2. *Semidefinite programming and Relaxation*

Lecturer: Massimo Lauria

Semidefinite programs is a valuable tools in approximation algorithms and in combinatorial optimization, since semidefinite relaxations are usually stronger than linear ones. In this lecture we describe what is a semidefinite program.

Positive semidefinite matrices

A fundamental concept in semidefinite programming is the one of *positive semidefinite matrices*.

Definition 1. *A square matrix A in $\mathbb{R}^{n \times n}$ is called positive semidefinite if*

- *A is symmetric*[1]*, i.e., $A = A^T$;*
- *for every $x \in R^n$ it holds that $x^T Ax \geq 0$.*

[1] the condition of symmetry is often forgotten, but it is required.

A matrix A is positive definite *if it is positive semidefinite and is also non-singular.*

We denote the set of positive semidefinite matrices in $\mathbb{R}^{n \times n}$ as $\mathsf{PSD_n}$ and we say that $A \succeq 0$ when $A \in \mathsf{PSD_n}$. We say that $A \succ 0$ when A is positive definite.

Fact 2. $\mathsf{PSD_n}$ *is closed under positive combinations (i.e., is a cone): for a sequence $M_1, M_2, \ldots, M_\ell$ of positive semidefinite matrices, the matrix*

$$\alpha_1 M_1 + \alpha_2 M_2 + \cdots + \alpha_\ell M_\ell \tag{1}$$

is positive semidefinite for $\alpha_1 \geq 0, \alpha_2 \geq 0, \ldots, \alpha_\ell \geq 0$.

Matrix decompositions

Positive definite matrices have a set of useful decomposition. If a matrix $A \succ 0$ then there are decompositions

$$A = U^T U \qquad \text{Cholesky decomposition} \tag{2}$$

$$A = LDL^T \qquad \text{LDL decomposition} \tag{3}$$

$$A = Q \Lambda Q^T \qquad \text{Spectral decomposition} \tag{4}$$

where U, L, D, Q, Λ are real matrices in $\mathbb{R}^{n \times n}$. Furthermore U, L, D are unique.

In *spectral decomposition* (also called *eigendecomposition*), matrix Q is an *orthogonal matrix*—it columns are unitary eigenvectors of A, i.e. $QQ^T = I$; and Λ is the diagonal matrix containing the corresponding positive eigenvalues. Consider the Euclidean geometry induced by the norm $x^T x$, then the geometry induced by the norm $x^T Ax$ is Euclidean, but the space is scaled along different axis.

Spectral decomposition: we have that A is symmetric and we need to show that eigenspaces with different eigenvalues are orthogonal: pick x with eigenvalue λ and y with eigenvalue μ,

$$x^T Ay = \lambda x^T y = \mu x^T y \tag{5}$$

If $\lambda \neq \mu$ then $x^T y = 0$.

Symmetric real matrices have real eigenvalues: since A is real, the eigenvectors can be real as well. For any eigenvector x with eigenvalue λ, we get that $(Ax)^T Ax = x^T A^T Ax = x^T AAx = \lambda^2 |x|^2$ is real and positive. Thus λ is real.

Positive semidefiniteness of A implies that λ is also positive.

In the *Cholesky decomposition* the matrix U is an upper triangular real matrix. This representation witnesses the fact that A is definite positive, since $x^T Ax = x^T U^T Ux = |Ux|^2 > 0$.

The *LDL decomposition* is a slight modification of Cholesky decomposition. The matrix L is *lower unit triangular*, which is a lower triangular matrix with unit diagonal, and D is a diagonal matrix. The decomposition exists for any symmetric matrix, but for positive definite matrices the diagonal matrix D has positive entries. The peculiarity of this decomposition is that the components of L and D are rational functions of the entries of A. Indeed the LDL decomposition is a partial version of Cholesky decomposition that avoid computing square roots. The decomposition for positive definite matrix A follows by induction from the equation (we have $\alpha > 0$)

$$\begin{bmatrix} \alpha & v^T \\ v & C \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ v/\alpha & I \end{bmatrix} \cdot \begin{bmatrix} \alpha & 0 \\ 0 & C - vv^T/\alpha \end{bmatrix} \cdot \begin{bmatrix} 1 & v^T/\alpha \\ 0 & I \end{bmatrix} \tag{6}$$

We need to show that $B = C - vv^T/\alpha$ is positive definite. Fix non zero $u \in \mathbb{R}^{n-1}$ then fix $x^T = [-u^T v/\alpha, u^T]$. Then $u^T Bu = x^T Ax > 0$.

