Convex Optimization

Prof. Nati Srebro

Lecture 9:
Matrix Inequalities and Semi-Definite Programming

Reading: Boyd and Vandenberghe Sections 2.2.5, 4.6.2, 4.6.3, 5.9.1-5.9.3

Matrix Inequalities (Semi Definite Constraints)

$$\boxed{\begin{array}{ll} \min\limits_{x\in\mathbb{R}^n} & f_0(x) \\ s.t. & f_i(x) \preccurlyeq 0 \\ & h_i(x) = 0 \end{array}}$$

$f_0: \mathbb{R}^n \to \mathbb{R}$ convex

$f_i: \mathbb{R}^n \to S^{k_i} \subset \mathbb{R}^{k_i \times k_i}$ **matrix-convex**:

$$f_i(\theta x + (1-\theta)x') \preccurlyeq \theta f_i(x) + (1-\theta) f_i(x')$$

$h_i: \mathbb{R}^n \to \mathbb{R}$ linear (or $\to S^k$, doesn't matter much)

Sometimes convenient to represent $x \in S^n$: doesn't matter much, its just a vector space of dimensionality $n(n+1)/2$

Semi-Definite Programming (SDP): f_0, f_i **linear**

Example: Closest Legal Covariance

$$\boxed{\begin{array}{ll} \min\limits_{A \in S^n} & \sum_{ij} \left| A_{ij} - \tilde{A}_{ij} \right| \\ s.t. & A \succcurlyeq 0 \end{array}}$$

Example: Fastest Mixing Markov Chain

$$\begin{array}{ll}
\min\limits_{\substack{P\in S^n \\ t\in\mathbb{R}}} & t \\
& \forall_i \sum_j P_{ij} = 1 \qquad (P\mathbf{1} = \mathbf{1}) \\
& \forall_{ij} P_{ij} \geq 0 \qquad (P \geq 0) \\
& \forall_{ij\notin E} P_{ij} = 0 \\
& \forall_{ij} P_{ij} = P_{ji} \qquad (P^\top = P) \\
& -tI \preccurlyeq P - \frac{1}{n}\mathbf{1}\mathbf{1}^\top \preccurlyeq tI
\end{array}$$

- Given undirected graph $G(V,E)$ on $V = \{1, \dots, n\}$ we want to construct random walk $X(t)$ on graph with symmetric transitions
$$P(X(t+1) = j | X(t) = i) = P_{ij} = P_{ji}$$
that minimizes the "mixing time", i.e. the time by which $X(t)$ is approximately uniform and independent of $X(0)$.

- Eigenvalues of P: $1 = \lambda_1 \geq \lambda_2 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq -1$

  $P\mathbf{1} = \mathbf{1}$, hence 1 is e.vec. with e.val. 1

- Claim: mixing time $\propto -\frac{1}{\log(\max(\lambda_2, -\lambda_n))}$

  ➔ we want to minimize $\max(\lambda_2, -\lambda_n)$

Example: Max-Cut Relaxation

- Given weighted undirected graph $G([n], E)$ on nodes $[n] = \{1, .., n\}$ with weights $w_{ij} > 0$, find maximal cut, i.e. partition $[n] = A \cup B$ maximizing:

$$\sum_{i \in A, j \in B} w_{ij}$$

- A quadratic integer program:

$$\begin{array}{ll} \min\limits_{x_i \in \mathbb{R}} & -\sum_{ij \in E} w_{ij} \left(\frac{(1 - x_i x_j)}{2} \right) \\ s.t. & x_i \in \pm 1 \end{array}$$

- Relaxing to vector "indicators":

$$\begin{array}{ll} \min\limits_{v_i \in \mathbb{R}^n} & -\sum_{ij \in E} w_{ij} \left(\frac{(1 - \langle v_i, v_j \rangle)}{2} \right) \\ s.t. & \|v_i\| = 1 \end{array}$$

- Equivalent program, $K = VV^\top$:

$$\begin{array}{ll} \min\limits_{K \in S^n} & -\sum_{ij \in E} w_{ij} \left(\frac{(1 - K_{ij})}{2} \right) \\ s.t. & K \succcurlyeq 0 \\ & K_{ii} = 1 \end{array}$$

