Semidefinite Programming

Fin Bauer and Stefan Heidekrüger

21. Mai 2015

2015-05-21

Semidefinite Programming

Semidefinite Programming

Fin Bauer and Stefan Heidekrüger

21. Mai 2015

Outline

1 Introduction
2 Some Theory
- Duality
- Optimality
3 Algorithms
- Interior Points
4 Applications
- The Lovász number
- Max Cut
- Machine Learning

What Is Semidefinite Programming?

$$
\begin{aligned}
\min_{X} \quad & \langle C, X\rangle := Tr(CX) = \sum_{i=1}^{n}\sum_{j=1}^{n} C_{ij}X_{ij} \\
\text{s.t.} \quad & \langle A_i, X\rangle = b_i, \quad i = 1, \dots, m \\
& X \succeq 0
\end{aligned}
$$

C, A_i, $X \in \mathbb{R}^{nxn}$ symmetric
$X \succeq 0 \triangleq$ X is positive semidefinite (p.s.d.)

$$
\begin{aligned}
\min_{X} \quad & \langle C, X\rangle := Tr(CX) = \sum_{i=1}^{n}\sum_{j=1}^{n} C_{ij}X_{ij} \\
\text{s.t.} \quad & \langle A_i, X\rangle = b_i, \quad i = 1, \dots, m \\
& X \succeq 0
\end{aligned}
$$

C, A_i, $X \in \mathbb{R}^{nxn}$ symmetric
$X \succeq 0 \triangleq$ X is positive semidefinite (p.s.d.)

An Easy Example

$$C = \begin{pmatrix} 1 & 2 \\ 2 & 9 \end{pmatrix}, \; A_1 = \begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix}, \; A_2 = \begin{pmatrix} 0 & 2 \\ 2 & 3 \end{pmatrix}, \; b_1 = 11, \; b_2 = 19$$

$C = \begin{pmatrix} 1 & 2 \\ 2 & 9 \end{pmatrix}, \; A_1 = \begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix}, \; A_2 = \begin{pmatrix} 0 & 2 \\ 2 & 3 \end{pmatrix}, \; b_1 = 11, \; b_2 = 19$

An Easy Example

$$C = \begin{pmatrix} 1 & 2 \\ 2 & 9 \end{pmatrix},\ A_1 = \begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix},\ A_2 = \begin{pmatrix} 0 & 2 \\ 2 & 3 \end{pmatrix},\ b_1 = 11,\ b_2 = 19$$

$$\begin{array}{ll} \min & x_{11} + 2x_{21} + 2x_{12} + 9x_{22} \\ \text{s.t.} & \qquad\qquad x_{11} + x_{22} = 11 \\ & 2x_{21} + 2x_{12} + 3x_{22} = 19 \\ & X = \begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{pmatrix} \succeq 0 \end{array}$$

$$C = \begin{pmatrix} 1 & 2 \\ 2 & 9 \end{pmatrix},\ A_1 = \begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix},\ A_2 = \begin{pmatrix} 0 & 2 \\ 2 & 3 \end{pmatrix},\ b_1 = 11,\ b_2 = 19$$

$$\begin{array}{ll} \min & x_{11} + 2x_{21} + 2x_{12} + 9x_{22} \\ \text{s.t.} & \qquad\qquad x_{11} + x_{22} = 11 \\ & 2x_{21} + 2x_{12} + 3x_{22} = 19 \\ & X = \begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{pmatrix} \succeq 0 \end{array}$$

Problem Session

Problem 1

Are Linear Programs (LP) and Convex Quadratically Constrained Quadratic Programs (CQCQP) Semidefinite Programs (SDP)?

Problem 2

max cut

Problem 3

lovasz

Problem 4

Is a certain optimization problem a SDP and what does it model?

Problem 1
Are Linear Programs (LP) and Convex Quadratically Constrained Quadratic Programs (CQCQP) Semidefinite Programs (SDP)?
Problem 2
max cut
Problem 3
lovasz
Problem 4
Is a certain optimization problem a SDP and what does it model?

Problem 1: Have I Ever Seen Semidefinite Programming Before?