Once you have LDL decomposition for positive definite matrices, the Cholesky decomposition follows immediately.

All the decomposition can be extended to **positive semidefinite** matrices, but then the entries of the diagonal of U and D can be zero; the matrices U, L, D are not uniquely defined; there are some zero eigenvectors in Λ. In particular in equation (6) we get that $\alpha \geq 0$ by positive semidefiniteness, but if α is zero, then v must be the zero vector too, otherwise[2] there would be y such that $y^T Ay < 0$.

[2] If the matrix has a minor $M = \begin{bmatrix} 0 & a \\ a & b \end{bmatrix}$ then pick vector $y = [t, 1]$. Then $y^T My = 2ta + b^2$, which is negative for the appropriate choice of t.

Some useful facts about matrices

Fact 3. *Let $A = (a_{ij})$, then $x^T Ax = (xx^T) \bullet A$*

Fact 4. *Let any two matrices $A \in \mathbb{R}^{n\times m}$ and $B \in \mathbb{R}^{m\times n}$. Then $\mathsf{Tr}(AB) = \mathsf{Tr}(BA)$.*

Fact 5. *Let any two* ***symmetric*** *A and B. It holds that $\mathsf{Tr}(AB) = A \bullet B$.*

Fact 6. *For any symmetric $n \times n$ matrix A, $A \in \mathsf{PSD_n}$ if and only if $A \bullet B \geq 0$ for all $B \in \mathsf{PSD_n}$.*

Proof. If A and B as positive semidefinite, then we represent A as $Q^T \Lambda Q$ using spectral decomposition and apply the chain of equations

$$A \bullet B = \mathsf{Tr}(AB) = \mathsf{Tr}(Q^T \Lambda QB) = \mathsf{Tr}(\Lambda QBQ^T). \tag{7}$$

Fix $B' = QBQ^T$, which is positive semidefinite and has non negative diagonal entries. Also Λ is a diagonal non negative matrix, thus $\mathsf{Tr}(AB) = \mathsf{Tr}(\Lambda B') = \sum_i \lambda_i b'_{ii} \geq 0$.

For the other direction is sufficient to consider y such that $y^T Ay < 0$. Then fix $B = yy^T$ and observe that $A \bullet B = y^T Ay$. B is semidefinite positive by construction. □

Fact 7. *For $A \succ 0$ it holds that $A \bullet B > 0$ for all $B \in \mathsf{PSD_n}$, unless $B = 0$.*

The proof of the latter fact is identical to the proof of Fact 6, with the remark that all eigenvalues of a positive definite matrix are positive.

Semidefinite programs

Semidefinite programs can be naturally intepreted are a relaxation of quadratic programs. Consider for example the MaxIndSet on a graph $G = ([n], E)$, which is expressed by the left quadratic program that follows.

$$\begin{aligned} \text{maximize} \quad & \sum_i x_i \\ \text{subject to} \quad & x_i x_j = 0 \qquad \text{if } \{i,j\} \in E \\ & x_i^2 - x_i = 0 \end{aligned} \tag{8}$$

$$\begin{aligned} \text{maximize} \quad & \sum_i x_i^2 \\ \text{subject to} \quad & x_i x_j = 0 \qquad \text{if } \{i,j\} \in E \\ & x_i^2 - x_0 x_i = 0 \\ & x_0^2 = 1 \end{aligned} \tag{9}$$

