Faster quantum and classical SDP approximations for quadratic binary optimization

Fernando G.S L. Brandão,[1,2,3] Richard Kueng[2,3] and Daniel Stilck França[4,5]

[1]Amazon Web Services, AWS Center for Quantum Computing, Pasadena, CA
[2]Institute for Quantum Information and Matter, California Institute of Technology, Pasadena, CA, US
[3]Department of Computing and Mathematical Sciences, California Institute of Technology, Pasadena, CA, US
[4]QMATH, Department of Mathematical Sciences, University of Copenhagen, Denmark
[5]Department of Mathematics, Technische Universität München, Germany

Abstract

We give a quantum speedup for solving the canonical semidefinite programming relaxation for binary quadratic optimization. The class of relaxations for combinatorial optimization has so far eluded quantum speedups. Our methods combine ideas from quantum Gibbs sampling and matrix exponent updates. A de-quantization of the algorithm also leads to a faster classical solver. For generic instances, our quantum solver gives a nearly quadratic speedup over state-of-the-art algorithms. We also provide an efficient randomized rounding procedure that converts approximately optimal SDP solutions into constant factor approximations of the original quadratic optimization problem.

1 Introduction

Quadratic optimization problems with binary constraints are an important class of of optimization problems. Given a (real-valued) symmetric $n \times n$ matrix A the task is to compute

$$\text{maximize} \quad \langle x|A|x\rangle \quad \text{subject to} \quad x \in \{\pm 1\}^n \qquad \text{(MaxQP)}. \tag{1}$$

This problem arises naturally in many applications across various scientific disciplines, e.g. image compression [OP83], latent semantic indexing [Kol98], correlation clustering [CW04, MMMO17] and structured principal component analysis, see e.g. [KT19a, KT19b] and references therein. Mathematically, MaxQPs (1) are closely related to computing the $\ell_\infty \to \ell_1$ norm of A. This norm, in turn, closely relates to the *cut norm* (replace $x \in \{\pm 1\}^n$ by $x \in \{0,1\}^n$), as both norms can only differ by a constant factor. These norms are an important concept in theoretical computer science [FK99, AFdlVKK03, AN06], since problems such as identifying the largest cut in a graph (MaxCut) can be naturally formulated as instances of these norms. This membership highlights that optimal solutions of (1) are NP-hard to compute in the worst case. Despite their intrinsic hardness, quadratic optimization

problems do admit a canonical *semidefinite programming* (SDP) relaxation[1] [GW95]:

$$\text{maximize} \quad \operatorname{tr}(AX) \quad \text{subject to} \quad \operatorname{diag}(X) = \mathbf{1},\ X \geq 0 \qquad \text{(MaxQP SDP)} \qquad (2)$$

Here, $X \geq 0$ indicates that the $n \times n$ matrix X is positive semidefinite (psd), i.e. $\langle y|X|y\rangle \geq 0$ for all $y \in \mathbb{R}^n$. SDPs comprise a rich class of convex optimization problems that can be solved efficiently, e.g. by using interior point methods [BV04].

Perhaps surprisingly, often the optimal value of the MaxQP relaxation provides a constant factor approximation to the optimal value of the original quadratic problem. However, the associated optimal matrix $X^\sharp$ is typically *not* in one-to one correspondence with an optimal feasible point $x^\sharp \in \{\pm 1\}^n$ of the original problem (1). Several randomized rounding procedures have been devised to overcome this drawback since the pioneering work of [GW95]. These transform $X^\sharp$ into a random binary vector $\tilde{x} \in \{\pm 1\}^n$ that achieves $\langle \tilde{x}|A|\tilde{x}\rangle \geq \gamma \max_{x\in\{\pm 1\}^n} \langle x|A|x\rangle$ in expectation for some constant γ depending on the structure of the input matrix.

Although tractable in a theoretical sense, the runtime associated with general purpose SDP solvers quickly becomes prohibitively expensive in both memory and time. This practical bottleneck has spurred considerable attention in the theoretical computer science community over the past decades [AHK05, BM05, BVB16, TYUC17]. (Meta) algorithms, like *matrix multiplicative weights* (MMW) [AHK05] solve the MaxQP SDP (2) up to multiplicative error ϵ in runtime $\mathcal{O}((n/\epsilon)^{2.5}s)$, where s denotes the column sparsity of A. Further improvements are possible, if the problem description A has additional structure [AK16].

Very recently, a line of works pointed out that quantum computers can solve certain SDPs even faster [BS17, vAG18b, BKL$^+$17, KP18]. However, current runtime guarantees depend on problem-specific parameters. These scale particularly poorly for most combinatorial optimization problems, including the MaxQP SDP, and negate any potential advantage.

In this work, we tackle this challenge and overcome shortcomings of existing quantum SDP solvers for the following variant of problem (2):

$$\begin{aligned} \text{maximize} \quad & \operatorname{tr}\left(\tfrac{1}{\|A\|}AX\right) && \text{(renormalized MaxQP SDP)} \qquad (3)\\ \text{subject to} \quad & \langle i|X|i\rangle = \tfrac{1}{n} \quad i \in [n]\,, \\ & \operatorname{tr}(X) = 1,\ X \geq 0. \end{aligned}$$

This renormalization of the original problem pinpoints connections to quantum mechanics: Every feasible point X obeys $\operatorname{tr}(X) = 1$ and $X \geq 0$, implying that it describes the state ρ of a n-dimensional quantum system. In turn, such quantum states can be represented approximately by a re-normalized matrix exponential $\rho = \exp(-H)/\operatorname{tr}(H)$, the *Gibbs state* associated with *Hamiltonian* H. We capitalize on this correspondence by devising a meta algorithm – *Hamiltonian Updates* (HU) – that is inspired by matrix exponentiated gradient updates [TRW05], see also [LSW15, BKL$^+$17, Haz06] for similar approaches.

Another key insight is that the diagonal constraints also have a clear quantum mechanical interpretation: the feasible states are those that are indistinguihable from the uniform

[1] Rewrite the objective function in (1) as $\operatorname{tr}(A|x\rangle\langle x|)$ and note that every matrix $X = |x\rangle\langle x|$ with $x \in \{\pm 1\}^n$ has diagonal entries equal to one and is psd with unit rank. Dropping the (non-convex) rank constraint produces a convex relaxation.

or maximally mixed state when measured in the computational basis. Adoopting this point of view, we consider the further relaxation of optimizing over quantum states that almost indistinguishable from such measurements. More precisely, we will call a solution approximately feasible for a paramater ϵ if we have $\sum_i \left|\langle i|X|i\rangle - \frac{1}{n}\right| \leq \epsilon$. This further relaxation will allow us to overcome shortcomings of previous solvers when dealing with SDPs of this form. Moreover, a significant part of this work will be to argue that, although this relaxation leads to faster quantum and classical algorithms, it only mildly changes the value of the program and still yields good solutions for randomized rounding techniques.

Although originally designed to exploit the fact that quantum architectures can sometimes create Gibbs states efficiently, it turns out that this approach also produces faster classical algorithms.

To state our results, we instantiate standard computer science notation. The symbol $\mathcal{O}(\cdot)$ describes limiting function behavior , while $\tilde{\mathcal{O}}(\cdot)$ hides poly-logarithmic factors in the problem dimension and polynomial dependencies on the inverse accuracy $1/\epsilon$. We are working with the adjacency list oracle model, where individual entries and location of nonzero entries of the problem description A can be queried at unit cost. We refer to Section 3.4 for a more detailed discussion.

Theorem I (Hamiltonian Updates: runtime)**.** *Let A be a (real-valued), symmetric $n \times n$ matrix with column sparsity s. Then, the associated renormalized* MaxQP SDP (3) *can be solved up to additive accuracy ϵ in runtime $\tilde{\mathcal{O}}\left(n^{1.5}s^{0.5+o(1)}\mathrm{poly}(1/\epsilon)\right)$ on a quantum computer and $\tilde{\mathcal{O}}\left(\min\{n^2 s, n^\omega\}\mathrm{poly}(1/\epsilon)\right)$ on a classical computer.*

Here ω is the matrix multiplication exponent. The polynomial dependency on inverse accuracy is rather high (e.g. $(1/\epsilon)^{12}$ for the classical algorithm) and we intend to improve this scaling in future work. We emphasize that the quantum algorithm also outputs a classical description of a solution that is approximately feasible in a sense that will be made precise below. Already the classical runtime improves upon the best known existing results and we refer to Section 2.5 for a detailed comparison. Access to a quantum computer would increase this gap further. However, it is important to point out that Theorem I addresses the renormalized MaxQP SDP (3). Converting it into standard form (2) results in an approximation error of order $n\|A\|\epsilon$. In contrast, MMW [AHK05] – the fastest existing algorithm – incurs an error proportional to $\epsilon\|A\|_{\ell_1}$. Importantly, this comparison is favorable for generic problem instances, see Section 2.5.

The quantum algorithm outputs a classical description of an optimal Hamiltonian $H^\sharp$ that encodes an approximately optimal, approximately feasible solution $\rho^\sharp = \exp(-H^\sharp)/\mathrm{tr}\left(\exp(-H^\sharp)\right)$ of the renormalized MaxQP SDP (3). This classical output can subsequently be used for randomized rounding for the $\ell_\infty \to \ell_1$ norm of a matrix A, $\|A\|_{\infty\to 1} = \max\limits_{x,y\in\{\pm 1\}^n} \langle x, A, y\rangle$.

Theorem II (Rounding)**.** *Suppose that $H^\sharp$ encodes an approximately optimal solution of the renormalized* MaxQP SDP (3) *with accuracy ϵ^4 for the target matrix*

$$A' = \begin{pmatrix} 0 & A \\ A^T & 0 \end{pmatrix},$$

where A is a $n \times n$ real matrix . Then, there is a classical $\tilde{\mathcal{O}}(ns)$-time randomized rounding

procedure that converts $H^\sharp$ into binary vectors $\tilde{x}, \tilde{y} \in \{\pm 1\}^n$ that obey

$$\gamma\left(\|A\|_{\infty\to 1} - \mathcal{O}(n\|A\|\epsilon)\right) \leq \mathbb{E}\left[\langle\tilde{x}|A|\tilde{y}\rangle\right] \leq \|A\|_{\infty\to 1}.,$$

where $\gamma = \frac{2}{\pi}$ if A is positive semidefinite and $\frac{4}{\pi} - 1$ else.

This result recovers the randomized rounding guarantees of [AN06] in the limit of perfect accuracy ($\epsilon = 0$). However, for $\epsilon > 0$ the error scales with $n\|A\|$. In turn, randomized rounding only provides a multiplicative approximation if $\|A\|_{\infty\to 1}$ is of the same order. This turns out to be the case for generic problem instances and we refer to Section 2.5 for a more detailed discussion.

2 Detailed summary of results

We present *Hamiltonian Updates* – a meta-algorithm for solving convex optimization problems over the set of quantum states based on quantum Gibbs sampling – in a more general setting, as we expect it to find applications to other problems. Throughout this work, $\|\cdot\|_{tr}$ and $\|\cdot\|$ denote the trace (Schatten-1) and operator (Schatten-∞) norms, respectively.

2.1 Convex optimization and feasibility problems

Most SDPs are a special instance of a more general class of convex optimization. For a bounded, convex function f from the set of symmetric matrices to the real numbers and closed convex sets $\mathcal{C}_1, \ldots, \mathcal{C}_n$, solve

$$\begin{aligned} \text{minimize} \quad & f(X) && \text{(CPOPT)} \\ \text{subject to} \quad & X \in \mathcal{C}_1 \cap \cdots \cap \mathcal{C}_m, \\ & \operatorname{tr}(X) = 1,\ X \geq 0. \end{aligned} \tag{4}$$

The constraint $\operatorname{tr}(X) = 1$ enforces normalization, while $X \geq 0$ is the defining structure constraint of semidefinite programming. Together, they restrict X to the set of n-dimensional quantum states $\mathcal{S}_n = \{X : \operatorname{tr}(X) = 1,\ X \geq 0\}$. We will now specialize to the case $f(A) = \operatorname{tr}(AX)$ for a symmetric matrix A, as this is our main case of interest, but remark that it is simple to generalize the discussion that follows for more general classes. This quantum constraint implies fundamental bounds on the optimal value: $|\operatorname{tr}(AX^\sharp)| \leq \|A\|\|X^\sharp\|_{tr} = \|A\|$, according to Matrix Hölder [Bha97, Ex. IV.2.12]. Binary search over potential optimal values $\lambda \in [-\|A\|, \|A\|]$ allows for reducing the convex optimization problem into a sequence of feasibility problems:

$$\begin{aligned} \text{find} \quad & X \in \mathcal{S}_n && \text{(CPFEAS}(\lambda)\text{)} \\ \text{subject to} \quad & \operatorname{tr}(AX) \leq \lambda, \\ & X \in \mathcal{C}_1 \cap \cdots \cap \mathcal{C}_m. \end{aligned} \tag{5}$$

The convergence of binary search is exponential. This ensures that the overhead is benign: a total of $\log(\|A\|/\epsilon)$ queries of CPFEAS(λ) suffice to determine the optimal solution of CPOPT (4) up to accuracy ϵ. In summary:

Fact 2.1. *Binary search reduces the task of solving convex optimization problems* (4) *to the task of solving convex feasibility problems* (5)*.*

Algorithm 1 *Meta-Algorithm for approximately solving convex feasibility problems* (5).

