A Note on Semidefinite Programming Relaxations For Polynomial Optimization Over a Single Sphere

Jiang Hu

BICMR,PKU

Thanks: Bo Jiang, Xin Liu, Zaiwen Wen

2016 Workshop on Stochastic Optimization and Tensor Analysis, Changsha

June 24, 2019

Outline

- Two instances of polynomial optimization problem
- Overview of the existing SDP relaxation
- Deterministic and randomized algorithm for our SDP relaxation
- Numerical experiments

Best rank-1 tensor approximations

- Given tensor $\mathcal{F} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_m}$, find

$$\min_{\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_m}} \|\mathcal{F} - \mathcal{X}\|^2 \quad \text{s.t.} \quad \mathrm{rank}(\mathcal{X}) = 1,$$

- $\mathrm{rank}(\mathcal{X}) = 1$ is equivalent to $\mathcal{X} = \lambda \cdot x^1 \otimes \cdots \otimes x^m$.
- The problem is equivalent to

$$\begin{cases} \max\limits_{x^1 \in \mathbb{R}^{n_1}, \cdots, x^m \in \mathbb{R}^{n_m}} & |F(x^1, \cdots, x^m)| \\ \text{s.t.} & \|x^1\| = \cdots = \|x^m\| = 1, \end{cases}$$

where $F(x^1, \cdots, x^m) = \sum_{i_1 i_2 \cdots i_m} \mathcal{F}_{i_1 i_2 \cdots i_m} x^{(1)}_{i_1} x^{(2)}_{i_2} \cdots x^{(m)}_{i_m}$.

Best rank-1 tensor approximations

- When $\mathcal{F}$ is symmetric, it always has an optimal symmetric tensor solution.
- Reduces to

$$\max_{x \in \mathbb{R}^n} \quad |f(x)| \quad \text{s.t.} \quad \|x\| = 1,$$

where $x = x^1 = \cdots = x^m$, and $f(x) = F(x, \cdots, x)$.

- $\lambda \cdot x \otimes \cdots \otimes x$ is the best symmetric tensor iff x is a global maximizer and $\lambda = f(x)$.
- Polynomial function optimization problem over a single sphere

Bose-Einstein condensates

Gross-Pitaevskii equation

$$i\hbar \frac{\partial \psi(\mathbf{x}, t)}{\partial t} = \left(-\frac{\hbar^2}{2m}\nabla^2 + V(\mathbf{x}) + NU_0|\psi(\mathbf{x}, t)|^2 - \Omega L_z\right)\psi(\mathbf{x}, t),$$

where $L_z = xp_y - yp_x = -i\hbar(x\partial_y - y\partial_x)$.

Minimization problem

$$\min_{\phi \in S} E(\phi),$$

$$E(\phi) = \int_{\mathbb{R}^d} \left[\frac{1}{2}|\nabla\phi(\mathbf{x})|^2 + V_d(\mathbf{x})|\phi(\mathbf{x})|^2 + \frac{\beta_d}{2}|\phi(\mathbf{x})|^4 - \Omega\bar{\phi}(\mathbf{x})L_z\phi(\mathbf{x})\right] d\mathbf{x}$$

$$S = \left\{\phi \mid E(\phi) < \infty, \int_{\mathbb{R}^d} |\phi(\mathbf{x})|^2 d\mathbf{x} = 1\right\}.$$

Discretized Bose-Einstein condensates

- Sine pseudospectral discretization
- Fourier pseudospectral discretization

$$\min_{X \in \mathbf{C}^N} \mathcal{F}(X) := \frac{1}{2} X^* A X + \alpha \sum_{i=1}^{N} |X_i|^4, \text{ s.t. } \|X\|_F = 1.$$

The first and second-order directional derivatives of $\mathcal{F}(X)$ along a direction $D \in \mathbf{C}^N$ are:

$$\begin{aligned} \nabla \mathcal{F}(X)[D] &= \Re(D^* A X) + 4\alpha \sum_{i=1}^{N} (X_i^* X_i) \Re(X_i^* D_i), \\ \nabla^2 \mathcal{F}(X)[D, D] &= D^* A D + \sum_{i=1}^{N} \left(4\alpha (X_i^* X_i)(D_i^* D_i) + 8\alpha \Re(X_i^* D_i)^2 \right). \end{aligned}$$

