Solving SDPs for synchronization and MaxCut problems via the Grothendieck inequality

Song Mei

Stanford University

July 8, 2017

Joint work with Theodor Misiakiewicz, Andrea Montanari, and Roberto I. Oliveira

The MaxCut SDP problem

- $A \in \mathbb{R}^{n\times n}$ symmetric.
- MaxCut SDP:

$$\begin{aligned} \underset{X\in\mathbb{R}^{n\times n}}{\text{maximize}} \quad & \langle A, X\rangle \\ \text{subject to} \quad & X_{ii} = 1, \quad i \in [n], \\ & X \succeq 0. \end{aligned} \tag{SDP}$$

- Applications: MaxCut problem, $\mathbb{Z}_2$ synchronization, Stochastic block model...

The MaxCut SDP problem

- $A \in \mathbb{R}^{n\times n}$ symmetric.
- MaxCut SDP:

$$\begin{aligned} \underset{X\in\mathbb{R}^{n\times n}}{\text{maximize}} \quad & \langle A, X\rangle \\ \text{subject to} \quad & X_{ii} = 1, \quad i \in [n], \\ & X \succeq 0. \end{aligned} \tag{SDP}$$

- Applications: MaxCut problem, $\mathbb{Z}_2$ synchronization, Stochastic block model...

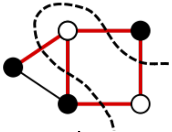
The MaxCut SDP problem

- $A \in \mathbb{R}^{n \times n}$ symmetric.
- MaxCut SDP:

$$\begin{aligned} \underset{X \in \mathbb{R}^{n\times n}}{\text{maximize}} \quad & \langle A, X \rangle \\ \text{subject to} \quad & X_{ii} = 1, \quad i \in [n], \\ & X \succeq 0. \end{aligned} \tag{SDP}$$

- Applications: MaxCut problem, $\mathbb{Z}_2$ synchronization, Stochastic block model...

MaxCut Problem



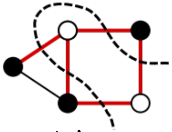
- G: a positively weighted graph. A_G: its adjacency matrix.
- MaxCut of G: known to be NP-hard

$$\underset{x\in\{\pm1\}^n}{\text{maximize}} \quad \frac{1}{4}\sum_{i,j=1}^{n} A_{G,ij}(1 - x_i x_j). \qquad \text{(MaxCut)}$$

- SDP relaxation: 0.878-approximate guarantee [Goemanns and Williamson, 1995]

$$\begin{aligned} \underset{X\in\mathbb{R}^{n\times n}}{\text{maximize}} \quad & \frac{1}{4}\sum_{i,j=1}^{n} A_{G,ij}(1 - X_{ij}), \\ \text{subject to} \quad & X_{ii} = 1, \\ & X \succeq 0. \end{aligned} \qquad \text{(SDPCut)}$$

MaxCut Problem



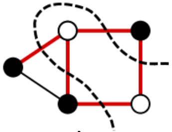
- G: a positively weighted graph. A_G: its adjacency matrix.
- MaxCut of G: known to be NP-hard

$$\underset{x \in \{\pm 1\}^n}{\text{maximize}} \quad \frac{1}{4} \sum_{i,j=1}^{n} A_{G,ij}(1 - x_i x_j). \qquad \text{(MaxCut)}$$

- SDP relaxation: 0.878-approximate guarantee [Goemanns and Williamson, 1995]

$$\begin{aligned} \underset{X \in \mathbb{R}^{n \times n}}{\text{maximize}} \quad & \frac{1}{4} \sum_{i,j=1}^{n} A_{G,ij}(1 - X_{ij}), \\ \text{subject to} \quad & X_{ii} = 1, \\ & X \succeq 0. \end{aligned} \qquad \text{(SDPCut)}$$

MaxCut Problem



- G: a positively weighted graph. A_G: its adjacency matrix.
- MaxCut of G: known to be NP-hard

$$\underset{x \in \{\pm 1\}^n}{\text{maximize}} \quad \frac{1}{4} \sum_{i,j=1}^{n} A_{G,ij}(1 - x_i x_j). \qquad \text{(MaxCut)}$$

- SDP relaxation: 0.878-approximate guarantee [Goemanns and Williamson, 1995]

$$\begin{aligned} \underset{X \in \mathbb{R}^{n \times n}}{\text{maximize}} \quad & \frac{1}{4} \sum_{i,j=1}^{n} A_{G,ij}(1 - X_{ij}), \\ \text{subject to} \quad & X_{ii} = 1, \\ & X \succeq 0. \end{aligned} \qquad \text{(SDPCut)}$$