- Linear Programming

Linear Program

$$\begin{aligned} \min_x \quad & b_0^T x + c_0 \\ \text{s.t.} \quad & b_i^T x + c_i \leq 0, \quad \text{i in 1,...,n} \\ & x \geq 0 \end{aligned}$$

Hint

Diagonal Matrix

- Convex Quadratically Constrained Quadratic Programming

CQCQP

$$\begin{aligned} \min \quad & x^T A_0 x + b_0^T x + c_0 \\ \text{s.t.} \quad & x^T A_i x + b_i^T x + c_i \leq 0, \quad \text{i in 1,...,n} \end{aligned}$$

Hint

$$\text{Given } A_i = M_i^T M_i$$

$$\text{then } x^T A_i x + b_i^T x + c_i \leq 0$$

$$\begin{pmatrix} I & M_i x \\ x^T M_i^T & -c_i - b_i^T x \end{pmatrix} \succeq 0$$

Optimization Hierarchy

LP < CQCQP < SDP < Convex Programming

LP < CQCQP < SDP < Convex Programming

What's the Dual?

Primal Problem in Standard Form

$$\inf_X \{ Tr(CX);\ Tr(A_i X) = b_i \ (i = 1, \dots, m),\ X \in \mathcal{S}_n^+ \}$$

Dual Problem in Standard Form

$$\sup_{y,S} \{ b^T y;\ \sum_{i=1}^{m} y_i A_i + S = C,\ S \in \mathcal{S}_n^+, y \in \mathbb{R}^m \}$$

Primal Problem in Standard Form
$\inf_X \{ Tr(CX);\ Tr(A_i X) = b_i \ (i = 1, \dots, m),\ X \in \mathcal{S}_n^+ \}$
Dual Problem in Standard Form
$\sup_{y,S} \{ b^T y;\ \sum_{i=1}^{m} y_i A_i + S = C,\ S \in \mathcal{S}_n^+, y \in \mathbb{R}^m \}$

Weak and Strong Duality

Example with Duality Gap

When is the Solution Optimal?

Optimality Conditions

$$
\begin{aligned}
Tr(A_i X) &= b_i, \quad X \succeq 0, \quad i = 1, \dots, m \\
\sum_{i=1}^{m} y_i A_i + S &= C, \quad S \succeq 0 \\
XS &= 0
\end{aligned}
$$

2015-05-21

Optimality Conditions

$$
\begin{aligned}
Tr(A_i X) &= b_i, \quad X \succeq 0, \quad i = 1, \dots, m \\
\sum_{i=1}^{m} y_i A_i + S &= C, \quad S \succeq 0 \\
XS &= 0
\end{aligned}
$$

bla

The Lovász number

Definition (SDP-variant)

Let $G = (V, E)$ be a graph.

Definition (SDP-variant)
Let $G = (V, E)$ be a graph.

Problem 2:
What is MAX CUT?

Group presentation time.

Group presentation time.

$$\max \frac{1}{2} \sum_{i<j} w_{ij}(1 - y_i y_j), \qquad s.t. y_i \in \{1, -1\}$$

Problem 2:
What is MAX CUT?

IQP Model

Problem (MC) should be on the board now.

IQP Model
Problem (MC) should be on the board now.

$$\max \frac{1}{2} \sum_{i<j} w_{ij}(1 - y_i y_j), \qquad s.t. y_i \in \{1, -1\}$$

An SDP-approximation algorithm (Goermans-Williamson)

Outline

- Relax (MC) into a QP (P)
- Find approximation bound of QP
- Show: equivalent SQP (SQ) to (P)
- Strong duality holds for (SQ) (ommitted)
- Solve SQP's dual in polynomial time

Outline
- Relax (MC) into a QP (P)
- Find approximation bound of QP
- Show: equivalent SQP (SQ) to (P)
- Strong duality holds for (SQ) (ommitted)
- Solve SQP's dual in polynomial time

- 1995
- Approx: 0.87856
- eq. algorithm already existed, but bound wasn't known.

QP Relaxation of (MC)

(P)

$$W_P^* := \max \frac{1}{2} \sum_{i<j} w_{ij}(1 - v_i^T v_j)$$

$$s.t.\ v_i \in \mathbb{S}^n \qquad \forall i \in V$$

(P)
$W_P^* := \max \frac{1}{2} \sum_{i<j} w_{ij}(1 - v_i^T v_j)$
$s.t.\ v_i \in \mathbb{S}^n \qquad \forall i \in V$

relaxation gives upper bound

A randomized algorithm using (P)

1. Solve (P) to get vectors v_i
2. Sample $r \sim \mathit{UNIFORM}(\mathbb{S}^n)$
3. Set $S := \{i | v_i^T r \geq 0\}$

1. Solve (P) to get vectors v_i
2. Sample $r \sim \mathit{UNIFORM}(\mathbb{S}^n)$
3. Set $S := \{i | v_i^T r \geq 0\}$

A randomized algorithm using (P)