- Randomized rounding: $\tilde{x}_i = sign(\langle u, v_i^* \rangle)$ for random vector u
- Theorem (Goemans Williamson '94): $\mathbb{E}_u[f_0(\tilde{x})] \leq 0.87 f_0(V^*) \leq 0.87 f_0(x^*)$

Duality with Matrix Inequalities

$$\boxed{\begin{array}{ll} \min\limits_{x \in \mathbb{R}^n} & f_0(x) \\ s.t. & f_i(x) \preccurlyeq 0 \\ & h_i(x) = 0 \end{array}} \quad \begin{array}{l} \\ \lambda_i \in S^{k_i} \\ \nu_i \in \mathbb{R} \end{array}$$

$f_0 \colon \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ convex, $f_i \colon \mathbb{R}^n \to S^{k_i}$ matrix convex, $h_j \colon \mathbb{R}^n \to \mathbb{R}$ linear

$$L\big(x, (\lambda, \nu)\big) = f_0(x) + \sum_{i=1}^{m} \langle \lambda_i, f_i\,(x) \rangle + \sum_{j=1}^{p} \nu_j h_j(x)$$

$$p^* = \inf_x \sup_{\lambda_i \succcurlyeq 0, \nu} L\big(x, (\lambda, \nu)\big)$$

$f_i(x) \preccurlyeq 0 \rightarrow$ for $\lambda_i \succcurlyeq 0$ we have $\langle f_i(x), \lambda_i \rangle \leq 0$ and $\sup = 0$

$f_i(x)$ has a eigenvalue $s > 0$ with eigenvector v

$\rightarrow$ for $\lambda_i = tvv^\top$ we have $\langle f_i(x), \lambda_i \rangle = ts \xrightarrow{t \to \infty} \infty$

Duality with Matrix Inequalities

$$\begin{array}{ll} \min\limits_{x \in \mathbb{R}^n} & f_0(x) \\ s.t. & f_i(x) \preccurlyeq 0 \quad \lambda_i \in S^{k_i} \\ & h_i(x) = 0 \quad \nu_i \in \mathbb{R} \end{array}$$

$$\begin{array}{ll} \max\limits_{\lambda_i \in S^{k_i}, \nu \in \mathbb{R}^p} & g(\lambda, \nu) \\ s.t. & \lambda_i \succcurlyeq 0 \end{array}$$

$f_0 \colon \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ convex, $f_i \colon \mathbb{R}^n \to S^{k_i}$ matrix convex, $h_j \colon \mathbb{R}^n \to \mathbb{R}$ linear

$$L\big(x, (\lambda, \nu)\big) = f_0(x) + \sum_{i=1}^{m} \langle \lambda_i, f_i\,(x) \rangle + \sum_{j=1}^{p} \nu_j h_j(x)$$

$$p^* = \inf_x \sup_{\lambda_i \succcurlyeq 0, \nu} L\big(x, (\lambda, \nu)\big) \geq \sup_{\lambda_i \succcurlyeq 0, \nu} \inf_x L\big(x, (\lambda, \nu)\big) = d^*$$

$$g(\lambda, \nu) = \inf_x L\big(x, (\lambda, \nu)\big)$$

Slater ($\exists$ feasible x s.t. $f_i(x) \prec 0$) ➔ Strong duality

Complementary Slackness

- With strong duality, for optimal x^*, (λ^*, ν^*):

$$f_0(x^*) = g(\lambda^*, \nu^*) = \inf_x L(x, (\lambda^*, \nu^*)) \leq L(x^*, (\lambda^*, \nu^*))$$

$$= f_0(x^*) + \sum_i \underbrace{\langle \underset{\succcurlyeq 0}{\lambda_i^*}, f_i(\underset{\preccurlyeq 0}{x^*}) \rangle}_{\leq 0} + \sum_i \nu_i^* h_i(x)$$