Require: Query access to m ϵ-separation oracles $O_{1,\epsilon}(\cdot), \ldots, O_{m,\epsilon}(\cdot)$

function HamiltonianUpdates(T, ϵ)
  $\rho = n^{-1} I$ and $H = 0$ ▷ initialize the maximally mixed state
  for $t = 1, \ldots, T$ **do**
    for $i = 1, \ldots, m$ **do** ▷ Query oracles and check feasibility
      if $O_{i,\epsilon}(\rho) = P$ **then**
        $H \leftarrow H - \frac{\epsilon}{8} P$ ▷ Penalize infeasible direction
        $\rho \leftarrow \exp(-H) / \mathrm{tr}(\exp(-H))$ ▷ Update quantum state
        break loop
      end if
    end for
    return (ρ, H) and **exit function** ▷ Current iterate is ϵ-feasible
  end for
end function

2.2 Meta-algorithm for approximately solving convex feasibility problems

We adapt a meta-algorithm developed by Tsuda, Rätsch and Warmuth [TRW05], see also [LRS15, AK16, Haz06, BKL+17] for similar ideas and [Bub15] for an overview of these techniques. All these algorithms, including the variation presented here, can be seen as instances of mirror descent with the mirror map given by the von Neumann entropy with adaptations tailored to the problem at hand. We believe our variation provides a path for also obtaining quantum speedups for nonlinear convex optimizations, so we state it in more detail.

For our algorithm, we require subroutines that allow for testing ϵ-closeness (in trace norm) to each convex set $\mathcal{C}_i$.

Definition 2.1 (ϵ-separation oracle)**.** *Let $\mathcal{C}$ be a closed, convex set. An ϵ-separation oracle (with respect to the trace norm) is a subroutine that either accepts a matrix X (if it is close to feasible), or provides a hyperplane P that separates ρ from the convex set:*

$$O_{\mathcal{C},\epsilon}(X) = \begin{cases} \text{accept } X \text{ if } \min_{Y \in \mathcal{C}} \|X - Y\|_{tr} \leq \epsilon, \\ \text{else: output } P \text{ s.t. } \|P\| \leq 1,\ \mathrm{tr}(P(Y - X)) \geq \epsilon \text{ for all } Y \in \mathcal{C}. \end{cases}$$

Hamiltonian Updates (HU) is based on a change of variables that automatically takes care of positive semidefiniteness and normalization: replace X in problem (5) by a Gibbs state $\rho_H = \exp(-H)/\mathrm{tr}(\exp(-H))$. At each iteration, we query ϵ-separation oracles. If they all accept, the current iterate is ϵ-close to feasible in the sense that there is a matrix in each C_i that is ϵ close in trace distance to the accepted state, and we are done. Otherwise, we update the matrix exponent to penalize infeasible directions: $H \to H - \frac{\epsilon}{8} P$, where P is a separating hyperplane that witnesses infeasibility. This process is visualized in Figure 1 and we refer to Algorithm 1 for a detailed description.

Theorem 2.1 (HU: convergence)**.** *Algorithm 1 requires at most $T = \lceil 16 \log(n)/\epsilon^2 \rceil + 1$ iterations to either certify that* (5) *is infeasible or output a state ρ satisfying:*

$$\forall i \exists Y_i \in C_i \text{ s.t. } \|Y_i - \rho\|_{tr} \leq \epsilon. \tag{6}$$

The proof follows from establishing constant step-wise progress in quantum relative entropy. The quantum relative entropy between *any* feasible state and the initial state $\rho_0 = n^{-1}I$ (maximally mixed state) is bounded by $\log(n)$. Therefore, the algorithm must terminate after sufficiently many iterations. Otherwise the problem is infeasible. We refer To Section 3.1 for details.

Theorem 2.1 has important consequences: The runtime of approximately solving quantum feasibility problems is dominated by the cost of implementing m separation oracles $O_{i,\epsilon}$ and the cost associated with matrix exponentiation. This reduces the task of efficiently solving convex feasibility problems to the quest of efficiently identifying separating hyperplanes and developing fast routines for computing Gibbs states.

The latter point already hints at a genuine quantum advantage: quantum architectures can efficiently prepare (certain) Gibbs states [CS17, Fra18, KBa16, PW09, TOV+09, TOV+09, YAG12, vAGGdW17]. Moreover, we will formulate the diagonal constraints for the problem at hand in a way that also allows for an easy quantum meachnical interpretation, overcoming hurdles of previous solvers. It should also be stressed that approximate feasibility guarantee in (6) is not very strong and a careful choice of the C_i and strong continuity bounds are required to ensure that it gives a good approximation.

2.3 Classical and quantum solvers for the renormalized MaxQP SDP

For fixed $\lambda \in [-1, 1]$ the (feasiblity) MaxQP SDP is equivalent to a quantum feasibility problem:

$$\begin{aligned} \text{find} \quad & \rho \in \mathcal{S}_n \cap \mathcal{A}_\lambda \cap \mathcal{D}_n \\ \text{where} \quad & \mathcal{A}_\lambda = \left\{X : \operatorname{tr}\left(A\|A\|^{-1}X\right) \geq \lambda\right\}, \\ & \mathcal{D}_n = \left\{X : \langle i|X|i\rangle = 1/n, i \in [n]\right\}. \end{aligned}$$

The set $\mathcal{A}_\lambda$ corresponds to a half-space, while $\mathcal{D}_n$ is an affine subspace with codimension n. The simple structure of both sets readily suggests two separation oracles:

$\mathcal{O}_{\mathcal{A}_\lambda}$: check $\operatorname{tr}(A\|A\|^{-1}\rho) \leq \lambda$ and output $P = A\|A\|^{-1}$ if this is not the case.

$\mathcal{O}_{\mathcal{D}_n}$: Check $\sum_i |\langle i|\rho|i\rangle - 1/n| \leq \epsilon$ and output $P = \sum_{i=1}^n \mathbb{I}\left\{\langle i|\rho|i\rangle > 1/n\right\} |i\rangle\langle i|$ if this is not the case.

The key insight to later obtain quantum speedups for the MaxQP SDP is that the second oracle can be interpreted as trying to distinguish the current state from the maximally mixed through computational basis measurements. This view is similar in spirit to [LRS15, Lemma 4.6], although here we focus on using this approach to construct solutions and to show that this notion of approximate feasbility is good enough for the MaxQP SDP.

2.3.1 Classical runtime

For fixed $\rho_H = \exp(-H)/\operatorname{tr}(-H)$ both separation oracles are easy to implement on a classical computer given access to ρ_H. Hence, matrix exponentiation is the only remaining bottleneck. This can be mitigated by truncating the Taylor series for $\exp(-H)$ after $l' = \mathcal{O}(\log(n)/\epsilon)$

many steps. Approximating ρ in this fashion only requires $\mathcal{O}(\min\left\{n^2 s, n^\omega\right\}\log(n)\epsilon^{-1})$ steps and only incurs an error of ϵ in trace distance. Moreover, it is then possible to convert an approximately feasible to point to a strictly feasible one with a similar value, see Section 3.3. The following result becomes an immediate consequence of Fact 2.1 and Theorem 2.1.

Corollary 2.1 (Classical runtime for the MaxQP SDP)**.** *Suppose that A has row-sparsity s. Then, the classical cost of solving the associated (renormalized)* MaxQP SDP *up to additive error ϵ is $\mathcal{O}(\min\{n^2 s, n^\omega\}\log(n)\epsilon^{-12})$.*

2.3.2 Quantum runtime

Quantum architectures can efficiently prepare (certain) Gibbs states and are therefore well suited to overcome the main classical bottleneck. In contrast, checking feasibility becomes more challenging, because information about ρ is not accessible directly. Instead, we must prepare multiple copies of ρ and perform quantum mechanical measurements to test feasibility:

- $\mathcal{O}(\epsilon^{-2})$ copies of ρ suffice to ϵ-approximate $\mathrm{tr}(A\|A\|^{-1}\rho)$ via phase estimation.
- $\mathcal{O}(n\epsilon^{-2})$ copies suffice with high probability to estimate the diagonal entries of ρ (up to accuracy ϵ in trace norm) via repeated computational basis measurements.

Combining this with the overall cost of preparing a single Gibbs state implies the following runtime for executing Algorithm 1 on a quantum computer. This result is based on the *sparse oracle input model* and we refer to Sec. 3.4 for details.

Corollary 2.2 (Quantum runtime for the MaxQP SDP)**.** *Suppose that A has row-sparsity s. Then, the quantum cost of solving the renormalized* MaxQP SDP *up to additive error ϵ is $\tilde{\mathcal{O}}(n^{1.5}s^{0.5+o(1)}\mathrm{poly}(1/\epsilon))$.*

The quantum algorithm also outputs a classical description of the Hamiltonian $H^\sharp$ corresponding to an approximately optimal, approximately feasible Gibbs state and its value. Moreover, we have the potential to produce samples from the associated approximately optimal Gibbs state $\rho^\sharp = \exp(-H^\sharp)/\mathrm{tr}\left(\exp(-H^\sharp)\right)$ in sub-linear runtime $\tilde{\mathcal{O}}(\sqrt{n})$ on a quantum computer. In principle it would require $\tilde{\mathcal{O}}(n^2 s)$ time to convert this approximately feasible solution to a strictly feasible one. But in the next section we show that the output of the algorithm is enough to give rise to good randomized roundings.

2.4 Randomized rounding

The renormalized MaxQP SDP (3) arises as a convex relaxation of an important quadratic optimization problem (1). However, the optimal solution $X^\sharp$ is typically not of the form $|x\rangle\langle x|$, with $x \in \{\pm 1\}^n$. Goemans and Williamson [GW95] pioneered randomized rounding techniques that allow for converting $X^\sharp$ into a cut $x^\sharp$ that is close-to optimal. However, their rounding techniques rely on the underlying matrix bein entrywhise positive and a more delicate analysis is required to derive analogous results forr broader classes of matrices. We will now follow the analysis of [AN06] to do the randomized rounding for the $\ell_\infty \to \ell_1$ norm.

Algorithm 2 *Randomized rounding based on optimal Hamiltonian* $H^\sharp$

function RandomizedRounding($H^\sharp$)
  Draw a random vector $g \in \mathbb{R}^n$ with i.i.d. $\mathcal{N}(0,1)$ entries.
  Compute $z = \sum_{k=0}^{l} \frac{(-H^\sharp)^k}{2^k k!} g$ for $l = \mathcal{O}(\epsilon^{-1}\log(n))$.
  output $x_i = \mathrm{sign}(z_i)$.
end function

First, let us make the connection between this norm and the MaxQP SDP clearer. Let A be a real matrix and define

$$A' = \begin{pmatrix} 0 & A \\ A^T & 0 \end{pmatrix}.$$

It is easy to see that for two binary vectors $x, y \in \{\pm 1\}^n$ we have $\langle x \oplus y, A'x \oplus y\rangle = 2\langle x, Ay\rangle$. This immediately shows that $2\|A\|_{1\to\infty} = \max_{z\in\{\pm1\}^{2n}} \langle z, A'z\rangle$, which is an instance of MaxQP SDP. We will now show that the rounding procedure is stable, i.e. randomized rounding of an approximately feasible, approximately optimal point, such as the ones outputted by the quantum algorithm, still result in a good binary vector for approximating this norm. We strengthen the stability of the rounding even further by showing that rounding with a truncated Taylor expansion of the solution is still good enough, saving runtime. The rounding procedure is described in Algorithm 2.

Proposition 2.1. *Let A be a real matrix and $H^\sharp$ be such that $\rho^\sharp = \exp(-H^\sharp)/\mathrm{tr}(\exp(-H^\sharp))$ is an ϵ-approximate solution to the renormalized* MaxQPSDP *for A' (3) with value $\alpha^\sharp = \mathrm{tr}\left(A\|A\|^{-1}\rho^\sharp\right)$. Then, the (random) output $x = (x_1 \oplus x_2) \in \{\pm1\}^{2n}$ of Algorithm 2 can be computed in $\tilde{\mathcal{O}}(ns)$-time and obeys*

$$\gamma n\|A\|\left(\alpha^\sharp - \mathcal{O}(\epsilon)\right) \leq \mathbb{E}\langle x_1|A|x_2\rangle \leq n\|A\|(\alpha^\sharp + \mathcal{O}(\epsilon)),$$

where $\gamma = 2/\pi$ for A p.s.d. and $4/\pi - 1$ else.

This rounding procedure is fully classical and can be executed in runtime $\tilde{\mathcal{O}}(ns)$. We refer to Sec. 3.5 for details. What is more, it applies to both quantum and classical solutions of the MaxQP SDP. Even the quantum algorithm provides $H^\sharp$ in classical form, while the associated $\rho^\sharp$ is only available as a quantum state. Rounding directly with $\rho^\sharp$ would necessitate a fully quantum rounding technique that, while difficult to implement and analyze, seems to offer no advantages over the classical Algorithm 2. Thus, it is possible perform the rounding even with the output of the quantum algorithm. We prove this theorem in two steps. First, we follow the proof technique of [AN06] to show that the our relaxed notion of approximately feasible is still good enough to ensure a good rounding in expectation. This shows that our notion of feasibility is strong enough for the problem at hand. The stability of the rounding w.r.t. to truncation of the Taylor series then follows by showing appropriate anticoncentration inequalities for the random vector.

Also note that in [AN06] the authors prove that the constant $\frac{2}{\pi}$ in Proposition 2.1 is optimal.

2.5 Comparison to existing work

The MaxQP SDP has already received a lot of attention in the literature. Table 1 contains a runtime comparison between the contributions of this work and the best existing classical results [AHK05, AK16]. This highlights regimes, where we obtain both classical and quantum speedups. In a nutshell, Hamiltonian Updates outperforms state of the art algorithms whenever the target matrix A has both positive and negative off-diagonal entries and the optimal value of the SDP scales as $n\|A\|$. It is worthwhile to explore the following examples.

Nearly quadratic quantum speedups and classical speedups for generic instances: The conditions under which Hamiltonian Updates offers speedups are generic for matrices that have both positive and negative entries, see Appendix A. More precisely, suppose that A is a random matrix whose entries are i.i.d samples of a centered random variable with bounded fourth moment. Then, $\|A\|_{\ell_1} = \Theta(n^{3/2}\|A\|)$ and $\|A\|_{1\to\infty} = \Theta(n\|A\|)$ in expectation . This implies that the runtime of Hamiltonian Updates improves upon MMW [AHK05]. For these dense instances, the MMW-runtime is $\tilde{\mathcal{O}}(n^{3.5})$ compared to $\tilde{\mathcal{O}}(n^3)$ provided in Corollary 2.1 and $\tilde{\mathcal{O}}(n^2)$ in the quantum case. This is almost a quadratic quantum improvement in n.