For convenience will consider a new dummy "x" variable x_0 to remove linear terms as in (9). The idea now is to relax the quadratic constraints by encoding the products $x_i x_j$ as new variables y_{ij}. Some of the identities implied by the interpretation $y_{ij} = x_i x_j$ are enforced by the program, but not all of them can —this is why it is a relaxation. One of the properties of matrix $Y = xx^T$ that we can enforce is positive semidefiniteness.

$$\begin{aligned} \text{maximize} \quad & \sum_i y_{ii} \\ \text{subject to} \quad & y_{ij} = 0 \qquad \text{if } \{i,j\} \in E \\ & y_{ii} = y_{0i} \\ & y_{00} = 1 \\ & Y \succeq 0 \end{aligned}$$

A semidefinite program is a refinement of a linear program in which the variables have a natural square matrix structure, and such matrix is positive semidefinite. Semidefinite programming has found many uses in combinatorics and approximation algorithms [3]. We denote semidefinite programs (SDP) in one of this equivalent ways

[3] Bernd Gartner and Jiri Matousek. *Approximation algorithms and semidefinite programming*. Springer, 2012

$$
\begin{aligned}
\text{maximize} \quad & C \bullet X \\
\text{subject to} \quad & A_1 \bullet X = b_1 \\
& A_2 \bullet X = b_2 \\
& \vdots \\
& A_\ell \bullet X = b_\ell \\
& X \succeq 0
\end{aligned}
\qquad
\begin{aligned}
\text{maximize} \quad & \sum_{i,j} c_{i,j}(v_i^T v_j) \\
\text{subject to} \quad & \sum_{i,j} a_{i,j}(v_i^T v_j) = b_1 \\
& \sum_{i,j} a^2_{i,j}(v_i^T v_j) = b_2 \\
& \vdots \\
& \sum_{i,j} a^\ell_{i,j}(v_i^T v_j) = b_\ell \\
& |v_0| = 1
\end{aligned}
$$

where A_i and C are symmetric matrices. A solution X is called *strictly feasible* if $X \succ 0$. It is possible write SDP in a more general fashion: we can have constraints on many different positive semidefinite matrices, since if $X_1 \succeq 0$ and $X_2 \succeq 0$ then $\begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix}$ is positive semidefinite. Furthermore we can enforce positive variables since $\begin{bmatrix} x \end{bmatrix} \succeq 0$ is equivalent to $x \geq 0$. We can translate inequality constraints into equality constraints using positive slack variables. Furthermore we can relax the condition of symmetry for matrices A_i and C, since X itself is required to be symmetric. Some forms of the semidefinite program are more convenient than others, when it comes to prove theorems. Another equivalent form of semidefinite program is the following:

$$
\begin{aligned}
\text{minimize} \quad & b^T y \\
\text{subject to} \quad & y_1 A_1 + y_2 A_2 + \cdots + y_m A_m \succeq C
\end{aligned}
$$

where A_i and C are symmetric square matrices and y_i are scalar variables. This form is particularly handy when we come to study duality.

Exercise 8. *Show that this latter form of semidefinite program is as expressive as the previous ones (i.e. it is possible to express the same problems).*

Semidefinite programming duality

The duality of semidefinite programming is a more general (and complex) version of duality of linear programs. We will discuss its proof here. Most of the proofs in this section come from Lovász lecture notes [4].

[4] László Lovász. Semidefinite programs and combinatorial optimization. In *Recent advances in algorithms and combinatorics*, pages 137–194. Springer, 2003

First we describe a form of "unsatisfiability" proof for SDP which is the key to semidefinite programming, as it was to linear programming.

Consider a semidefinite program for an optimization problem.

$$
\begin{aligned}
\text{maximize} \quad & C \bullet X \\
\text{subject to} \quad & A_1 \bullet X = b_1 \\
& A_2 \bullet X = b_2 \\
& \vdots \\
& A_m \bullet X = b_m \\
& X \succeq 0
\end{aligned}
\tag{P}
$$

Consider a combination of proof lines with coefficient $y^T = [y_1 \ldots y_m]$, such that $Y = \sum_i y_i A_i - C \succeq 0$. Then

$$Y \bullet X = \sum_i y_i b_i - C \bullet X. \tag{10}$$

If $X \succeq 0$ then $Y \bullet X \geq 0$ by Fact 6 which means that $C \bullet X \leq \sum_i y_i b_i$. This proves that there is an upper bound to the maximization problem, and the smaller is the upper bound the better.

