Semidefinite Programming - Quick Introduction

Source: Matoušek semidefinite programming

Recall: Let $A \in \mathbb{R}^{n\times n}$. The *trace* of A is $Tr(A) = \sum_{i=1}^{n} a_{i,i}$.

Let $SYM_n \subseteq \mathbb{R}^{n\times n}$ be the set of all symmetric $n \times n$ real-valued matrices.

For $X, Y \in \mathbb{R}^{n\times n}$, let the dot product of X and Y be $X \bullet Y = Tr(X^T Y)$.

We say $X \in SYM_n$ is *positive semidefinite* if $v^T X v \geq 0$ for all $v \in \mathbb{R}^n$, denoted by $X \succeq 0$.

1: Show that if $X \succeq 0$, then $X_{i,i} \geq 0$ for all i.

Solution: If e_i is the ith basis vector, then $0 \leq e_i^T X e_i = x_{i,i}$.

$$(LP)\begin{cases} \text{maximize} & \mathbf{c}^T\mathbf{x} \\ \text{subject to} & A\mathbf{x} = \mathbf{b} \\ & \mathbf{x} \geq 0 \end{cases} \quad \text{is equivalent to } (LP)\begin{cases} \text{maximize} & \mathbf{c}\bullet\mathbf{x} \\ \text{subject to} & \mathbf{a}_1\bullet\mathbf{x} = b_1 \\ & \mathbf{a}_2\bullet\mathbf{x} = b_2 \\ & \vdots \\ & \mathbf{a}_m\bullet\mathbf{x} = b_m \\ & \mathbf{x} \geq 0 \end{cases}$$

where $\mathbf{c}, \mathbf{x} \in \mathbb{R}^n$, $\mathbf{b}, \in \mathbb{R}^m$, $A \in \mathbb{R}^{m\cdot n}$, and $\mathbf{a}_i$ is the ith row of A.

A semidefinite program (SDP) is a generalization of a linear program, with matrices instead of vectors.

$$(SDP)\begin{cases} \text{maximize} & C\bullet X \\ \text{subject to} & A_1\bullet X = b_1 \\ & A_2\bullet X = b_1 \\ & \vdots \\ & A_m\bullet X = b_m \\ & X \succeq 0 \end{cases}$$

Where $C, X, A_i \in SYM_n$ and $b_i \in \mathbb{R}$.

2: Compute

$$Tr\left(\begin{pmatrix} c_{11} & c_{12} \\ c_{12} & c_{22}\end{pmatrix}^T \begin{pmatrix} x_{11} & x_{12} \\ x_{12} & x_{22}\end{pmatrix}\right) = \begin{pmatrix} c_{11} & c_{12} \\ c_{12} & c_{22}\end{pmatrix} \bullet \begin{pmatrix} x_{11} & x_{12} \\ x_{12} & x_{22}\end{pmatrix} =$$

Solution: $= c_{11}x_{11} + 2c_{12}x_{12} + c_{22}x_{22}$

3: Show that the following is an equivalent form of (SDP) up to some scaling.

$$(SDP)\begin{cases} \text{maximize} & \sum_{i\leq j} c_{i,j}x_{i,j} \\ \text{subject to} & \sum_{i\leq j} a_{i,j,k}x_{i,j} = b_k \quad \text{for } k = 1 \dots m \\ & X \succeq 0 \end{cases}$$

Hint: How about the diagonal terms?

Solution: In the original original problem, the diagonal elements contribute once whereas the upper triangle contributes twice. Simply halve the diagonal coefficients of the matrix C.

4: Write the following linear program as a semidefinite program (use matrices and their dot product).

$$(LP)\begin{cases} \text{maximize} & 2x_1 + 3x_2 \\ \text{subject to} & x_1 + 2x_2 = 1 \\ & x_1 - x_2 \geq 2 \\ & x_1, x_2 \geq 0 \end{cases}$$

Solution: One needs to add one slack variable for equality.

$$(SDP)\begin{cases} \text{maximize} & \begin{pmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 0 \end{pmatrix} \bullet X \\ \text{subject to} & \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{pmatrix} \bullet X = 1 \\ & \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \bullet X = 2 \\ & X \succeq 0 \end{cases}$$

5: Write the following general linear program as a semidefinite program.

