15.093 Optimization Methods

Lecture 24: Semidefinite Optimization

1 Outline

SLIDE 1

1. SDO formulation
2. The Maximum cut problem
3. Minimizing Polynomials as an SDP
4. Linear Difference Equations and Stabilization
5. Barrier Algorithm for SDO

2 SDO formulation

2.1 Primal and dual

SLIDE 2

- $$\begin{aligned} (P): \quad & \min \quad \boldsymbol{C} \bullet \boldsymbol{X} \\ & \text{s.t.} \quad \boldsymbol{A}_i \bullet \boldsymbol{X} = b_i \quad i = 1, \ldots, m \\ & \qquad \boldsymbol{X} \succeq \boldsymbol{0} \end{aligned}$$

- $$\begin{aligned} (D): \quad & \max \quad \sum_{i=1}^{m} y_i b_i \\ & \text{s.t.} \quad \boldsymbol{C} - \sum_{i=1}^{m} y_i \boldsymbol{A}_i \succeq \boldsymbol{0} \end{aligned}$$

3 MAXCUT

SLIDE 3

- Given $G = (N, E)$ undirected graph, weights $w_{ij} \geq 0$ on edge $(i, j) \in E$
- Find a subset $S \subseteq N$: $\sum_{i \in S, j \in \bar{S}} w_{ij}$ is maximized
- $x_j = 1$ for $j \in S$ and $x_j = -1$ for $j \in \bar{S}$

$$\begin{aligned} MAXCUT: \quad & \max \quad \frac{1}{4} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}(1 - x_i x_j) \\ & \text{s.t.} \quad x_j \in \{-1, 1\}, \quad j = 1, \ldots, n \end{aligned}$$

3.1 Reformulation

SLIDE 4

- Let $\boldsymbol{Y} = \boldsymbol{x}\boldsymbol{x}'$, i.e., $Y_{ij} = x_i x_j$
- Let $\boldsymbol{W} = [w_{ij}]$

- Equivalent Formulation

$$
\begin{aligned}
MAXCUT: \quad \max \quad & \frac{1}{4}\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} - \boldsymbol{W} \bullet \boldsymbol{Y} \\
\text{s.t.} \quad & x_j \in \{-1, 1\}, \quad j = 1, \ldots, n \\
& Y_{jj} = 1, \quad j = 1, \ldots, n \\
& \boldsymbol{Y} = \boldsymbol{x}\boldsymbol{x}'
\end{aligned}
$$

3.2 Relaxation

SLIDE 5

- $\boldsymbol{Y} = \boldsymbol{x}\boldsymbol{x}' \succeq \boldsymbol{0}$
- Relaxation

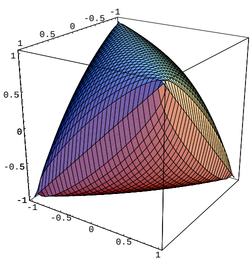
$$
\begin{aligned}
RELAX: \quad \max \quad & \frac{1}{4}\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} - \boldsymbol{W} \bullet \boldsymbol{Y} \\
\text{s.t.} \quad & Y_{jj} = 1, \quad j = 1, \ldots, n \\
& \boldsymbol{Y} \succeq \boldsymbol{0}
\end{aligned}
$$

3.3 Feasible set

SLIDE 6

- For $n = 3$, we have

$$
\begin{bmatrix} 1 & Y_{12} & Y_{13} \\ Y_{12} & 1 & Y_{23} \\ Y_{13} & Y_{23} & 1 \end{bmatrix} \succeq 0
$$



An outer approximation to the true feasible set.

SLIDE 7

- $$MAXCUT \leq RELAX$$
- It turns out that:

$$0.87856 \; RELAX \leq MAXCUT \leq RELAX$$

- The value of the SDO relaxation is guaranteed to be no more than 12% higher than the value of the very difficult to solve problem MAXCUT

4 Minimizing Polynomials

4.1 Sum of squares

SLIDE 8

- A polynomial $f(x)$ is a **sum of squares** (SOS) if

$$f(x) = \sum_j g_j^2(x)$$

  for some polynomials $g_j(x)$.
- A polynomial satisfies $f(x) \geq 0$ for all $x \in \mathbb{R}$ if and only if it is a sum of squares.
- **Not** true in more than one variable!

4.2 Proof

SLIDE 9

- ($\Leftarrow$) Obvious. If $f(x) = \sum_j g_j^2(x)$ then $f(x) \geq 0$.
- ($\Rightarrow$) Factorize $f(x) = C \prod_j (x - r_j)^{n_j} \prod_k (x - a_k + ib_k)^{m_k} (x - a_k - ib_k)^{m_k}$. Since $f(x)$ is nonnegative, then $C \geq 0$ and all the n_j are even. Then, $f(x) = f_1(x)^2 + f_2(x)^2$, where

$$\begin{aligned} f_1(x) &= C^{\frac{1}{2}} \prod_j (x - r_j)^{\frac{n_j}{2}} \prod_k (x - a_k)^{m_k} \\ f_2(x) &= C^{\frac{1}{2}} \prod_j (x - r_j)^{\frac{n_j}{2}} \prod_k b_k^{m_k} \end{aligned}$$

4.3 SOS and SDO

SLIDE 10

- Let $\tilde{x} = (1, x, x^2, \ldots, x^k)'$.
- $f(x) = \tilde{x}'\boldsymbol{Q}\tilde{x}$ is a sum of squares if and only if

$$f(x) = \tilde{x}'\boldsymbol{Q}\tilde{x},$$

  where $\boldsymbol{Q} \succeq \mathbf{0}$, i.e., $\boldsymbol{Q} = \boldsymbol{L'L}$.
- Then, $f(x) = \tilde{x}'\boldsymbol{L'L}\tilde{x} = ||L\tilde{x}||^2$.