➔ $\langle \lambda_i^*, f_i(x^*) \rangle = 0$

- Does this mean that $\lambda_i^* = 0$ or $f_i(x^*) = 0$?
  - No: e.g. $\lambda_i^* = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$, $f_i(x^*) = \begin{bmatrix} 0 & 0 \\ 0 & -1 \end{bmatrix}$
- But: since $0 = tr\left(\lambda_i^{*\top} f_i(x^*)\right)$ and $\lambda_i^{*\top} f_i(x^*) \preccurlyeq 0$, we must have:

$$\lambda_i^{*\top} f_i(x^*) = 0 \in S^{k_i}$$

- Also: $x^* = \arg\min L(x, (\lambda^*, \nu^*))$

Optimality Condition (assuming Strong Duality)

x^* optimal for (P) $\Rightarrow$

λ^*, ν^* optimal for (D) $\Leftarrow$

KKT Conditions

$f_i(x^*) \preccurlyeq 0 \; \forall_{i=1\ldots m}$

$h_j(x^*) = 0 \; \forall_{j=1\ldots p}$

$\lambda_i^* \succcurlyeq 0 \; \forall_{i=1..m}$

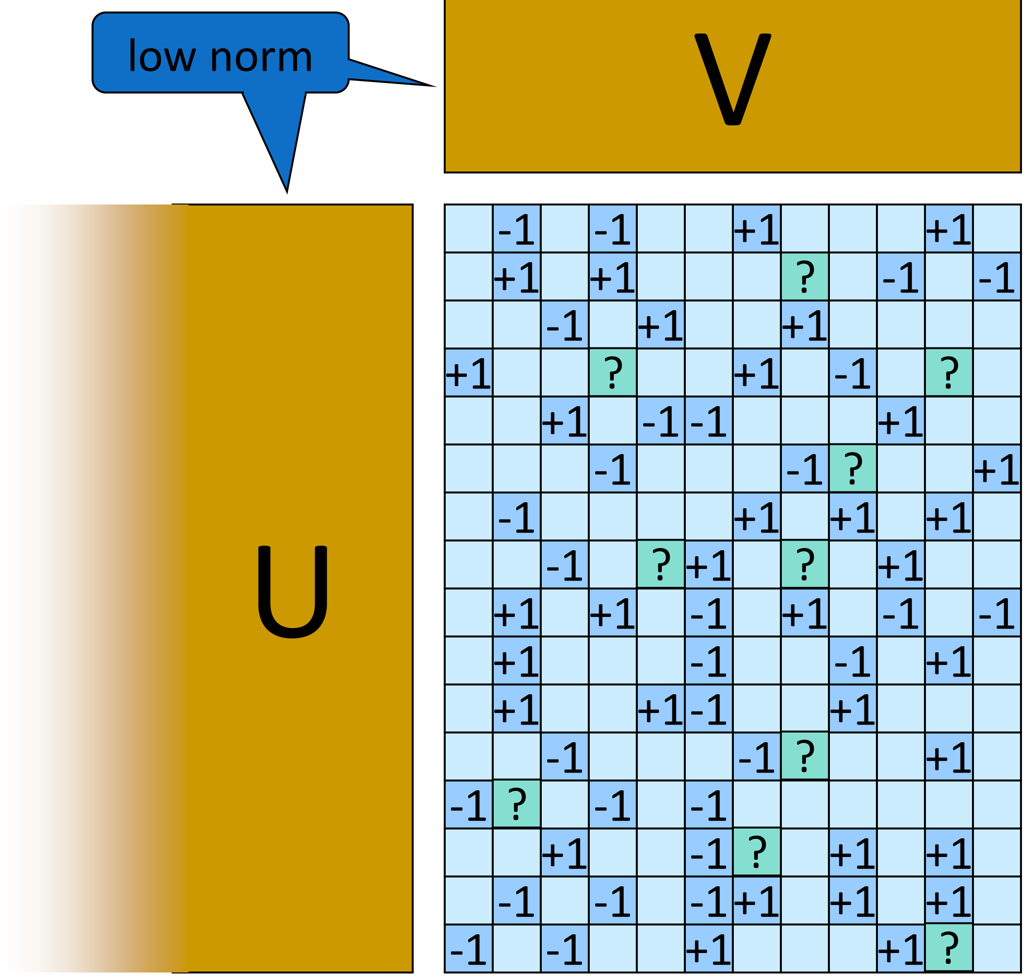
$\nabla_x L\left(x^*, (\lambda^*, \nu^*)\right) = 0$

or $0 \in \partial_x L\left(x^*, (\lambda^*, \nu^*)\right)$

$\langle \lambda_i^*, f_i\,(x^*)\rangle = 0 \; \forall_{i=1..m}$

$\boxed{x^* \text{ optimal for (P)} \Leftrightarrow \exists_{\lambda^*,\nu^*} \text{ s.t. } x^*, \lambda^*, \nu^* \text{ satisfy KKT}}$