$$(LP)\begin{cases} \text{maximize} & \mathbf{c}^T\mathbf{x} \\ \text{subject to} & A\mathbf{x} = \mathbf{b} \\ & \mathbf{x} \geq 0 \end{cases}$$

Solution: We will make $\mathbf{x}$ correspond to the diagonal of $X \succeq 0$. Denote ith row of A by $\mathbf{a}_i$. Suppose $A \in \mathbb{R}^{m\times n}$. Create matrices C and A_i, where

$$C_{k,\ell} = \begin{cases} \mathbf{c}_k & \text{if } k = \ell \\ 0 & \text{otherwise} \end{cases} \qquad (A_i)_{k,\ell} = \begin{cases} (\mathbf{a}_i)_k & \text{if } k = \ell \\ 0 & \text{otherwise} \end{cases}$$

That is

$$C = \begin{pmatrix} \mathbf{c}_1 & 0 & \cdots & 0 \\ 0 & \mathbf{c}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{c}_n \end{pmatrix} \qquad A_i = \begin{pmatrix} (\mathbf{a}_i)_1 & 0 & \cdots & 0 \\ 0 & (\mathbf{a}_i)_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & (\mathbf{a}_i)_n \end{pmatrix}$$

The we put them to (SDP). Notice that off diagonal entries of X do not matter and $X \succeq 0$ means that all entries on the diagonal of X are ≥ 0.

$$(SDP)\begin{cases} \text{maximize} & C \bullet X \\ \text{subject to} & A_i \bullet X = b_i \quad \text{for all } 1 \leq i \leq m \\ & X \succeq 0 \end{cases}$$

Dual form of (SDP) is

$$(DSDP)\begin{cases} \text{minimize} & \mathbf{b}^T\mathbf{y} \\ \text{subject to} & y_1A_1 + y_2A_2 + \cdots + y_mA_n - C \succeq 0 \end{cases}$$

(SDP) is *strictly feasible* if exists feasible X which is positive definite ($X \succ 0$).

$(DSDP)$ is *strictly feasible* if exists feasible $\mathbf{y}$ such that $(\sum_i \mathbf{y}A_i) - C \succ 0$.

Theorem: Strong duality of (SDP)
If (SDP) is strictly feasible and has an optimal solution of value γ, then $(DSDP)$ is feasible and has an optimal solution of value γ.

If $(DSDP)$ is strictly feasible and has an optimal solution of value γ, then $(DSDP)$ is feasible and has an optimal solution of value γ.

Theorem: Solvability of (SDP) in polynomial time
Let (SDP) be feasible, set of feasible solutions F bounded. Let $R \in \mathbb{N}$ be such that $R \geq \sqrt{Tr(X^TX)}$ for all $X \in F$ and $\varepsilon > 0$ be constants. Let n be the size of a binary encoding of (SDP). Then in polynomial time in n we can compute $X' \in F$ of value at least $optimum - \varepsilon$.

In other words, if no solution is not too big (R) and we are happy with ε precision, we have a polynomial time algorithm.

A solution can be obtained using interior point methods. There exist free and efficient implementations CSDP and SDPA.

6: Let $A \in SYM_n$. A principal minor (of order k) of A is a determinant of a $k \times k$ submatrix that is obtained by picking k rows and k columns. A theorem is saying that A is positive semidefinite if and only if all of its principal minors are nonnegative.

What does it mean for a 2×2 matrix A?

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} \\ a_{1,2} & a_{2,2} \end{pmatrix} \succeq 0$$

Solution: Principal minors of size 1 are saying $a_{1,1} \geq 0$ and $a_{2,2} \geq 0$. The principal minor of size 2×2 is just the determinant of A, which is

$$\begin{vmatrix} a_{1,1} & a_{1,2} \\ a_{1,2} & a_{2,2} \end{vmatrix} = a_{1,1}a_{2,2} - a_{1,2}^2.$$

Notice that we got non-negativity constraints for the diagonal and some kind of a quadratic constraint! for the off diagonal entries.

The following exercise will demonstrate that semidefinite programming can contain some quadratic terms.

7: Write the following program (P) as $(DSDP)$

$$(P)\begin{cases} \text{minimize} & \frac{(\mathbf{c}^T\mathbf{x})^2}{\mathbf{d}^T\mathbf{x}} \\ \text{subject to} & A\mathbf{x}+\mathbf{b} \geq 0 \end{cases}$$

where $\mathbf{d}^T\mathbf{x} \geq 0$ whenever $A\mathbf{x}+\mathbf{b} \geq 0$. (So the objective function is always ≥ 0 and we do not have to worry about division by zero.)