Homogeneous Quadratic and Quartic poly. (HQQS)

- Consider

$$\begin{aligned} \min_x \quad & f(x) := \frac{1}{2} x^{\mathrm{T}} A x + \frac{\beta}{2} \sum_{i=1}^n x_i^4 \\ \text{s.t.} \quad & \|x\|_2 = 1, \end{aligned}$$

- It is equivalent to

$$\begin{cases} \min\limits_x & f(x) = \frac{1}{2} x^\top A x x^\top x + \frac{\beta}{2} \sum_{i=1}^n x_i^4 \\ \text{s.t.} & \|x\|_2 = 1. \end{cases} .$$

- Best rank-1 tensor approximation to a 4-order symmetric tensor $\mathcal{F}$.

$$\mathcal{F}_{\pi(i,j,k,l)} = \begin{cases} a_{kl}/4, & i = j = k \neq l, \\ a_{kl}/12, & i = j, i \neq k, i \neq l, k \neq l, \\ (a_{ii} + a_{kk})/12, & i = j \neq k = l, \\ a_{ii}/2 + \beta/4, & i = j = k = l, \\ 0, & \text{others}. \end{cases}$$

Finding Global Optimal Solutions by SDP Relaxation

- Nie and Wang consider

$$f_{\max} := \max_{x \in \mathbb{R}^n} \quad f(x) \quad \text{s.t.} \quad x^\top x = 1.$$

- Let $m = 2d$, $c := \{c_\alpha\}$ and $g := \{g_\alpha\}$ be the coefficients of the polynomial functions $f(x)$ and $x^\top x$ such that

$$f(x) := \sum_{\alpha \in \mathbb{N}_m^n} c_\alpha x^\alpha, \quad x^\top x := \sum_{\alpha \in \mathbb{N}_m^n} g_\alpha x^\alpha.$$

- Introducing $[x^d] := [x_1^d, \;\; x_1^{d-1}x_2, \cdots, x_1^{d-1}x_n, \cdots, x_n^d]^\top$ of length $\binom{n+d-1}{d}$, we obtain a square matrix

$$M \;\; := \;\; [x^d][x^d]^\top = \sum_{\alpha \in \mathcal{N}_m^n} A_\alpha x^\alpha,$$

where A_α is a symmetric matrix with dimensions $\binom{n+d-1}{d} \times \binom{n+d-1}{d}$.

Finding Global Optimal Solutions by SDP Relaxation

- For each x^α, we assign a linear variable y_α to replace x^α.
- Given $y \in \mathbb{R}^{N_m^n}$, we define linear functions

$$\langle c, y \rangle = \sum_{\alpha \in \mathbb{N}_m^n} c_\alpha y_\alpha, \quad \langle g, y \rangle := \sum_{\alpha \in \mathbb{N}_m^n} g_\alpha y_\alpha, \quad M_1(y) := \sum_{\alpha \in N_m^n} A_\alpha y_\alpha.$$

- Therefore, the optimization is equivalent to

$$\max_{y \in \mathbb{R}^{N_m^n}} \langle c, y \rangle \quad \text{s.t.} \quad M_1(y) \succeq 0, \ \langle g, y \rangle = 1, \ \operatorname{rank}(M_1(y)) = 1.$$

- Removing the rank-1 constraint yields a SDP relaxation

$$f_{\max}^{\text{sdp}} := \max_{y \in \mathbb{R}^{N_m^n}} \langle c, y \rangle \quad \text{s.t.} \quad M_1(y) \succeq 0, \quad \langle g, y \rangle = 1.$$

- Global optimal solution of poly. opt. if $M_1(y)$ is rank one.

Finding Global Optimal Solutions by SDP Relaxation

- Jiang, Ma and Zhang consider

$$f_{\max} := \max_{x \in \mathbb{R}^n} \quad f(x) \quad \text{s.t.} \quad x^\top x = 1.$$

- Matricization of symmetric tensor of order $2d$

$$M_2(\mathcal{F})_{kl} := \mathcal{F}_{i_1 \cdots i_d i_{d+1} \cdots i_{2d}}, 1 \leqslant i_1, \cdots i_d, i_{d+1}, \cdots i_{2d} \leqslant n,$$

where

$$k = \sum_{j=1}^{d} (i_j - 1) n^{d-j} + 1, \text{and } l = \sum_{j=d+1}^{2d} (i_j - 1) n^{2d-j} + 1.$$

- Removing the rank-1 constraint yields a SDP relaxation

$$\begin{cases} \max & \langle F, X \rangle \quad \boxed{F = M_2(\mathcal{F})} \\ \text{s.t.} & \operatorname{trace}(X) = 1, \quad M_2^{-1}(X) \in S^{n^{2d}}, \\ & X \in S^{n^d \times n^d}, \quad \boxed{\operatorname{rank}(X) = 1 \to X \succeq 0}, \end{cases}$$

- Global optimal solution of poly. opt. if X is rank one.

