多项式优化入门

王杰

中国科学院数学与系统科学研究院

北京师范大学, 2024 年秋季

课程内容

1. 半定规划
2. 平方和理论
3. 测度和矩
4. 矩-平方和松弛分层
5. 变量稀疏（CS）
6. 项稀疏（TS）
7. 扩展与应用
8. 软件与实验

线性规划（LP）

原问题：

$$p_{\mathrm{lp}} := \inf_{\mathbf{x}\in\mathbb{R}^n} \quad \mathbf{c}^\intercal \mathbf{x}$$
$$\text{s.t.} \quad A\mathbf{x} = \mathbf{b},$$
$$\mathbf{x} \geq 0$$

对偶问题：

$$d_{\mathrm{lp}} := \sup_{\mathbf{y}\in\mathbb{R}^m} \quad \mathbf{b}^\intercal \mathbf{y}$$
$$\text{s.t.} \quad A^\intercal \mathbf{y} \leq \mathbf{c}$$

线性规划的基本性质

- 弱对偶性：$p_{\mathrm{lp}} \geq d_{\mathrm{lp}}$
- 强对偶性：$p_{\mathrm{lp}} = d_{\mathrm{lp}}$
- 互补松弛：$\mathbf{x}^* \circ (\mathbf{c} - A^\intercal \mathbf{y}^*) = \mathbf{0}$

基本定义

- $\mathbb{S}_n$: n 阶对称矩阵
- $\mathbb{S}_n^+$: n 阶半正定矩阵
- $A \in \mathbb{S}_n^+ \iff A \succeq 0$
- $\langle A, B \rangle = \mathrm{tr}(AB)$ for $A, B \in \mathbb{S}_n$

半定规划（SDP）

原问题：

$$
\begin{aligned}
p_{\mathrm{sdp}} := \inf \quad & \langle C, X\rangle \\
\text{s.t.} \quad & \langle A_i, X\rangle = b_i, \quad i = 1, \ldots, m, \\
& X \succeq 0
\end{aligned}
$$

对偶问题：

$$
\begin{aligned}
d_{\mathrm{sdp}} := \sup \quad & \mathbf{b}^\intercal \mathbf{y} \\
\text{s.t.} \quad & \sum_{i=1}^{m} A_i y_i \preceq C
\end{aligned}
$$

- 线性规划的推广
- 最优电力流、组合优化、计算机视觉、量子信息、信号处理...

半定规划的弱对偶性

- 弱对偶性： $p_{\mathrm{sdp}} \geq d_{\mathrm{sdp}}$

设 X 和 $\mathbf{y}$ 分别是原问题和对偶问题的可行解：

$$\langle C, X\rangle - \mathbf{b}^{\mathsf{T}}\mathbf{y} = \langle C, X\rangle - \sum_{i=1}^{m}\langle A_i, X\rangle y_i = \left\langle C - \sum_{i=1}^{m} A_i y_i, X \right\rangle \geq 0$$

半定规划的最优性条件

- X 和 $\mathbf{y}$ 分别是原问题和对偶问题的可行解
- 最优性条件：

$$\left(C-\sum_{i=1}^{m} A_i y_i\right) X = 0 \iff \left\langle C-\sum_{i=1}^{m} A_i y_i, X\right\rangle = 0 \iff \langle C, X\rangle = \mathbf{b}^{\intercal}\mathbf{y}$$

$\Longrightarrow$ X 和 $\mathbf{y}$ 分别是原问题和对偶问题的最优解

半定规划的强对偶性

- 若原问题或对偶问题**严格可行**，则强对偶性成立：$p_{\text{sdp}} = d_{\text{sdp}}$
- 若原问题和对偶问题**都是可行的**，则原问题和对偶问题均有最优解

多块半定规划

原问题：

$$
\begin{aligned}
p_{\mathrm{sdp}} := \inf \quad & \sum_{k=1}^{t} \langle C_k, X_k \rangle \\
\text{s.t.} \quad & \sum_{k=1}^{t} \langle A_i^k, X_k \rangle = b_i, \quad i = 1, \ldots, m \\
& X_k \succeq 0, \quad k = 1, \ldots, t
\end{aligned}
$$

