SEMIDEFINITE OPTIMIZATION

ZHIRONG LI ZHIRONG.LEE@GMAIL.COM

Abstract. This is the course review on Semidefinite Programming taught by Prof. Levent Tuncel. I will follow the structure of Levent's monograph on SDP optimization. As well, I will put my understanding and relevant materials to help understand the contents. In the first part, I will show some background knowledge on linear programming and some properties of semidefinite matrix. Then I will give the cannonical format of SDP and associated weak and strong duality theorem. Ellipsoid and primal-dual interior point method will be given in much detail to show how to solve a general SDP problem with Slater condition satisfied. Finally, some applications based on SDP in approximation algorithms will be presented to give a better look and feel how powerful SDP is in many research areas.

1. Preliminary Knowledge Requirements

In order to keep the learning curve smoothly, we will recap some background knowledge on linear programming (LP), polyhedral theory and linear algebra related to SDP. It serves the purpose to see the similarities and differences between LP and SDP.

1.1. **Linear Programming.** In linear programming, we are exposed to the most important geometric object, a polyhedron $P \subseteq \mathbb{R}^n$ which is defined as the intersection of finitely many half spaces in $\mathbb{R}^n$. The linear programming problem is defined in a way such that we want to minimize or maximize a linear function of n variables over a polyhedron in $\mathbb{R}^n$. One great thing differentiating LP from other nonlinear programming problems lies in that the both primal optimum and dual optimum are attained and there is no duality gap between primal optimum and dual optimum (Strong duality theorem always holds). I will conform to the symbolic conventions used in Levent's monograph on SDP. Suppose $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$ and $c \in \mathbb{R}^n$ are given, we have the primal problem given by

$$\begin{aligned} (LP) \quad & \min c^T x \\ st. \quad & Ax = b \\ & x \geq 0 \end{aligned}$$

and the dual problem as framed as

$$\begin{aligned} (LD) \quad & \max b^T y \\ st. \quad & A^T y + s = c \\ & s \geq 0 \end{aligned}$$

where y is the free dual variable, s is the slack variable to transform inequaltiy form LD to its equality form. s or y is uniquely determined once the other is known. Keep in mind we have the following mnemonic rules in formulating a primal-dual problem.

The feasible region of the primal problem is $\{x \in \mathbb{R}^n : Ax = b, x \geq 0\}$ and the polyhedron defining the dual problem is given by $\left\{y \in \mathbb{R}^m : A^T y \leq c\right\}$. If we take the slack variable s into consideration as well, we have

primal problem (max)	dual problem (min)
equality constraint	unrestricted variable
$\leq$constraint	nonnegative variable
nonnegative variable	$\geq$constraint
unrestricted variable	equality constraint

TABLE 1. Relationships between primal and dual problem

row(dual min problem)\column(primal max problem)			
	optimal solution	unbounded	infeasible
optimal solution	can occur	impossible	impossible
unbounded	impossible	impossible	can occur
infeasible	impossible	can occur	can occur

TABLE 2. Possibilities for primal-dual pair

$\left\{(y,s) \in \mathbb{R}^m \bigoplus \mathbb{R}^n_+ : A^T y + s = c\right\}$ and $\{s \in \mathbb{R}^n : Fs = Fc, s \geq 0\}$ where the row vectors of $F \in \mathbb{R}^{(n-m)\times n}$ form a basis for the null space of A.

We should be very familiar with the Fundamental Theorem of LP, and it can be stated as: Exactly one of the following three cases can occur for an LP problem:

(1) LP problem is infeasible. There is no solution for $\{x \in \mathbb{R}^n : Ax = b, x \geq 0\}$;
(2) LP problem is unbounded. We can always find a sequence $\left\{x_{(k)} \in \{x \in \mathbb{R}^n : Ax = b, x \geq 0\}\right\}$ such that $c^T x_{(k)} \to -\infty$.
(3) LP problem has optimal solution(s). We can find x_{opt} such that $c^T x_{opt} = \min\left\{c^T x : Ax = b, x \geq 0\right\}$ and $x_{opt} \in \{x \in \mathbb{R}^n : Ax = b, x \geq 0\}$.

We can classify the possibilities for a primal-dual pair in a table. The dual of the dual problem is the primal problem so it is quite natural to find the accurate correspondence between primal and dual problems from above table. For example, if primal (maximization problem) is unbounded, from weak duality, dual problem can not be unbounded (otherwise, since every dual value is the upper bound for the primal value, if dual is unbounded which means the dual value can go down to minus infinity, the primal can not be unbounded). The only possible situation which can happen to dual problem is infeasible. Similarly, if the primal problem is infeasible, dual problem could be either infeasible or unbounded. If the feasible region is a pointed polyhedron (contains no line in it) and the LP problem has optimal solution(s), there exists an extreme point that is the optimal solution.

The weak duality theorem of LP says that for every feasible $\bar{x}$ in (LP) and every feasible $\bar{y}$ in (LD), $c^T\bar{x}$ is greater than or equal to $b^T\bar{y}$. This follows directly from the definitions $c^T\bar{x} \geq (A^T\bar{y})^T\bar{x} = \bar{y}^T A\bar{x} = b^T\bar{y}$. If both (LP) and (LD) are feasible, then they both have optimal solutions and the optimum values coincide. To prove this, we need to refer to the very fundamental fact known as Farkas' Lemma. Exactly one of the following systems has a solution:

(I) $Ax = b, x \geq 0$

(II) $A^T y \leq 0, b^T y > 0$

If (I) has a feasible solution $\bar{x}$, then we can construct an (LP) as $\min\left\{0^T x : Ax = b, x \geq 0\right\}$. From weak duality, we know for the (LD) $\max\left\{b^T y : A^T y \leq 0\right\}$, we can get $b^T y \leq \mathbf{0}^T\bar{x} = 0$.