Definition 9. *A semidefinite program in the form* (P) *is called the* primal *program, and its* dual *is the program*

$$\begin{aligned} \textit{minimize} \quad & b^T y \\ \textit{subject to} \quad & y_1 A_1 + \cdots + y_m A_m - C \succeq 0 \end{aligned} \tag{D}$$

A solution is strictly feasible *if* $y_1 A_1 + \cdots + y_m A_m - C \succ 0$ *holds.*

The primal and the dual programs bound each other, assuming they both have feasible solutions. We know that in linear programming the optimum values for the two programs are the same. In semidefinite programming it gets a little bit more tricky. We start by proving a version of Farkas' Lemma for this framework.

Lemma 10 (Farkas' Lemma for semidefinite programs)**.** *Let be* $A_1, A_2, \ldots A_m$ *and* C *some* $n \times n$ *symmetric matrices over the reals. Then the inequality*

$$y_1 A_1 + y_2 A_2 + \cdots + y_m A_m - C \succ 0 \tag{11}$$

has no solutions $y_1, y_2, \ldots, y_m$ *if and only if there exists a symmetric matrix* $X \neq 0$

$$\begin{aligned} A_1 \bullet X &= 0 \\ A_2 \bullet X &= 0 \\ &\vdots \\ A_m \bullet X &= 0 \\ C \bullet X &\geq 0 \\ X &\succeq 0. \end{aligned}$$

Proof. **If part:** A solution X for the second system implies

$$Y \bullet X \leq 0 \tag{12}$$

for any $Y = y_1 A_1 + y_2 A_2 + \cdots + y_m A_m - C$. It follows that Y is not positive definite[5] because $X \succeq 0$, $X \neq 0$ and Fact 7.

Only if: We first prove this part with $C = 0$.

Consider the linear subspace $L \subseteq \mathbb{R}^{n \times n}$ generated by $y_1 A_1 + y_2 A_2 + \cdots + y_m A_m$. If equation (11) has no solution, then L is disjoint from the interior of the convex cone $\mathsf{PSD_n}$ and there is an hyperplane H separating L and the interior of $\mathsf{PSD_n}$[6]. Assume H is characterized by some non trivial

[5] We will see that this part cannot be improved to provide proof of unsatisfiability in case the constraint is relaxed to $Y \succeq 0$.

[6] By the hyperplane separator theorem, which we won't prove here.

equation $X \bullet Y = 0$ where X are the coefficients and Y the variables. Matrix X can be made symmetric, and since the equation is non trivial, $X \neq 0$. L is in H otherwise L would not be in one of the halfspaces. It follows that $X \bullet A_i = A_i \bullet X = 0$ for every i. For every $Y \in \mathsf{PSD_n}$ the separator X has $X \bullet Y \geq 0$ which implies $X \succeq 0$ because of fact 6.

Now we go back to $C \neq 0$. We consider the new constraint

$$y_1 \begin{bmatrix} A_1 & 0 \\ 0 & 0 \end{bmatrix} + y_2 \begin{bmatrix} A_2 & 0 \\ 0 & 0 \end{bmatrix} + \cdots + y_m \begin{bmatrix} A_m & 0 \\ 0 & 0 \end{bmatrix} + z \begin{bmatrix} -C & 0 \\ 0 & 1 \end{bmatrix} \succ 0 \tag{13}$$

which is equivalent to (11): if $z = 0$ then the sum is not positive definite, so every solution must have $z \neq 0$.