Moreover, we expect it to be possible to obtain quadratic quantum speedups. It is easy to see that for a s sparse matrix A we have $\|A\|_{\ell_1} \leq ns\|A\|$. For s-sparse matrices such that $\|A\|_{1\to\infty} = \Theta(n\|A\|)$ and $\|A\|_{\ell_1} = \Theta(ns\|A\|)$ we then have the runtime $\tilde{\mathcal{O}}(\min\{(ns)^{2.5}, n^3 s\})$ for MMW. It is then not difficult to see that identifying instances with this scaling and $s = \Omega(n^{1/3})$ would lead to quadratic quantum speedups.

To the best of our knowledge, the quantum implementation of Hamiltonian updates establishes the first quantum speedup for problems of this type. Corollary 2.2 establishes a nearly quadratic speedup for generic MaxQP SDP instances compared to current state of the art.

No speedups for MaxCut: Additional structure can substantially reduce the runtime of existing MMW solvers [AK16]. For weighted MaxCut, in particular, A is related to the adjacency matrix of a graph and has non-negative entries. This additional structure facilitates the use of powerful dimensionality reduction and sparsification techniques that outperform Hamiltonian Updates. However, these techniques do not readily apply to general problem instances, where the entries of A can be both positive and negative (sign problem). We refer to Appendix B for a more detailed discussion.

Sherrington-Kirkpatrick Model: It is also interesting to take generic problem instances more seriously. Sampling A from the Gaussian ensemble (i.e. each entry is an i.i.d. standard normal random variable) produces an ensemble of MaxQP SDPs that is closely related to the Sherringkton-Kirkpatrick (SK) model [Pan13]. This problem has received considerable attention in the statistical physics literature. In particular, recent work [Mon18] shows that, under some unproven conjectures, it is possible to solve the quadratic optimization in (1) directly in time $\tilde{\mathcal{O}}(n^2)$. Furthermore, there is an integrability gap for the SDP relaxation of this problem in the Gaussian setting [KB19].

Previous quantum SDP solvers: previous quantum SDP solvers [BS17, vAG18b, BKL$^+$17] with inverse polynomial dependence on the errror do not provide speedups for solving the MaxQP SDP, as their complexity depends on a problem specific parameter, the width of the SDP. We refer to the aforementioned references for a definition of this parameter and for the complexity of the solvers under different input models and only focus on why none of them readily gives speedups for the problem at hand. As shown in [vAGGdW17, Theorem 24], the width parameter scales at least linearly in the dimension n for the MaxQP SDP. To the best of our knowledge, the solvers mentioned above have a dependence that is at least quadratic in the width and at least a $n^{\frac{1}{2}}$ dependence on the dimension. Thus, the combination of the term stemming from the width and the dimension already gives a higher complexity than our solver. One reason why we bypass these restrictions is that we do not use the primal-dual approach to solve the SDP from the aforementioned references. Another, and arguably conceptually more interesting, reason why our algorithm outperforms other solvers is how we enforce the diagonal constraint.

Enforcing that each diagonial constraint of the renormalized MAXQP SDP in Eq. 3 is satisfied up to an additive error, i.e.

$$|\langle i|\rho|i\rangle - 1/n| \leq \epsilon$$

would require an error ϵ that is at most of order n^{-1} to render the feasible set different from all quantum states. Given that previous solvers only handle linear constraints like the one above with complexity scaling at least quadratically in ϵ^{-1}, previous quantum methods do not readily apply and have worse runtimes than available classical algorithms.

In [KP18], the authors give a quantum SDP solver whose complexity is $\widetilde{O}\left(\frac{n^{2.5}}{\xi^2}\mu\kappa^3 \log\left(\epsilon^{-1}\right)\right)$. Here κ and μ are again problem specific parameters and ξ is precision to which each constraint is satisfied. As noted before, a straightforward implementation of the MAXQP SDP requires ξ to be at most of order n^{-1}, which establishes a runtime of order at least $n^{4.5}$ using those methods. Thus, we conclude that all current quantum SDP solvers do not offer speedups over state of the art classical algorithms, see Table 1 for more details.

This discussion showcases that our technique to relax the diagonal constraints gives rise to a novel way of enforcing constraints that allows for a better control of errors in quantum SDP solvers and could be used for other relevant SDPs. Moreover, the fact that the approximate solution can still be used to obtain good roundings highlights the fact that our notion of approximate feasbility renders the problem artificially easy.

Finally, we want to point out that subtleties regarding error scaling do not arise for MaxCut. If A is the adjacency matrix of a d-regular graph on n vertices, then $n\|A\|_\infty = nd = \|A\|_{\ell_1}$ and the different errors in Table 1 all coincide.

3 Technical details and proofs

3.1 Proof of Theorem 2.1

By construction, Algorithm 1 (Hamiltonian Updates) terminates as soon as it has found a quantum state ρ that is ϵ-close to being feasible. Correctly flagging infeasibility is the more interesting aspect of Theorem 2.1 (convergence to feasible point).

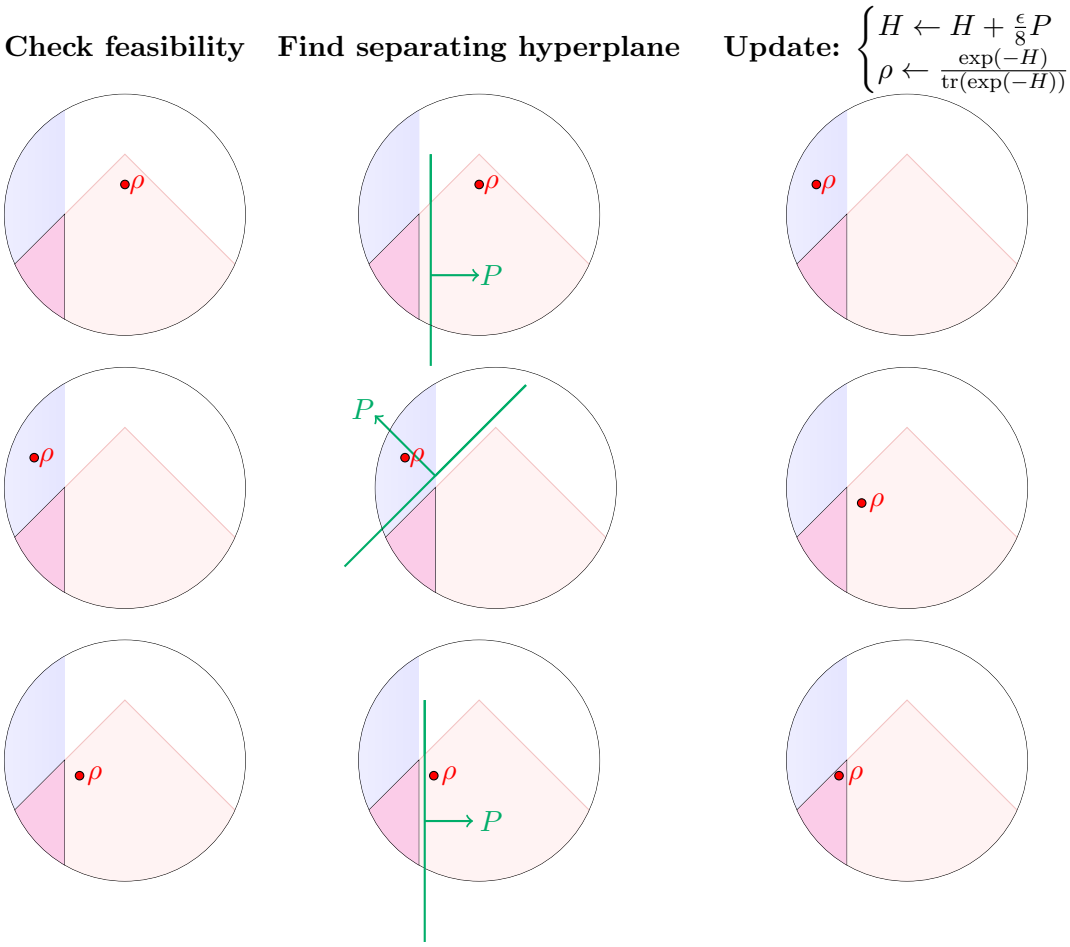

Figure 1: *Caricature of Hamiltonian Update iterations in Algorithm 1:* Schematic illustration of the intersection of three convex sets (i) a halfspace (blue), (ii) a diamond shaped convex set (red) and (iii) the set of all quantum states (clipped circle). Algorithm 1 (Hamiltonian Updates) approaches a point in the convex intersection (magenta) of all three sets by iteratively checking feasibility (left column), identifying a separating hyperplane (central column) and updating the matrix exponent to penalize infeasible directions (right column).

Algorithm	Runtime	Error	Speedup
This work (Classical)	$\tilde{\mathcal{O}}(\min\{n^2 s, n^\omega\}\epsilon^{-12})$	$\epsilon n\|A\|$	-
This work (Quantum)	$\tilde{\mathcal{O}}(n^{1.5}s^{0.5+o(1)}\epsilon^{-12})$	$\epsilon n\|A\|$	-
MMW [AHK05]	$\tilde{\mathcal{O}}(\min\{(n/\epsilon)^{2.5}s, n^3\beta^{-1}\})$	$\epsilon\|A\|_{\ell_1}$	$\|A\|_{\ell_1} \geq n\|A\|$, $\epsilon = \Theta(1)$
Interior Point [LSW15]	$\mathcal{O}(n^{\omega+1}\log(\epsilon^{-1}))$	ϵ	$\epsilon = \Theta(1)$
MMW for MaxCut [AK16] (non-negative entries only)	$\tilde{\mathcal{O}}(ns)$	$\epsilon\|A\|_{\ell_1}$	-

Table 1: comparison of different classical algorithms to solve the original MaxQP SDP (2). The speedup column clarifies in which regimes we obtain speedups and ω denotes the exponent of matrix multiplication. Here β corresponds to the value of MAXQP SDP multiplied by $n\|A\|/\|A\|_{\ell_1}$.

Lemma 3.1. *Suppose Algorithm 1 does not terminate after $T = \lceil 16\log(n)/\epsilon^2\rceil + 1$ steps. Then, the feasibility problem (5) is infeasible.*

Proof. By contradiction. Suppose there exists a feasible point ρ^* in the intersection of all $m+1$ sets and we ran the algorithm for T steps. Instantiate the short-hand notation $\rho_t = \rho_{H_t} = \exp(-H_t)/\mathrm{tr}(\exp(-H_t))$ for the t-th state and Hamiltonian in Algorithm 1. Initialization with $H_0 = 0$ and $\rho_0 = I/n$ is crucial, as it implies that the quantum relative entropy between ρ^* and ρ_0 is bounded:

$$S(\rho^*\|\rho_0) = \mathrm{tr}(\rho^*(\log\rho^* - \log\rho_0)) \leq \log(n).$$

We will now show that the relative entropy between successive (infeasible) iterates ρ_{t+1}, ρ_t and the feasible state ρ^* necessarily decreases by a finite amount. Let P_t be the hyperlane that separates ρ_t from the feasible set. The update rule $H_{t+1} = H_t - \frac{\epsilon}{8}P_t$ then asserts

$$\begin{aligned} S(\rho^*\|\rho_{t+1}) - S(\rho^*\|\rho_t) =& \mathrm{tr}(\rho^*(H_t - H_{t_1})) + \log\left(\frac{\mathrm{tr}(\exp(-H_{t+1}))}{\mathrm{tr}(\exp(-H_t))}\right) \\ =& -\tfrac{\epsilon}{8}\mathrm{tr}(P_t\rho^*) - \log\left(\frac{\mathrm{tr}\left(\exp\left(-H_{t+1} + \frac{\epsilon}{8}P_t\right)\right)}{\mathrm{tr}(\exp(-H_{t+1}))}\right). \end{aligned}$$

The logarithmic ratio can be bounded using the Peierls-Bogoliubov inequality [AL70, Lemma 1]: $\log(\mathrm{tr}(\exp(F+G))) \geq \mathrm{tr}(F\exp(G))$ provided that $\mathrm{tr}(\exp(G)) = 1$. This implies

$$\begin{aligned} \log\left(\frac{\mathrm{tr}\left(\exp\left(-H_{t+1} + \frac{\epsilon}{8}P_t\right)\right)}{\mathrm{tr}(\exp(-H_{t+1}))}\right) =& -\log\left(\mathrm{tr}\left(\exp\left(H_{t+1} + \tfrac{\epsilon}{8}P_t - \log(\mathrm{tr}(\exp(-H_{t_1})))\, I\right)\right)\right) \\ \leq& \mathrm{tr}\left(\tfrac{\epsilon}{8}P_t\exp(H_{t+1} - \log(\mathrm{tr}(\exp(H_{t+1})))I)\right) \\ =& \tfrac{\epsilon}{8}\mathrm{tr}(P_t\exp(-H_{t+1})/\mathrm{tr}(\exp(-H_{t+1}))) = \tfrac{\epsilon}{8}\mathrm{tr}(P_t\rho_{t+1}). \end{aligned}$$

Next, note that the updates are mild in the sense that ρ_{t+1} and ρ_t are close in trace distance. [BS17, Lem. 16] implies $\|\rho_{t_1} - \rho_t\|_{tr} \leq 2\left(\exp(\frac{\epsilon}{8}\|P_t\|) - 1\right) \leq \frac{\epsilon}{2}$, because $\|P_t\| \leq 1$ by construction and we can also assume $\frac{\epsilon}{8} \leq \log(2)$. Combining these insights with Matrix Hölder

[Bha97, Ex. IV.2.12] ensures

$$\begin{aligned} S(\rho^*\|\rho_{t+1}) - S(\rho^*\|\rho_t) \leq & -\tfrac{\epsilon}{8}\mathrm{tr}(P_t\rho^*) + \tfrac{\epsilon}{8}\mathrm{tr}(P_t\rho_{t+1}) \\ = & \tfrac{\epsilon}{8}\left(\mathrm{tr}(P_t(\rho_{t+1}-\rho_t)) - \mathrm{tr}(P_t(\rho_t-\rho^*))\right) \\ \leq & \tfrac{\epsilon}{8}\left(\|P_t\|\|\rho_{t+1}-\rho_t\|_{tr} + \mathrm{tr}(P_t(\rho_t-\rho^*))\right). \end{aligned}$$