1.2. **Semidefinite Programming.** In semidefinite programming, the variables take form as of the matrices instead of the column vectors. The inner product form of the object function for SDP is changed to use the trace operator on symmetric matrices $\langle X, S\rangle := tr(X^T S) = \Sigma_{i=1}^n \Sigma_{j=1}^n X_{x,j} S_{i,j} = tr(SX^T)$. Trace has the cyclic property which means $tr(ABC) = tr(CAB) = tr(BCA)$. For every nonsingular $P \in \mathbb{R}^{n\times n}$, $tr(PXP^{-1}) = tr\left(XP^{-1}P\right) = tr(X)$. For given $X \in \mathbb{R}^{n\times n}$, eigenvalues of X satisfies the polynomial equation $\det(X - \lambda I) = 0$. We order the eigenvalues in this way $\lambda_1(X) \geq \lambda_2(X) \cdots \geq \lambda_n(X)$. It is obvious $tr(X) = \sum_{i=1}^{i=n} X_{i,i} = \sum_{i=1}^{i=n} \lambda_i(X)$ since X is always similar to its Jordan form with eigenvalues on the main diagonal. The Frobenius norm for X is defined as

$$|X|_F \triangleq \langle X, X\rangle^{1/2} = \sqrt{\sum_{i=1}^{n} (\lambda_i(x))^2}$$

In general, let $\mathcal{A} : \Sigma^n \to \mathbb{R}^m$ be a linear transformation, $\mathcal{A}$'s adjoint $\mathcal{A}^\star$ is given by $\langle \mathcal{A}^\star(y), X\rangle_{(\Sigma^n)} = \langle y, \mathcal{A}(X)\rangle_{(\mathbb{R}^m)}$ for $\forall y \in \mathbb{R}^m, X \in \Sigma^n$. We can also define $\mathcal{A}$ as $[\mathcal{A}(X)]_i = \langle A_i, X\rangle, \forall i \in \{1, 2, \cdots m\}$ where $A_i \in \Sigma^n, i \in \{1, 2, \cdots m\}$. Now we can rewrite $\mathcal{A}^\star(y)$ in terms of those A_i's by $\mathcal{A}^\star(y) = \sum_{i=1}^m y_i A_i$. We can verify the inner product equality $\langle \mathcal{A}^\star(y), X\rangle_{(\Sigma^n)} = \langle y, \mathcal{A}(X)\rangle_{(\mathbb{R}^m)}$ by substituting the new form of $\mathcal{A}^\star$ into the left hand side:

$$\langle \mathcal{A}^\star(y), X\rangle_{(\Sigma^n)} = \left\langle \sum_{i=1}^m y_i A_i, X \right\rangle = \sum_{i=1}^m y_i \langle A_i, X\rangle = \sum_{i=1}^m y_i [\mathcal{A}(X)]_i = \langle y, \mathcal{A}(X)\rangle$$

For every symmetric matrix $X \in \Sigma^n$, it can be decomposed as

$$X = QDiag(\lambda(X))Q^T$$

where $Q \in \Sigma^n$ and $Q^T Q = I$. An $n \times n$ matrix is diagonalizable if and only if the sum of the dimensions of the eigenspaces is n. In general, a square complex matrix is similar to a block diagonal matrix $J = \begin{bmatrix} J_1 & & \\ & \ddots & \\ & & J_p \end{bmatrix}$ where each $J_i = \begin{bmatrix} \lambda_i & 1 & \\ & \ddots & 1 \\ & & \lambda_i \end{bmatrix}$ is called a Jordan block.

If above X is positive definite, we can define the square root of X as $X^{1/2} = Q\left[Diag(\lambda(X))\right]^{1/2} Q^T$. There is a necessary and sufficient condition to decide whether or not X is positive semidefinite by Cholesky Decomposition. The theorem says $X \in \Sigma_+^n$ iff there exists a lower triangle matrix $B \in \mathbb{R}^{n\times n}$ such that $X = BB^T$.

How do we know $X \in \Sigma^n$ is positive semidefinite? The following statements are equivalent to each other.

(1) X is positive semidefinite;
(2) All eigenvalues of X are nonnegative: $\forall j,\ \lambda_j(X) \geq 0$;
(3) X can be decomposed as the NONNEGATIVE linear combination of n outer product of column vectors. That is, $\exists \mu \in \mathbb{R}_+^n$ such that $X = \sum_{i=1}^n \mu_i h^{(i)} {h^{(i)}}^T$ where $h^{(i)} \in \mathbb{R}^n$;
(4) X can be Cholesky decomposed, B can be singular;
(5) The determinant of any sub-matrix along the main diagonal (indices could be noncontinuous) is nonnegative. This can be formulated as $\forall J \subseteq \{1, 2, \cdots n\}$, $\det(X_J) \geq 0$ where $X_J \triangleq \{[X_{i,j}] : i, j \in J\}$;
(6) $\forall S \in \Sigma_+^n$, the inner product of X and S is nonnegative.

Similarly, the following are also equivalent (TFAE) in determining a symmetric matrix X is positive definite or not:

(1) X is positive definite;

(2) All eigenvalues of X are positive: $\forall j,\ \lambda_j(X) > 0$;
(3) X can be decomposed as the POSITIVE linear combination of n outer product of column vectors. That is, $\exists \mu \in \mathbb{R}^n_{++}$ such that $X = \sum_{i=1}^{n} \mu_i h^{(i)} h^{(i)^T}$ where $h^{(i)} \in \mathbb{R}^n$;
(4) X can be Cholesky decomposed, B is nonsingular;
(5) $\forall J \subseteq \{1, 2, \cdots n\}$, $\det(X_J) > 0$ where $X_J \triangleq \{[X_{i,j}] : i, j \in J\}$;
(6) $\forall S \in \Sigma^n_+ \backslash \{0\}\ \langle X, S\rangle > 0$;
(7) $X \succeq 0$ and $rank(X) = n$.

It is worth pointing out that every positive semidefinite matrix has the property that if $X_{i,i} = 0$, then the whole column and the whole row at index i must be zero as well. That is, if $X \succeq 0$ and $X_{ii} = 0$, then $X_{ij} = X_{ji} = 0, \forall j \in x\,\{1, \cdots n\}$.

Gersgorin Disk Theorem gives the boundary condition for the eigenvalues of the matrix. Let $A \in \mathbb{C}^{n\times n}$ with entries (A_{ij}). For each $i \in \{1, \cdots n\}$, we denote $R_i = \sum_{j\neq i} |a_{ij}|$. Let $D(a_{ii}, R_i)$ be the closed disk centered at a_{ii} with radius R_i, those disks are called Gersgorin Disks. We have the fact that every eigenvalue of A lies within at least one of the Gersgorin Disks. Let λ be an eigenvalue of A and $x = (x_j)^T$ be the corresponding eigenvector. Let i be chose such that $|x_i| = \max_j |x_j|$. $|x_i|$ can not be zero otherwise the eigenvector is NULL. We have $AX = \lambda X$ or in component-wise equality $\sum_j a_{ij}x_j = \lambda x_i$ so we get $\lambda x_i - a_{ii}x_i = \sum_{j\neq i} a_{ij}x_j$. We can divide both sides by $|x_i|$, then we arrive at

$$|\lambda - a_{ii}| = \left|\frac{\sum_{j\neq i} a_{ij}x_j}{x_i}\right| \leq \sum_{j\neq i} |a_{ij}| = R_i$$

We say a matrix X is strictly diagonal dominant if $X_{ii} \geq \sum_{j\neq i} |X_{ij}| \geq |\lambda - X_{ii}|$. Directly from Gersgorin disk theorem, we can safely draw the conclusion for $X \in \Sigma^n$ and X is (strictly) diagonal dominant, X is positive semidefinite (definite). This is because for every eigenvalue λ, it lies within a Gersgorin Disk. From the definition of the diagonal dominant, we can conclude every $\lambda \geq 0$ (X is symmetric thus all eigenvalues are real).