If there is no such solution then by reducing to the homogeneous case, there must be a non zero $n+1 \times n+1$ matrix $X' \succeq 0$ that witnesses it. Let us call X the submatrix made by its first n columns and n rows. Then $A_i \bullet X = 0$ and $\begin{bmatrix} -C & 0 \\ 0 & 1 \end{bmatrix} \bullet X' = -C \bullet X + x'_{n+1,n+1} = 0$. By semidefiniteness of X' then $x'_{n+1,n+1} \geq 0$ and thus $C \bullet X \geq 0$. □

We are now set to prove the duality theorem. In a way similar to linear programming duality, the Farkas' lemma provides a general way to witness unsatisfiability of the primal by the means of the dual, and vice versa. Unfortunately for semidefinite programming there are some conditions in order to get strong duality[7].

[7] This is called *Slater condition* and apply in general to convex programming. If both the primal and the dual have strictly feasible solutions, i.e. solutions which are in the interior of the convex body, then strong duality in convex programming follows. Notice that the two strictly feasible solutions do not need to be optimal.

Theorem 11. *Assume that both primal and dual have feasible solutions. Then $v_P \geq v_D$, where v_P and v_D are the optimal values to the primal and dual, respectively. Moreover, if the dual has a strictly feasible solution then*

1. *the primal optimum is attained;*

2. $v_P = v_D$.

Similarly, if the primal is strictly feasible, then the dual optimum is attained, and it is equal to the dual optimum. Hence, if both the primal and dual have strictly feasible solutions, then both v_P and v_D are attained.

Proof. Let be X any solution for program (P) and y a solution for program (D). By Fact 6 we get that

$$0 \leq \left(\sum_i y_i A_i - C \right) \bullet X = y^T b - C \bullet X \tag{14}$$

which means that $v_P \leq v_D$[8].

[8] This part of the theorem is often called weak duality theorem.

Furthermore the constraints $b^T y < v_D$ and $\sum_i y_i A_i \succeq C$ have no feasible solution by definition of the dual. Thus consider the matrices

$$A'_i = \begin{bmatrix} -b_i & 0 \\ 0 & A_i \end{bmatrix} \qquad C' = \begin{bmatrix} -v_D & 0 \\ 0 & C \end{bmatrix} \tag{15}$$

then the system $\sum_i y_i A'_i \succ C'$ is unsatisfiable and then by Farkas' Lemma there is a non zero positive semidefinite matrix X' such that $A'_i \bullet X' = 0$ for all $1 \leq i \leq m$ and $C' \bullet X' \geq 0$. Let us write

$$X' = \begin{bmatrix} x_{00} & x^T \\ x & X \end{bmatrix} \tag{16}$$

so we get that

$$A_i \bullet X = x_{00} b_i \quad \text{for } i = 1 \dots m \qquad C \bullet X \geq x_{00} v_d. \tag{17}$$

We claim that $x_{00} \neq 0$ since otherwise X would be a witness (by Farkas' Lemma) of the unsolvability of program (D) by strict feasible solutions, and this contradicts our assumption. Since $x_{00} \neq 0$ we can assume $x_{00} = 1$ by scaling the solution in (17).

After rescaling, X is feasible a solution of program (P) with value $C \bullet X \geq v_D$. Thus we get that $v_P \leq v_D$. □

Exercise 12. *In the proof of Theorem 11 we just showed that if the dual has strictly feasible solution then the primal optimum is attained. Please prove that if the primal has strictly feasible solution then the dual optimum is attained as well.*

Examples of strong duality failing

In the proof of the strong duality theorem (Theorem 11) the optimum is attained by assuming that the there is a strictly feasible dual solution. Furthermore we used that to prove that primal and dual optimum are the same. We will see examples from[9] in which the two optimum are not the same (i.e. there is a *duality gap*) or in which the optimum is only reached in the limit.