Burer-Monteiro approach

- Convex formulation: n up to 10^3 using interior point method

$$\begin{aligned} \underset{X \in \mathbb{R}^{n\times n}}{\text{maximize}} \quad & \langle A, X \rangle \\ \text{subject to} \quad & X_{ii} = 1, \quad i \in [n], \\ & X \succeq 0. \end{aligned} \tag{SDP}$$

- Change variable $X = \sigma \cdot \sigma^{\mathsf{T}}$, $\sigma \in \mathbb{R}^{n\times k}$, $k \ll n$.
- Non-convex formulation: n up to 10^5

$$\begin{aligned} \underset{\sigma \in \mathbb{R}^{n\times k}}{\text{maximize}} \quad & \langle \sigma, A\sigma \rangle \\ \text{subject to} \quad & \sigma = [\sigma_1, \dots, \sigma_n]^{\mathsf{T}}, \\ & \|\sigma_i\|_2 = 1, \quad i \in [n]. \end{aligned} \tag{k-Ncvx-SDP}$$

Burer-Monteiro approach

- Convex formulation: n up to 10^3 using interior point method

$$\begin{aligned} \underset{X \in \mathbb{R}^{n\times n}}{\text{maximize}} \quad & \langle A, X \rangle \\ \text{subject to} \quad & X_{ii} = 1, \quad i \in [n], \\ & X \succeq 0. \end{aligned} \tag{SDP}$$

- Change variable $X = \sigma \cdot \sigma^{\mathsf{T}}$, $\sigma \in \mathbb{R}^{n\times k}$, $k \ll n$.

- Non-convex formulation: n up to 10^5

$$\begin{aligned} \underset{\sigma \in \mathbb{R}^{n\times k}}{\text{maximize}} \quad & \langle \sigma, A\sigma \rangle \\ \text{subject to} \quad & \sigma = [\sigma_1, \ldots, \sigma_n]^{\mathsf{T}}, \\ & \|\sigma_i\|_2 = 1, \quad i \in [n]. \end{aligned} \tag{k-Ncvx-SDP}$$

Burer-Monteiro approach

- Convex formulation: n up to 10^3 using interior point method

$$\begin{aligned} \underset{X \in \mathbb{R}^{n \times n}}{\text{maximize}} \quad & \langle A, X \rangle \\ \text{subject to} \quad & X_{ii} = 1, \quad i \in [n], \\ & X \succeq 0. \end{aligned} \tag{SDP}$$

- Change variable $X = \sigma \cdot \sigma^{\mathsf{T}}$, $\sigma \in \mathbb{R}^{n \times k}$, $k \ll n$.

- Non-convex formulation: n up to 10^5

$$\begin{aligned} \underset{\sigma \in \mathbb{R}^{n \times k}}{\text{maximize}} \quad & \langle \sigma, A\sigma \rangle \\ \text{subject to} \quad & \sigma = [\sigma_1, \dots, \sigma_n]^{\mathsf{T}}, \\ & \|\sigma_i\|_2 = 1, \quad i \in [n]. \end{aligned} \tag{k-Ncvx-SDP}$$

Related literatures

- As $k \geq \sqrt{2n}$, the global maxima of the Non Convex formulation coincide with the global maximizer of the Convex formulation [Pataki, 1998], [Barviok, 2001], [Burer and Monteiro, 2003].

- As $k \geq \sqrt{2n}$, Non Convex formulation has no spurious local maxima [Boumal, *et al.*, 2016].

- What if k remains of order 1, as $n \to \infty$? Is there spurious local maxima?

- How is these local maxima?

Related literatures

- As $k \geq \sqrt{2n}$, the global maxima of the Non Convex formulation coincide with the global maximizer of the Convex formulation [Pataki, 1998], [Barviok, 2001], [Burer and Monteiro, 2003].
- As $k \geq \sqrt{2n}$, Non Convex formulation has no spurious local maxima [Boumal, *et al.*, 2016].
- What if k remains of order 1, as $n \to \infty$? Is there spurious local maxima?
- How is these local maxima?

Related literatures

- As $k \geq \sqrt{2n}$, the global maxima of the Non Convex formulation coincide with the global maximizer of the Convex formulation [Pataki, 1998], [Barviok, 2001], [Burer and Monteiro, 2003].
- As $k \geq \sqrt{2n}$, Non Convex formulation has no spurious local maxima [Boumal, *et al.*, 2016].
- What if k remains of order 1, as $n \to \infty$? Is there spurious local maxima?
- How is these local maxima?