Having above knowledge, we can define the semidefinite programming problem in standard form and its dual. Suppose $C \in \Sigma^n$, $b \in \mathbb{R}^m$ and a linear operation $\mathcal{A} : \Sigma^n \to \mathbb{R}^m$ are given, the primal of the SDP problem is given by

$$\begin{aligned} (P) \quad & \inf \langle C, X\rangle \\ st. \quad & \mathcal{A}(X) = b \\ & X \succeq 0 \end{aligned}$$

and the dual of the SDP problem is as follows

$$\begin{aligned} (D) \quad & \sup b^T y \\ st. \quad & \mathcal{A}^\star(y) + S = C \\ & S \succeq 0 \end{aligned}$$

We can make reference to the previous more explicit form of $\mathcal{A}$. Let $A_1, A_2, \cdots A_m \in \Sigma^n$ and for every $X \in \Sigma^n$, $|\mathcal{A}(X)|_i = \langle A_i, X\rangle$. Rewrite the above (P) and (D) as:

$$\begin{aligned} (P) \quad & \inf \langle C, X\rangle \\ st. \quad & \langle A_i, X\rangle = b_i \quad \forall i \in \{1, \cdots m\} \\ & X \succeq 0 \end{aligned}$$

and

$$\begin{array}{rc} (D) & \sup b^T y \\ st. & \sum_{i=1}^m y_i A_i + S = C \\ & S \succeq 0 \end{array}$$

Notice that we use inf and sup instead of min and max respectively since for nonlinear programming problems, optimal solution may not be attained. Here I will bring in some concepts concerning convex cones and we will be able to find that above primal-dual pair SDP problem actually is a special case of more general convex optimization problems in conic form. I will give more details about conic and convex programming in the next tutorial. A cone $\mathcal{K} \in \mathbb{R}^d$ is a convex cone if $\mathcal{K} + \mathcal{K} \in \mathcal{K}$ and $\forall \alpha > 0, \alpha\mathcal{K} \subset \mathcal{K}$. A dual of a $\mathcal{K}$ is $\mathcal{K}^* \triangleq \left\{ s \in \mathbb{R}^d : \langle x, s\rangle \geq 0, \forall x \in \mathcal{K} \right\}$.

For SDP programming problems, we also have weak duality relation. Let $\bar{X}$ be feasible in (P) and $(\bar{y}, \bar{S})$ be feasible in (D). Then $\left\langle C, \bar{X}\right\rangle - b^T \bar{y} = \left\langle \bar{X}, \bar{S}\right\rangle \geq 0$. The proof is very similar to the proof in LP and we know

$$\left\langle C, \bar{X}\right\rangle - b^T\bar{y} = \left\langle C, \bar{X}\right\rangle - \mathcal{A}(\bar{X})^T\bar{y} = \left\langle C, \bar{X}\right\rangle - \left\langle \bar{X}, \mathcal{A}^\star(\bar{y})\right\rangle = \left\langle C - \mathcal{A}^\star(\bar{y}), \bar{X}\right\rangle = \left\langle \bar{X}, \bar{S}\right\rangle \geq 0$$

We have the similar corollary about the relationships between primal and dual for SDP problems as well. If (P) is unbounded then (D) is infeasible and if (D) is unbounded then (P) is infeasible. If we find feasible solution $\bar{X}$ of (P) and $(\bar{y}, \bar{S})$ of (D) and the duality gap $\left\langle C, \bar{X}\right\rangle - b^T\bar{y} = \left\langle \bar{X}, \bar{S}\right\rangle = 0$, then $\bar{X}$ is optimal in (P) and $\left(\bar{y}, \bar{S}\right)$ is optimal in (D).

Let $X, S \in \Sigma_+^n$ then $\langle X, S\rangle = 0$ iff. $XS = 0$. Sufficiency is trivial since $\langle X, S\rangle = tr(X^T S) = tr(XS) = 0$. If $\langle X, S\rangle = 0$ and they both are positive semidefinite, by cyclic property of trace operator, we have $tr(XS) = tr(X^{1/2}X^{1/2}S) = tr(X^{1/2}SX^{1/2}) = 0$. Notice that $X^{1/2}SX^{1/2}$ is also positive semidefinite so all its eigenvalues are nonnegative and the summation of all the eigenvalues should be equal to its trace hereby 0. This implies $0 = X^{1/2}SX^{1/2} = \left(X^{1/2}S^{1/2}\right)\left(X^{1/2}S^{1/2}\right)^T$. Thus $X^{1/2}S^{1/2} = 0$ and we get $XS = X^{1/2}\left(X^{1/2}S^{1/2}\right)S^{1/2} = 0$.

Rayleigh quotient is defined as $\rho(h) = \rho(h; X) = \frac{h^T X h}{h^T h}$. The Rayleigh quotient enjoys the following properties:

(1) Homogeneity: $\rho(\alpha h) = \rho(h)$, $\alpha \neq 0$
(2) Boundedness: $\rho(h)$ ranges over the interval $[\lambda_n(X), \lambda_1(X)]$ as h ranges over non-zero $n-$vectors.
(3) Stationarity: The gradient of ρ is 0 is at and only at the eigenvectors of X.

Courant-Fischer min-max theorem related to Rayleigh quotient says: Let $X \in \Sigma^n$ be Hermitian, then

$$\lambda_k(X) = \min_{L \subseteq \mathbb{R}^n, \dim(L) = n-k+1} \max_{h \in L\setminus\{0\}} \frac{h^T X h}{h^T h} = \max_{L \subseteq \mathbb{R}^n, \dim(L) = k} \max_{h \in L\setminus\{0\}} \frac{h^T X h}{h^T h}, \; \forall k \in \{1, 2, \cdots, n\}$$

This theorem also gives the well known cases $\lambda_1(X) = \max_{h \in \mathbb{R}^n\setminus\{0\}} \frac{h^T X h}{h^T h}$ and $\lambda_n(X) = \min_{h \in \mathbb{R}^n\setminus\{0\}} \frac{h^T X h}{h^T h}$. The proof of the above theorem can be referenced to Golub/van Loan.