4.4 Formulation

SLIDE 11

- Consider $\min f(x)$.
- Then, $f(x) \geq \gamma$ if and only if $f(x) - \gamma = \tilde{x}'\boldsymbol{Q}\tilde{x}$ with $\boldsymbol{Q} \succeq \mathbf{0}$. This implies linear constraints on γ and $\boldsymbol{Q}$.
- Reformulation

$$\max \gamma$$
$$\text{s.t.} \left\{ \begin{array}{rcl} f(x) - \gamma &=& \tilde{x}'\boldsymbol{Q}\tilde{x} \\ \boldsymbol{Q} &\succeq& \mathbf{0} \end{array} \right.$$

4.5 Example

4.5.1 Reformulation

Slide 12

$$\min\ f(x) = 3 + 4x + 2x^2 + 2x^3 + x^4.$$

$$f(x) - \gamma = 3 - \gamma + 4x + 2x^2 + 2x^3 + x^4 = (1, x, x^2)' \boldsymbol{Q} (1, x, x^2).$$

$$\begin{array}{rl} \max & \gamma \\ \text{s.t.} & 3 - \gamma = q_{11} \\ & 4 = 2q_{12}, \quad 2 = 2q_{13} + q_{22} \\ & 2 = 2q_{23}, \quad 1 = q_{33} \\ & \boldsymbol{Q} = \begin{bmatrix} q_{11} & q_{12} & q_{13} \\ q_{12} & q_{22} & q_{23} \\ q_{13} & q_{23} & q_{33} \end{bmatrix} \succeq \boldsymbol{0} \end{array}$$

Extensions to multiple dimensions.

5 Stability

Slide 13

- A linear difference equation

$$x(k+1) = \boldsymbol{A} x(k), \qquad x(0) = x_0$$

- $x(k)$ converges to zero iff $|\lambda_i(\boldsymbol{A})| < 1, i = 1, \dots n$
- Characterization:

$$|\lambda_i(\boldsymbol{A})| < 1 \quad \forall i \iff \exists \boldsymbol{P} \succ 0 \quad \boldsymbol{A}'\boldsymbol{P}\boldsymbol{A} - \boldsymbol{P} \prec 0$$

5.1 Proof

Slide 14

- ($\Longleftarrow$) Let $\boldsymbol{A} v = \lambda v$. Then,

$$0 > v'(\boldsymbol{A}'\boldsymbol{P}\boldsymbol{A} - \boldsymbol{P})v = (|\lambda|^2 - 1) \underbrace{v'\boldsymbol{P}v}_{>0},$$

  and therefore $|\lambda| < 1$

- ($\Longrightarrow$) Let $\boldsymbol{P} = \sum_{i=0}^{\infty} \boldsymbol{A}^{i\prime} \boldsymbol{Q} \boldsymbol{A}^i$, where $\boldsymbol{Q} \succ 0$. The sum converges by the eigenvalue assumption. Then,

$$\boldsymbol{A}'\boldsymbol{P}\boldsymbol{A} - \boldsymbol{P} = \sum_{i=1}^{\infty} \boldsymbol{A}^{i\prime} \boldsymbol{Q} \boldsymbol{A}^i - \sum_{i=0}^{\infty} \boldsymbol{A}^{i\prime} \boldsymbol{Q} \boldsymbol{A}^i = -\boldsymbol{Q} \prec 0$$

5.2 Stabilization

SLIDE 15

- Consider now the case where $\boldsymbol{A}$ is not stable, but we can change some elements, e.g., $\boldsymbol{A}(L) = \boldsymbol{A} + \boldsymbol{LC}$, where $\boldsymbol{C}$ is a fixed matrix.
- Want to find an $\boldsymbol{L}$ such that $\boldsymbol{A} + \boldsymbol{LC}$ is stable.
- Use Schur complements to rewrite the condition:

$$(\boldsymbol{A}+\boldsymbol{LC})'\boldsymbol{P}(\boldsymbol{A}+\boldsymbol{LC}) - \boldsymbol{P} \prec 0, \qquad \boldsymbol{P} \succ 0$$

$$\Updownarrow$$

$$\begin{bmatrix} \boldsymbol{P} & (\boldsymbol{A}+\boldsymbol{LC})'\boldsymbol{P} \\ \boldsymbol{P}(\boldsymbol{A}+\boldsymbol{LC}) & \boldsymbol{P} \end{bmatrix} \succ 0$$

