A Quantum Interior Point Method for LPs and SDPs

Iordanis Kerenidis [1] Anupam Prakash [1]

[1]CNRS, IRIF, Université Paris Diderot, Paris, France.

September 26, 2018

Semi Definite Programs

- A Semidefinite Program (SDP) is an optimization problem with inputs a vector $c \in \mathbb{R}^m$ and matrices $A^{(1)}, \ldots, A^{(m)}, B$ in $\mathbb{R}^{n \times n}$.
- The primal and dual SDP are given by,

$$Opt(P) = \min_{x \in \mathbb{R}^m} \{ c^t x \mid \sum_{k \in [m]} x_k A^{(k)} \succeq B \}.$$

$$Opt(D) = \max_{Y \succeq 0} \{ Tr(BY) \mid Y \succeq 0, Tr(YA^{(j)}) = c_j \}.$$

- We will be working in the case where P, D are strictly feasible, in this case strong duality holds and $Opt(P) = Opt(D)$.

Semi Definite Programs

- A linear programs (LP) is a special cases of an SDP where $(A^{(i)}, B)$ are diagonal matrices,

$$Opt(P) = \min_{x \in \mathbb{R}^m} \{c^t x \mid \sum_{i \in [m]} x_i a_i \geq b, a_i \in \mathbb{R}^n\}$$

$$Opt(D) = \max_{y \geq 0} \{b^t y \mid y^t a_j = c_j\}$$

- SDPs are one of the most general class of optimization problems for which we have efficient algorithms.
- SDPs capture a large class of convex optimization problems, they are also used for approximate solutions to NP hard problems.

SDP algorithms

- The running time for SDP algorithms will be given in terms of m, n and ϵ, here we consider $m = O(n^2)$.
- The first polynomial time algorithms for LPs and SDPs were obtained using the ellipsoid method and the interior point method.
- The best known LP and SDP algorithms have complexity $O(n^{2.5} \log(1/\epsilon))$ and $O((m^3 + mn^\omega + mn^2 s) \log(mn/\epsilon))$.
- In addition there is the Arora Kale method, whose complexity is upper bounded by,

$$\tilde{O}\left(nms\left(\frac{Rr}{\epsilon}\right)^4 + ns\left(\frac{Rr}{\epsilon}\right)^7\right)$$

Quantum SDP algorithms

- Quantum SDP algorithms based on Arora-Kale framework were proposed by [Brandao-Svore 17] and subsequently improved by [van Appeldoorn-Gribling-Gilyen-de Wolf 17].
- These algorithms were recently improved even further in [BKLLSW18] and [AG18].
- The best known running time for a quantum SDP algorithm using the Arora-Kale framework is,

$$\tilde{O}\left(\left(\sqrt{m} + \sqrt{n}\left(\frac{Rr}{\epsilon}\right)\right)\left(\frac{Rr}{\epsilon}\right)^4 \sqrt{n}\right).$$

- For Max-Cut and scheduling LPs , the complexity is at least $O(n^6)$ [AGGW17, Theorem 20].

Our Results

- We provide a quantum interior point method with complexity $\widetilde{O}(\frac{n^{2.5}}{\xi^2}\mu\kappa^3 \log(1/\epsilon))$ for SDPs and $\widetilde{O}(\frac{n^{1.5}}{\xi^2}\mu\kappa^3 \log(1/\epsilon))$ for LPs .
- The output of our algorithm is a pair of matrices (S, Y) that are ϵ-optimal ξ-approximate SDP solutions.
- The parameter μ is at most $\sqrt{2}n$ for SDPs and $\sqrt{2n}$ for LPs .
- The parameter κ is an upper bound on the condition number of the intermediate solution matrices.
- If the intermediate matrices are 'well conditioned', the running time scales as $\widetilde{O}(n^{3.5})$ and $\widetilde{O}(n^2)$.

Input models

- Sparse oracle model [BS16, AGGW17]: The input matrices $A^{(i)}$ are assumed to be s-sparse and we have access to $O_A : |i, k, l, 0\rangle \rightarrow |i, k, l, index(i, k, l)\rangle$.
- Quantum state model [BKLLSW17]: $A^{(i)} = A^{(i)}_{+} + A^{(i)}_{-}$ and we have access to the purifications of the corresponding density matrices for all $i \in [m]$.
- Operator model [AG18]: Access to unitary block encodings of the $A^{(i)}$, that is implementations of:

$$U_j = \begin{pmatrix} A^{(j)}/\alpha_j & . \\ . & . \end{pmatrix}$$

.