The first contribution is bounded by $\frac{\epsilon}{2}\|P_t\| \leq \frac{\epsilon}{2}$, while Definition 2.1 ensures $\mathrm{tr}(P_t(\rho_t-\rho^*)) \leq -\epsilon$ (ρ^* is feasible and P_t is an ϵ-separation oracle for the infeasible point ρ_t). In summary,

$$S(\rho^*\|\rho_{t+1}) - S(\rho^*\|\rho_t) \leq \tfrac{\epsilon}{8}\left(\tfrac{\epsilon}{2}-\epsilon\right) = -\tfrac{\epsilon^2}{16} \quad \text{for all iterations} \quad t=0,\ldots,T$$

and we conclude

$$S(\rho^*\|\rho_T) = \sum_{t=0}^{T}\left(S(\rho^*\|\rho_{t+1}) - S(\rho^*\|\rho_t)\right) + S(\rho^*\|\rho_0) \leq -T\tfrac{\epsilon^2}{16} + \log(n).$$

This expression becomes negative as soon as the total number of steps T surpasses $16\log(n)/\epsilon^2$. A contradiction, because quantum relative entropy is always non-negative. □

3.2 Stability of the relaxed MaxQP SDP

Note that even if Algorithm 1 accepts a candidate point, it does not necessarily mean that this point is exactly feasible. Theorem 2.1 only asserts that this point is ϵ-close to all sets of interest. For the MaxQP SDP (3), this means that the outputs of the algorithm will only satisfy the diagonal constraints approximately and, in principle, the value of this further relaxed problem could differ significantly from the original value. In the next proposition we show that this is not the case:

Proposition 3.1. *Let α_ϵ be the value attained by an approximately optimal, approximately feasible – up to accuracy ϵ – solution to the* MaxQP SDP (3) *with input matrix A. Then,*

$$|\alpha_{\epsilon^4} n\|A\| - \alpha| = \mathcal{O}(\epsilon n\|A\|), \tag{7}$$

where α is the true optimal value of the original SDP (2)*. Moreover, it is possible to construct a feasible point of* (2) *whose value satisfies* (7) *from the approximate solution in time $\mathcal{O}(n^2)$.*

Proof. Let ρ be a solution to the relaxed MaxQP SDP (3) with relaxation parameter ϵ^4. We will now construct an exactly feasible point $\rho^\sharp$ of the MaxQP SDP (3). This modifications are mild enough to ensure that the associated SDP value will only change by $\mathcal{O}(\epsilon n\|A\|)$. We proceed in two steps: (i) $\rho \mapsto \rho'$: Identify diagonal entries that substantially deviate from $1/n$ in the sense that $|\langle i|\rho|i\rangle - 1/n| > \epsilon^2/n$. Subseqently, replace ρ_{ii} by $1/n$ and set all entries in the i-th row and i-th column to zero. This ensures that ρ' remains positive semidefinite. (ii) $\rho' \to R$: Replace all remaining entries by $1/n$. This may thwart positive semidefiniteness, but the following convex combination restores this feature:

$$\rho^\sharp = \tfrac{1}{1+\epsilon^2}\left(R + \tfrac{\epsilon^2}{n}I\right).$$

By construction, this matrix is both psd and obeys $\langle i|\rho^\sharp|i\rangle = 1/n$ for all $i \in [n]$. In words: it is a feasible point of the renormalized MaxQP SDP (3).

We now show that these reformulations are mild. To this end, let $B = \{i : \langle i|\rho|i\rangle - 1/n\} \subset [n]$ be the indices associated with large deviations. Without loss of generality, we can assume that these are the first $|B|$ indices. Then,

$$\begin{aligned}\|\rho' - \rho\|_1 =& \left\| \begin{pmatrix} n^{-1}I_B & 0 \\ 0 & \rho_{22} \end{pmatrix} - \begin{pmatrix} \rho_{11} & \rho_{12} \\ \rho 21 & \rho_{22} \end{pmatrix} \right\|_1 = \left\| \begin{pmatrix} n^{-1}I_B - \rho_{11} & -\rho_{12} \\ -\rho_{12} & 0 \end{pmatrix} \right\|_{tr} \\ \leq& \|\rho_{11}\|_{tr} + 2\|\rho_{12}\|_{tr} + \|n^{-1}I_B\|_{tr}. \end{aligned} \tag{8}$$

Next, note that ϵ^4-approximate feasibility implies $\sum_{i=1}^n (|\langle i|\rho|i\rangle - 1/n|) \leq \epsilon^4$. This, in turn, demands $|B|\frac{\epsilon^2}{n} \leq \epsilon^4$ or, equivalently $|B| \leq n\epsilon^2$. The definition of B moreover asserts

$$\|\rho_{22}\|_{tr} \geq (n - |B|)\tfrac{1-\epsilon^2}{n} \geq (1-\epsilon^2)^2.$$

Moreover, as shown in [Kin03], we have

$$\left\| \begin{bmatrix} \|\rho_{11}\|_{tr} & \|\rho_{12}\|_{tr} \\ \|\rho_{12}^T\|_{tr} & \|\rho_{22}\|_{tr} \end{bmatrix} \right\|_{tr} \leq \left\| \begin{bmatrix} \rho_{11} & \rho_{12} \\ \rho_{12}^T & \rho_2 \end{bmatrix} \right\|_{tr} = \|\rho\|_{tr} = \mathrm{tr}(\rho) = 1.$$

As $\|\cdot\|_{tr} \geq \|\cdot\|_2$ (the Frobenius, or Schatten-2 norm), it follows from the last equation that

$$\|\rho_{11}\|_{tr}^2 + 2\|\rho_{12}\|_{tr}^2 + \|\rho_{22}\|_{tr}^2 \leq 1.$$

And, as $\|\rho_{22}\|_{tr}^2 \geq (1-\epsilon^2)^4$, we conclude $\|\rho_{11}\|_{tr}^2 + 2\|\rho_{12}\|_{tr}^2 = \mathcal{O}(\epsilon^2)$. which in turn implies $\|\rho_{11}\|_{tr} + 2\|\rho_{12}\|_{tr} = \mathcal{O}(\epsilon)$. Inserting this relation into Eq. (8) yields

$$\|\rho' - \rho\|_{tr} = \mathcal{O}(\epsilon).$$

Next, note that we obtain R from ρ' by just replacing all diagonal entries of ρ' by $1/n$. This matrix is not necessarily positive semidefnite, but as all diagonals of ρ' are in the range $(1 \pm \epsilon^2)/n$, it is easy to see that the matrix $R + \frac{\epsilon^2}{n} I$ is psd and has diagonal entries equal to $\left(1+\epsilon^2\right)/n$. Thus, $\rho^\#$ is a feasible point of the renormalized MaxQP SDP (3). It is easy to see that

$$\mathrm{tr}\left(An\rho^\sharp\right) = \tfrac{n}{1+\epsilon^2}\left(\mathrm{tr}\left(AR\right) + \tfrac{\epsilon^2}{n}\mathrm{tr}\left(A\right)\right).$$

Now note that

$$\left|\mathrm{tr}\left(AR\right) - \mathrm{tr}\left(A\rho'\right)\right| \leq \tfrac{\epsilon^2}{n}\left(\sum_i |A_{ii}|\right),$$

as these two matrices only differ on the diagonal, and there by at most ϵ^2/n. Since $\frac{1}{n}\sum_i |A_{ii}| \leq \|A\|$, we conclude that

$$\left|\mathrm{tr}\left(An\rho^\sharp\right) - \mathrm{tr}\left(An\rho'\right)\right| = \mathcal{O}\left(\epsilon^2(1+\epsilon^2)^{-1}n\|A\|\right) = \mathcal{O}(\epsilon n\|A\|).$$

The claim then follows from combining triangle and (matrix) Hölder inequality:

$$\left|\mathrm{tr}\left(nA\rho\right) - \mathrm{tr}\left(nA\rho'\right)\right| \leq n\|A\|\|\rho - \rho'\|_{tr} = \mathcal{O}(n\|A\|\epsilon).$$

Notet that the proof technique above is constructive and allows us to construct a feasible point from an approximately feasible one in $\mathcal{O}(n^2)$ time by manipulating the entries. □

3.3 Approximately solving to the MaxQP SDP on a classical computer

We will now show how to use Hamiltonian Updates (Algorithm 1) to solve the MaxQP SDP (3) on a classical computer. It turns out that the main classical bottleneck is the cost of computing matrix exponentials $\rho = \exp(-H)/\mathrm{tr}\left(\exp(-H)\right)$. The following result,also observed in [LRS15], asserts that coarse truncations of the matrix exponential already yield accurate approximations.

Lemma 3.2. *Fix a Hermitian $n \times n$ matrix H, an accuracy ϵ and let l be the smallest even number that obeys $(l+1)(\log(l+1)-1) \geq 2\|H\| + \log(n) + \log(1/\epsilon)$. Then, the truncated matrix exponential $T_l = \sum_{k=0}^{l} \frac{1}{k!}(-H)^k$ is guaranteed to obey*

$$\left\| \frac{\exp(-H)}{tr\left(\exp(-H)\right)} - \frac{T_l}{tr\left(T_l\right)} \right\|_{tr} \leq \epsilon.$$

Proof. First note, that truncation at an even integer l ensures that T_l is positive semidefinite. This is an immediate consequence of the fact that even-degree Taylor expansions of the (scalar) exponential are non-negative polynomials. In particular, $\|T_l\|_{tr} = \mathrm{tr}\left(T_l\right)$. Combine this with $\mathrm{tr}\left(X\right) \leq \|X\|_{tr} \leq n\|X\|$ for all Hermitian $n \times n$ matrices to conclude

$$\begin{aligned}\left\| \frac{\exp(-H)}{\mathrm{tr}\left(\exp(-H)\right)} - \frac{T_l}{\mathrm{tr}\left(T_l\right)} \right\|_{tr} \leq & \frac{1}{\mathrm{tr}\left(\exp(-H)\right)} \left\|\exp(-H) - T_l\right\|_{tr} + \frac{\left|\mathrm{tr}\left(\exp(-H)\right) - \mathrm{tr}\left(T_l\right)\right|}{\mathrm{tr}\left(T_l\right)\mathrm{tr}\left(\exp(-H)\right)}\|T_l\|_{tr} \\ \leq & \frac{2\|\exp(-H) - T_l\|_{tr}}{\mathrm{tr}\left(\exp(-H)\right)} \leq 2n\exp(\|H\|)\|\exp(-H) - T_l\|,\end{aligned}$$

where we have also used $\mathrm{tr}\left(\exp(-H)\right) \geq \|\exp(-H)\| \geq \exp(-\|H\|)$. By construction, both $\exp(-H)$ and T_l commute and are diagonal in the same eigenbasis. Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of H. Then, Taylor's remainder theorem asserts

$$\|\exp(-H) - T_l\| = \max_{1 \leq i \leq n} \left| \exp(-\lambda_i) - \sum_{k=0}^{l} \tfrac{1}{k!}(-\lambda)^k \right| \leq \frac{\max_i \exp\left(-\lambda_i\right)}{(l+1)!} \leq \frac{\exp(\|H\|)}{(l+1)!}.$$

The value of l is chosen such that

$$\frac{2n\exp(2\|H\|)}{(l+1)!} \leq \exp(2\|H\| + \log(2) + \log(n) - 1 - (l+1)(\log(l+1)-1)) \leq \epsilon,$$

because $(l+1)! \geq \mathrm{e}\left((l+1)/\mathrm{e}\right)^{l+1}$. □

Corollary 3.1. *Given an s sparse, symmetric $n \times n$ matrix A and $\epsilon > 0$, we can solve the* MaxQP SDP (3) *up to an additive error $\mathcal{O}(\epsilon n\|A\|)$ in time $\tilde{\mathcal{O}}(\min\{n^2 s\}, n^\omega\}\epsilon^{-12})$ in a classical computer.*

Although the dependency in ϵ for our algorithm is high, we expect that a more refined analysis of the error could improve this significantly. This is because the approximately feasible to feasible conversion behind Proposition 3.1 requires ϵ^4 accuracy.

Proof. As each run of Algorithm 1 takes time $\tilde{\mathcal{O}}(1)$, we only need to implement the required oracles in time $\tilde{\mathcal{O}}(n^2 s\epsilon^{-1})$ to establish the advertised runtime for an approximate solution. First, note that the operator norm $\|H_t\|$ only grows modestly with the number of iterations $t = 0, \ldots, T$. This readily follows from $H_0 = 0$, and $\|H_{t+1} - H_t\| \leq \frac{\epsilon}{8}\|P_t\| \leq \frac{\epsilon}{8}$. What is more, the maximal number of steps is $T = \lceil 16 \log(n)/\epsilon^2 \rceil$, implying $\|H_t\| \leq 2\log(n)\epsilon$ for all t.

In turn, Lemma 3.2 implies that computing the Taylor series of $\exp(-H_t)$ up to a term of order $\mathcal{O}(\log(n)/\epsilon)$ suffices to compute a matrix $\tilde{\rho}_t$ that is ϵ-close to the true iterate $\rho_t = \exp(-H_t)/\mathrm{tr}\left(\exp(-H_t)\right)$ in trace distance. Now note that the complexity of multiplying any matrix with H_t is $\mathcal{O}(\min\{n^2 s\}, n^\omega\})$, as H_t is a linear combination of a diagonal matrix and A. Thus, we conclude that computing $\tilde{\rho}_t$ takes time $\mathcal{O}(n^2 s \log(n)\epsilon^{-1})$. Checking the diagonal constraints then takes time $\mathcal{O}(n)$ and computing $\mathrm{tr}\left(A\|A\|^{-1}\tilde{\rho}_t\right)$ takes time $\mathcal{O}(ns)$. This suffices to implement both ϵ-separation oracle and highlights that the runtime is dominated by computing approximations of the matrix exponential.

