & \vdots \\
\frac{1}{2^{1/p}} & 0 & \cdots & 0 & \frac{1}{2^{1/p}} & 0 & \cdots & 0
\end{pmatrix}.
\end{align*}$$  \hfill (46)

We denote the maximin value in \((43)\) of these cases as $\sigma_1$ and $\sigma_2$, respectively. We have:

- $\sigma_2 = \frac{1}{2^{1/p}}$.

On the one hand, consider $\bar{x} = \bar{e}_l \in \mathbb{B}_d^q$ (the vector in $\mathbb{R}^d$ with the $l$th coordinate being 1 and all other coordinates being 0). Then $A_i \bar{x} = \frac{1}{2^{1/p}}$ for all $i \in [n]$, and hence $\sigma_2 \geq \min_{i \in [n]} A_i \bar{x} = \frac{1}{2^{1/p}}$. On the other hand, for any $x = (x_1, \ldots, x_d) \in \mathbb{B}_q^d$, we have

$$\min_{i \in [n]} A_i x = \min \left\{ -\frac{1}{2^{1/p}} x_1 + \frac{1}{2^{1/p}} x_l, \frac{1}{2^{1/p}} x_1 + \frac{1}{2^{1/p}} x_l \right\} \leq \frac{1}{2^{1/p}} x_l \leq \frac{1}{2^{1/p}},$$

where the first inequality comes from the fact that $\min\{a, b\} \leq \frac{a+b}{2}$ for all $a, b \in \mathbb{R}$ and the second inequality comes from the fact that $x \in \mathbb{B}_q^d$ and $|x_l| \leq 1$. As a result, $\sigma_2 = \max_{x \in \mathbb{B}_q^d} \min_{i \in [n]} A_i x \leq \frac{1}{2^{1/p}}$. In conclusion, we have $\sigma_2 = \frac{1}{2^{1/p}}$.

- $\sigma_1 = \frac{1}{(1+(2^{-1/q}+1))^q}$.

On the one hand, consider $\bar{x} = \frac{1}{(1+(2^{-1/q}+1))^q} \bar{e}_1 + \frac{2^{1-1/q}+1}{(1+(2^{-1/q}+1))^q} \bar{e}_l \in \mathbb{B}_d$. It can be seen that $\bar{x} \in \mathbb{B}_q^d$, moreover, since $\frac{1}{p} + \frac{1}{q} = 1$,

$$\begin{align*}
A_i \bar{x} &= \frac{1}{2^{1/p}} - \frac{1}{(1+(2^{-1/q}+1))^q} \frac{1}{(1+(2^{-1/q}+1))^q} x_l \\
A_l \bar{x} &= \frac{1}{(1+(2^{-1/q}+1))^q} \frac{1}{(1+(2^{-1/q}+1))^q} x_l + \frac{2^{1-1/q}+1}{(1+(2^{-1/q}+1))^q} = \frac{1}{(1+(2^{-1/q}+1))^q} \forall i \in [n]/\{1, k\};
\end{align*}$$

$$A_k \bar{x} = 1 - \frac{2^{1-1/q}+1}{(1+(2^{-1/q}+1))^q} = \frac{1}{(1+(2^{-1/q}+1))^q}.$$  \hfill (47)

In all, $\sigma_1 \geq \min_{i \in [n]} A_i \bar{x} = \frac{1}{(1+(2^{-1/q}+1))^q}$.

On the other hand, for any $x = (x_1, \ldots, x_d) \in \mathbb{B}_d$, we have

$$\min_{i \in [n]} A_i x = \min \left\{ -\frac{1}{2^{1/p}} x_1 + \frac{1}{2^{1/p}} x_l, \frac{1}{2^{1/p}} x_1 + \frac{1}{2^{1/p}} x_l, 0 \right\}. $$

If $x_1 \leq \frac{1}{(1+(2^{-1/q}+1))^q}$, then $A_i x \leq \frac{1}{(1+(2^{-1/q}+1))^q}$ for all $i \in [n]$; if $x_1 \geq \frac{1}{(1+(2^{-1/q}+1))^q}$, then $x_l \leq (1 - x_l^q)^{1/q} = \frac{2^{1-1/q}+1}{(1+(2^{-1/q}+1))^q}$.  \hfill (48)

\[ x_l \leq (1 - x_l^q)^{1/q} = \frac{2^{1-1/q}+1}{(1+(2^{-1/q}+1))^q}. \]  \hfill (49)
and hence by \( (48) \) we have

\[
\min_{i \in [n]} A_i x \leq -\frac{1}{2^{1/p} x_1} + \frac{1}{2^{1/p} x_1} \leq -\frac{1}{2^{1/p}} \cdot \frac{1}{(1 + (2^{1-1/q} + 1)^q)^{1/q}} + \frac{1}{2^{1/p}} \cdot \frac{2^{1-1/q} + 1}{(1 + (2^{1-1/q} + 1)^q)^{1/q}}
\]

\[
= \frac{1}{(1 + (2^{1-1/q} + 1)^q)^{1/q}}.
\]