对偶问题：

$$
\begin{aligned}
d_{\mathrm{sdp}} := \sup \quad & \mathbf{b}^{\intercal} \mathbf{y} \\
\text{s.t.} \quad & \sum_{i=1}^{m} A_i^k y_i \preceq C_k, \quad k = 1, \ldots, t
\end{aligned}
$$

应用：二元二次规划（BQP）

$$\begin{cases} \inf\limits_{\mathbf{x}} & \mathbf{x}^\intercal Q \mathbf{x} \\ \text{s.t.} & x_i^2 = 1, \quad i = 1, \ldots, n \end{cases} \Longleftrightarrow \begin{cases} \inf\limits_{X} & \langle Q, X \rangle \\ \text{s.t.} & X = \mathbf{x}\mathbf{x}^\intercal \succeq 0 \\ & X_{ii} = 1, \quad i = 1, \ldots, n \end{cases}$$

$$\Longleftrightarrow \begin{cases} \inf\limits_{X} & \langle Q, X \rangle \\ \text{s.t.} & X \succeq 0, \quad \operatorname{rank}(X) = 1 \\ & X_{ii} = 1, \quad i = 1, \ldots, n \end{cases} \xRightarrow{\text{relax}} \begin{cases} \inf\limits_{X} & \langle Q, X \rangle \\ \text{s.t.} & X \succeq 0 \\ & X_{ii} = 1, \quad i = 1, \ldots, n \end{cases}$$

Goemans-Williamson rounding

- **Step 1**：将 X 分解成 $X = V^\intercal V$，其中 $V = [v_1, \ldots, v_n] \in \mathbb{R}^{r \times n}$，$r = \mathrm{rank}(X)$. 由于 $v_i^\intercal v_i = X_{ii} = 1$, 我们得到 r 维单位球面上的 n 个点 $v_1, \ldots, v_n$.
- **Step 2**：随机选取 $\mathbb{R}^r$ 中的一个超平面（经过原点），根据 v_i 位于超平面的哪一侧给 x_i 赋值 $+1$ 或 -1.

最大割问题的近似比

- 对角占优矩阵：Q 是对称矩阵且对于所有的 i，$Q_{ii} \geq \sum_{j \neq i} |Q_{ij}|$
- 最大割问题：$\max\limits_{x_i = \pm 1} \frac{1}{4} \sum_{(i,j) \in E} (x_i - x_j)^2$

定理

假设 X^* 是 *SDP* 松弛问题的最优解，$\mathbf{x}^{\mathrm{sdp}}$ 是 *Goemans-Williamson rounding* 给出的近似解. 那么，

$$\alpha_{GW} \cdot \langle Q, X^* \rangle \leq \mathbf{E}\left((\mathbf{x}^{\mathrm{sdp}})^\intercal Q \mathbf{x}^{\mathrm{sdp}} \right) \leq \langle Q, X^* \rangle,$$

其中，$\alpha_{GW} \approx 0.878.$

应用：低秩矩阵补全

- 低秩矩阵补全问题：给定低秩矩阵 $M \in \mathbb{R}^{s\times t}$ 的一组矩阵元

$\{M_{ij}\}_{(i,j)\in\Omega}$，恢复出低秩矩阵 M

- 可建模为凸优化问题：

$$\begin{cases} \inf\limits_{Z\in\mathbb{R}^{s\times s}} & \|Z\|_* \\ \text{s.t.} & Z_{ij} = M_{ij}, \quad \forall (i,j)\in\Omega \end{cases}$$

其中，$\|Z\|_* := \mathrm{Tr}(Z^\intercal Z)^{\frac{1}{2}}$ 是 Z 的核范数

应用：低秩矩阵补全

$$
\begin{cases}
\inf\limits_{Z\in\mathbb{R}^{s\times s}} & \|Z\|_* \\
\text{s.t.} & Z_{ij} = M_{ij}, \forall (i,j)\in\Omega
\end{cases}
\iff
\begin{cases}
\inf\limits_{X\in\mathbb{S}_{2s}} & \operatorname{Tr}(X) \\
\text{s.t.} & \begin{bmatrix} 0_{s\times s} & E_{ij}^{\intercal} \\ E_{ij} & 0_{s\times s} \end{bmatrix} X = 2M_{ij}, \forall (i,j)\in\Omega \\
& X = \begin{bmatrix} U & Z^{\intercal} \\ Z & V \end{bmatrix} \succeq 0
\end{cases}
$$