- How to construct block encodings for A and what α can be achieved?

QRAM data structure model

- QRAM data structure model: Access to efficient data structure storing $A^{(i)}, i \in [m]$ in a QRAM (Quantum Random Access Memory).
- Given $d_i, i \in [N]$ stored in the QRAM, the following queries require time polylog(N),

$$|i, 0\rangle \rightarrow |i, x_i\rangle$$

Definition

A QRAM data structure for storing a dataset D of size N is efficient if it can be constructed in a single pass over the entries (i, d_i) for $i \in [N]$ and the update time per entry is $O(\text{polylog}(N))$.

- Introduced to address the state preparation problem in Quantum Machine Learning, to prepare arbitrary vector states $|x\rangle$ without incurring an $O(\sqrt{n})$ overhead.

Block encodings using QRAM

- The optimal value of $\alpha = \|A\|$, any minor of a unitary matrix has spectral norm at most 1.
- The quantum linear system problem with $A \in \mathbb{R}^{n \times n}$ is to produce states $|Ab\rangle, |A^{-1}b\rangle$, it is scale invariant so we assume $\|A\| = 1$.
- Define $s_p(A) = \max_{i \in [n]} \sum_{j \in [n]} A_{ij}^p$ and let $\mu(A) = \min_{p \in [0,1]}(\|A\|_F, \sqrt{s_{2p}(A) s_{(1-2p)}(A^T)})$.

Theorem (KP16, KP17, CGJ18)

There are efficient QRAM data structures, that allow a block encodings for $A \in \mathbb{R}^{n \times n}$ with $\alpha = \mu(A)$ to be implemented in time $O(polylog(n))$.

- Notice that $\mu(A) < \sqrt{n}$ is sub-linear, it can be $O(\sqrt{n})$ in the worst case.

Quantum linear system solvers

- Given an efficient block encoding for A, there is a quantum linear system solver with running time $\widetilde{O}(\mu(A)\kappa^2(A)/\epsilon)$ [KP16, KP17].
- Given an efficient block encoding for A, there is a quantum linear system solver with running time $\widetilde{O}(\mu(A)\kappa(A)\log(1/\epsilon))$. [CGJ18, GSLW18].
- Composing block encodings: Given block encodings for M_1, M_2 with paramaters μ_1, μ_2, the linear system in $M = M_1 M_2$ can be solved in time $O((\mu_1 + \mu_2)\kappa(M)\log(1/\epsilon))$.
- Can quantum linear systems be leveraged for optimization using iterative methods? Gradient descent with affine update rules. [KP17].
- This work: Quantum LP and SDP solvers are not much harder than quantum linear systems!

Interior Point Method overview

- The classical IPM starts with feasible solutions (S, Y) to the SDP and updates them $(S, Y) \rightarrow (S + dS, Y + dY)$ iteratively.
- The updates (dS, dY) are obtained by solving a $n^2 \times n^2$ linear system called the Newton linear system.
- The matrix for the Newton linear system is not explicit and is expensive to compute from the data.
- After $O(\sqrt{n} \log(1/\epsilon))$ iterations, the method converges to feasible solutions (S, Y) with duality gap at most ϵ

Classical Interior Point Method

Algorithm 1 Classical interior point method.

Require: Matrices $A^{(k)}$ with $k \in [m]$, $B \in \mathbb{R}^{n \times n}, c \in \mathbb{R}^m$ in memory, precision $\epsilon > 0$.

1. Find feasible initial point (S_0, Y_0, ν_0) close to the analytic center.
2. Starting with (S_0, Y_0, ν_0) repeat the following steps $O(\sqrt{n} \log(1/\epsilon))$ times.
    1. Solve the Newton linear system to get (dS, dY).
    2. Update $S \leftarrow S + dS$, $Y \leftarrow Y + dY$, $\nu \leftarrow Tr(SY)/n$.
3. Output (S, Y).