Solution: First we introduce dummy variable t to make the objective function linear:

$$(P')\begin{cases} \text{minimize} & t \\ \text{subject to} & A\mathbf{x}+\mathbf{b} \geq 0 \\ & \frac{(\mathbf{c}^T\mathbf{x})^2}{\mathbf{d}^T\mathbf{x}} \leq t \end{cases}$$

Now $\frac{(\mathbf{c}^T\mathbf{x})^2}{\mathbf{d}^T\mathbf{x}} \leq t$ is same as $(\mathbf{c}^T\mathbf{x})^2 \leq t\cdot\mathbf{d}^T\mathbf{x}$ and hence $0 \leq t\cdot\mathbf{d}^T\mathbf{x} - (\mathbf{c}^T\mathbf{x})^2$. Notice this corresponds to

$$\begin{vmatrix} t & \mathbf{c}^T\mathbf{x} \\ \mathbf{c}^T\mathbf{x} & \mathbf{d}^T\mathbf{x} \end{vmatrix} \geq 0$$

This gives a program

$$(DSDP)\begin{cases} \text{minimize} & t \\ \text{subject to} & \begin{pmatrix} \mathbf{a}_1\cdot\mathbf{x}+b_1 & & & & 0 \\ & \ddots & & & \\ & & \mathbf{a}_m\cdot\mathbf{x}+b_m & & \\ & & & t & \mathbf{c}^T\mathbf{x} \\ 0 & & & \mathbf{c}^T\mathbf{x} & \mathbf{d}^T\mathbf{x} \end{pmatrix} \succeq 0 \end{cases}$$

It is indeed $(DSDP)$ since it can be written as

$$(DSDP)\begin{cases} \text{minimize} & t \\ \text{subject to} & \sum_i x_i \begin{pmatrix} a_{1,i} & & & & 0 \\ & \ddots & & & \\ & & a_{m,i} & & \\ & & & 0 & c_i \\ 0 & & & c_i & d_i \end{pmatrix} + t\begin{pmatrix} 0 & & 0 \\ & 0 & \\ & & 1 \\ 0 & & 0 \end{pmatrix} - \begin{pmatrix} -b_1 & & & & 0 \\ & \ddots & & & \\ & & -b_m & & \\ & & & 0 & \\ 0 & & & & 0 \end{pmatrix} \succeq 0 \end{cases}$$

8: Now we show that the requirement of R, the value of largest solution, for polynomial time solvability is indeed necessary. Consider the following constraint for $(DSDP)$. Show that x_n is HUGE in *any* feasible solution.

$$\begin{pmatrix} 1 & 2 & & & & & & & \\ 2 & x_1 & & & & & & & \\ & & 1 & x_1 & & & & & \\ & & x_1 & x_2 & & & & & \\ & & & & 1 & x_2 & & & \\ & & & & x_2 & x_3 & & & \\ & & & & & & \ddots & & \\ & & & & & & & 1 & x_{n-1} \\ & & & & & & & x_{n-1} & x_n \end{pmatrix} \succeq 0$$

Use that the matrix is positive semidefinite if each block is positive semidefinite and derive what constraints it brings.

Solution: So we see that

$$\begin{pmatrix} 1 & 2 \\ 2 & x_1 \end{pmatrix} \succeq 0 \Rightarrow \begin{vmatrix} 1 & 2 \\ 2 & x_1 \end{vmatrix} \geq 0 \Rightarrow x_1 - 4 \geq 0$$

and

$$\begin{pmatrix} 1 & x_i \\ x_i & x_{i+1} \end{pmatrix} \succeq 0 \Rightarrow \begin{vmatrix} 1 & x_i \\ x_i & x_{i+1} \end{vmatrix} \geq 0 \Rightarrow x_{i+1} - x_i^2 \geq 0$$

So we get $x_1 \geq 2^2$, $x_2 \geq (2^2)^2 = 2^4$, $x_3 \geq ((2^2)^2)^2 = 2^8$. By induction, $x_n \geq 2^{2^n}$. Therefore, just writing x_n will take time at least $O(\log 2^{2^n}) = O(2^n)$.