2. Duality Theory for SDP

Duality is very import in LP theory. Many wonderful aspects of duality can not be applied to SDP and general convex optimization problems. We will try to give explanations of duality in Geometry therefore the following will focus on conic programming problems. We know that the dual cone of $\mathcal{K} \subseteq \mathbb{R}^d$ is defined as $\mathcal{K}^* \triangleq \left\{ s \in \mathbb{R}^d : \langle x, s\rangle \geq 0, \forall x \in \mathcal{K} \right\}$. The polar set of $s \subseteq \mathbb{E}, s \neq \phi$ is $s^o \triangleq \{a : \langle a, x\rangle \leq 1, \forall x \in s\}$. The negative polar of $\mathcal{K}$ is defined as $\mathcal{K}^- \triangleq \{a : \langle a, k\rangle \leq 0, \forall k \in \mathcal{K}\}$. If $\mathcal{K}$ is a convex cone, then its polar cone and negative polar coincide. This follows for $k \neq 0, \alpha > 0$, $\alpha k \in \mathcal{K}$, $0 < \langle a, \alpha k\rangle = \alpha \langle a, k\rangle \to \infty$ as $\alpha \to \infty$,

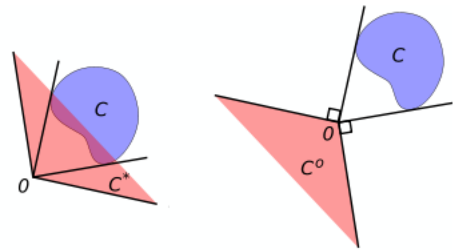

FIGURE 2.1. Dual Cone and Polar Cone wrt. Set C

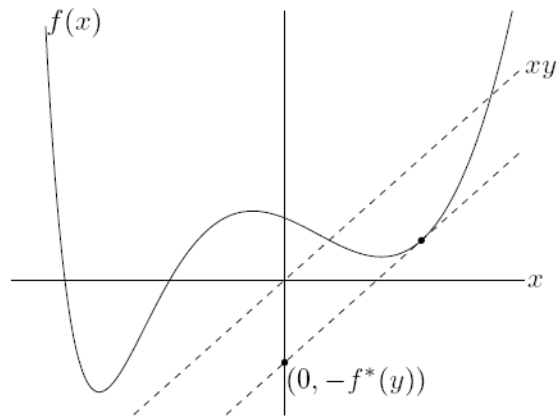

FIGURE 2.2. Fenchel conjugate function $f^*(y)$ is the maximum gap between yx and $f(x)$ maximum occurs at x where $f^{'}(x) = y$

which contradicts ≤ 1 constraint. Therefore for convex cone $\mathcal{K}$, we have $\mathcal{K}^o = \mathcal{K}^-$. The following diagram shows the dual cone and polar cone in geometry.

For a given function $f : \mathbb{R}^d \mapsto \mathbb{R} \cup \{\infty\}$, we define its Legendre-Fenchel conjugate as

$$f_*(s) \triangleq \sup \left\{ - \langle s, x \rangle - f(x) : x \in \mathbb{R}^d \right\}$$

In other literature, Fenchel conjugate function is also defined as $f^*(s) = f_*(-s)$. The Fenchel conjugate's geometric meaning is shown in the following figure.

The epigraph of a function f is defined as

$$epi(f) \triangleq \left\{ \left(\begin{array}{c} t \\ x \end{array} \right) \in \mathbb{R} \oplus \mathbb{R}^d : f(x) \leq t \right\}$$

and it is obvious to see that a function is convex if and only if its associated epigraph is a convex set.

$$\left(\begin{array}{c} y \\ t \end{array} \right) \in epi(f) \Rightarrow \left[\begin{array}{c} \nabla f(x) \\ -1 \end{array} \right]^T \left(\left[\begin{array}{c} y \\ t \end{array} \right] - \left[\begin{array}{c} x \\ f(x) \end{array} \right] \right) \leq 0$$

Hahn-Banach Theorem in finite dimension version servers a central role in optimization. If $S \subseteq \mathbb{E}$ is convex closed set and $\bar{x} \notin S$, then there exists $a \in \mathbb{E}, b \in \mathbb{R}$ such that $\langle a, \bar{x} \rangle > b \geq \langle a, x \rangle, \forall x \in S$. This is

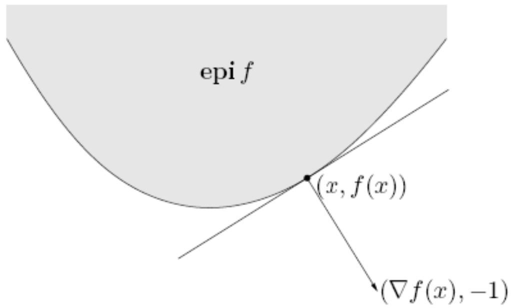

FIGURE 2.3. Vector $(-\nabla f(x), -1)$defines the supporting hyperplane to the *epi*(f) at x

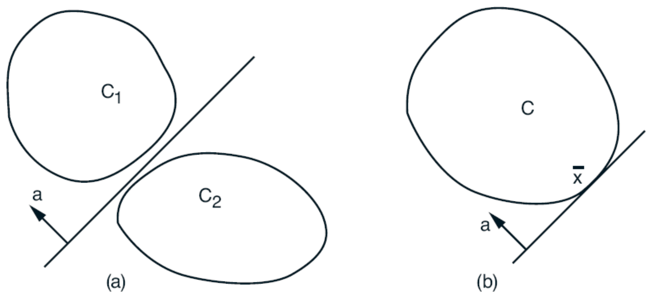

FIGURE 2.4. Hyperplane Separation Theorem

given another well-known name as Hyperplane Separation Theorem. We can prove the corollary that let $C_1, C_2 \subset \mathbb{R}^d$ be disjoint, nonempty closed convex set. If C_1 or $C_2$is bounded then there exists $a \in \mathbb{R}^d \backslash \{0\}$ such that

$$\inf \left\{a^T x : x \in C_1\right\} > \sup \left\{a^T x : x \in C_2\right\}.$$

The proof is not very hard when we consider the distance defined by inner product in Euclidean space. Figure 2.2 shows the geometry for this hyperplane separation theorem, and supporting hyperplane for a convex set. Notice that a is the normal to the hyperplane which separates the two convex sets.

Definition 2.1. We say (P) satisfies the Slater condition, or (P) has a Slater point if there exists a symmetric matrix $\bar{X}$ such that it is feasible and positive definite ($\mathcal{A}(\bar{X}) = b, \bar{X} \succ 0$). Similarly for the dual problem, (D) has a Slater point if there exists $\bar{S} \in \Sigma^n$ and $\bar{y} \in \mathbb{R}^m$ such that $\mathcal{A}^\star(\bar{y}) + \bar{S} = C$ and $\bar{S} \succ 0$.

We can move to prove a Strong Duality Theorem for SDP problems.