Quantum Interior Point Method overview

- We construct block encodings for the Newton linear system matrix which allows us to solve this linear system with low cost in the quantum setting.
- We need to find (dS, dY) classically to write the Newton linear system for the next iteration.
- We give a linear time tomography algorithm that reconstructs d-dimensional state to error δ with complexity $O(\frac{d \log d}{\delta^2})$.
- We show that with tomography precision $\delta = O(\frac{1}{\kappa})$ the method converges at the same rate as the classical IPM.
- The solutions output by the QIPM are ξ-approximately feasible as (dS, dY) are reconstructed by tomography.

Quantum Interior Point Method

Algorithm 2 Quantum interior point method.

Require: Same as classical algorithm with inputs stored in QRAM.

1. Find feasible initial point $(S, Y, \nu) = (S_0, Y_0, \nu_0)$ and store in QRAM.
2. Repeat the for T iterations and output the final (S, Y).
    1. Solve Newton linear system to obtain state close to $|dS \circ dY\rangle$ to error δ^2/n^3.
    2. Estimate $\|dS \circ dY\|$, perform tomography and use the norm estimate to obtain,

    $$\left\|\overline{dS \circ dY} - dS \circ dY\right\|_2 \leq 2\delta \left\|dS \circ dY\right\|_2 .$$

    3. Update $Y \leftarrow Y + \overline{dY}$ and $S \leftarrow S + \overline{dS}$ and store in QRAM. Update $\nu \leftarrow Tr(SY)/n$.

Running time overview

- Running time for SDPs is $\widetilde{O}(\frac{n^{2.5}}{\xi^2}\mu\kappa^3 \log(1/\epsilon))$.
- There are $O(n^{0.5}\log(1/\epsilon))$ iterations.
- In each iteration, we solve Newton linear system having size $O(n^2)$ in time $O(\mu\kappa \log(1/\epsilon))$.
- We then perform tomography in time $O(\frac{n^2\kappa^2}{\xi^2})$.
- Running time for LPs is $\widetilde{O}(\frac{n^{1.5}}{\xi^2}\mu\kappa^3 \log(1/\epsilon))$, the linear system has size $O(n)$.

Talk Overview

- Tomography for efficient vector states.
- Classical interior point method.
- Analysis of the approximate interior point method.
- Quantum interior point method.

Vector state tomography

- The vector state for $x \in \mathbb{R}^d$ is defined as $|x\rangle = \sum_i x_i |i\rangle$.
- We assume that $|x\rangle$ can be prepared efficiently, that is we have have access to a unitary U that prepares copies of $|x\rangle$.
- We need to learn $sgn(x_i)$ for vector state tomography, we would need to learn a phase $e^{-2\pi\theta_i}$ for pure state tomography.

Theorem

There is an algorithm with time and query complexity $O(\frac{d \log d}{\delta^2})$ that produces an estimate $\widetilde{x} \in \mathbb{R}^d$ with $\|\widetilde{x}\|_2 = 1$ such that $\|\widetilde{x} - x\|_2 \leq \delta$ with probability at least $(1 - 1/poly(d))$.

Vector state tomography algorithm

- Measure $N = \frac{36 d \ln d}{\delta^2}$ copies of $|x\rangle$ in the standard basis to obtain estimates $p_i = \frac{n_i}{N}$ where n_i is the number of times outcome i is observed.
- Store $\sqrt{p_i}$ for $i \in [d]$ in QRAM data structure so that $|p\rangle = \sum_{i \in [d]} \sqrt{p_i} |i\rangle$ can be prepared efficiently.
- Create N copies of $\frac{1}{\sqrt{2}} |0\rangle \sum_{i \in [d]} x_i |i\rangle + \frac{1}{\sqrt{2}} |1\rangle \sum_{i \in [d]} \sqrt{p_i} |i\rangle$ using a control qubit.
- Apply a Hadamard gate on the first qubit of each copy of the state to obtain $\frac{1}{2} \sum_{i \in [d]} [(x_i + \sqrt{p_i}) |0, i\rangle + (x_i - \sqrt{p_i}) |1, i\rangle]$.
- Measure each copy in the standard basis and maintain counts $n(b, i)$ of the number of times outcome $|b, i\rangle$ is observed for $b \in 0, 1$. Set $\sigma_i = 1$ if $n(0, i) > 0.4 p_i N$ and -1 otherwise.
- Output the unit vector $\widetilde{x}$ with $\widetilde{x}_i = \sigma_i \sqrt{p_i}$.