Condition is nonlinear in $(\boldsymbol{P}, \boldsymbol{L})$

5.3 Changing variables

SLIDE 16

- Define a new variable $\boldsymbol{Y} := \boldsymbol{P}L$

$$\begin{bmatrix} \boldsymbol{P} & \boldsymbol{A}'\boldsymbol{P} + \boldsymbol{C}'\boldsymbol{Y}' \\ \boldsymbol{PA} + \boldsymbol{YC} & \boldsymbol{P} \end{bmatrix} \succ 0$$

- This is linear in $(\boldsymbol{P}, \boldsymbol{Y})$.
- Solve using SDO, recover $\boldsymbol{L}$ via $\boldsymbol{L} = \boldsymbol{P}^{-1}\boldsymbol{Y}$

6 Primal Barrier Algorithm for SDO

SLIDE 17

- $\boldsymbol{X} \succeq \boldsymbol{0} \Leftrightarrow \lambda_1(\boldsymbol{X}) \geq 0, \ldots, \lambda_n(\boldsymbol{X}) \geq 0$
- Natural barrier to repel $\boldsymbol{X}$ from the boundary $\lambda_1(\boldsymbol{X}) > 0, \ldots, \lambda_n(\boldsymbol{X}) > 0$:

$$-\sum_{j=1}^{n} \log(\lambda_i(\boldsymbol{X})) =$$

$$-\log(\prod_{j=1}^{n} \lambda_i(\boldsymbol{X})) = -\log(\det(\boldsymbol{X}))$$

SLIDE 18

- Logarithmic barrier problem

$$\begin{array}{ll} \min & B_\mu(\boldsymbol{X}) = \boldsymbol{C} \bullet \boldsymbol{X} - \mu \log(\det(\boldsymbol{X})) \\ \text{s.t.} & \boldsymbol{A}_i \bullet \boldsymbol{X} = b_i \quad, i = 1, \ldots, m, \\ & \boldsymbol{X} \succ \boldsymbol{0} \end{array}$$

- Derivative: $\nabla B_\mu(\boldsymbol{X}) = \boldsymbol{C} - \mu \mathbf{X}^{-1}$
  Follows from

$$\log\det(\boldsymbol{X} + \boldsymbol{H}) \approx \log\det(\boldsymbol{X}) + \text{trace}(\boldsymbol{X}^{-1}\boldsymbol{H}) + \cdots$$

- KKT conditions

$$\begin{aligned}&\mathbf{A}_i \bullet \mathbf{X} = \mathbf{b}_i \quad, i = 1, \dots, m,\\ &\boldsymbol{C} - \mu \mathbf{X}^{-1} = \sum_{i=1}^{m} y_i \mathbf{A}_i.\\ &\mathbf{X} \succ \mathbf{0},\end{aligned}$$

- Given μ, need to solve these nonlinear equations for X, C, y_i
- Apply Newton's method until we are "close" to the optimal
- Reduce value of μ, and iterate until the duality gap is small

6.1 Another interpretation

SLIDE 19

- Recall the optimality conditions:

$$\begin{aligned}\mathbf{A}_i \bullet \mathbf{X} &= \mathbf{b}_i \quad, i = 1, \dots, m,\\ \sum_{i=1}^{m} y_i \mathbf{A}_i + \mathbf{S} &= \mathbf{C}\\ \mathbf{X}, \mathbf{S} &\succeq \mathbf{0},\\ \mathbf{X}\,\mathbf{S} &= \mathbf{0}\end{aligned}$$

- Cannot solve directly. Rather, perturb the complementarity condition to $\mathbf{X}\,\mathbf{S} = \mu \mathbf{I}$.
- Now, unique solution for every $\mu > 0$ (the "central path")
- Solve using Newton, for decreasing values of μ.

7 Differences with LO

SLIDE 20

- Many different ways to linearize the nonlinear complementarity condition

$$\mathbf{X}\,\mathbf{S} = \mu \mathbf{I}$$

- Want to preserve symmetry of the iterates
- Several search directions, most common is Nesterov-Todd.

8 Convergence

8.1 Stopping criterion

SLIDE 21

- The point $(\mathbf{X}, y_i)$ is feasible, and has duality gap:

$$\mathbf{C} \bullet \mathbf{X} - \sum_{i=1}^{m} y_i b_i = \mu \mathbf{X}^{-1} \bullet \mathbf{X} = n\mu$$

- Therefore, reducing μ always decreases the duality gap
- Barrier algorithm needs $O\left(\sqrt{n} \log \frac{\epsilon_0}{\epsilon}\right)$ iterations to reduce duality gap from ϵ_0 to ϵ

9 Conclusions

Slide 22

- SDO is a very powerful modeling tool
- SDO represents the present and future in continuous optimization
- Barrier and primal-dual algorithms are very powerful
- Many good solvers available: SeDuMi, SDPT3, SDPA, etc.
- Pointers to literature and solvers:
  `www-user.tu-chemnitz.de/~helmberg/semidef.html`