Finally, we show in Proposition 3.1 that in order to ensure an additive error of order $\mathcal{O}(\epsilon n\|A\|)$ for the MaxQP SDP, it suffices to solve the relaxed one up to an error ϵ^4, from which the claim follows and we can then convert the approximately feasible solution to a feasible solution in time $\mathcal{O}(n^2)$. □

3.4 Approximately solving to the MaxQP SDP on a quantum computer

We will now show how to implement ϵ-separation oracles on a quantum computer. As discussed before, implementing the oracle requires us to evaluate the diagonal of the sequence of Gibbs states $\rho = \exp(-H)/\mathrm{tr}\left(\exp(-H)\right)$ and the value of $\mathrm{tr}\left(\rho A\|A\|^{-1}\right)$. These two tasks can be performed easily on a quantum computer given the ability to prepare copies of the quantum state ρ.

Lemma 3.3. *We can implement ϵ-separation oracles for the* MaxQP SDP (3) *on a quantum computer given access to $\mathcal{O}(n\epsilon^{-2})$ copies of the the input state ρ and the ability to measure $tr(A\rho)\,\|A\|^{-1}$. Moreover, the classical postprocessing time needed to implement the oracle is $\mathcal{O}(n\epsilon^{-2})$.*

Proof. We implement the oracle by first measuring $\mathcal{O}(n\epsilon^{-2})$ copies of the input ρ in the computational basis. This is enough to ensure that with probability of failure at most $\mathcal{O}(\mathrm{e}^{-n})$ the resulting empirical distribution of the measurement outcomes, $\hat{p} = \sum_i \hat{p}(i)|i\rangle\langle i|$, satisfies

$$\|\sum_i \langle i|\rho|i\rangle |i\rangle\langle i| - \hat{p}\|_{tr} \leq \frac{\epsilon}{2}.$$

If $\|I/n - \hat{p}\|_{tr} \leq \frac{\epsilon}{2}$, then the oracle for the diagonal constraints accepts the current state. If not, we output $P = \sum_i \mathbb{I}\left\{\hat{p}_i > 1/n\right\}|i\rangle\langle i|$. This step requires a classical postprocessing time of order $\mathcal{O}(n\epsilon^{-2})$. For implementing the second oracle, we simply measure $A\|A\|^{-1}$ directly. A total of $\mathcal{O}(\epsilon^{-2})$ copies of ρ suffice to determine $\mathrm{tr}\left(A\|A\|^{-1}\rho\right)$ up to precision ϵ via phase estimation [NC00]. □

Lemma 3.3 reduces the task of implementing separation oracles to the task of preparing independent copies of a fixed Gibbs state. There are many different proposals for prepar-

ing Gibbs states on quantum computers [CS17, Fra18, KBa16, PW09, TOV$^+$09, TOV$^+$09, YAG12, vAGGdW17]. Here, we will follow the algorithm proposed in [PW09]. This approach allows us to reduce the problem of preparing $\rho_H = \exp(-H)/\mathrm{tr}\left(\exp(-H)\right)$ to the task of simulating the Hamiltonian H. More precisely,[PW09, Appendix] highlights that $\tilde{\mathcal{O}}\left(\sqrt{n}\epsilon^{-3}\right)$ invocation of a controlled U, where U satisfies

$$\|U - e^{it_0 H}\| \leq \mathcal{O}(\epsilon^3) \quad \text{where} \quad t_0 = \pi/(4\|H\|)$$

suffice to produce a state that is ϵ close in trace distance to ρ_H. The probability of failure is at most $n^{-1/\epsilon}\epsilon^2$. We expect that a more refined analysis can lead to a better dependence on the error ϵ. The methods presented in [vAGGdW17] seem like a good starting point for such future improvements. Here, however, we prioritize the scaling in the problem dimension n only.

By construction, the Hamiltonians we wish to simulate are all of the form $H = aA\|A\|^{-1} + bD$, where $a, b = \mathcal{O}(\log(n)\epsilon^{-1})$ and D is a diagonal matrix with bounded operator norm $\|D\| \leq 1$. It follows from [CW12, Theorem 1] that $\tilde{\mathcal{O}}\left(t\exp(1.6\sqrt{\log\log(n)t\epsilon^{-1}})\right)$ separate simulations of aA and bD suffice to simulate H for time t up to an error ϵ. Thus, we further reduce the problem of simulating H to simulating A and D separately.

At this point it is important to specify input models for the matrix A, the problem description of the MaxQP SDP. We will work in the *sparse oracle input model*. That is, we assume to have access to an oracle O_{sparse} that gives us the position of the nonzero entries. Given indices i for a column of A and a number $1 \leq j \leq s$, where A is s-sparse, the oracle acts as:

$$O_{\mathrm{sparse}}\,|i,j\rangle = |i, f(i,j)\rangle\,.$$

Here $f(i,j)$ is the $j-$th nonzero element of the $i-$th column of A. Moreover, we assume that the magnitude of individual entries are also accessible by means of another oracle:

$$O_A\,|i,j,z\rangle = |i,j,z \oplus (A_{ij}\|A\|^{-1})\rangle\,,$$

Here, the entry $\left[A\|A\|^{-1}\right]_{ij}$ is represented by a bit string long enough to ensure the desired precision. The results of [Low19] then highlight that it is possible to simulate $\exp(itA\|A\|^{-1})$ in time $\mathcal{O}\left(\left(t\sqrt{s}\|H\|\right)^{1+o(1)}\epsilon^{o(1)}\right)$.

Let us now turn to the task of simulating diagonal Hamiltonians D. Let O_D be the matrix entry oracle for D. We suppose that it acts on $\mathbb{C}^n \otimes \left(\mathbb{C}^2\right)^{\otimes m}$, where m is large enough to represent the diagonal entries to desired precision in binary, as

$$O_D\,|i,z\rangle \mapsto |i, z \oplus D_{ii}\rangle\,. \tag{9}$$

It is then possible to simulate $H = D$ for times $t = \tilde{\mathcal{O}}(\epsilon^{-1})$ with $\tilde{\mathcal{O}}(1)$ queries to the oracle O_D and elementary operations [BACS07]. Thus, efficient simulation of e^{-iDt} follows from an efficient implementation of the oracle O_D. The latter can be achieved with a quantum RAM [GLM08]. We consider the quantum RAM model from [Pra14]. There, it is possible to make insertions in time $\tilde{\mathcal{O}}\,(1)$. Thus, given a classical description of a diagonal matrix D, we may update the quantum RAM in time $\tilde{\mathcal{O}}\,(n)$. After we have updated the quantum RAM, we may implement the oracle O_D in time $\tilde{\mathcal{O}}(1)$. Combining all these subroutines establishes the second main result of this work.

Corollary 3.2. *Given an s-sparse, symmetric $n \times n$ matrix A (with appropriate oracle access) and $\epsilon > 0$, we can solve the renormalized* MaxQP SDP (3) *up to an additive error ϵ in time $\tilde{\mathcal{O}}\left(n^{1.5}\left(\sqrt{s}\right)^{1+o(1)}\right)$ on a quantum computer.*

Proof. It follows from Theorem 3.3 that producing $\tilde{\mathcal{O}}(n)$ copies of Gibbs states suffices to implement the oracle. The results of [PW09] then imply that this can be done with $\tilde{\mathcal{O}}(\sqrt{n})$ Hamiltonian simulation steps, which, as discussed above, can be done in time

$$\tilde{\mathcal{O}}\left(n^{0.5}\left(\sqrt{s}\right)^{1+o(1)}\right).$$

□

3.5 Randomized rounding

As pioneered by the seminal work of Goemans and Williamson [GW95], it is possible to use randomized rounding techniques to obtain an approximate solution to the original quadratic optimization problem for certain instances (1). These solutions are in expectation withinin a multiplicative factor of the value of the SDP relaxation (3) and the exact constant depends on the structure of the matrix A . We will explore Rietz's method, as in [AN06], to show that it is possible to perform the rounding on a classical computer to approximate $\|A\|_{\infty \to 1}$ with our approximately feasible solutions to MaxQP SDP and still obtain good approximations.

First, recall that the rounding algorithms usually work by first multiplying a random Gaussian vector by the square root of the solution. The approximate solution is then given by the signs of this random vector. Note that both classical and quantum algorithms output a classical description of the Hamiltonian $H^\sharp$ associated with an approximately optimal, approximately feasible Gibbs state $\rho^\sharp$ to (3). Pseudocode for the rounding algorithm is provided in Algorithm 2. The first important proof ingredient is an adaptation of [AN06, Eq. (4.1)].

Lemma 3.4. *Fix $v, w \in \mathbb{R}^n$ (non-zero) and let $g \in \mathbb{R}^n$ be a random vector with standard normal entries. Then,*

$$\begin{aligned} &\tfrac{\pi}{2}\mathbb{E}\left[sign(\langle v, g\rangle) sign(\langle w, g\rangle)\right] \\ =&\langle \tfrac{v}{\|v\|}, \tfrac{w}{\|w\|}\rangle + \mathbb{E}\left[\left(\langle \tfrac{v}{\|v\|}, g\rangle - \sqrt{\tfrac{\pi}{2}} sign\left(\langle \tfrac{v}{\|v\|}, g\rangle\right)\right)\left(\langle \tfrac{w}{\|w\|}, g\rangle - \sqrt{\tfrac{\pi}{2}} sign\left(\langle \tfrac{w}{\|w\|}, g\rangle\right)\right)\right]. \end{aligned} \tag{10}$$

Proof. In [AN06, Eq. (4.1)] the authors use rotation invariance to establish this identity for two unit vectors. The claim then follows from observing that the distribution of $\text{sign}(\langle v, g\rangle)\text{sign}(\langle w, g\rangle)$ is invariant under scaling both v and w by non-negative numbers. In particular, $v \mapsto v/\|v\|$ and $w \mapsto w/\|w\|$ does not affect the distribution. □

The next step involves a technical continuity argument.

Lemma 3.5. *Fix $\epsilon > 0$ and let ρ be a quantum state s.t.:*

$$\|\sum_i \langle i|\rho|i\rangle |i\rangle\langle i| - I/n\|_{tr} \leq \epsilon^4$$

Define the set $B = \{i \in [i] : \left|\rho_{ii} - \frac{1}{n}\right| > \frac{\epsilon^2}{n}\}$ *and let* $\rho_{\bar{B}}$ *be the submatrix with indices in the complement* $\bar{B}$ *of* B*. Then, the matrix* σ *with entries* $\sigma_{ij} = \frac{\rho_{ij}}{n\sqrt{\rho_{ii}\rho_{jj}}}$ *is a quantum state that obeys* $\|\rho_{\bar{B}} - \sigma_{\bar{B}}\|_{tr} \leq 3\epsilon$*.*

Proof. Note that $\sigma_{\bar{B}} = \mathcal{D}\left(\rho_{\bar{B}}\right)$, where $\mathcal{D}$ is the linear map given by $\mathcal{D}(X) = D_{\bar{B}} X D_{\bar{B}}$ and $D_{\bar{B}}$ is a $|\bar{B}| \times |\bar{B}|$ diagonal matrix with entries $\sqrt{n\rho_{ii}}^{-1}$ for $i \in \bar{B}$. This implies

$$\|\rho_{\bar{B}} - \sigma_{\bar{B}}\|_{tr} = \| \left(\mathrm{id} - \mathcal{D}\right)\left(\rho_{\bar{B}}\right) \|_{tr} \leq \|\mathrm{id} - \mathcal{D}\|_{tr\to tr} \|\rho_{\bar{B}}\|_{tr} \leq \|\mathrm{id} - \mathcal{D}\|_{\infty\to\infty},$$

because $\|\rho_{\bar{B}}\|_{tr} \leq \|\rho\|_{tr} = \mathrm{tr}(\rho) = 1$. Duality of norms and the fact that both id and $\mathcal{D}$ are self-adjoint with respect of the Frobenius inner product $\mathrm{tr}\left(X^T Y\right)$ implies $\|\mathrm{id} - \mathcal{D}\|_{\infty\to\infty} = \|\mathrm{id} - \mathcal{D}\|_{tr\to tr}$. This allows us to bound $\|\mathrm{id} - \mathcal{D}\|_{\infty\to\infty}$ instead. By construction we have that all the entries of $D_{\bar{B}}$ are in $1 \pm \epsilon$. Write $D_{\bar{B}} = I + D_\epsilon$, where D_ϵ is a diagonal matrix with entries that are bounded by ϵ in absolute value. Then,

$$\mathrm{id} - \mathcal{D}(X) = D_\epsilon X + X D_\epsilon + D_\epsilon X D_\epsilon \quad \text{for any matrix} \quad X.$$

Submultiplicativity of the operator norm then implies

$$\|D_\epsilon X + X D_\epsilon + D_\epsilon X D_\epsilon\|_\infty \leq 2\|D_\epsilon\|_\infty \|X\| + \|D_\epsilon\|_\infty^2 \|X\|_\infty \leq 3\epsilon \|X\|_\infty.$$

and, in turn, $\|\mathrm{id} - \mathcal{D}\|_{\infty\to\infty} \leq 3\epsilon$. □

We are now ready to prove the main stability result required for randomized rounding.

Theorem 3.1. *Let* $\rho^\sharp$ *be an approximately feasible, optimal point of* (3) *with accuracy* $\epsilon^4 > 0$ *and input matrix* A' *with*

$$A' = \begin{pmatrix} 0 & A \\ A^T & 0 \end{pmatrix},$$

where A *is a real* $n \times n$ *matrix. Let* $v_1, \ldots, v_{2n}$ *be the columns of* $\sqrt{\rho^\sharp}$*, sample* $g \in \mathbb{R}^{2n}$ *with i.i.d. Gaussian entries and set* $x_i = sign(\langle v_i, g\rangle)$ *and* $y = (x_1, \ldots, x_n), z = (x_{n+1}, \ldots, x_{2n})$*. Then,*

$$tr\left(\rho^\sharp A\right) n + \mathcal{O}(\epsilon n \|A\|) \geq \sum_{i,j} A_{ij} \mathbb{E}(y_i z_j) \geq (4/\pi - 1) tr\left(\rho^\sharp A\right) n - \mathcal{O}(\epsilon n \|A\|).$$

Proof. The upper bound follows immediately from the fact MaxQP SDP (2) relaxations (renormalized or not) provide upper bounds to the original problem (1). The factors $n\|A\|$ is an artifact of the renormalization (3).