[9] László Lovász. Semidefinite programs and combinatorial optimization. In *Recent advances in algorithms and combinatorics*, pages 137–194. Springer, 2003

Example 13. Consider the program

$$\begin{aligned} \text{maximize} \quad & -x_1 \\ \text{subject to} \quad & \begin{bmatrix} x_1 & 1 & 0 \\ 1 & x_2 & 0 \\ 0 & 0 & x_1 \end{bmatrix} \succeq 0 \end{aligned}$$

Because of the positive semidefinite requirement, we get that $x_1, x_2 \geq 0$ and that $x_1 x_2 \geq 1$. The objective function $-x_1$ can get arbitrarily close to 0 but will stay negative, thus $v_P = 0$ and there is no solution which actually get to that value.

So in the previous program we have an optimum, which is reached only in the limit.

Example 14. Consider the primal program

$$\text{maximize} \quad -x_{33} \tag{18}$$
$$\text{subject to} \quad x_{12} + x_{21} + x_{33} = 1 \tag{19}$$
$$x_{22} = 0 \tag{20}$$
$$X \succeq 0 \tag{21}$$

The feasible solutions for this programs are matrices of the form

$$\begin{bmatrix} a & 0 & b \\ 0 & 0 & 0 \\ b & 0 & 1 \end{bmatrix} \tag{22}$$

which implies $a \geq b^2$ and $v_P = -1$. Now we consider its dual program

$$\text{minimize} \quad y_1 \tag{23}$$

$$\text{subject to} \quad y_1 \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} + y_2 \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \succeq \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix} \tag{24}$$

Dual solutions have $y_1 = 0$ and $y_2 \geq 0$, thus $v_D = 0$. Here we have a duality gap of -1 between v_P and v_D.

Example 15. Consider a variant of the program in Example 14 where x_{11} is involved in the optimized function. This example comes from[10].

$$\text{maximize} \quad -\epsilon x_{11} - x_{33} \tag{25}$$

$$\text{subject to} \quad x_{12} + x_{21} + x_{33} = 1 \tag{26}$$

$$x_{22} = 0 \tag{27}$$

$$X \succeq 0 \tag{28}$$

[10] Anupam Gupta and Ryan O'Donnell. Linear and semidefinite programming (advanced algorithms) fall 2011. http://www.cs.cmu.edu/afs/cs.cmu.edu/academic/class/15859-f11/www/notes/lecture12.pdf, 2011. Lecture 12

The feasible solutions for this programs do not change, and are still matrices of the form

$$\begin{bmatrix} a & 0 & b \\ 0 & 0 & 0 \\ b & 0 & 1 \end{bmatrix} \tag{29}$$

where $a \geq b^2$ and so $v_P = -1$ as in the previous case, since it pays to make a as small as possible and we can fix it to be 0. Now we consider its dual program

$$\text{minimize} \quad y_1 \tag{30}$$

$$\text{subject to} \quad y_1 \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} + y_2 \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \succeq \begin{bmatrix} -\epsilon & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix} \tag{31}$$

Now see that with $y_1 = -1$ and $y_2 = \frac{1}{\epsilon}$ the constrain matrix in (31) is

$$\begin{bmatrix} \epsilon & -1 & 0 \\ -1 & \frac{1}{\epsilon} & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} \sqrt{\epsilon} \\ -\frac{1}{\sqrt{\epsilon}} \\ 0 \end{bmatrix} \cdot \begin{bmatrix} \sqrt{\epsilon} & -\frac{1}{\sqrt{\epsilon}} & 0. \end{bmatrix} \tag{32}$$

Since the matrix is positive semidefinite then the dual solution is feasible and $v_D = -1$, equal to the primal optimum.

References

[GM12] Bernd Gartner and Jiri Matousek. *Approximation algorithms and semidefinite programming*. Springer, 2012.

[GO11] Anupam Gupta and Ryan O'Donnell. Linear and semidefinite programming (advanced algorithms) fall 2011. http://www.cs.cmu.edu/afs/cs.cmu.edu/academic/class/15859-f11/www/notes/lecture12.pdf, 2011. Lecture 12.

[Lov03] László Lovász. Semidefinite programs and combinatorial optimization. In *Recent advances in algorithms and combinatorics*, pages 137–194. Springer, 2003.