1. Solve (P) to get vectors v_i
2. Sample $r \sim UNIFORM(\mathbb{S}^n)$
3. Set $S := \{i | v_i^T r \geq 0\}$

For cut W obtained this way:

$$E[W] = \sum_{i<j} w_{ij} \frac{\arccos(v_i^T v_j)}{\pi}$$

1. Solve (P) to get vectors v_i
2. Sample $r \sim UNIFORM(\mathbb{S}^n)$
3. Set $S := \{i | v_i^T r \geq 0\}$

For cut W obtained this way:

$$E[W] = \sum_{i<j} w_{ij} \frac{\arccos(v_i^T v_j)}{\pi}$$

A bound given by $E[W]$

Theorem

$$E[W] \geq \alpha \frac{1}{2} \sum_{i<j} w_{ij}(1 - v_i^T v_j)$$

with

$$\alpha := \min_{0 \leq \Theta \leq \pi} \frac{2}{\pi} \frac{\Theta}{1 - \cos \Theta} > .87856\ldots$$

Theorem

$$E[W] \geq \alpha \frac{1}{2} \sum_{i<j} w_{ij}(1 - v_i^T v_j)$$

with

$$\alpha := \min_{0 \leq \Theta \leq \pi} \frac{2}{\pi} \frac{\Theta}{1 - \cos \Theta} > .87856\ldots$$

A bound given by $E[W]$

Theorem

$$E[W] \geq \alpha \frac{1}{2} \sum_{i<j} w_{ij}(1 - v_i^T v_j)$$

with

$$\alpha := \min_{0 \leq \Theta \leq \pi} \frac{2}{\pi} \frac{\Theta}{1 - \cos\Theta} > .87856\ldots$$

Corollary

$$E[W] \geq \alpha W_P^* \geq \alpha W_{MC}^*$$

Theorem
$$E[W] \geq \alpha \frac{1}{2} \sum_{i<j} w_{ij}(1 - v_i^T v_j)$$
with
$$\alpha := \min_{0 \leq \Theta \leq \pi} \frac{2}{\pi} \frac{\Theta}{1 - \cos\Theta} > .87856\ldots$$
Corollary
$$E[W] \geq \alpha W_P^* \geq \alpha W_{MC}^*$$

SDP formulation of (P)

(SD)

$$
\begin{aligned}
W_P^* := \max \ & \frac{1}{2} \sum_{i<j} w_{ij}(1 - y_{ij}) \\
s.t. \ & y_{ii} = 1 \qquad \forall i \in V \\
& Y \text{ sym. pos. sem. def.}
\end{aligned}
$$

(SD)

$W_P^* := \max \frac{1}{2} \sum_{i<j} w_{ij}(1 - y_{ij})$ s.t. $y_{ii} = 1 \quad \forall i \in V$, Y sym. pos. sem. def.

SDP formulation of (P)

(SD)

$$W_P^* := \max \frac{1}{2}\sum_{i<j} w_{ij}(1-y_{ij})$$
$$s.t.\ y_{ii} = 1 \qquad \forall i \in V$$
$$Y \text{ sym. pos. sem. def.}$$

How is this an SDP? Rewrite the objective!

$$= \frac{1}{4}\sum_{i\in V}\sum_{j\in V} w_{ij}(1-y_{ij})$$
$$= \frac{1}{2}W_{tot} - \frac{1}{4}\langle W, Y\rangle$$

(SD)
$$W_P^* := \max \frac{1}{2}\sum_{i<j} w_{ij}(1-y_{ij})$$
$$s.t.\ y_{ii} = 1 \qquad \forall i \in V$$
$$Y \text{ sym. pos. sem. def.}$$
How is this an SDP? Rewrite the objective!
$$= \frac{1}{4}\sum_{i\in V}\sum_{j\in V} w_{ij}(1-y_{ij})$$
$$= \frac{1}{2}W_{tot} - \frac{1}{4}\langle W, Y\rangle$$

How is $(P) \Leftrightarrow (SP)$?