Equivalence of The Two SDP Relaxations

- Define two sets

$$\Phi := \left\{ Y \in S^{\binom{n+d-1}{d} \times \binom{n+d-1}{d}} \mid Y = \sum_{\alpha \in \mathcal{N}_m^n} y_\alpha A_\alpha,\ y \in \mathbb{R}^{\binom{n+2d-1}{2d}} \right\},$$

$$\Psi := \left\{ Y \in S^{n^d \times n^d} \mid Y = \sum_{\alpha \in \mathcal{N}_m^n} y_\alpha B_\alpha,\ y \in \mathbb{R}^{\binom{n+2d-1}{2d}} \right\}.$$

- Introduce a map $\tau : \ \Phi \to \Psi$:

$$\bar{X} = \tau(X) = \sum_{\alpha \in \mathcal{N}_m^n} y_\alpha B_\alpha \left(X = \sum_{\alpha \in \mathcal{N}_m^n} y_\alpha A_\alpha \in \Phi\right).$$

- Key bijective map $\sigma : \ y \to Y$ as:

$$Y = \tau(M_1(y)) = \tau\left(\sum_{\alpha \in \mathcal{N}_m^n} y_\alpha A_\alpha \right).$$

Equivalence of The Two SDP Relaxations

- Lemma: For any SDP matrix $X \in \Phi$, if $\bar{X} = \tau(X)$, then $\bar{X}$ is positive semidefinite and $\mathrm{rank}(\bar{X}) = \mathrm{rank}(X)$. For any matrix $\bar{X}$, it holds $\tau^{-1}(\bar{X})$ is semidefinite and $\mathrm{rank}(\bar{X}) = \mathrm{rank}(\tau^{-1}(\bar{X}))$.
- Theorem: Let y and Y be the optimal solutions of NW and JMZ respectively. Then there exists a bijective map σ, such that $\sigma^{-1}(Y)$ and $\sigma(y)$ are the optimal solutions of NW and JMZ respectively.

Quadratic SDP Relaxation

- Let $X = xx^\top$, the HQQS problem is equivalent to

$$\begin{cases} \min\limits_X & \frac{1}{2}\langle A, X\rangle + \frac{\beta}{2}\sum_{i=1}^n X_{ii}^2 \\ \text{s.t.} & \mathrm{tr}(X) = 1, \quad \mathrm{rank}(X) = 1. \end{cases}$$

- Remove the rank-one constraint, we obtain

$$\text{(SDR)} \quad \begin{cases} \min\limits_X & \frac{1}{2}\langle A, X\rangle + \frac{\beta}{2}\sum_{i=1}^n X_{ii}^2 \\ \text{s.t.} & \mathrm{tr}(X) = 1, \quad X \succeq 0. \end{cases}$$

- Let X^* be a solution with $\mathrm{rank}(X^*) = r$. Let $X^* = \sum_{k=1}^r x^k (x^k)^\top$

$$\mathrm{tr}(x^k (x^k)^\top) = (x^k)^\top x^k = \mathrm{tr}(X^*)/r = 1/r, \ \forall\ k = 1, \dots, r.$$

Quadratic SDP Relaxation

- Then we compute

$$\hat{x} = \arg \min_{x^k, k=1,\dots,r} \frac{1}{2}(x^k)^\top A x^k + \frac{\beta}{2}\sum_{i=1}^n (x_i^k)^4 \quad \text{and} \quad x^* = \sqrt{r}\hat{x}.$$

  Obviously $(x^*)^\top x^* = r(\hat{x})^\top \hat{x} = 1$ which means that x^* is feasible.