Theorem 2.2. *Suppose (D) has a Slater point. If the objective value of (D) is **bounded from above** then (P) **attains** its optimum value and the optimum values of (P) and (D) **coincide**.*

Proof. This a little bit intense proof is also referring to Levent's monograph. We need some real analysis, linear algebra, weak duality of SDP to complete the proof. After this, we will give some concrete examples of SDP and see the difference between LP and SDP in terms of their dualities. Let $\bar{S} \in \Sigma^n_{++}, \bar{y} \in \mathbb{R}^m$

be the Slater point satisfying $\mathcal{A}(\bar{y})+\bar{S}=C$, we define $z^* \triangleq \sup\left\{b^T y : \mathcal{A}^\star(y) \preceq C\right\}$ to be the optimum value of the dual problem. We will deal with two cases for b. If $b=0$, then $(\bar{y},\bar{S})$ is the optimal solution to the dual problem since every feasible solution to (D) is optimal and the optimum value is equal to 0. By weak duality theorem of SDP, the smallest value the primal problem can choose is 0, and at $\bar{X}=0\in\Sigma_+^n$ this value is attained. Therefore (P) attains its optimum value and the optimum values of (P) and (D) coincide as stated above. If $b\neq 0$, we can construct two separable closed convex sets G_1 and G_2: $G_1 \triangleq \left\{S\in\Sigma^n : S=C-\mathcal{A}^\star(y), y\in\mathbb{R}^m, b^T y\geq z^*\right\}$and $G_2\triangleq\Sigma_{++}^n$. First of all, $G_1$is not empty. By the definition of supremum, there exists a sequence (dual problem is feasible as given in the theorem) $\left\{y^{(k)},S^{(k)}\right\}\subset\mathbb{R}^m\oplus\Sigma_+^n$ such that $\mathcal{A}^\star(y^{(k)})+S^{(k)}=C$ and $b^T y^{(k)}\to z^*$. A linear function $b^T y$ over an affine subspace $\mathcal{A}^\star(y)+S=C, \mathcal{A}^\star(y)=\sum_{i=1}^m y_i A_i$ attains its limit. That is, $\lim_{k\to\infty}\left(y^{(k)},S^{(k)}\right)\to\left(\hat{y},\hat{S}\right)$ such that $\mathcal{A}^\star(\hat{y})+\hat{S}=C$ and $b^T\hat{y}=z^*$. Therefore $\hat{S}$ is in the set $G_1$and we need to show G_1 is also a convex set. By definition, suppose $S_1=C-\mathcal{A}^\star(y_1)\in G_1$ and $S_2=C-\mathcal{A}^\star(y_2)\in G_1$, for any $\lambda\in[0,1]$, the convex combination

$$\lambda S_1+(1-\lambda)S_2\triangleq\tilde{S}=C-\lambda\mathcal{A}^\star(y_1)-(1-\lambda)\mathcal{A}^\star(y_2)$$

is symmetric. In addition, since

$$\lambda\mathcal{A}^\star(y_1)+(1-\lambda)\mathcal{A}^\star(y_2)=\lambda\sum_{i=1}^m y_{1,i}A_i+(1-\lambda)\sum_{i=1}^m y_{2,i}A_i=\sum_{i=1}^m(\lambda y_{1,i}+(1-\lambda)y_{2,i})A_i=\mathcal{A}^\star(\lambda y_1+(1-\lambda)y_2)$$

we can say $\tilde{S}=C-\mathcal{A}^\star(\tilde{y})$ where $\tilde{y}=\lambda y_1+(1-\lambda)y_2$ and $b^T\tilde{y}=b^T\left(\lambda y_1+(1-\lambda)y_2\right)\geq\lambda z^*+(1-\lambda)z^*=z^*$. The convexity of G_1 is obvious since we the objective value $b^T y$ is defined on an affine subspace. Σ_{++}^n is a convex cone and $int(\Sigma_+^n)=\Sigma_{++}^n$. Now we will prove $G_1\cap G_2=\phi$ by contradiction. Suppose $\bar{y}$ is positive definite ($\in G_2=\Sigma_{2++}^n$) and $b^T\bar{y}\geq z^*$, then we can find a neighbour $\hat{y}\triangleq\bar{y}+\epsilon b$ for some $\epsilon>0$ such that $\mathcal{A}^\star(\hat{y})\prec C$ (Σ_{++}^n is continuous and $\mathcal{A}^\star$ is an affine transform, thus preserves the property of positive definiteness) and $b^T\hat{y}=b^T(\bar{y}+\epsilon b)=z^*+\underbrace{\epsilon b^T b}_{>0(b\neq 0)}>z^*$. This contradicts the fact that z^* is the optimal solution of the (D) hence $G_1\cap G_2=\phi$ as desired. Now we can apply the variant version of the separation theorem to disjoint convex set G_1 and G_2. There exists $\tilde{X}\in\Sigma^n\backslash\{0\}$ such that

$$\sup_{S\in G_1}\left\langle\tilde{X},S\right\rangle\leq\inf_{S\in\Sigma_{++}^n}\left\langle\tilde{X},S\right\rangle$$

From the LHS of the inequality we arrive at a conclusion that $\inf_{S\in G_2}\left\langle\tilde{X},S\right\rangle$ is bounded since $G_1$is not empty (every feasible point in G_1 gives a lower bound to the RHS infimum). Keep in mind $G_2=\Sigma_{++}^n$is a convex cone, we must have $\left\langle\tilde{X},S\right\rangle\geq 0$. Otherwise, if $\left\langle\tilde{X},S\right\rangle<0$,then for arbitrary $\alpha>0$, $\alpha S\in\Sigma_{++}^n$ and $\left\langle\tilde{X},\alpha S\right\rangle=\alpha\left\langle\tilde{X},S\right\rangle\to-\infty$ as $\alpha\to+\infty$, which contradicts the infimum is bounded from below. This implies for every $S\in cl(\Sigma_{++}^n)=\Sigma_+^n$, $\left\langle\tilde{X},S\right\rangle\geq 0$. Hence $\tilde{X}$ is positive semidefinite. Therefore $\inf_{S\in\Sigma_{++}^n}\left\langle\tilde{X},S\right\rangle\geq 0$ and we can construct a sequence $\left\{S^{(k)}\right\}\subset\Sigma_{++}^n\to 0$ to conclude the infimum is 0 indeed. Thus $\sup_{S\in G_1}\left\langle\tilde{X},S\right\rangle\leq 0$ for every $y\in\mathbb{R}^m, b^T y\geq z^*$. This is equivalent to saying $\left\langle\tilde{X},C\right\rangle-\left\langle\tilde{X},\mathcal{A}^\star(y)\right\rangle\leq 0\Rightarrow\mathcal{A}(\tilde{X})^T y\geq\left\langle C,\tilde{X}\right\rangle$ over $\left\{y\in\mathbb{R}^m : b^T y\geq z^*\right\}$.$LP$ duality implies here $\mathcal{A}(\tilde{X})=\alpha b$ for some $\alpha\geq 0$ (holds at $\tilde{X}$). We will prove α can not be zero. Otherwise, $\mathcal{A}(\tilde{X})=0\Rightarrow\left\langle C,\tilde{X}\right\rangle\leq 0$. From separation theorem we know $\tilde{X}\in\Sigma^n, \tilde{X}\neq 0$, and we have proved $\tilde{X}\in\Sigma_+^n$. We are also given $(\bar{y},\bar{S})\in\mathbb{R}^m\oplus\Sigma^n$

such that $\mathcal{A}^\star(\bar{y}) + \bar{S} = C$ and $\bar{S} \succ 0$. Based on those facts, we have arrived at a contradiction

$$0 \geq \left\langle C, \tilde{X} \right\rangle = \left\langle \mathcal{A}^\star(\bar{y}) + \bar{S}, \tilde{X} \right\rangle = \underbrace{\mathcal{A}(\tilde{X})^T y}_{=0} + \underbrace{\left\langle \tilde{X}, \bar{S} \right\rangle}_{>0} > 0$$