(50)

In all, we always have \( \min_{i \in [n]} A_i x \leq \frac{1}{(1 + (2^{1-1/q} + 1)^q)^{1/q}} \). As a result, \( \sigma_1 = \max_{x \in \mathbb{B}^d} \min_{i \in [n]} A_i x \leq \frac{1}{(1 + (2^{1-1/q} + 1)^q)^{1/q}} \). In conclusion, we have \( \sigma_1 = \frac{1}{(1 + (2^{1-1/q} + 1)^q)^{1/q}} \).

Now, we prove that an \( \bar{x} \in \mathbb{B}^d \) satisfying \( (44) \) would simultaneously reveal whether \( A \) is from Case 1 or Case 2 as well as the value of \( l \in \{2, \ldots, d\} \), by the following algorithm:

1. Check if one of \( \bar{x}_2, \ldots, \bar{x}_d \) is at least \( 1 - 0.04 \cdot 2^{1/p} \); if there exists an \( l' \in \{2, \ldots, d\} \) such that \( \bar{x}_{l'} \geq 1 - 0.04 \cdot 2^{1/p} \), return ‘Case 2’ and \( l = l' \);
2. Otherwise, return ‘Case 1’ and \( l = \arg \max_{i \in \{2, \ldots, d\}} \bar{x}_i \).

We first prove that the classification of \( A \) between Case 1 and Case 2 is correct. On the one hand, assume that \( A \) comes from Case 1. If we wrongly classified \( A \) as from Case 2, we would have \( \bar{x}_{l'} \geq 1 - 0.04 \cdot 2^{1/p} \) and \( \bar{x}_l \leq (1 - (1 - 0.04 \cdot 2^{1/p})^q)^{1/q} \). We denote

\[
f_1(q) := 1 - (1 - 0.04 \cdot 2^{1-1/q})^q - \frac{1}{(1 + (2^{1-1/q} + 1)^q)^{1/q}} - 0.04 \cdot 2^{1-1/q}.
\]

(51)

It can shown that \( f_1 \) is a decreasing function on \( [1, 2] \); furthermore, \( f_1(2) > 0 \). See Figure 1. As a result, \( f_1(q) > 0 \), which implies

\[
\min_{i \in [n]} A_i \bar{x} = \min \left\{ -\frac{1}{2^{1/p}} \bar{x}_1 + \frac{1}{2^{1/p}} \bar{x}_1, \frac{1}{2^{1/p}} \bar{x}_1 + \frac{1}{2^{1/p}} \bar{x}_1, \bar{x}_1 \right\} \leq \bar{x}_1 < \sigma_1 - \epsilon.
\]

(52)

However, this contradicts with \( (44) \). Therefore, for this case we must make the correct classification that \( A \) comes from Case 1.

On the other hand, assume that \( A \) comes from Case 2. If we wrongly classified \( A \) as from Case 1, we would have \( \bar{x}_l \leq \max_{i \in \{2, \ldots, d\}} \bar{x}_i < 1 - 0.04 \cdot 2^{1/p} \); this would imply

\[
\min_{i \in [n]} A_i \bar{x} = \min \left\{ -\frac{1}{2^{1/p}} \bar{x}_1 + \frac{1}{2^{1/p}} \bar{x}_1, \frac{1}{2^{1/p}} \bar{x}_1 + \frac{1}{2^{1/p}} \bar{x}_1 \right\} \leq \frac{1}{2^{1/p}} \bar{x}_l < \sigma_2 - \epsilon,
\]

(53)

which contradicts with \( (44) \). Therefore, for this case we must make the correct classification that \( A \) comes from Case 2. In all, our classification is always correct.