- Suppose that A is positive definite. Let X^* be optimal with $\mathrm{rank}(X^*) = r$ and x^* be constructed as above. Then x^* is an approximate solution to the HQQS problem with an approximation ratio $r \leq n$, i.e.,

$$\frac{1}{2}(x^*)^\top A x^* + \frac{\beta}{2}\sum_{i=1}^n (x_i^*)^4 \leq r\, f_{\min}(HQQS),$$

Randomized SDP Relaxation

- We still consider the SDP relaxation:

$$\begin{cases} \min\limits_X & \frac{1}{2}\langle A, X\rangle + \frac{\beta}{2}\sum_{i=1}^n X_{ii}^2 \\ \text{s.t.} & \mathrm{tr}(X) = 1, \qquad X \succeq 0. \end{cases}$$

- Let X^* be optimal solu. to (SDR) and ξ is a random vector from a Gaussian distribution $\mathcal{N}(0, X^*)$, then $\mathrm{E}[\xi^\top \xi] = \mathrm{tr}(X^*) = 1$ and

$$\begin{aligned} \mathrm{E}\left[f(\xi)\right] = \mathrm{E}\left[\frac{1}{2}\xi^\top A\xi + \frac{\beta}{2}\sum_{i=1}^n \xi_i^4\right] &= \frac{1}{2}\langle A, X^*\rangle + \frac{3\beta}{2}\sum_{i=1}^n (X_{ii}^*)^2 \\ &\le 3\left(\frac{1}{2}\langle A, X^*\rangle + \frac{\beta}{2}\sum_{i=1}^n (X_{ii}^*)^2\right) \\ &\le 3\left(\frac{1}{2}(x^*)^\top A x^* + \frac{\beta}{2}\sum_{i=1}^n (x_i^*)^4\right) \\ &= 3f(x^*). \end{aligned}$$

Randomized SDP Relaxation

- Based on the above observations, we are interested in the following event:

$$\left\{ f(\xi) < \alpha\, \mathsf{E}[f(\xi)],\ \xi^\top \xi \geq \mathrm{tr}(X^*) \geq 1 \right\}.$$

- Once the above event does occur, we construct $\hat{x} = \xi / \|\xi\|_2$ and obtain

$$\begin{aligned} f(\hat{x}) = \frac{1}{2\,\|\xi\|_2^2}\xi^\top A\xi + \frac{\beta}{2\,\|\xi\|_2^4}\sum_{i=1}^n \xi_i^4 \quad &\leq \quad \frac{1}{2}\xi^\top A\xi + \frac{\beta}{2}\sum_{i=1}^n \xi_i^4 \\ &\leq \quad \alpha\left(\frac{1}{2}\langle A, X^*\rangle + \frac{3\beta}{2}\sum_{i=1}^n (X_{ii}^*)^2\right) \\ &\leq \quad 3\alpha\, f(x^*). \end{aligned}$$

Thus $\hat{x}$ is an approximate solution to problem HQQS with approximation ratio 3α.

Randomized SDP Relaxation

- Lemma: Let $Q \in S_+^{n\times n}$. Suppose $\eta \in \mathbb{R}^n$ is a random vector generated from Gaussian distribution $\mathcal{N}(0, I)$. Then,

$$\text{prob}\left(\eta^\top Q\eta < \mathrm{E}[\eta^\top Q\eta]\right) \leq 1 - \theta$$

  where $\theta := 1/960$.

- Suppose that A is positive definite, X^* is a solution to (SDR) and random variable ξ is drawn from $\mathcal{N}(0, X^*)$. Construct $\hat{x} = \xi/\|\xi\|_2$, then we have

$$\text{prob}\left(f(\hat{x}) \leq 3\sqrt{2/\theta}\, f_{\min}(HQQS)\right) \geq \theta/2$$

  with $\theta := 1/960$.

A Feasible Method for Solving (SDR)

- Suppose that the solution $\bar{X}$ of (SDR) is rank p. Then $\bar{X}$ can be decomposed as $\bar{X} = V^\top V$ with $V = [V_1, \cdots, V_n] \in \mathbb{R}^{p \times n}$.
- Consequently, we convert (SDR) into an equivalent problem

$$\begin{aligned} \max_{V=[V_1,\cdots,V_n]} \quad & f(V) := \frac{1}{2}\left\langle A, V^\top V \right\rangle + \frac{\beta}{2}\|\mathrm{diag}(V^\top V)\|_2^2, \\ \text{s.t.} \quad & \|V\|_F = 1, \quad i = 1, \cdots, n, \end{aligned}$$

  where $\mathrm{diag}(V^\top V)$ is the diagonals of $V^\top V$.
- Solve the above problem by a feasible gradient method.