- Recall that a symmetric matrix $A \in \mathbb{R}^n$ is positive semidefinite iff for some $m \leq n$

$$\exists B \in \mathbb{R}^{m \times n} : A = B^T B$$

- Given pos. semidef. A, such a B can be found in $\mathcal{O}(n^3)$ using incomplete Cholesky decomposition.
- Interpret Y in (SP) as the Gram-Matrix of vectors v_i in (P)

- Recall that a symmetric matrix $A \in \mathbb{R}^n$ is positive semidefinite iff for some $m \leq n$

$$\exists B \in \mathbb{R}^{m \times n} : A = B^T B$$

- Given pos. semidef. A, such a B can be found in $\mathcal{O}(n^3)$ using incomplete Cholesky decomposition.
- Interpret Y in (SP) as the Gram-Matrix of vectors v_i in (P)

(Almost) Solving (SQ) in polynomial time

- For this particular Problem, strong duality holds.
- Using the dual, a cut with weight at least $W^*_{SQ} - \varepsilon$ can be found in $\mathcal{O}(\sqrt{n}(\log W_{tot} + \log \frac{1}{\varepsilon})$ iterations using an interior point algorithm. Each iteration can be implemented in $\mathcal{O}(n^3)$
- This cut is a 0.878 approximation to W^*_{MC}

- For this particular Problem, strong duality holds.
- Using the dual, a cut with weight at least $W^*_{SQ} - \varepsilon$ can be found in $\mathcal{O}(\sqrt{n}(\log W_{tot} + \log \frac{1}{\varepsilon})$ iterations using an interior point algorithm. Each iteration can be implemented in $\mathcal{O}(n^3)$
- This cut is a 0.878 approximation to W^*_{MC}

Algo: Alizadeh 1995

'Quality' of the approximation

Can $\alpha > 0.87856$ be improved?

No! The relaxation is tight.

- For $C_5 : E[W] \approx .884 W^*_{MC}$
- For Peterson graph $\approx .8787$
- Examples are known such that $E[W] < .8786 W^*_{MC}$

How does the algorithm do in practice?

- Usually within 4% of $W^*_M C$
- 'Almost always' within 9%

Can $\alpha > 0.87856$ be improved?
No! The relaxation is tight.
- For $C_5 : E[W] \approx .884 W^*_{MC}$
- For Peterson graph $\approx .8787$
- Examples are known such that $E[W] < .8786 W^*_{MC}$
How does the algorithm do in practice?
- Usually within 4% of $W^*_M C$
- 'Almost always' within 9%

Problem 4: What is modeled here and is it a SDP?

$$
\begin{aligned}
\max \quad & \rho \\
\text{s.t.} \quad & (a_i - c)^T E (a_i - c) \leq 1 \ \forall i \\
& (b_j - c)^T E (b_j - c) \geq \rho^2 \ \forall j \\
& E \in \mathbb{S}^n_+
\end{aligned}
$$

$$
\begin{aligned}
\max \quad & \rho \\
\text{s.t.} \quad & (a_i - c)^T E (a_i - c) \leq 1 \ \forall i \\
& (b_j - c)^T E (b_j - c) \geq \rho^2 \ \forall j \\
& E \in \mathbb{S}^n_+
\end{aligned}
$$

Classification - Using SDP to tell two things apart (1)

Ellipsoid

$$\mathcal{E} = \{x \in \mathbb{R}^n;\ (x-c)^T E(x-c) <= 1,\ E \text{ is p.s.d.}\}$$

First idea for SDP

$$\begin{aligned}
\max \quad & \rho \\
\text{s.t.} \quad & (a_i - c)^T E(a_i - c) \leq 1\ \forall i \\
& (b_j - c)^T E(b_j - c) \geq \rho^2\ \forall j \\
& E \in \mathbb{S}^n_+
\end{aligned}$$

Ellipsoid
$\mathcal{E} = \{x \in \mathbb{R}^n;\ (x-c)^T E(x-c) <= 1,\ E \text{ is p.s.d.}\}$
First idea for SDP
max ρ
s.t. $(a_i - c)^T E(a_i - c) \leq 1\ \forall i$
$(b_j - c)^T E(b_j - c) \geq \rho^2\ \forall j$
$E \in \mathbb{S}^n_+$

Classification - Using SDP to tell two things apart (2)

$$
\begin{aligned}
\max \quad & \rho \\
\text{s.t.} \quad & (1, a_i)^T \bar{E} (1, a_i) \leq 1 \ \forall i \\
& (1, b_j)^T \bar{E} (1, b_j) \geq \rho^2 \ \forall j \\
& E \in \mathbb{S}_+^{n+1}
\end{aligned}
$$

$$
\begin{aligned}
\max \quad & \rho \\
\text{s.t.} \quad & (1, a_i)^T \bar{E} (1, a_i) \leq 1 \ \forall i \\
& (1, b_j)^T \bar{E} (1, b_j) \geq \rho^2 \ \forall j \\
& E \in \mathbb{S}_+^{n+1}
\end{aligned}
$$