Therefore α must be a positive number. We can define $\bar{X} \triangleq \frac{1}{\alpha}\tilde{X} \in \Sigma_+^n$ and $\mathcal{A}(\bar{X}) = \mathcal{A}(\frac{1}{\alpha}\tilde{X}) = \frac{1}{\alpha}\mathcal{A}(\tilde{X}) = b$. We should be aware of that $\bar{X}$ is the feasible solution in (P). Because of $\mathcal{A}(\tilde{X})^T y \geq \left\langle C, \tilde{X} \right\rangle$ for all $y \in \mathbb{R}^m$ such that $b^T y \geq z^*$, we can conclude $\langle C, \bar{X} \rangle \leq z^*$ and by weak duality theorem we have $\langle C, \bar{X} \rangle \geq z^*$ thus $\langle C, \bar{X} \rangle = z^* = \sup\left\{ b^T y : \mathcal{A}^\star(y) \preceq C \right\}$. □

The strong duality theorem implies that the optimum value of (D) may not be attained even though it does exist. There are two corollaries to the strong duality theorem.

Corollary 2.3. *If (P) and (D) both have Slater points, then they both attain their optimal values and the optimal values of (P) and (D) are the same.*

The proof simply follows if both primal and dual have Slater point (feasible solution as well), by weak duality theorem, (P) is bounded from below and (D) is bounded from above. Then by applying the strong duality theorem, both dual and primal problems attain their optimum values and the optimum values coincide.

Corollary 2.4. *If (P) has a feasible solution and (D) has a Slater point then (P) attains its optimal value and the optimal value of (P) and (D) are the same.*

The proof is trivial (similar to above corollary).

Let's do some examples to show the duality theorem in practice. I have to take the examples from Levent's lecture notes because I think those examples are really very elegant and self-explained to many concepts we have just introduced.

Example 2.5. Suppose we have the primal problem stated as:

$$\begin{aligned} (P) \quad & \inf \left\langle \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, X \right\rangle \\ st. \quad & \left\langle \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, X \right\rangle = 2 \\ & X \succeq 0 \end{aligned}$$

and dual problem (dual constraint is $C \left(= \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \right) - \mathcal{A}^\star(y) \left(= yA_1 = \begin{pmatrix} 0 & y \\ y & 0 \end{pmatrix} \right) = S \succeq 0$) given by:

$$\begin{aligned} (D) \quad & \sup 2y \\ st. \quad & \begin{pmatrix} 1 & -y \\ -y & 0 \end{pmatrix} \succeq 0 \end{aligned}$$

Suppose $X = \begin{pmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{pmatrix}$ is the feasible solution of (P). By the constraint we can establish the relationships between the component of X. From the equality constraint we get $X_{12} = X_{21} = 1$. From positive semidefinite constraint, we should guarantee $X_{11}X_{22} \geq 1, X_{11} \geq 0, X_{22} \geq 0$. Notice that $\begin{pmatrix} \epsilon & 1 \\ 1 & \frac{1}{\epsilon} \end{pmatrix}$ is feasible in (P) for every $\epsilon > 0$ and $\langle C, X(\epsilon) \rangle = X_{11} = \epsilon \to 0$ as $\epsilon \to 0$. Since $\langle C, X \rangle = X_{11} \geq 0$ and we have constructed $X(\epsilon)$ such that $\inf \langle C, X(\epsilon) \rangle = 0$, thus the primal has the optimal value equal to 0. The optimum value 0 is not attained otherwise $X_{11} = 0$ will contradicts $X_{11}X_{22} \geq 1$. Or since X is positive semidefinite, the diagonal element $X_{11} = 0$ implies $X_{21} = X_{12} = 0$. This also violates the equality constraint. For the dual problem, in order to keep $\begin{pmatrix} 1 & -y \\ -y & 0 \end{pmatrix} \succeq 0$, $y = 0$ is the only feasible point in (D) and thus optimal as well. The dual problem's optimal value is attained at $y = 0$ and there is no duality gap between the primal and the dual. This suggests primal problem should have a slater point since dual problem's optimal value coincides with primal's and is attained as well. Actually $\bar{X} = I + ee^T = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$ is a Slater point in (P), which justifies our conjecture. Let's do another example.

Example 2.6. $C \triangleq \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$, $A_1 \triangleq \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$, $A_2 \triangleq \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}$, $A_3 \triangleq \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$, $A_4 \triangleq \begin{pmatrix} 0 & -1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 2 \end{pmatrix}$, $b \triangleq \begin{pmatrix} 0 \\ 0 \\ 0 \\ 10 \end{pmatrix}$

We can infer the dimension of the constraints by counting the row number of b or the number of A_i's. The dimension of X must be the same to C or any A_i. From the first equality constraint we can conclude X must take form as $X = \begin{pmatrix} & 0 & \\ 0 & 0 & 0 \\ & 0 & \end{pmatrix}$(main diagonal element is 0, then the elements sitting at the same row and column must be 0 as well). Finally we get $X = \begin{pmatrix} * & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 5 \end{pmatrix}$ is the feasible solution in (P). Every feasible solution in (P) is the optimal solution with the optimal value 0. The dual variable $y = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} \in \mathbb{R}^4$ have to satisfy

$$C - y_1A_1 - y_2A_2 - y_3A_3 - y_4A_4 = \begin{pmatrix} 0 & 1+y_4 & -y_2 \\ 1+y_4 & -y_1 & -y_3 \\ -y_2 & -y_3 & -2y_4 \end{pmatrix} \succeq 0$$

We can easily work out $y_4 = -1, y_2 = 0$ thus the dual problem's objective value $b^Ty = 10 \times (-1) = -10$ for all feasible solutions in (D). There is a duality gap of $\langle C, X_{opt} \rangle - b^T y_{opt} = 10$, which implies neither problem has a Slater point.