Related literatures

- As $k \geq \sqrt{2n}$, the global maxima of the Non Convex formulation coincide with the global maximizer of the Convex formulation [Pataki, 1998], [Barviok, 2001], [Burer and Monteiro, 2003].

- As $k \geq \sqrt{2n}$, Non Convex formulation has no spurious local maxima [Boumal, *et al.*, 2016].

- What if k remains of order 1, as $n \to \infty$? Is there spurious local maxima? Sadly, yes.

- How is these local maxima?

Related literatures

- As $k \geq \sqrt{2n}$, the global maxima of the Non Convex formulation coincide with the global maximizer of the Convex formulation [Pataki, 1998], [Barviok, 2001], [Burer and Monteiro, 2003].

- As $k \geq \sqrt{2n}$, Non Convex formulation has no spurious local maxima [Boumal, *et al.*, 2016].

- What if k remains of order 1, as $n \to \infty$? Is there spurious local maxima? Sadly, yes.

- How is these local maxima?

Related literatures

- As $k \geq \sqrt{2n}$, the global maxima of the Non Convex formulation coincide with the global maximizer of the Convex formulation [Pataki, 1998], [Barviok, 2001], [Burer and Monteiro, 2003].

- As $k \geq \sqrt{2n}$, Non Convex formulation has no spurious local maxima [Boumal, *et al.*, 2016].

- What if k remains of order 1, as $n \to \infty$? Is there spurious local maxima? Sadly, yes.

- How is these local maxima? Empirically, they are good!

Geometry

$$\left.\begin{aligned} &\underset{\sigma \in \mathbb{R}^{n\times k}}{\text{maximize}} \quad \langle \sigma, A\sigma \rangle &&:= f(\sigma) \\ &\text{subject to} \quad \|\sigma_i\|_2 = 1. && \end{aligned}\right\} \mathcal{M}_k = \{\sigma \in \mathbb{R}^{n\times k} : \|\sigma_i\|_2 = 1\}.$$

Definition (ε-approximate concave point)

We call $\sigma \in \mathcal{M}_k$ an ε-approximate concave point of f on $\mathcal{M}_k$, if for any tangent vector $u \in T_\sigma \mathcal{M}_k$, we have

$$\langle u, \mathrm{Hess} f(\sigma)[u] \rangle \le \varepsilon \langle u, u \rangle. \tag{1}$$

Remark

A local maximizer is 0-approximate concave. An ε-approximate concave point is nearly locally optimal.

Geometry

$$\begin{aligned} \underset{\sigma \in \mathbb{R}^{n\times k}}{\text{maximize}} \quad & \langle \sigma, A\sigma \rangle && := \; f(\sigma) \\ \text{subject to} \quad & \|\sigma_i\|_2 = 1. && \Big\} \; \mathcal{M}_k = \{\sigma \in \mathbb{R}^{n\times k} : \|\sigma_i\|_2 = 1\}. \end{aligned}$$

Definition (ε-approximate concave point)

We call $\sigma \in \mathcal{M}_k$ an ε-approximate concave point of f on $\mathcal{M}_k$, if for any tangent vector $u \in T_\sigma \mathcal{M}_k$, we have

$$\langle u, \operatorname{Hess} f(\sigma)[u] \rangle \le \varepsilon \langle u, u \rangle. \tag{1}$$

Remark

A local maximizer is 0-approximate concave. An ε-approximate concave point is nearly locally optimal.

Geometry

$$\begin{aligned} \underset{\sigma \in \mathbb{R}^{n\times k}}{\text{maximize}} \quad & \langle \sigma, A\sigma \rangle \quad := f(\sigma) \\ \text{subject to} \quad & \|\sigma_i\|_2 = 1. \end{aligned} \Bigg\} \; \mathcal{M}_k = \{\sigma \in \mathbb{R}^{n\times k} : \|\sigma_i\|_2 = 1\}.$$

Definition (ε-approximate concave point)

We call $\sigma \in \mathcal{M}_k$ an ε-approximate concave point of f on $\mathcal{M}_k$, if for any tangent vector $u \in T_\sigma \mathcal{M}_k$, we have

$$\langle u, \mathrm{Hess} f(\sigma)[u] \rangle \leq \varepsilon \langle u, u \rangle. \tag{1}$$

Remark

A local maximizer is 0-approximate concave. An ε-approximate concave point is nearly locally optimal.