Regularized Newton Method

- using the exact Hessian:

$$\begin{aligned}\tilde{m}^k(X) \quad &:= \nabla \mathcal{F}(X^{(k)})[X - X^{(k)}] + \frac{1}{2}\nabla^2 \mathcal{F}(X^{(k)})[X - X^{(k)}, X - X^{(k)}] \\ &\qquad + \frac{\tau^k}{2}\|X - X^{(k)}\|_F^2,\end{aligned}$$

- The proximal term: trust region like strategy for ensuring global convergence.
- Compute the regularized Newton step:

$$\begin{aligned}\min \quad & \tilde{m}^k(X) \\ \text{s.t.} \quad & \|X\|_F = I.\end{aligned}$$

- Cartis, Gould, Toint '10, '11, '12 on cubic regularization

Numerical Results

- Example 1: Consider a BEC problem with $d = 1$, $V(x) = \frac{1}{2}x^2$ and $\beta = 400$. The problems are discretized by the finite difference scheme on a domain $U = (-16, 16)$ with different mesh sizes $h = 2, 1, \frac{1}{2}, \frac{1}{4}, \frac{1}{6}, \frac{1}{8}$.
- Example 2: Consider a BEC problem with $d = 1$, $V(x) = \frac{1}{2}x^2 + 25\sin^2(\frac{\pi x}{4})$ and $\beta = 250$. We discretize the problem by the finite difference scheme on a domain $U = (-16, 16)$ with different mesh sizes $h = 2, 1, \frac{1}{2}, \frac{1}{4}, \frac{1}{6}, \frac{1}{8}$.

Numerical Results: Example 1

h	time		λ	
	NW	JMZ	NW	JMZ
2	5.5782	22.3343	21.3773	21.3773
1	110.7321	1483.9856	21.3592	21.3592
1/2	—	—	—	—
1/4	—	—	—	—
1/6	—	—	—	—
1/8	—	—	—	—

Table: Computational results on a 1D BEC problem. "—" means that the computational time is more than 30 minutes.

h	λ			rank (ratio)	
	RN	SDR1	SDR2	SDR1	SDR2
2	21.3773	21.5434	22.5696	5(1.0077)	5(18.3641,4.9596,1.0558)
1	21.3592	21.3896	21.5076	5(1.0014)	5(43.2324,5.9762,1.0069)
1/2	21.3598	21.3858	21.7901	5(1.0012)	5(22.6133,4.8358,1.0201)
1/4	21.3600	21.3886	21.5701	5(1.0013)	5(35.6182,5.8311,1.0098)
1/6	21.3600	21.3952	21.7423	5(1.0016)	5(21.3470,4.6358,1.0179)
1/8	21.3601	21.3786	21.4868	5(1.0009)	5(41.1597,6.3262,1.0059)

Table: Computational results on a 1D BEC problem

Numerical Results: Example 2

h	time		λ	
	NW	JMZ	NW	JMZ
2	5.0423	26.0845	23.8977	23.8976
1/2	—	—	—	—
1/4	—	—	—	—
1/6	—	—	—	—
1/8	—	—	—	—

Table: Computational results on a BEC problem. "—" means that the computational time is more than 30 minutes.

h	λ			rank (ratio)	
	RN	SDR1	SDR2	SDR1	SDR2
2	23.8977	24.3978	24.7984	5(1.0209)	5(37.0336,6.0068,1.0377)
1	26.0573	26.0932	27.3979	5(1.0014)	5(10.6342,3.0240,1.0514)
1/2	26.0755	26.0865	26.4186	5(1.0004)	5(21.2076,5.3864,1.0132)
1/4	26.0818	26.1140	26.9123	5(1.0012)	5(50.9194,4.9634,1.0318)
1/6	26.0830	26.1078	26.6175	5(1.0009)	5(16.3740,3.9630,1.0215)
1/8	26.0834	26.1119	27.8197	5(1.0011)	5(35.7704,6.1598,1.0666)

Table: Computational results of on a 1D BEC problem

Many thanks for your attention!