Let's study the following example in examination of the $LP-like$ unboundedness proof in SDP.

Example 2.7. Suppose primal and dual are given by

$$\begin{aligned} (P) \quad & \inf \left\langle \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, X \right\rangle \\ st. \quad & \left\langle \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, X \right\rangle = 1 \\ & X \succeq 0 \end{aligned}$$

and

$$\begin{array}{ll} (D) & \sup y \\ st. & \begin{pmatrix} y & 0 \\ 0 & 0 \end{pmatrix} \preceq \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \end{array}$$

Let $X(\alpha) \triangleq \begin{pmatrix} 1 & \alpha \\ \alpha & \alpha^2 \end{pmatrix}$ and $X(\alpha)$ is feasible for every $\alpha \in \mathbb{R}$. $\langle C, X(\alpha) \rangle = 2\alpha$ goes to minus infinity when α goes to minus infinity. So the primal problem is unbounded and the dual problem is infeasible ($\begin{pmatrix} -y & 1 \\ 1 & 0 \end{pmatrix}$ can not be positive semidefinite as expected). In LP, when we prove unboundedness, we can start from a feasible point and always find a search direction such that $Ad = 0, d \geq 0$ with $c^T d < 0$ ($x_0 \in \{x \in \mathbb{R}^n : A(x) = b\}$, a new point $x_1 = x_0 + d$ with the objective value at x_1 being $c^T(x_0 + d) = c^T x_0 + \underbrace{c^T d}_{<0} \to -\infty$ as $d \to \infty$). In this example, the feasible region is parabolic since $\begin{pmatrix} 1 & X_{12} \\ X_{12} & X_{22} \end{pmatrix} \succeq 0 \Rightarrow \left\{ \begin{pmatrix} X_{12} \\ X_{22} \end{pmatrix} \in \mathbb{R}^2 : X_{22} \geq X_{12}^2 \right\}$.We can not find a line search direction from starting point $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$ such that all points along that direction are feasible and keep the objective value increasing for max/deceasing for min to infinity. This is the difference between LP and SDP: we may go along a parabolic curve ($X_{22}(= \alpha^2) = X_{12}^2(X_{12} = \alpha)$ as given in the example) to prove the unboundedness of the problem in question instead of a line. With SDP terminology, it is equivalent to say if there does not exist a search direction $D \in \Sigma^n_+$ such that $\mathcal{A}(D) = 0$ and $tr(CD) < 0$, the problem is unbounded. However, we say (D) is almost feasible if we perturb C a little bit for every $\epsilon > 0$, there exists $C' \in \Sigma^n$ such that $||C - C'|| < \epsilon$ and $\mathcal{A}^\star(y) \preceq C'$ feasible.

Theorem 2.8. *Suppose $\mathcal{A} : \Sigma^n \mapsto \mathbb{R}^m$ and $C \in \Sigma^n$ are given. Then*

(a) if there exists $D \in \Sigma^n$ such that $D \succeq 0$, $\mathcal{A}(D) = 0$, $tr(CD) < 0$, then there does not exist $y \in \mathbb{R}^m$ such that $\mathcal{A}^\star(y) \preceq C$ (unboundedness$\Rightarrow$infeasible)

(b) if there does not exist $D \in \Sigma^n$ such that $D \succeq 0$, $\mathcal{A}(D) = 0$, $tr(CD) < 0$ then (D) is almost feasible.

Proof. Part (a) is an easy part which can be proved by contradiction. $0 \underbrace{\leq}_{C - \mathcal{A}^\star(y), D \in \Sigma^n_+} \langle C, D \rangle - \langle \mathcal{A}^\star(y), D \rangle = \langle C, D \rangle - y^T \underbrace{\mathcal{A}(D)}_{=0} = \langle C, D \rangle < 0$. In order to prove part (b), we construct a primal-dual pair as follows:

$$\begin{array}{ll} (D_1) & \sup \eta \\ st. & \mathcal{A}^\star(y) + \eta I \preceq C \\ & \eta \leq 0 \end{array}$$

and its dual is given by

$$\begin{array}{ll} (P_1) & \inf \langle C, X \rangle \\ st. & \mathcal{A}(X) = 0 \\ & tr(X) \leq 1 \\ & X \succeq 0 \end{array}$$

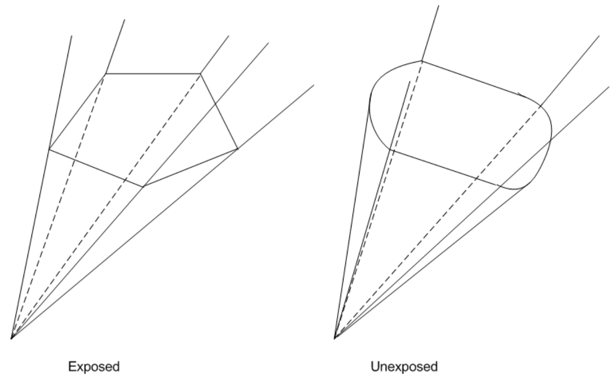

FIGURE 3.1. Face Exposure

Let $\bar{y} \triangleq 0, \bar{\eta} \triangleq -||C||_2 - 1$. The way we choose η is to guarantee $\mathcal{A}^\star(\bar{y}) + \bar{\eta} I \preceq C$ by introducing diagonal dominant matrix to

$$C - \bar{\eta} I = \begin{pmatrix} C_{11} + ||C||_2 + 1 & C_{12} & \cdots & C_{1n} \\ C_{21} & C_{22} + ||C||_2 + 1 & & C_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ C_{n1} & C_{n2} & \cdots & C_{nn} + ||C||_2 + 1 \end{pmatrix}$$

where $||C||_2 = \langle C, C \rangle^{1/2} = \sqrt{\left(\sum_{j=1}^n \sum_{i=1}^n C_{ij}^2 \right)}$. Thus the dual problem (D_1) has a Slater point and is bounded from 0. Notice that $\bar{X} \triangleq 0$ is feasible in (P_1) therefore by corollary of the strong duality theorem, (P_1) has attained its optimal value and there is no duality gap. Suppose we don't have $D \in \Sigma^n$ such that $D \succeq 0$, $\mathcal{A}(D) = 0$, $tr(CD) < 0$, this implies the optimal objective value of (P_1) is zero. By strong duality theorem, the optimal objective value of (D_1) is also 0. Therefore either (D_1) attains its optimal value ($\eta = 0$, and $\mathcal{A}^\star(y) \preceq C$) or there exists a sequence $\left\{ y^{(k)}, \eta_k \right\}$ such that

$$\mathcal{A}^\star \left(y^{(k)} \right) + \eta_k I \preceq C$$

and $\eta_k \to 0^-$. In both cases, $\mathcal{A}^\star(y) \preceq C$ is almost feasible. □