For the lower bound, we once more define $B = \{i \in [i] : |\rho_{ii} - 1/2n| \geq \epsilon^2/2n\} \subset [2n]$. Plugging in v_i and v_j in (10), multiplying both sides by A'_{ij} and summing over i, j implies

$$\frac{\pi}{2} \sum_{i,j} A'_{ij} \mathbb{E}(x_i x_j) = 2n \sum_{i,j} A'_{ij} \left(\sigma_{ij} + \tau_{ij}\right) \quad \text{with} \quad \sigma_{ij} = \frac{\rho_{ij}}{2n\sqrt{\rho_{ii}\rho_{jj}}} \quad \text{and}$$

$$\tau_{ij} = \mathbb{E}\left[\left(\langle \tfrac{v_i}{\|v_i\|}, g\rangle - \sqrt{\tfrac{\pi}{2}} \mathrm{sign}\left(\langle \tfrac{v_i}{\|v_i\|}, g\rangle\right)\right)\left(\langle \tfrac{v_j}{\|v_j\|}, g\rangle - \sqrt{\tfrac{\pi}{2}} \mathrm{sign}\left(\langle \tfrac{v_j}{\|v_j\|}, g\rangle\right)\right)\right].$$

Following the same proof strategy as in [AN06, Sec. 4.1], we note that the matrix T defined by $[T]_{ij} = \tau_{ij}$ is a Gram matrix and, thus, psd. Moreover, in [AN06, Sec. 4.1] the author also shows that $\tau_{ii} = \frac{\pi}{2} - 1$. These two properties imply that $\left(\frac{\pi}{2} - 1\right)^{-1} (2n)^{-1} T$ is a feasible point of (3). Moreover, because of the structure of the matrix A', we have that

$$|\mathrm{tr}\left(TA'\right)| \leq \left(\frac{\pi}{2} - 1\right) \mathrm{tr}\left(\rho^{\sharp} A'\right) n - \mathcal{O}(\epsilon n \|A\|) \tag{11}$$

To see this, consider the block unitary

$$U = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}.$$

Then for any psd matrix X we have that $\mathrm{tr}\left(A'UXU^{\dagger}\right) = -\mathrm{tr}\left(A'X\right)$ and so $\mathrm{tr}\left(A'U\rho^{\sharp}U^{\dagger}\right)$ provides a lower bound to the value over the approximately feasible set . Thus,

$$\begin{aligned} &\frac{\pi}{2} \sum_{i,j} A'_{ij} \mathbb{E}(y_i z_j) = 2n \sum_{i,j} A'_{ij} \left(\sigma_{ij} + \tau_{ij}\right) \geq \\ &2n \sum_{i,j} A'_{ij} \sigma_{ij} - \left(\frac{\pi}{2} - 1\right) \mathrm{tr}\left(\rho^{\sharp} A'\right) n - \mathcal{O}(\epsilon n \|A\|) \end{aligned}$$

We now have to relate $\mathrm{tr}\left(\rho^{\sharp} A'\right)$ with $\mathrm{tr}\left(\sigma A'\right)$. To do so, we can argue like in Proposition 3.1 and see that $\mathrm{tr}\left(\sigma_{11}\right), \mathrm{tr}\left(\rho_{11}\right) = \mathcal{O}(\epsilon^2)$ (these correspond to the $|B| \times |B|$ psd submatrices with entries in B only). As both σ and ρ are states, we conclude

$$\|\rho_{12}\|_{tr}, \|\sigma_{12}\|_{tr} = \mathcal{O}(\epsilon)$$

by reusing the analysis provided in the proof of Proposition 3.1. Thus, it follows from Hölders inequality and Lemma 3.5 that

$$\begin{aligned} \mathrm{tr}\left(A"(\rho - \sigma)\right) =& \mathrm{tr}\left(A'(\rho_{22} - \sigma_{22})\right) + \mathrm{tr}\left(A'(\rho_{11} - 2\rho_{12} - \sigma_{11} - 2\sigma_{12})\right) \\ =& \|A\| \left(\|\sigma_{22} - \rho_{22}\|_{tr} + \|\rho_{11}\|_{tr} + 2\|\rho_{12}\|_{tr} + 2\|\sigma_{11}\|_{tr} + 2\|\sigma_{12}\|_{tr}\right) = \mathcal{O}(\|A\|\epsilon), \end{aligned}$$

from which the claim follows. □

Proposition 3.1 highlights that performing the rounding with approximate solutions to the MaxQP SDP (3) still ensures a good approximate solution in expectation for the $\|A\|_{\infty \to 1}$ norm. In the case of matrices A that are psd it also possible to improve the constant in the rounding and we do not to resort to lifting the problem to a matrix with double the dimension:

Corollary 3.3. *Let $\rho^{\sharp}$ be an approximately feasible, optimal point of (3) with accuracy $\epsilon^4 > 0$ and psd input matrix A. Let $v_1, \ldots, v_n$ be the columns of $\sqrt{\rho^{\sharp}}$, sample $g \in \mathbb{R}^n$ with i.i.d. Gaussian entries and set $x_i = sign(\langle v_i, g \rangle)$. Then,*

$$tr\left(\rho^{\sharp} A\right) n + \mathcal{O}(\epsilon n \|A\|) \geq \sum_{i,j} A_{ij} \mathbb{E}(x_i x_j) \geq (2/\pi) tr\left(\rho^{\sharp} A\right) n - \mathcal{O}(\epsilon n \|A\|).$$

Proof. The proof follows by following the same proof as above but noting that we may use the estimate $\operatorname{tr}(TA) \geq 0$ instead of (11), as both A and T are psd. Optimality of the constant was shown in [AN06]. □

As Alon [AN06] also shows that for psd matrices A we have

$$\|A\|_{\infty\to 1} = \max_{x\in\{\pm 1\}^n} \langle x, A\rangle,$$

i.e. we may restrict to the same vector on the left and right, it follows that Corollary 3.3 gives almost optimal rounding guarantees. These two statements certify that, as longs as $\|A\|_{\infty\to 1} = \Theta(n\|A\|)$, performing the rounding with our approximately feasible solutions gives rise to approximations of the $\ell_1 \to \ell_\infty$ norm that are almost as good the strictlty feasible solutions.

But computing $\sqrt{\rho^\sharp}g = \exp(-H/2)g/\sqrt{\operatorname{tr}(\exp(-H))}$ directly stilll remains expensive because of matrix exponentiation. We will surpass this bottleneck by truncating the Taylor series of the matrix exponential in a fashion similar to Lemma 3.2. The following standard anti-concentration result for Gaussian random variables will be essential for this argument.

Fact 3.1. *Let X be a $\mathcal{N}(0,\sigma^2)$ random variable. Then $\mathbb{P}(|X| \leq \sigma\epsilon) = \mathcal{O}(\epsilon)$.*

Lemma 3.6. *Let $\rho^\sharp$ with associated Hamiltonian $H^\sharp$ be an approximately optimal solution to the MAQP SDP (3) with $\|H^\sharp\| = \mathcal{O}(\log(n)/\epsilon)$. Set $S_l = \sum_{k=0}^{l} \frac{1}{k!}(-H^\sharp/2)^k$ with $l = \mathcal{O}(\log(n)/\epsilon)$. Then, a random vector $g \in \mathbb{R}^n$ with standard normal entries obeys*

$$\textit{sign}\left[\left(e^{H^\sharp/2}g\right)_i\right] = \textit{sign}\left[(S_l g)_i\right] \quad \textit{for all} \quad i \in [n] \quad \textit{such that} \quad \left|\rho^\sharp_{ii} - \tfrac{1}{n}\right| < \tfrac{\epsilon}{n}$$

with probability at least $1 - \mathcal{O}(\epsilon^{-1})$.

Note that the design of Algorithm 1 ensures that optimal Hamiltonians always obey $\|H^\sharp\| = \mathcal{O}(\log(n)/\epsilon)$.

Proof. Define $h = \exp(-H^\sharp/2)g$ and note that this is a Gaussian random vector with covariance matrix $\exp(-H^\sharp)$. Let $B = \left\{i : \ |\rho_{ii} - 1/n| > \frac{\epsilon}{n}\right\} \subset [n]$ denote the set of indices for which ρ_{ii} deviates substantially from $1/n$. Then, every entry of h that is not contained in this index set obeys

$$[h]_i = [\exp(-H/2)g]_i \sim \mathcal{N}\left(0, \tfrac{c}{n}\operatorname{tr}(\exp(-H))\right) \quad \text{with} \quad c \in (1-\epsilon, 1+\epsilon).$$

The assumption $\|H^\sharp\| = \mathcal{O}(\log(n)/\epsilon)$ ensures $\operatorname{tr}(\exp(-H^\sharp))/n \geq n^{-c'/\epsilon - 1}$ for some constant c'. We can combine this with Fact 3.1 (Gaussian anti-concentration) to conclude

$$\mathbb{P}\left[|\,[h]_i\,| \leq n^{-2-c'/(2\epsilon)}\right] = \mathcal{O}(1/n^2) \quad \text{for all} \quad i \in \bar{B} = [n] \setminus B.$$

A union bound then asserts

$$\mathbb{P}\left[\exists i \in \bar{B} : \ |\,[h]_i\,| \leq n^{-2-c'/\epsilon}\right] = \mathcal{O}(1/n).$$

Moreover, it follows from standard concentration arguments that

$$\mathbb{P}\left[n - n^{\frac{1}{4}} \leq \|g\|^2 \leq n + n^{\frac{1}{4}}\right] \geq 1 - 2e^{-\sqrt{n}/8}.$$

Thus, with probability at least $1 - \mathcal{O}(n^{-1})$, we have that $\|g\|^2 \leq n + n^{\frac{1}{4}}$ and $|\,[h]_i\,| \geq n^{-2-c'/\epsilon}$ for every entry $i \in \bar{B}$. Following the same proof strategy as in Lemma 3.2, it is easy to see that by picking $l = \mathcal{O}(\epsilon^{-1}\log(n))$ suffices to ensure that

$$\|S_l - \exp(-H/2)\| \leq n^{-4-\frac{c'}{2\epsilon}}$$

Conditioning on the events emphasized above, implies

$$\max_{i\in[n]} |[(\exp(-H/2) - S_l)\, g]_i| \leq \|\,(\exp(-H/2) - S_k)\, g\| \leq \|\exp(-H/2) - S_k\|\|g\| \leq n^{-4-\frac{c'}{2\epsilon}}\|g\|.$$

This in turn ensures $\max_{i\in\bar{B}} |[(\exp(-H/2) - S_l)\, g]_i| \leq n^{-3-\frac{c'}{2\epsilon}}$, and in turn

$$\operatorname{sign}\left([h]_i\right) = \operatorname{sign}\left([(\exp(-H/2)g)]_i\right) = \operatorname{sign}\left([S_k g]_i\right) \quad \text{for all} \quad i \in \bar{B},$$

because conditioning ensures $|[\exp(-H/2)g]_i| \geq n^{-2-\frac{c'}{2\epsilon}}$. □

Combining the statements we just proved we conclude that:

Proposition 3.2 (Restatement of Proposition 2.1)**.** *Let $\epsilon > 0$ and A a real, psd matrix be given. Moreover, let H be the solution Hamiltonian to the relaxed* MaxQP SDP (3) *with error paramater ϵ^4 and α^* its value. Then, with probability at least $1 - n^{-1}$, the output x of Algorithm 2 satisfies:*

$$n\|A\|(\alpha^* + \mathcal{O}(\epsilon)) \geq \mathbb{E}\big[\sum_{ij} A_{ij}x_i x_j\big] \geq \frac{2}{\pi} n\|A\|(\alpha^* - \mathcal{O}(\epsilon)), \tag{12}$$

Proof. It follows from Lemma 3.6 that the output of Algorithm 2 will only differs from the vector obtained by performing the rounding with the approximate solution on a set of size $\mathcal{O}(n\epsilon^2)$ with probability at least $1 - n^{-1}$. This is because, as argued before, by picking ϵ^4 we have at most $\mathcal{O}(\epsilon^2 n)$ diagonal entries that do not satsify $|\rho_{ii} - 1/n| \leq \epsilon/n$. We will now argue that sign vectors that differ at $\mathcal{O}(n\epsilon^2)$ position can differ in value by at most $\mathcal{O}(\epsilon n\|A\|)$. Let x be the vector obtained by the ideal rounding and x' the one with the truncated Taylor series. Then there exists a vector e with at most $\mathcal{O}(n\epsilon^2)$ nonzero entries bounded by 2 such that $x = x + e$ by our assumption. By Cauchy-Schwarz:

$$|\langle x, Ax\rangle - \langle x', Ax'\rangle| \leq |\langle e, Ax\rangle| + |\langle x, Ae\rangle| + |\langle e, Ae\rangle| \leq \|A\|\left(2\|x\|\|e\| + \|e\|^2\right).$$

Now, as x is a binary vector, $\|x\| = \sqrt{n}$ and, as e has at most $\mathcal{O}(\epsilon^2 n)$ nonzero entries, it follows that $\|e\| = \mathcal{O}(\epsilon\sqrt{n})$ and we conclude that:

$$|\langle x, Ax\rangle - \langle x', Ax'\rangle| = \mathcal{O}(\epsilon n)$$

As Theorem 3.1 asserts that performing the rounding with the approximate solution is enough to produce a sign vector that satisfies (12) in expectation, this yields the claim. □

The analogous claim, i.e. that truncating still gives rise good solutions, clearly also holds in the setting of Proposition 3.1.

Thus, we conclude that the rounding can be performed in time $\tilde{\mathcal{O}}(ns)$ on a classical computer, as multiplying a vector with H takes time $\tilde{\mathcal{O}}(ns)$ and we only need to perform this operations a logarithmic number of steps. As $ns \leq n^{1.5}\sqrt{s}$ for $s \leq n$, we conclude that the cost of solving the relaxed MaxQP SDP (3) dominates the cost of rounding.