定理 (Candes, Recht, and Tao, 2010)

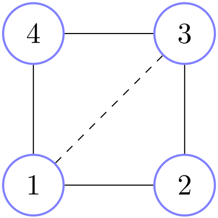
假设矩阵 M 低秩且 *incoherent*, 样本个数满足 $|\Omega| \geq Cn(\log n)^2$, $n = s + t$, C 是常数. 则 M 可通过求解半定规划问题被准确恢复.

图的基本定义

- 图 $G(V,E)$：顶点集 V，边集 $E \subseteq \{\{v_i, v_j\} \mid v_i \neq v_j, (v_i, v_j) \in V \times V\}$
- 团（clique）：完全子图
- 极大团（maximal clique）：不包含于更大的团
- 最大团：包含顶点个数最多的团
- 团数：最大团包含顶点的个数

弦图（chordal graph）

- 弦：连接圈中非邻接顶点的边
- 弦图：长度大于等于 4 的圈必有一条弦



RIP 条件

子集 $I_1, \ldots, I_p \subseteq [n]$ 满足 RIP（running intersection property）：对每个 $k \in [p-1]$，存在 $i \leq k$ 使得

$$\left(I_{k+1} \cap \bigcup_{j \leq k} I_j \right) \subseteq I_i$$

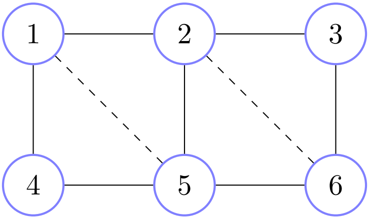
定理

一个连通图是弦图当且仅当它的极大团集满足 *RIP.*

弦扩张

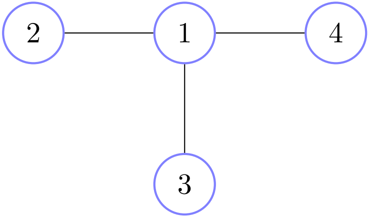
- 弦扩张：任何非弦图通过添加合适的边可以变成弦图
  - 最大弦扩张：将连通分支变成完全图
  - （近似）最小弦扩张：团数尽可能小

图：弦扩张



稀疏矩阵

- $\mathbb{S}(G,0)$: $Q \in \mathbb{S}_n$, $Q_{ij} = Q_{ij} = 0, \forall i \neq j$ 且 $\{i,j\} \notin E$
- $\mathbb{S}^+(G,0) := \mathbb{S}(G,0) \cap \mathbb{S}_n^+$
- $\mathbb{S}(G,0)$ 中的矩阵具有块稀疏结构：每个块对应 G 的一个极大团



$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}$$

稀疏矩阵

- 对 $G(V,E)$ 的任一极大团 C，定义矩阵 $R_C \in \mathbb{R}^{|C|\times|V|}$:

$$[R_C]_{ij} = \begin{cases} 1, & \text{如果} C(i) = j \\ 0, & \text{其他} \end{cases}$$

- $Q_C = R_C Q R_C^\intercal \in \mathbb{S}_{|C|}$：提取 Q 的子矩阵 Q_C
- $Q = R_C^\intercal Q_C R_C$：将 Q_C 扩充成稀疏矩阵 Q

稀疏半正定矩阵分解

定理

设 G 是弦图，$C_1, \ldots, C_p$ 是 G 的极大团集. 则 $Q \in \mathbb{S}^+(G, 0)$ 当且仅当存在 $Q_k \in \mathbb{S}^+_{|C_k|}$，$k \in [p]$ 使得 $Q = \sum_{k=1}^{p} R_{C_k}^{\intercal} Q_k R_{C_k}$.

$$Q = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 1 \end{bmatrix} \succeq 0 \iff Q = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} (\succeq 0) + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix} (\succeq 0)$$