Analysis of tomography algorithm

- Let $S = \{i \in [d] \mid x_i^2 \geq \delta^2/d\}$, note that $\overline{S}$ has at most a δ^2 fraction of the ℓ_2 norm.
- Claim1: The sign is estimated correctly for all $i \in S$ with probability at least $1 - 1/poly(d)$.
- Claim2: For all $i \in S$ we have that $|x_i - \sqrt{p_i}| \leq \delta/\sqrt{d}$ with probability at least $1 - 1/poly(d)$.
- Claim3: $\sum_{i \in \overline{S}} p_i \leq 2\delta^2$ with probability at least $1 - 1/poly(d)$.
- Error analysis using the three claims:

$$\sum_{i \in [d]} (x_i - \sigma(i)\sqrt{p_i})^2 = \sum_{i \in S} (|x_i| - \sqrt{p_i})^2 + \sum_{i \in \overline{S}} (|x_i| + \sqrt{p_i})^2$$

$$\leq \delta^2 + 2\sum_{i \in \overline{S}} (x_i^2 + p_i) \leq 3\delta^2 + 2\sum_{i \in \overline{S}} p_i$$

Classical Interior Point Method

- Recall the primal and dual SDP ,

$$Opt(P) = \min_{x \in \mathbb{R}^m} \{c^t x \mid \sum_{k \in [m]} x_k A^{(k)} \succeq B\}.$$

$$Opt(D) = \max_{Y \succeq 0} \{Tr(BY) \mid Y \succeq 0, Tr(YA^{(j)}) = c_j\}.$$

- Define $L = \text{Span}_{k \in [m]}(A^{(k)})$, let $L^{\perp}$ be the orthogonal complement.
- Let C be an arbitrary dual feasible solution and $S = \sum_{k \in [m]} x_k A^{(k)} - B$.
- The SDP pair can be written in a more symmetric form,

$$Opt(P') = \min_{S \succeq 0} \{Tr(CS) + Tr(BC) \mid S \in (L - B)\}$$

$$Opt(D) = \max_{Y \succeq 0} \{Tr(BY) \mid Y \in (L^{\perp} + C)\}$$

Classical Interior Point Method

- As $Tr((S+B)(Y-C)) = 0$ we have that the duality gap is $Tr(SY)$.
- The logarithmic barrier is defined as $K(X) = -\log(\det(X))$, it is defined on the interior of the psd cone.
- The central path is parametrized by ν and is given by the solutions to,

$$Opt(P_\nu) = \min_{S \succeq 0} \{ Tr(CS) + \nu K(S) \mid S \in (L - B) \}$$

$$Opt(D_\nu) = \max_{Y \succeq 0} \{ Tr(BY) - \nu K(Y) \mid Y \in (L^\perp + C) \}$$

- As $\nu \to 0$ we recover solutions to the original SDP .
- Theorem: The optimal solutions (S_ν, Y_ν) on the central path satisfy $S_\nu Y_\nu = Y_\nu S_\nu = \nu I$.

Classical Interior Point Method

- An ideal interior point method would follow the central path in the direction $\nu \to 0$. The actual method stays close to the path.
- Define the distance $d(S, Y, \nu) = \left\| I - \nu^{-1} S^{1/2} Y S^{1/2} \right\|_F^2$.
- Theorem: The duality gap and distance from central path are related as,

$$\nu(n - \sqrt{n} d(S, Y, \nu)) \leq Tr(SY) \leq \nu(n + \sqrt{n} d(S, Y, \nu))$$

- It suffices to stay close to the central path, if $d(S, Y, \nu)) \leq \eta$ for $\eta \in [0, 1]$ then $Tr(SY) \leq 2\nu n$.

Classical Interior Point Method

- The interior point method starts with a pair of feasible solutions (S, Y) with duality gap $Tr(SY) = \nu n$ and $d(S, Y, \nu) \leq \eta$ for a constant $\eta \leq 0.1$.
- A single step of the method updates the solution to $(S' = S + dS, Y' = Y + dY)$ such that $Tr(S'Y') = \nu' n$ for $\nu' = (1 - \chi/\sqrt{n})\nu$ where $\chi \leq \eta$ is a positive constant.
- The updates are found by solving the Newton linear system,

$$dS \in L, \quad dY \in L^{\perp}$$
$$dSY + SdY = \nu' I - SY$$

- The classical analysis shows that: (i) The Newton linear system has a unique solution. (ii) The updated solutions (S', Y') are positive definite. (iii) The distance $d(S, Y, \nu') \leq \eta$.