4 Conclusion and Outlook

By adapting ideas from [TRW05, Haz06, LSW15, BKL$^+$17], we have provided a general meta-algorithm for approximately solving convex feasibility problems with psd constraints. *Hamiltonian Updates* is an iterative procedure based on a simple change of variables: represent a trace-normalized, positive semidefinite matrix as $X = \exp(-H)/\mathrm{tr}\left(\exp(-H)\right)$. At each step, infeasible directions are penalized in the matrix exponent until an approximately feasible point is reached. This procedure can be equipped with rigorous convergence guarantees and lends itself to quantum improvements: $X = \exp(-H)\mathrm{tr}\left(\exp(-H)\right)$ is a *Gibbs state* and H is the associated *Hamiltonian*. Quantum architectures can produce certain Gibbs states very efficiently.

We have demonstrated the viability of this approach by considering semidefinite programming relaxations of quadratic problems with binary constraints (MaxQP SDP) (2). The motivation for considering this practically important problem class was two-fold: (i) MaxQP SDPs have received a lot of attention in the (classical) computer science community. Powerful meta-algorithms, like matrix multiplicative weights [AK16], have been designed to solve these SDPs very quickly. (ii) Existing quantum SDP solvers [BS17, vAG18b, BKL$^+$17, KP18] do not provide *any* speedups for MaxQP SDPs. The quantum runtime associated with these solvers depends on problem-specific parameters that scale particularly poorly for MaxQP SDPs. Moreover, the notions of approximate feasibility championed in these other works are too loose for this class of problem.

The framework developed in this paper has allowed us to address these points. Firstly, we showed that a classical implementation of Hamiltonian Updates already improves upon the best existing results. A runtime of $\tilde{\mathcal{O}}(n^2 s)$ suffices to find an approximately optimal solution. Secondly, we have showed that quantum computers do offer additional speedups. A quantum runtime of $\tilde{\mathcal{O}}(n^{1.5}s^{0.5+o(1)})$ is sufficient. We emphasize that this is the first quantum speedup for the practically important class of MaxQP SDP relaxations. Subsequently, we have also devised a classical randomized rounding procedure that converts both quantum and classical solutions into close-to-optimal solutions of the original quadratic problem.

We note in passing that our algorithm is very robust, in the sense that it only requires the preparation of Gibbs states up to a constant precision, computational basis measurements and the ability to estimate the expectation value of the target matrix on states. Although the subroutines used in this work to perform these tasks certainly require nontrivial quantum circuits, it would be interesting to identify classes of target matrices A for which preparing the corresponding Gibbs state and estimating the expectation values is feasible on near-term devices.

We believe that the framework presented here lends itself to further applications.

One concrete application of Hamiltonian Updates, in particular the idea to treat constraints as the statistics of measurements, would be quantum speedups for *quantum state tomography*, see e.g. [BCG13] and references therein. Sample-optimal tomography protocols have revealed that classical post-processing is the main bottleneck for reconstructing density matrices [FGLE12, OW16, HHJ$^+$17, GKKT18]. A natural starting point for establishing quantum speedups is tomography via low-rank matrix reconstruction [GLF$^+$10]. There, reconstruction is achieved by solving an SDP that penalizes the rank of the density matrix. While solving the SDP classically is rather expensive, this approach allows for considerably reducing the number of measurement settings. For instance, an order of $\mathrm{rank}(\rho)$ sufficiently random basis measurements suffice to accurately reconstruct ρ [Vor13, Kue15, CHK$^+$16]. The basis structure of these measurements equips the resulting reconstruction SDP with a structure that closely resembles the MaxQP SDP (3). We will address quantum speedups for such SDPs in upcoming work.

Another promising and practically relevant application is *binary matrix factorization*. A recent line of works [KT19a, KT19b] reduces this problem to a sequence of SDPs. Importantly, each SDP corresponds to a MAXQP SDP (2) with a random rank-one objective $A = |a\rangle\langle a|$ and an additional affine constraint $\mathrm{tr}\,(PX) = n$. Here, P is a fixed low-rank orthoprojector. This application, however, is likely going to be more demanding in terms of approximation accuracy. Hence, improving the runtime scaling in inverse accuracy will constitute an important first step that is also of independent interest.

5 Acknowledgments

We would like to thank Aram Harrow for inspiring discussions. Our gratitude extends, in particular, to Ronald de Wolf, András Gilyén and Joran van Apeldoorn who provided valuable feedback regarding an earlier version of this draft. D.S.F. would like to thank the hospitality of Caltech's Institute for Quantum Information and Matter, where the main ideas in this paper were conceived during a visit. F.G.L.S.B and R.K. acknowledge funding provided by the Institute for Quantum Information and Matter, an NSF Physics Frontiers Center (NSF Grant PHY-1733907), as well as financial support from Samsung. R.K's work is also supported by the Office of Naval Research (Award N00014-17-1-2146) and the Army Research Office (Award W911NF121054). D.S.F. acknowledges financial support from VILLUM FONDEN via the QMATH Centre of Excellence (Grant no. 10059), the graduate program TopMath of the Elite Network of Bavaria, the TopMath Graduate Center of TUM Graduate School at Technische Universität München and by the Technische Universität München – Institute for Advanced Study, funded by the German Excellence Initiative and the European Union Seventh Framework Programme under grant agreement no. 291763.

References

[AFdlVKK03] N. Alon, W. Fernandez de la Vega, R. Kannan, and M. Karpinski. Random sampling and approximation of MAX-CSPs. volume 67, pages 212–243. 2003. Special issue on STOC2002 (Montreal, QC).

[AHK05] S. Arora, E. Hazan, and S. Kale. Fast algorithms for approximate semidefinite

programming using the multiplicative weights update method. In *46th Annual IEEE Symposium on Foundations of Computer Science (FOCS'05)*, pages 339–348, Oct 2005.

[AK16] S. Arora and S. Kale. A combinatorial, primal-dual approach to semidefinite programs. *J. ACM*, 63(2):Art. 12, 35, 2016.

[AL70] H. Araki and E. H. Lieb. Entropy inequalities. *Comm. Math. Phys.*, 18:160–170, 1970.

[AN06] N. Alon and A. Naor. Approximating the cut-norm via Grothendieck's inequality. *SIAM J. Comput.*, 35(4):787–803, 2006.

[BACS07] D. W. Berry, G. Ahokas, R. Cleve, and B. C. Sanders. Efficient quantum algorithms for simulating sparse Hamiltonians. *Comm. Math. Phys.*, 270(2):359–371, 2007.

[BCG13] K. Banaszek, M. Cramer, and D. Gross. Focus on quantum tomography. *New Journal of Physics*, 15(12):125020, dec 2013.

[Bha97] R. Bhatia. *Matrix analysis*, volume 169 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1997.

[BKL+17] F. G. Brandão, A. Kalev, T. Li, C. Y.-Y. Lin, K. M. Svore, and X. Wu. Exponential quantum speed-ups for semidefinite programming with applications to quantum learning. *arXiv preprint arXiv:1710.02581*, 2017.

[BM05] S. Burer and R. D. C. Monteiro. Local minima and convergence in low-rank semidefinite programming. *Math. Program.*, 103(3, Ser. A):427–444, 2005.

[BS17] F. G. S. L. Brandao and K. M. Svore. Quantum speed-ups for solving semidefinite programs. In *58th Annual IEEE Symposium on Foundations of Computer Science—FOCS 2017*, pages 415–426. IEEE Computer Soc., Los Alamitos, CA, 2017.

[Bub15] S. Bubeck. Convex Optimization: Algorithms and Complexity. *Foundations and Trends® in Machine Learning*, 8(3-4):231–357, 2015.

[BV04] S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge University Press, Cambridge, 2004.

[BVB16] N. Boumal, V. Voroninski, and A. Bandeira. The non-convex burer-monteiro approach works on smooth semidefinite programs. In D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett, editors, *Advances in Neural Information Processing Systems 29*, pages 2757–2765. Curran Associates, Inc., 2016.

[CHK+16] C. Carmeli, T. Heinosaari, M. Kech, J. Schultz, and A. Toigo. Stable pure state quantum tomography from five orthonormal bases. *EPL (Europhysics Letters)*, 115(3):30001, aug 2016.

[CS17] A. N. Chowdhury and R. D. Somma. Quantum algorithms for Gibbs sampling and hitting-time estimation. *Quantum Inf. Comput.*, 17(1-2):41–64, 2017.

[CW04] M. Charikar and A. Wirth. Maximizing Quadratic Programs: Extending Grothendieck's Inequality. In *45th Annual IEEE Symposium on Foundations of Computer Science*, pages 54–60. IEEE, 2004.

[CW12] A. M. Childs and N. Wiebe. Hamiltonian simulation using linear combinations of unitary operations. *Quantum Inf. Comput.*, 12(11-12):901–924, 2012.

[FGLE12] S. T. Flammia, D. Gross, Y.-K. Liu, and J. Eisert. Quantum tomography via compressed sensing: error bounds, sample complexity and efficient estimators. *New Journal of Physics*, 14(9):095022, sep 2012.

[FK99] A. Frieze and R. Kannan. Quick approximation to matrices and applications. *Combinatorica*, 19(2):175–220, 1999.

[Fra18] D. S. França. Perfect sampling for quantum Gibbs states. *Quantum Inf. Comput.*, 18(5-6):361–388, 2018.

[Git13] A. Gittens. *Topics in Randomized Numerical Linear Algebra*. ProQuest LLC, Ann Arbor, MI, 2013. Thesis (Ph.D.)–California Institute of Technology.

[GKKT18] M. Guta, J. Kahn, R. Kueng, and J. A. Tropp. Fast state tomography with optimal error bounds. *arXiv preprint arXiv:1809.11162*, 2018.

[GLF+10] D. Gross, Y.-K. Liu, S. T. Flammia, S. Becker, and J. Eisert. Quantum state tomography via compressed sensing. *Phys. Rev. Lett.*, 105:150401, Oct 2010.

[GLM08] V. Giovannetti, S. Lloyd, and L. Maccone. Quantum random access memory. *Phys. Rev. Lett.*, 100(16):160501, 4, 2008.

[GW95] M. X. Goemans and D. P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *J. Assoc. Comput. Mach.*, 42(6):1115–1145, 1995.

[Haz06] E. Hazan. *Efficient algorithms for online convex optimization and their applications*. 2006.

[HHJ+17] J. Haah, A. W. Harrow, Z. Ji, X. Wu, and N. Yu. Sample-optimal tomography of quantum states. *IEEE Transactions on Information Theory*, 63(9):5628–5641, 2017.

[KB19] D. Kunisky and A. S. Bandeira. A tight degree 4 sum-of-squares lower bound for the sherrington-kirkpatrick hamiltonian. *arXiv preprint arXiv:1907.11686*, 2019.

[KBa16] M. J. Kastoryano and F. G. S. L. Brandão. Quantum Gibbs samplers: the commuting case. *Comm. Math. Phys.*, 344(3):915–957, 2016.

[Kin03] C. King. Inequalities for trace norms of 2×2 block matrices. *Comm. Math. Phys.*, 242(3):531–545, 2003.

[KLP+16] R. Kyng, Y. T. Lee, R. Peng, S. Sachdeva, and D. A. Spielman. Sparsified Cholesky and multigrid solvers for connection Laplacians. In *STOC'16—Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing*, pages 842–850. ACM, New York, 2016.

[Kol98] T. G. Kolda. *Limited-memory matrix methods with applications*. PhD thesis, University of Michigan, 1998.

[KP18] I. Kerenidis and A. Prakash. A quantum interior point method for LPs and SDPs. *arXiv preprint arXiv:1808.09266*, 2018.

[KT19a] R. Kueng and J. A. Tropp. Binary component decomposition Part I: the positive-semidefinite case. *arXiv preprint arXiv:1907.13603*, 2019.

[KT19b] R. Kueng and J. A. Tropp. Binary component decomposition Part II: the asymmetric case. *arXiv preprint arXiv:1907.13602*, 2019.

[Kue15] R. Kueng. Low rank matrix recovery from few orthonormal basis measurements. In *2015 International Conference on Sampling Theory and Applications (SampTA)*, pages 402–406, May 2015.

[Lat05] R. Latala. Some estimates of norms of random matrices. *Proc. Amer. Math. Soc.*, 133(5):1273–1282, 2005.

[Low19] G. H. Low. Hamiltonian simulation with nearly optimal dependence on spectral norm. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing - STOC 2019*, pages 491–502, New York, New York, USA, 2019. ACM Press.

[LRS15] J. R. Lee, P. Raghavendra, and D. Steurer. Lower bounds on the size of semidefinite programming relaxations. In *STOC'15—Proceedings of the 2015 ACM Symposium on Theory of Computing*, pages 567–576. ACM, New York, 2015.

[LSW15] Y. T. Lee, A. Sidford, and S. C.-W. Wong. A faster cutting plane method and its implications for combinatorial and convex optimization. In *2015 IEEE 56th Annual Symposium on Foundations of Computer Science—FOCS 2015*, pages 1049–1065. IEEE Computer Soc., Los Alamitos, CA, 2015.

[MMMO17] S. Mei, T. Misiakiewicz, A. Montanari, and R. I. Oliveira. Solving sdps for synchronization and maxcut problems via the grothendieck inequality. *arXiv preprint arXiv:1703.08729*, 2017.

[Mon18] A. Montanari. Optimization of the sherrington-kirkpatrick hamiltonian. *arXiv preprint arXiv:1812.10897*, 2018.

[NC00] M. A. Nielsen and I. L. Chuang. *Quantum computation and quantum information*. Cambridge University Press, Cambridge, 2000.

[Nik09] V. Nikiforov. Cut-norms and spectra of matrices. *arXiv preprint arXiv:0912.0336*, 2009.

[OP83] D. O'Leary and S. Peleg. Digital image compression by outer product expansion. *IEEE Transactions on Communications*, 31(3):441–444, March 1983.

