LECTURE: Q. ALGORITHMS FOR OPTIMIZATION

Base problem:

<u>Binary quadratic optimization</u> For A a symmetric, $n \times n$ matrix,

(1) $$\underset{x \in \mathbb{R}^n}{\text{maximize}} \quad \langle x, Ax \rangle = \operatorname{tr}(Axx^*)$$

$$\text{s.t.} \quad x \in \{\pm 1\}^n$$

Captures many important problems:

Maxcut (largest cut of a graph)

Community detection
└ divide a network into sets of nodes belonging to two communities.

<u>Strategy</u>

Phase I: Relax the problem (NP-hard in worst-case so relax to something easy in theory.)

Phase II: optimization problem $\Rightarrow$ feasibility problem

Phase III: develop quantum-inspired meta-algorithms

quantum boost: use quantum subroutines.

MMW but without dual $\rightarrow$ mirror descent.

Based on arXiv 1909.04613

Phase I

S^n = set of $n \times n$ PSD matrices

SDP relaxation

② $xx^* \to X$

$$\max_{X \in S^n} \ \mathrm{tr}(AX)$$

s.t. $\mathrm{diag}(X) = \vec{1}$

$X \succeq 0$

~~$\mathrm{rank}(X) = 1$~~

③ Rescaling

$$\max_{X \in S^n} \ \mathrm{tr}(\tilde{A} X) \qquad \tilde{A} = \frac{1}{\|A\|} A$$

s.t. $\mathrm{diag}(X) = \frac{1}{n} \vec{1}$

$\mathrm{tr}(X) = 1 \longrightarrow$ superfluous constraint

$X \succeq 0$

REMARK:

Special case of a convex optimization problem

$\max \quad f(X) = \mathrm{tr}(\tilde{A}X)$

$X \in C_1 \cap C_2$ where

$C_1 = \{X : \mathrm{diag}(X) = \frac{\vec{1}}{n}\}$ affine subspace

$C_2 = \{X : X \succeq 0\}$ convex cone

But algorithm will work for general cvx problems:

$\max f(X)$, f bdd, concave

s.t. $X \in C_1 \cap \cdots \cap C_m$ where $C_1 \cdots C_m$ are convex sets.

$\mathrm{Tr}(X) = 1$, $X \succeq 0$.

Phase II

Suffices to solve the following feasibility problem:

notice diff!

Find $X \in S_n$ s.t. $\mathrm{tr}(\tilde{A}X) \leq \lambda$

$\mathrm{diag}(X) = \frac{1}{n}\vec{1}$

$\mathrm{tr}(X) = 1$

$X \succeq 0$

Assume $\tilde{A}$ has bounded norm.

$\rightarrow \mathrm{tr}(\tilde{A}X) \in$ Interval

By wrapping this in an outer loop where we binary search the interval to choose val of λ only need

$\log(1/\epsilon)$

queries to get multiplicative ϵ-approximation

Phase III

Quantum-inspired change of variables

$$X = \rho_H = \frac{\exp(-H)}{\mathrm{tr}(\exp(-H))} \in S_n \quad \text{(Gibbs state)}$$

- ensures X is PSD, trace 1.

New problem:

Let $\tilde{A} = \frac{A}{\|A\|}$,

find $H \in S^n$ s.t. $\mathrm{tr}(\tilde{A}\rho_H) \leq \lambda$ $\quad (\rho_H \in A_\lambda)$

$\mathrm{diag}(\rho_H) = \frac{I}{n}$ $\quad (\rho_H \in D_n)$

Again: can solve this for any # convex constraints.

The algorithm

WILL ASSUME ACCESS TO ORACLE:

Def (ϵ-separation oracle)

contains every line segment btw 2 points in the set

Let $\mathcal{C} \subset S_n$ be a closed, convex subset of q. states

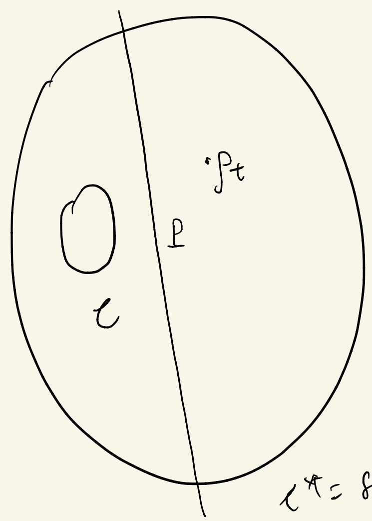
$\mathcal{C}^* \subset \{X = X^\dagger \in \mathbb{C}^{n \times n} : \|X\| \le 1\}$ closed, convex subset of observables. "tests"

$$O_{\mathcal{C},\epsilon}(\rho) = \begin{cases} \underline{\text{accept}}\ \rho \text{ if } \min\limits_{Y \in \mathcal{C}} \max\limits_{L \in \mathcal{C}^*} \operatorname{tr}(L(\rho - Y)) \le \epsilon \\ \text{else: output } L \in \mathcal{C}^* \text{ s.t. } \operatorname{tr}(L(\rho - Y)) \ge \frac{\epsilon}{2} \ \forall Y \in \mathcal{C}. \end{cases}$$

Interpretation: observables from $\mathcal{C}^*$ cannot distinguish ρ from elements of $\mathcal{C}$

Interpretation: there is an observable to which ρ looks different from all states in Y

Intuition:



$\mathcal{C}^*$ = set of planes.

ANIMATION:

If an oracle told me L, I could always improve my guess to push towards $\mathcal{C}$.

Algorithm "Hamiltonian updates" for convex feasibility problem

Start w/ $H=0$ ("infinite temperature"), $\rho_H = I/n$

For $t = 1, \cdots T$,

- check if $\rho_H \in A_\lambda$ and $\rho_H \in D_n$ by querying $O_{A_\lambda, \epsilon}$, $O_{D_n, \epsilon}$

  if true we are done

  else update H to penalize infeasible directions.

  ↳ given separating hyperplane P, update

$$H \leftarrow H + \frac{\epsilon}{16} P$$

- $\rho_H \leftarrow \exp(-H)/\mathrm{tr}(\exp(-H))$

Convergence proof

Theorem (Brandão, Kueng, França).

Hamiltonian updates finds an approximately feasible point after

(at most) $T = \lceil 16 \log(n)/\epsilon^2 \rceil + 1 = \tilde{O}(1)$ steps.

Otherwise, the problem is infeasible.

Proof idea:

starting state

- Relative entropy between $\rho_0 = \frac{I}{n}$ and any feasible point ρ^* is bdd:

$$S(\rho^* \| \rho_0) \leq \log(n).$$

- Show that each iteration makes constant progress in relative entropy: (let ρ^* = feasible point)

$$\boxed{S(\rho^* \| \rho_{t+1}) - S(\rho^* \| \rho_t) \leq -\frac{\epsilon^2}{16}}$$

$\Rightarrow$ convergence after at most T steps, or $S(\rho^* \| \rho_T) < 0$. impossible by defn of rel. ent.

Remark:

= mirror descent with vN entropy potential function.

↳ Bregman divergence wrt VN entropy

= KL divergence

↳ $D_h(y\|x) := h(y) - h(x) - \langle \nabla h(x), y-x \rangle$

for h any strictly convex function.

Proof of (1)

Suppose $\exists$ feasible point ρ^*, let $\rho_t = \dfrac{\exp(-H_t)}{Tr(\exp(-H_t))}$

Distance @ time = 0

$$S(\rho^*\|\rho_0) = Tr(\rho^*(\log \rho^* - \log \rho_0)) \leq \log(n)$$

improvement at every step:

$$S(\rho^*\|\rho_{t+1}) - S(\rho^*\|\rho_t) = tr(\rho^*(\log \rho_t - \log \rho_{t+1}))$$

$$= tr(\rho^*(-H_t - \log Tr(\exp(-H_t)) + H_{t+1} + \log Tr(\exp(-H_{t+1}))))$$

$$= tr(\rho^*(H_{t+1} - H_t)) + \log\left(\frac{Tr(\exp(-H_{t+1}))}{Tr(\exp(-H_t))}\right)$$

Recall update step:

$H_{t+1} = H_t + \frac{\epsilon}{16}\rho_t$

$$= \frac{\epsilon}{16} tr(\rho^* \rho_t) - \underbrace{\log\left(\frac{Tr(\exp(-H_{t+1} + \frac{\epsilon}{16}\rho_t))}{Tr(\exp(-H_{t+1}))}\right)}_{\text{bad boi}}$$

(*)

(a bookmark!)

useful facts to analyze bad boi

(i) Peierls – Bogoliubov inequality : $\log(\mathrm{tr}(\exp(F+G))) \geq \mathrm{tr}(F\exp(G))$

provided $\mathrm{tr}(\exp(G)) = 1$

(ii) $\mathrm{Tr}(e^{-H}/c) = \mathrm{Tr}(e^{-H} \cdot e^{-\log c} I)$

scalar

$= \mathrm{Tr}(\exp(-H - (\log c) I))$

"putting scalars in the exponent".

By (ii), bad boi becomes

$$= \log\Big(\mathrm{Tr}\big(\exp\big(\overbrace{-H_{t+1} - \log(\mathrm{Tr}(\exp(-H_{t+1})))I}^{G} + \overbrace{\tfrac{\epsilon}{16}\rho_t}^{F}\big)\big)\Big)$$

by (i)

$$\geq \mathrm{Tr}\left(\frac{\epsilon}{16}\rho_t \exp(-H_{t+1} - \log(\mathrm{Tr}(\exp(-H_{t+1})))I)\right)$$

$$= \frac{\epsilon}{16}\mathrm{Tr}\left(\rho_t \frac{\exp(-H_{t+1})}{\mathrm{Tr}(\exp(-H_{t+1}))}\right) = \frac{\epsilon}{16}\mathrm{tr}(\rho_t \rho_{t+1})$$

\end{badboi analysis}

Continuing from (*), $S(\rho^* \| \rho_{t+1}) - S(\rho^* \| \rho_t) = \frac{\epsilon}{16}\mathrm{tr}(\rho_t(\rho^* - \rho_{t+1}))$

$$\leq \frac{\epsilon}{16}\left(\mathrm{tr}(\rho_t(\rho_t - \rho_{t+1})) - \mathrm{tr}(\rho_t(\rho_t - \rho^*))\right)$$

$$\leq \frac{\epsilon}{16}\Big(\|\rho_t\| \underbrace{\|\rho_t - \rho_{t+1}\|_{tr}} - \frac{\epsilon}{2}\Big)$$ by defn of ϵ-separation oracle.

$$\leq 2(\exp(\tfrac{\epsilon}{16}\|\rho_t\|) - 1)$$

$$\leq \frac{\epsilon}{4} \quad \because \|\rho_t\| < 1.$$

Fact. Let H, H' be Hermitian matrices. Then

$$\left\| \frac{e^H}{\mathrm{Tr}(e^H)} - \frac{e^{H'}}{\mathrm{Tr}(e^{H'})} \right\|_1 \leq 2(e^{\|H - H'\|} - 1)$$

(see arXiv 1609.05537 for proof)

$$\le \frac{\epsilon}{16}\left(\frac{\epsilon}{4} - \frac{\epsilon}{2}\right) = -\frac{\epsilon^2}{64}.$$

Runtime of algorithm

Computational cost

Total # iterations of Hamiltonian updates:

$$T = \lceil 16 \log(n)/\epsilon^2 \rceil + 1 = \tilde{O}(1)$$

Per-iteration runtime:

CLASSICAL:

rate-limiting step!

Each step requires:

matrix exponentiation

i) compute Gibbs state: $O(n^3)$ (n = matrix dimension)

ii) compute trace (check if $\rho \in A_\lambda$): $O(ns)$ (s = row sparsity)

iii) check diagonal $\sim \vec{1}/n$ ($\rho \in D_n$): $O(n)$.

SIDENOTE: they obtain a better classical algorithm for quadratic SDP relaxations too, by approximating step i).

$\rightarrow \tilde{O}(n^2 s)$ runtime prev: $\tilde{O}(n^{2.5} s)$

Key idea: truncate Taylor expansion of $\exp(-H)$.

$$\exp(-H) \simeq \sum_{k=0}^{l} \frac{H^k}{k!}, \quad l = O(\log(n)/\epsilon) = \tilde{O}(1)$$

+ proof of convergence is robust to approximation of ρ_H.