The approximate interior point method

- Approximate interior point method analysis:

Theorem

Let $\left\|dY - \overline{dY}\right\|_F \leq \frac{\xi}{\|Y^{-1}\|_2}$ *and* $\left\|dS - \overline{dS}\right\|_F \leq \frac{\xi}{\|Y\|_2}$ *be approximate solutions to the Newton linear system and let* $(\overline{S} = S + \overline{dS}, \overline{Y} = Y + \overline{dY})$ *be the updated solution. Then,*

• *The updated solution is positive definite, that is* $\overline{S} \succ 0$ *and* $\overline{Y} \succ 0$.

• *The updated solution satisfies* $d(\overline{S}, \overline{Y}, \overline{\nu}) \leq \eta$ *and* $Tr(\overline{S}\ \overline{Y}) = \overline{\nu}n$ *for* $\overline{\nu} = (1 - \frac{\alpha}{\sqrt{n}})\nu$ *for a constant* $0 < \alpha \leq \chi$.

The quantum interior point method

- How to solve the Newton linear system? If we use the variables (dS, dY) then $dS \in L$ is hard to express.
- With variables (dx, dY) the Newton linear system is given by $M(dx, dY) = (\nu' I - SY, 0^m)$.

$$M = \begin{bmatrix} (A^{(1)}Y)_{11} & \dots & (A^{(m)}Y)_{11} & (1 \otimes S_1)^T \\ \vdots & \vdots & \vdots & \ddots \\ (A^{(1)}Y)_{ij} & \dots & (A^{(m)}Y)_{ij} & (j \otimes S_i)^T \\ \vdots & \vdots & \vdots & \ddots \\ (A^{(1)}Y)_{nn} & \dots & (A^{(m)}Y)_{nn} & (n \otimes S_n)^T \\ 0 & \dots & 0 & (vec(A^{(1)}))^T \\ \vdots & \vdots & \vdots & \ddots \\ 0 & \dots & 0 & (vec(A^{(m)}))^T \end{bmatrix}$$

The quantum interior point method

- Let $Z \in \mathbb{R}^{n \times n}$, then define matrix $\widetilde{Z}$ with rows $\widetilde{Z}_{ij} = (i \otimes Z_j)^T$ and $\widehat{Z}$ with rows $\widehat{Z}_{ij} = (j \otimes Z_i)^T$.
- $\widetilde{Z}$ is a block diagonal matrix with n copies of Z on diagonal blocks while $\widehat{Z}$ is obtained by permuting the rows of $\widetilde{Z}$.
- The Newton linear system matrix can be factorized as a product of matrices,

$$M = M_1 M_2 = \begin{pmatrix} \widetilde{Y} & 0 \\ 0 & I_m \end{pmatrix} \cdot \begin{pmatrix} \mathcal{A}^T & \widetilde{Y^{-1}} \widehat{S} \\ 0 & \mathcal{A} \end{pmatrix}$$

- Block encoding for M_1: Is the same as constructing a block encoding for Y.

The quantum interior point method

- Block encoding for M_2: Rows of $\widetilde{Y^{-1}}\widehat{S}$ are tensor products of the rows of Y^{-1} and S, that is $\widetilde{Y^{-1}}\widehat{S} = (Y_j^{-1} \otimes S_i)^T$.
- It suffices to prepare the rows and columns of M_2 efficiently, if we pre-compute Y^{-1} the rows and columns can be prepared efficiently.
- In addition we provide a procedure for preparing $|a \circ b\rangle$ given unitaries for preparing $|a\rangle, |b\rangle$ in time $O(T(U_a) + T(U_b))$.
- Further technical details: Recovery of dS, the precision $O(\frac{1}{\kappa})$ for tomography, linear programs.
- Guarantees: $Tr(SY) \leq \epsilon$ and $(S, Y) \in (L - B', L^{\perp} + C')$ with $\|B \oplus C - B' \oplus C'\| \leq \xi \|B \oplus C\|_F$.

Open questions

- For what optimization problems can one get a polynomial quantum speedup?
- Find quantum analogs of interior point methods for the case of sparse SDPs .
- Improve the classical step in the IPM, find better quantum algorithms for LPs .
- Quantum algorithms for convex optimization with provable polynomial speedups?