[OW16] R. O'Donnell and J. Wright. Efficient quantum tomography. In *Proceedings of the Forty-eighth Annual ACM Symposium on Theory of Computing*, STOC '16, pages 899–912, New York, NY, USA, 2016. ACM.

[Pan13] D. Panchenko. *The Sherrington-Kirkpatrick model*. Springer Monographs in Mathematics. Springer, New York, 2013.

[Pra14] A. Prakash. *Quantum algorithms for linear algebra and machine learning*. PhD thesis, University of California, Berkeley, 2014.

[PW09] D. Poulin and P. Wocjan. Sampling from the thermal quantum Gibbs state and evaluating partition functions with a quantum computer. *Phys. Rev. Lett.*, 103(22):220502, 4, 2009.

[RV18] E. Rebrova and R. Vershynin. Norms of random matrices: local and global problems. *Adv. Math.*, 324:40–83, 2018.

[ST11] D. A. Spielman and S.-H. Teng. Spectral sparsification of graphs. *SIAM J. Comput.*, 40(4):981–1025, 2011.

[TOV+09] K. Temme, T. J. Osborne, K. G. Vollbrecht, D. Poulin, and F. Verstraete. Quantum metropolis sampling. *Nature*, 471:87,2011, November 2009.

[TRW05] K. Tsuda, G. Rätsch, and M. K. Warmuth. Matrix exponentiated gradient updates for on-line learning and Bregman projection. *J. Mach. Learn. Res.*, 6:995–1018, 2005.

[TYUC17] J. A. Tropp, A. Yurtsever, M. Udell, and V. Cevher. Practical sketching algorithms for low-rank matrix approximation. *SIAM J. Matrix Anal. Appl.*, 38(4):1454–1485, 2017.

[vAG18a] J. van Apeldoorn and A. Gilyén. Private Communication, 2018.

[vAG18b] J. van Apeldoorn and A. Gilyén. Improvements in quantum SDP-solving with applications. *arXiv preprint arXiv:1804.05058*, 2018.

[vAGGdW17] J. van Apeldoorn, A. Gilyén, S. Gribling, and R. de Wolf. Quantum SDP-solvers: better upper and lower bounds. In *58th Annual IEEE Symposium on Foundations of Computer Science—FOCS 2017*, pages 403–414. IEEE Computer Soc., Los Alamitos, CA, 2017.

[Vor13] V. Voroninski. Quantum tomography from few full-rank observables. *arXiv preprint arXiv:1309.7669*, 2013.

[YAG12] M.-H. Yung and A. Aspuru-Guzik. A quantum–quantum metropolis algorithm. *Proceedings of the National Academy of Sciences*, 109(3):754–759, 2012.

A Norms of random matrices

There is an interesting discrepancy in the error scaling between the methods presented here and existing ones by Arora et al. [AHK05]: $\|A\|_{\ell_1}$ [AHK05] vs $n\|A\|$ (here). The following fundamental relations relate these norms [Nik09]:

$$\|A\|_{\infty\to 1} \leq n\|A\|, \quad \|A\|_{\infty\to 1} \leq \|A\|_{\ell_1}, \quad \|A\| \leq \|A\|_{\ell_1} \leq n\sqrt{\operatorname{rank}(A)}\|A\|.$$

All inequalities are tight up to constants. The above inequalities highlight that it is a priori not clear what the correct scaling for errors approximating the cut norm should be. The goal of this section will be to show that for random matrices A with independent, standardized entries that have bounded fourth moment $n\|A\|$ reproduces the correct error behavior, while $\|A\|_{\ell_1}$ does not.

Proposition A.1 (Cut norm of random matrices)**.** *Let A be a $n \times n$ random matrix whose entries are sampled independently from a real-valued distribution α that obeys $\mathbb{E}[\alpha] = 0$, $\mathbb{E}\left[\alpha^2\right] = 1$ and $\mathbb{E}\left[\alpha^4\right] = \mathcal{O}(1)$. Then,*

$$\mathbb{E}\left[\|A\|_{\ell_1}\right] = \Theta(n^2), \quad \mathbb{E}\left[\|A\|_{\infty\to 1}\right] = \Theta(n^{1.5}), \quad \mathbb{E}\left[\|A\|\right] = \mathcal{O}\left(\sqrt{n}\right).$$

Proof. We refer to Latala's work for the third claim [Lat05]. A key ingredient for establishing the second claim is [Git13, Corollary 3.10]:

$$\tfrac{1}{\sqrt{2}}\mathbb{E}\left(\|A\|_{\text{col}}\right) \leq \mathbb{E}\left(\|A\|_{\infty\to 1}\right) \leq 4\mathbb{E}\left(\|A\|_{\text{col}}\right),$$

where $\|A\|_{\text{col}} = \sum_i \sqrt{\sum_j [A]_{ij}^2}$ is the sum of the Euclidean norms of the columns of A. Now, note that the entries of A are i.i.d. copies of the random variable α. In turn, the expected column norm of A is just n times the expected Euclidean norm of the random vector $a = (a_1, \ldots, a_n)^T$, where each a_i is an independent copy of α. Jensen's inequality then asserts

$$\mathbb{E}\left[\|a\|_2\right] \leq \left(\mathbb{E}\left[\sum_{i=1}^n a_i^2\right]\right)^{1/2} = \sqrt{n\mathbb{E}\left[\alpha^2\right]} = \sqrt{n},$$

while a matching lower bound follows from $\sqrt{x} \geq \frac{1}{2}(1 + x - (x-1)^2)$. Indeed, define $y = \|a\|_2^2/n = \frac{1}{n}\sum_{i=1}^n a_i^2$ and note that this new random variable obeys $\mathbb{E}[y] = 1$ and $\mathbb{E}[(y-1)^2] = \mathcal{O}(1/n)$ by assumption. This ensures a matching lower bound:

$$\mathbb{E}\left[\|a\|_2\right] = \sqrt{n}\mathbb{E}\left[\sqrt{y}\right] \geq \tfrac{\sqrt{n}}{2}\left(1 + \mathbb{E}\left[y\right] - \mathbb{E}\left[(y-1)^2\right]\right) = \Omega(\sqrt{n}),$$

This ensures $\mathbb{E}\left[\|A\|_{\infty\to 1}\right] = n\mathbb{E}\left[\|a\|_2\right] = \Theta(n^{3/2})$ and establishes the second claim.

The first claim follows from the fact that the fourth-moment bound $\mathbb{E}\left[\alpha^4\right] = \mathcal{O}(1)$ demands $\mathbb{E}\left[|\alpha|\right] = \Theta(1)$. Combine this with i.i.d. entries of the random matrix A to conclude

$$\mathbb{E}\left[\|A\|_{\ell_1}\right] = n^2\mathbb{E}\left[|\alpha|\right] = \Theta(n^2).$$

□

Another family of random matrices for which we expect that $n\|\cdot\|$ provides the correct error scaling for cut norms are matrices of the form $B = A^*A$, where A again has i.i.d. entries of mean 0 and unit variance. Indeed, in [RV18] the authors show that

$$\mathbb{E}\left(\|A\|_{\infty\to 2}\right) \leq \mathcal{O}\left(\sqrt{n}\mathbb{E}\left(\|A\|_{2\to\infty}\right)\right).$$

with high probability. One can combine these recent results with more standard relations, like $\|A\|_{2\to\infty}^2 = \|B\|_{1\to\infty}$, $\|A\|_{2\to\infty} \leq \|A\|$ and $\|B\| = \|A\|^2$. This asserts $\mathbb{E}\left[\|B\|_{1\to\infty}\right] \leq n\mathbb{E}\left[\|B\|\right] = \mathcal{O}(n^2)$, while $\mathbb{E}\left[\|B\|_{\ell_1}\right] = \Omega(n^{2.5})$.

B Comparison to previous work and techniques for further improvement

This section is devoted to give a brief overview over some promising proposals for speeding up SDP solvers for problems with similar structure. The main message is that these unfortunately do not immediately apply to the general MaxQP SDP setting.

The main classical bottleneck behind Algorithm 1 is computing matrix exponentials. Dimension reduction techniques, like Johnson-Lindenstrauss, can sometimes considerably speed up this process, see e.g. [AK16]. There, Arora and Kale apply this idea to solve the MAXCUT SDP up to a multiplicative error of $\mathcal{O}(\epsilon nd)$ in time $\tilde{\mathcal{O}}(nd)$ for a d regular graph on n vertices. Moreover, sparsification techniques [vAG18a] can be used to bring this complexity down to $\tilde{\mathcal{O}}(n)$ in the adjacency list model and $\tilde{\mathcal{O}}(\min(nd, n^{1.5}d^{-1}))$ in the adjacency matrix input model. Note that the MaxCut SDP is just an instance of the MaxQP SDP, as both have the same constraints. The only difference is that the MaxQP SDP has the further constraint that the target matrix is the weighted adjacency matrix of a graph and, thus, has positive entries. The extra assumption of non-negative entries is a key ingredient behind the fastest approximate MaxCUT SDP solvers which would outperform the main results of this work. It is therefore worthwhile to discuss why these ideas do not readily extend to more general problem instances.

First, note that the fact that the entries of the target matrix has positive entries is crucial for the soundness of the oracle presented in [AK16, Theorem 5.2]. This already rules out the possibility of directly applying their methods to MAXQP if the matrix A has negative entries. The second crucial observation of [AK16] is that it is possible to rewrite the MaxCut SDP as:

$$\begin{aligned} \text{maximize} \quad & \operatorname{tr}\sum_{i,j}[A]_{ij}\|v_i - v_j\|^2 \\ \text{subject to} \quad & \|v_i\|^2 = 1,\ v_i \in \mathbb{R}^n,\ i \in [n] \end{aligned} \tag{13}$$

In this reformulation, the vectors v_i correspond to columns of a Cholesky-decomposition associated with feasible points: $[X]_{ij} = \langle v_i, v_j\rangle$. Next, recall the following variation of the polarization identity:

$$\langle u, v\rangle = \frac{1}{2}\left(\|u\|^2 + \|v\|^2 - \|u - v\|^2\right).$$

This allows us to rewrite the original objective function as

$$\operatorname{tr}(AX) = \sum_{i,j} [A]_{ij} \langle v_i, v_j \rangle = \frac{1}{2} \sum_{i,j} [A]_{ij} \left(\|v_i\|^2 + \|v_j\|^2 - \|v_i - v_j\|^2 \right).$$

Feasibility of X then demands $1 = \langle i|X|i\rangle = \langle v_i, v_i\rangle = \|v_i\|^2$ and we, thus, only need to optimize over $\|v_i - v_j\|^2$. Subsequently, Arora and Kale apply dimensionality reduction techniques to compute approximate vectors v_i', v_j' that satisfy:

$$\left| \|v_i - v_j\|^2 - \|v_i' - v_j'\|^2 \right| \leq \epsilon \|v_i - v_j\|^2. \tag{14}$$

in time $\mathcal{O}(ns)$. A priori, similar techniques can be applied to the more general MaxQP SDP (3). However, sign problems can substantially affect the approximation error. Pointwise estimates like the one in (14) only suffice to estimate $\operatorname{tr}(XA)$ up to an error of order $\mathcal{O}(\epsilon \|A\|_{\ell_1})$. This is fine for matrices with non-negative entries, where this error scaling is comparable to the size of the optimal SPD solution. Matrix entries with different signs, however, may lead to cancellations that result in a much smaller size of the optimal SDP solution. In summary: adapting the ideas of Arora and Kale [AK16] is advisable in situations, where the problem matrix obeys $\|A\|_{\ell_1} = \Theta(n\|A\|)$. This ensures a correct error behavior and dimension reduction allows for reducing the classical runtime to $\tilde{\mathcal{O}}(ns)$.

Another important technique for complexity reduction in SDPs is sparsification. Once again, one seminal example is MAXCUT, where spectral sparsification methods can be used to reduce the complexity [ST11, KLP+16]. Here, the idea is to find a (usually random) sparser matrix B that has approximately the same cut value as A and then run the algorithm on B instead. Unfortunately, once again signed matrix entries render this approach problematic. Up to our knowledge, the best current sparsification results available for the $\infty \to 1$ are those of [Git13, Chapter 3]. There, the author shows in Corollary 3.9 that if we let B be a random matrix with independent random entries s.t. $\mathbb{E}(B_{ij}) = A_{ij}$, then

$$\mathbb{E}\left[\|A - B\|_{\infty \to 1}\right] \leq 2 \sum_i \sqrt{\sum_j \operatorname{Var}[B_{ij}]}.$$

A necessary pre-requisite for accurate sparsification is therefore $2 \sum_i \sqrt{\sum_j \operatorname{Var}[B_{ij}]} = \mathcal{O}(\epsilon n \|A\|)$ It seems unlikely that it is possible to obtain good and general sparsification bounds from this result. To see why this is the case, note that in order for B to be sparse in expectation, we require that $\mathbb{P}(B_{ij} = 0) = p_{ij}$ for suitably large p_{ij}. This will result in a matrix that has, in expectation, $\sum_{ij}(1 - p_{i,j})$ nonzero entries. To make sure that the number of nonzero entries is not $\mathcal{O}(n^2)$, we need to set many $1 - p_{ij} = o(1)$. Now note that $\mathbb{P}[B_{ij} = 0] = p_{ij}$ and $\mathbb{E}[B_{ij}] = A_{ij}$ necessarily enforce $\mathbb{E}[(B_{ij}^2)] \geq \frac{p_{ij}}{1-p_{ij}} A_{ij}^2$. Thus, we see that we expect this technique to only work in the regime where A has many columns with entries that are $o(1)$ and can be neglected with high probability. Roughly speaking, this corresponds to the regime in which $\|A\|_{\text{col}} \ll n\|A\|$. It is then easy to see that the random matrices considered before do not satisfy this and, thus, we do not expect that those instances can be sparsified.

Last, but not least, we emphasize that it also easy to construct examples where the error term $\|A\|_{\ell_1}$ conveys the right scaling, not $n\|A\|$. A concrete example are extremely sparse matrices, where all but $s \ll n$ of the entries are zero.