Matrix Completion



Given partial observations Y_{ij} for $(i, j) \in S$, reconstruct factors U, V s.t.

$$Y_{ij}\langle u_i, v_j\rangle \quad \geq 1$$

Matrix Completion

$$\boxed{\begin{array}{ll} \min\limits_{u_i, v_j \in \mathbb{R}^k} & \frac{1}{2}\left(\sum_{i=1}^{n} \|u_i\|^2 + \sum_{j=1}^{m} \|v_i\|^2\right) \\ s.t. & \forall_{ij \in S} Y_{ij} \langle u_i, v_j \rangle \geq 1 \end{array}}$$

- If $k \geq n + m$, can express as:

$$\boxed{\begin{array}{ll} \min\limits_{\substack{A \in S^n, B \in S^m \\ X \in \mathbb{R}^{n \times m}}} & \frac{1}{2}\left(tr(A) + tr(B)\right) \\ s.t. & \begin{array}{c|cc} & U & V^\top \\ \hline U^\top & A & X \\ V & X' & B \end{array} \quad \begin{bmatrix} A & X \\ X' & B \end{bmatrix} \succcurlyeq 0 \\ & \forall_{ij \in S} Y_{ij} X_{ij} \geq 1 \end{array}}$$

- $(n + m)^2$ variables
- $(n + m) \times (n + m)$ matrix inequality + $|S|$ scalar inequalities

Dual?

Dual of Matrix Completion

$$\min_{\substack{A\in S^n, B\in S^m\\ X\in\mathbb{R}^{n\times m}}} \frac{1}{2}(tr(A)+tr(B))$$

$$s.t.\quad \begin{bmatrix} A & X \\ X' & B \end{bmatrix} \succcurlyeq 0 \qquad \begin{bmatrix} \lambda_A & \lambda_X \\ \lambda_X^\top & \lambda_B \end{bmatrix}$$

$$\forall_{ij\in S} Y_{ij}X_{ij} \geq 1 \qquad \alpha_{ij}$$

$$\max_{\alpha\in\mathbb{R}^S} \sum_{ij\in S}\alpha_{ij}$$

$$s.t.\quad \alpha \geq 0$$

$$2\lambda = \begin{bmatrix} I_n & Q(\alpha) \\ Q(\alpha)^\top & I_m \end{bmatrix} \succcurlyeq 0$$

$Q_{ij} = Y_{ij}\alpha_{ij}$ or
$Q_{ij} = 0$ if $ij \notin S$

$\|Q(\alpha)\|_2 \leq 1$

$$L\big((X,A,B),(\lambda,\alpha)\big) = \frac{1}{2}(trA + trB) - \left\langle \begin{bmatrix} \lambda_A & \lambda_X \\ \lambda_X^\top & \lambda_B \end{bmatrix}, \begin{bmatrix} A & X \\ X & B \end{bmatrix} \right\rangle + \sum_{ij\in S}\alpha_{ij}(1 - Y_{ij}X_{ij})$$

$$= -2\langle \lambda_X, X\rangle - \sum_{ij\in S}\alpha_{ij}Y_{ij}X_{ij} + \langle \frac{1}{2}I_n - \lambda_A, A\rangle + \langle \frac{1}{2}I_m - \lambda_B, B\rangle + \sum_{ij\in S}\alpha_{ij}$$

$trA = \langle I, A\rangle$ $\qquad$ $trB = \langle I, B\rangle$

$$g(\lambda,\alpha) = \inf_{X,A,B} L\big((X,A,B),(\lambda,\alpha)\big) = \sum\alpha_{ij}$$

$$\text{s.t. } \lambda_A = \frac{1}{2}I_n,\ \lambda_B = \frac{1}{2}I_m \text{ and } (\lambda_X)_{ij} = \begin{cases} 0 & ij \notin S \\ -\frac{1}{2}Y_{ij}\alpha_{ij} & ij \in S \end{cases}$$