半正定补全矩阵

- 对图 G 定义投影 $\Pi_G : \mathbb{S}_n \to \mathbb{S}(G,0)$：

$$[\Pi_G(Q)]_{ij} = \begin{cases} Q_{ij}, & \text{如果} i = j \text{ 或} \{i,j\} \in E \\ 0, & \text{其他} \end{cases}$$

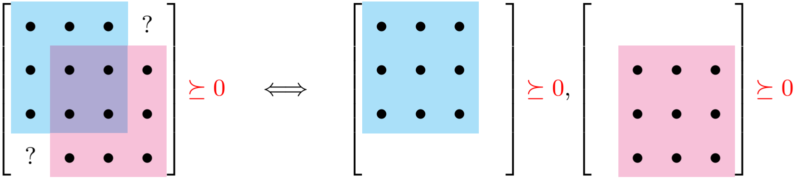
- 半正定补全矩阵：

$$\mathbb{S}^+(G,?) := \Pi_G(\mathbb{S}_n^+) = \{\Pi_G(Q) \mid Q \in \mathbb{S}_n^+\}$$

半正定补全矩阵刻画

定理

设 G 是弦图，$C_1, \ldots, C_p$ 是 G 的极大团集. 则 $Q \in \mathbb{S}^+(G, ?)$ 当且仅当 $Q_{C_i} \succeq 0,\ i = 1, \ldots, p.$



SDP 的主要算法

➤通用算法

- 内点法：高精度解，内存占用大
- ALM+半光滑牛顿法：中/高精度解，内存占用较小
- ADMM：低精度解，内存占用较小

➤利用常数迹性质

- CGAL：低精度解，内存占用小

➤利用低秩性质（$m \leq n$）

- Burer-Monteiro 分解 (local search/流形优化)：高精度解，内存占用小

➤利用秩一性质（$m \gg n$）

- Local search + L-BFGS：高精度解，内存占用小

SDP 的主要算法

➤通用算法

- 内点法：高精度解，内存占用大
- ALM+半光滑牛顿法：中/高精度解，内存占用较小
- ADMM：低精度解，内存占用较小

➤利用常数迹性质

- CGAL：低精度解，内存占用小

➤利用低秩性质（$m \leq n$）

- Burer-Monteiro 分解 (local search/流形优化)：高精度解，内存占用小

➤利用秩一性质（$m \gg n$）

- Local search + L-BFGS：高精度解，内存占用小

SDP 的主要算法

➤通用算法

- 内点法：高精度解，内存占用大
- ALM+半光滑牛顿法：中/高精度解，内存占用较小
- ADMM：低精度解，内存占用较小

➤利用常数迹性质

- CGAL：低精度解，内存占用小

➤利用低秩性质（$m \leq n$）

- Burer-Monteiro 分解 (local search/流形优化)：高精度解，内存占用小

➤利用秩一性质（$m \gg n$）

- Local search + L-BFGS：高精度解，内存占用小

SDP 的主要算法

➤通用算法

- 内点法：高精度解，内存占用大
- ALM+ 半光滑牛顿法：中/高精度解，内存占用较小
- ADMM：低精度解，内存占用较小

➤利用常数迹性质

- CGAL：低精度解，内存占用小

➤利用低秩性质（$m \leq n$）

- Burer-Monteiro 分解 (local search/流形优化)：高精度解，内存占用小

➤利用秩一性质（$m \gg n$）

- Local search + L-BFGS：高精度解，内存占用小

SDP 软件

- 商业软件：MOSEK，COPT
- 开源软件
  - 内点法：SDPT3，SEDUMI，SDPA
  - 半光滑牛顿 ALM：SDPNAL+
  - ADMM：COSMO，CDCS
  - 低秩 ALM：SDPLR，ManiSDP
  - 秩一 L-BFGS：STRIDE

下次课

- 平方和理论

参考文献

- Jie Wang and Victor Magron, *Sparse Polynomial Optimization: Theory and Practice*, World Scientific Publishing, 2023.

更多信息见个人主页

https://wangjie212.github.io/jiewang