![Figure 1: The plot of \( f_1 \), where the x-axis represents \( q \) and the y-axis represents \( f_1(q) \).](image-url)
It remains to prove that the value of \( l \) is correct. If \( A \) is from Case 1, we have
\[
\sigma_1 - \epsilon \leq \min_{i \in [n]} A_i \bar{x} = \min \left\{ -\frac{1}{2^{1/p}} \bar{x}_1 + \frac{1}{2^{1/p}} \bar{x}_l, \frac{1}{2^{1/p}} \bar{x}_1 + \frac{1}{2^{1/p}} \bar{x}_l, \frac{1}{2^{1/p}} \bar{x}_1, \frac{1}{2^{1/p}} \bar{x}_l \right\};
\]
as a result, \( \bar{x}_1 \geq \sigma_1 - \epsilon \) and \( -\frac{1}{2^{1/p}} \bar{x}_1 + \frac{1}{2^{1/p}} \bar{x}_l > \sigma_1 - \epsilon \), which imply
\[
\bar{x}_l > (2^{1/p} + 1)(\sigma_1 - \epsilon) > (2^{1-1/q} + 1)\left(\frac{1}{1 + (2^{1-1/q} + 1)^{1/q}} - 0.04\right).
\]
We denote
\[
f_2(q) := 2(2^{1-1/q} + 1)^q \left(\frac{1}{1 + (2^{1-1/q} + 1)^{1/q}} - 0.04\right)^q.
\]
It can shown that \( f_2 \) is an increasing function on \([1, 2]\); furthermore, \( f_1(1) > 1 \). See Figure 2. As a result, \( f_2(q) > 1 \), which implies \( |\bar{x}_l|^q > 1/2 \). Therefore, \( \bar{x}_l \) must be the largest among \( \bar{x}_2, \ldots, \bar{x}_d \) (otherwise \( \ell' = \arg \max_{i \in \{2, \ldots, d\}} \bar{x}_i \) and \( \ell \neq \ell' \) would imply \( ||\bar{x}||_q^q = \sum_{i \in [d]} |\bar{x}_i|_q^q \geq |\bar{x}_l|_q^q + |\bar{x}_l|_q^q \geq 2|\bar{x}_l|_q > 1 \), contradiction). Therefore, Line 2 of the algorithm correctly returns the value of \( \ell \).

If \( A \) is from Case 2, we have
\[
\sigma_2 - \epsilon \leq \min_{i \in [n]} A_i \bar{x} = \min \left\{ -\frac{1}{2^{1/p}} \bar{x}_1 + \frac{1}{2^{1/p}} \bar{x}_l, \frac{1}{2^{1/p}} \bar{x}_1 + \frac{1}{2^{1/p}} \bar{x}_l \right\} \leq \frac{1}{2^{1/p}} \bar{x}_l,
\]
and hence \( \bar{x}_l \geq 2^{1/p}(\sigma_2 - \epsilon) \geq 2^{1/p}(\frac{1}{2^{1/p}} - 0.04) > 1 - 0.04 \cdot 2^{1/p} \). Since \( \frac{1}{p} + \frac{1}{q} = 1 \), we have
\[
1 - 0.04 \cdot 2^{1/p} = 1 - 0.08 \cdot 2^{-1/q} > 2^{-1/q},
\]
where the last inequality comes from the fact that \( 2^{-1/q} \leq 2^{-1/2} = \frac{1}{\sqrt{2}} < \frac{1}{1.08} \) for any \( q \in [1, 2] \). Therefore, \( 2(1 - 0.04 \cdot 2^{1/p})^q > 1 \), and only one coordinate of \( \bar{x} \) could be at least \( 1 - 0.04 \cdot 2^{1/p} \) and we must have \( l = \ell' \). Therefore, Line 1 of the algorithm correctly returns the value of \( \ell \).

In all, we have proved that an \( \epsilon \)-approximate solution \( \bar{x} \in \mathbb{B}_q^d \) for (44) would simultaneously reveal whether \( A \) is from Case 1 or Case 2 as well as the value of \( \ell \). As a result:

- Classically: On the one hand, notice that distinguishing these two cases requires \( n - 2 \) classical queries to the entries of \( A \) for searching the position of \( k \); therefore, it gives an \( \Omega(n) \) classical query lower bound for returning an \( \bar{x} \) that satisfies (44). On the other hand, finding the value of \( \ell \) is also a search problem on the entries of \( A \), which requires \( d - 1 = \Omega(\sqrt{d}) \) queries. These observations complete the proof of the classical lower bound in Theorem 5.
Quantumly: On the one hand, notice that distinguishing these two cases requires $\Omega(\sqrt{n} - 2) = \Omega(\sqrt{n})$ quantum queries to $O_A$ for searching the position of $k$ because of the quantum lower bound for search (Bennett et al. 1997); therefore, it gives an $\Omega(\sqrt{n})$ quantum lower bound on queries to $O_A$ for returning an $\bar{x}$ that satisfies (44). On the other hand, finding the value of $l$ is also a search problem on the entries of $A$, which requires $\Omega(\sqrt{d-1}) = \Omega(\sqrt{d})$ quantum queries to $O_A$ also due to Bennett et al. (1997). These observations complete the proof of the quantum lower bound in Theorem 5.