Landscape Theorem

Theorem (A Grothendieck-type inequality)

For any ε-approximate concave point $\sigma \in \mathcal{M}_k$ of the rank-k non-convex problem, we have

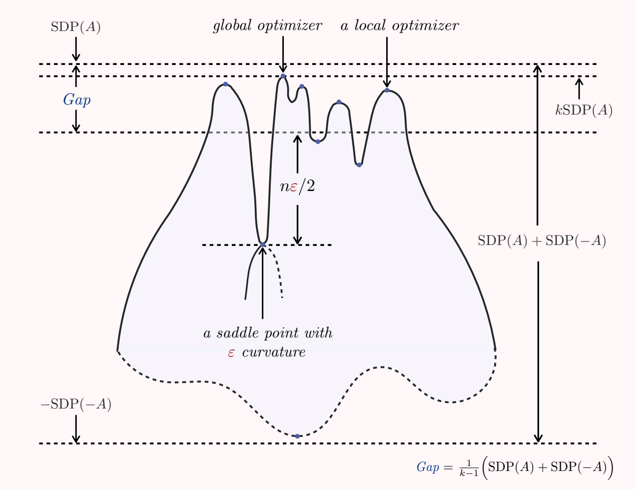
$$f(\sigma) \geq \mathrm{SDP}(A) - \frac{1}{k-1}(\mathrm{SDP}(A) + \mathrm{SDP}(-A)) - \frac{n}{2}\varepsilon. \tag{2}$$

SDP(A): the maximum value of SDP with input matrix A.

Geometric iterpretation: the function value for all local maxima are within a gap of order $O(1/k)$ within the global maximum.

Landscape of non-convex SDP

- $f(\sigma) \geq \mathrm{SDP}(A) - \frac{1}{k-1}(\mathrm{SDP}(A) + \mathrm{SDP}(-A)) - \frac{n}{2}\varepsilon.$



Efficient Algorithm

- Guaranteed converge rate using Riemannian trust region method.
- Getting absolute error $n\varepsilon + (\mathrm{SDP}(A) + \mathrm{SDP}(-A))/(k-1)$ within $c \cdot n||A||_1^2/\varepsilon^2$ trust region steps.
- Empirically, gradient descent converges faster than what is guaranteed.

Approximate MaxCut Guarantee

Theorem (Approximate MaxCut Guarantee)

For any $k \geq 3$, if $\sigma^\star$ is a local maximizer of corresponding rank-k non-convex problem, then we can use $\sigma^\star$ to find a $0.878 \times (1 - 1/(k-1))$-approximate MaxCut.

The global maximizer: 0.878-approximate MaxCut.

Any Local maximizers: $0.878 \times (1 - 1/(k-1))$-approximate MaxCut.

Approximate MaxCut Guarantee

Theorem (Approximate MaxCut Guarantee)

For any $k \geq 3$, if $\sigma^\star$ is a local maximizer of corresponding rank-k non-convex problem, then we can use $\sigma^\star$ to find a $0.878 \times (1 - 1/(k-1))$-approximate MaxCut.

The global maximizer: 0.878-approximate MaxCut.

Any Local maximizers: $0.878 \times (1 - 1/(k-1))$-approximate MaxCut.

Group Synchronization

- SO(d) synchronization, Orthogonal Cut SDP

$$\begin{aligned} \underset{X \in \mathbb{R}^{nk \times nk}}{\text{maximize}} \quad & \langle A, X \rangle \\ \text{subject to} \quad & X_{ii} = \mathbf{I}_k, \\ & X \succeq 0. \end{aligned} \tag{3}$$

- Similar guarantee.

Conclusion

- Non-convex MaxCut SDP: all local maxima are near global maxima.
- An alternate algorithm for approximating MaxCut.
- Conclusion generalizable to general SDP problem.

What I did not go into detail

- $\mathbb{Z}_2$ synchronization and $\mathrm{SO}(d)$ synchronization.
- The one page proof for the Grothendieck-type inequality.

Questions?

$$f(\sigma) \geq \mathrm{SDP}(A) - \frac{1}{k-1}(\mathrm{SDP}(A) + \mathrm{SDP}(-A)) - \frac{n}{2}\varepsilon.$$

- $\mathrm{SDP}(-A)$? Typically has no relationship with $\mathrm{SDP}(A)$. You can think of it has the same order as $\mathrm{SDP}(A)$. Fit well in the MaxCut problem.
- $1/k$ tight? We believe Yes.

Questions?

$$f(\sigma) \geq \mathrm{SDP}(A) - \frac{1}{k-1}(\mathrm{SDP}(A) + \mathrm{SDP}(-A)) - \frac{n}{2}\varepsilon.$$

- $\mathrm{SDP}(-A)$? Typically has no relationship with $\mathrm{SDP}(A)$. You can think of it has the same order as $\mathrm{SDP}(A)$. Fit well in the MaxCut problem.
- $1/k$ tight? We believe Yes.