Result: Quantum SDP solver via Gibbs sampling

Classical bottleneck: compute Gibbs states

Quantum speedup:

prepare copies of ρ_H on q. computer. $\tilde{O}(\sqrt{ns}\, s^{o(1)})$

Estimate $tr(\tilde{A}\rho_H)$ via phase estimation $O(1/\epsilon^2)$ copies

Estimate $diag(\rho_H)$ via computational basis meas. $O(n/\epsilon^2)$ copies.

distribution learning lower bound.

Theorem:

$\boxed{\text{Hamiltonian updates app. solves binary quadratic SDP relaxations in q. runtime } \tilde{O}(n^{1.5}(\sqrt{s})^{1+o(1)})}$

Details about q. subroutine:

Hamiltonians are very structured:

$$H = \alpha\tilde{A} + \beta D, \quad \alpha, \beta = O(\log(n)/\epsilon)$$

1. [Poulin, Wocjan] preparing ρ_H reduces to simulating time evolutions
2. [Childs, Wiebe, 2012]: split up time evolutions
3. [Low 2019]: implementing $\exp(it\alpha\tilde{A})$ costs $\tilde{O}(\sqrt{s}^{1+o(1)})$
4. [Prakash, 2014] implementing $\exp[it\beta D]$ w/ QRAM costs $\tilde{O}(n)$

$\Rightarrow$ total cost: $\tilde{O}(n^{1.5}\sqrt{s}^{1+o(1)})$

Limitations of optimization algs (Bonus application of rel. ent convergence, from 2009.05532)

Fix Hamiltonian H.

Let $\tilde{\rho}$ be "close" to m.m. state $\mathbb{1}/2^n$.

$$D(\tilde{\rho} \| \mathbb{1}/2^n) \leq \beta \epsilon \|H\|.$$

Then

$$\mathrm{Tr}(H\tilde{\rho}) \geq \mathrm{tr}(\sigma_\beta H) - \epsilon \|H\|. \qquad \sigma_\beta := \frac{e^{-\beta H}}{\mathrm{Tr}(e^{-\beta H})}$$

$$\Leftrightarrow \quad |\mathrm{Tr}(H(\tilde{\rho} - \sigma_\beta))| \leq \epsilon \|H\|.$$

Pf

$$f(\beta) = \log(\overbrace{\mathrm{Tr}(e^{-\beta H})}^{Z_\beta}) \text{ is differentiable}$$

$$\rightarrow \text{MVT}: \quad \exists \tilde{\beta} \in (0, \beta) \text{ s.t. } \frac{f(0) - f(\beta)}{\beta} = -f'(\tilde{\beta}) = \mathrm{Tr}(\sigma_{\tilde{\beta}} H)$$

$$\underbrace{\mathrm{Tr}(\sigma_{\tilde{\beta}} H)}_{\langle E \rangle_{\tilde{\beta}}} \geq \underbrace{\mathrm{Tr}(\sigma_\beta H)}_{\langle E \rangle_\beta} \qquad \text{b/c } \tilde{\beta} \text{ at higher temperature than } \beta.$$

Fact: Variational formulation of rel. ent. For any σ

$$Z_{\beta,\sigma} := \mathrm{Tr}(e^{-\beta H + \log(\sigma)})$$

$$\Rightarrow \underbrace{\mathrm{Tr}(\Phi(\rho) H)}_{\text{Let: } \tilde{\rho}} \geq \sup_{\beta \geq 0} \beta^{-1} (\underbrace{\log(Z_{\beta,\sigma})}_{\downarrow \text{ For } \sigma := \frac{\mathbb{1}}{2^n} \; n - \log(Z_\beta)} - D(\Phi(\rho) \| \underbrace{\sigma}_{\frac{\mathbb{1}}{2^n}}))$$

$$\to \mathrm{Tr}(H\tilde{\rho}) \geq \mathrm{Tr}(G_{\tilde{\beta}} H) - \beta^{-1} D(\tilde{\rho} \| \mathbb{I}/2^n) \geq \mathrm{Tr}(G_{\tilde{\beta}} H) - \epsilon \|H\|.$$

$\uparrow$

from Fact