3. SLATER CONDITION AND BORWEIN-WOLKOWICZ APPROACH

In many cases for SDP problems, both primal and dual problem have feasible solutions but their optimal values do not coincide. There are ways to define an appropriate dual problem such that strong duality theorem holds with the same optimal objective value. First let's see a definition of face of a convex cone. Let $\mathcal{K} \subset \mathbb{R}^d$ be a convex cone, $G \subseteq \mathcal{K}$ is called a face of $\mathcal{K}$ if for every $u, v \in \mathcal{K}$ such that $u + v \in G$, we have both end points $u, v \in G$. A face G of $\mathcal{K}$ is called exposed if there exists a normal vector $a \in \mathbb{R}^d$ such that $G = \{x \in \mathcal{K} : \langle a, x \rangle = 0\}$ and $\mathcal{K} \subseteq \{x : \langle a, x \rangle \geq 0\}$, which is equivalent to saying exposed face G is the intersection of $\mathcal{K}$ with one of its supporting hyperplanes. An exposed face and unexposed face of a convex cone is shown in the following figure:

The LHS figure shows the fact that every polyhedral cones are facially exposed. The RHS figure show that the two sided faces are not exposed since we can not find a normal a satisfying above both contraints. For positive semidefinite cone Σ_+^n, the face of it has very special and interesting properties.

Theorem 3.1. *(a) Every proper face of Σ_+^n is characterized by a linear subspace $\mathcal{L}$ such that for every face G, $G = \{x \in \Sigma_+^n : Null(X) \supseteq \mathcal{L}\}$ and $relint(G) = \{X \in \Sigma_+^n : Null(X) = \mathcal{L}\}$.*

(b) Every proper face is exposed.

(c) Every proper face is projectionally exposed. In particular, $\exists Q$ such that $G = (I-Q)\Sigma_+^n(I-Q)$ where $Q \in \Sigma_+^n$ is the projection onto the unique subspace $\mathcal{L}$ defining G. In other words, every proper face of Σ_+^n is isomorphic to Σ_+^k for $k < n$. $\exists W \in \mathbb{R}^{n\times n}$, nonsingular such that $G = W\left\{\begin{pmatrix} X & 0 \\ 0 & 0 \end{pmatrix} : X \in \Sigma_+^k\right\}W^T$. $\exists T \in Aut(\Sigma_+^n)$ such that $G = T\begin{pmatrix} \Sigma_+^k & 0 \\ 0 & 0 \end{pmatrix}$.

If we can find a minimum face of Σ_+^n such that it contains the feasible region, we can apply linear isomorphism so that the image of the proper face is $\left\{\begin{pmatrix} X & 0 \\ 0 & 0 \end{pmatrix} \in \Sigma^n : X \in \Sigma_+^k\right\}$. Let's consider the original SDP problem:

$$\begin{aligned} (P) \quad & \inf \langle C, X\rangle \\ st. \quad & \mathcal{A}(X) = b \\ & X \succeq 0 \end{aligned}$$

Suppose the primal has a finite optimal objective value, it implies that the feasible region is nonempty. Let $\bar{G}$ denote the minimal face of Σ_+^n containing the feasbile region, we can rewrite the primal problem in a new form $(\bar{P})$:

$$\begin{aligned} (\bar{P}) \quad & \inf \langle C, X\rangle \\ st. \quad & \mathcal{A}(X) = b \\ & X \in \bar{G} \end{aligned}$$

where $X \in \bar{G}$ if and only if $X \in \mathcal{L}\left\{\begin{matrix} \Sigma_+^k & 0 \\ 0 & 0 \end{matrix}\right\}$.

In general, if we add redudant constraints to the original primal problem, the duality gap can be closed. However, the dual feasible region could be potentially larger by doing so. Let's do an example to see how this happens.

The revised primal-dual pair can be written as:

$$\begin{aligned} (\tilde{P}) \quad & \inf \langle C, X\rangle \\ st. \quad & \mathcal{A}(X) = b \\ & \underbrace{\tilde{\mathcal{A}}(X) = 0}_{redundent} \\ & X \succeq 0 \end{aligned}$$

and

$$\begin{aligned} (\tilde{D}) \quad & \sup b^T y \\ st. \quad & \mathcal{A}^\star(y) + \tilde{\mathcal{A}}^\star(v) \preceq C \end{aligned}$$

The feasible region y is larger than the original one.

Example 3.2. i.e, $C = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ $A_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ $A_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$ $A_3 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$ $b = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$. Suppose the third constraint is newly added redundant constraint (By redundant we mean from A_1 and A_2 we are able to decide $X_{23} = X_{32} = 0$ of X already). We know the matrix variable X should take form as $\begin{pmatrix} 1 & 0 & * \\ 0 & 0 & 0 \\ * & 0 & * \end{pmatrix}$ from above contraints and the primal problem has the optimal value 1 and the optimal value is attained. In the revised dual problem

$$\begin{array}{ll} (\tilde{D}) & \sup y_2 \\ st. & \begin{pmatrix} 1-y_2 & 0 & 0 \\ 0 & -y_1 & -y_2-y_3 \\ 0 & -y_2-y_3 & 0 \end{pmatrix} \succeq 0 \end{array}$$

we can deduce that $y_2 = -y_3$, $y_2 \leq 1$ and $y_1 \leq 0$. The optimal value of $(\tilde{P})$ is 1 and there is no duality gap. In the original problem without the third constaint in $(\tilde{P})$, the dual problem (D) is given by

$$\begin{array}{ll} (D) & \sup y_2 \\ st. & \begin{pmatrix} 1-y_2 & 0 & 0 \\ 0 & -y_1 & -y_2 \\ 0 & -y_2 & 0 \end{pmatrix} \succeq 0 \end{array}$$

where we have the following contraints on y_1, y_2 such that $y_2 = 0, y_1 \leq 0$. Notice that the duality gap is 1 and it is closed by adding the redundant constraint.

Borwein and Wolkowicz gives an algorithm on how to get to $(\bar{P})$ from (P) in finitely many steps. Since $(\bar{P})$ considers minimal face of Σ^n_+, Slater condition holds for $(\bar{P})$ thus strong duality theorem can be applied to $(\bar{P})$. Ramana proposed another way to arrive from (P) to $(\bar{P})$. For the original dual problem

$$\begin{array}{ll} (D) & \sup b^T y \\ st. & \mathcal{A}^\star(y) \preceq C \end{array}$$

Ramana gives the so called Extended Lagrangian-Slater Dual

$$\begin{array}{ll} (ELSD) & \inf \langle C, U + W + W^T \rangle \\ st. & \mathcal{A}(U + W + W^T) = b \\ & \mathcal{A}(V) = 0 \\ & U \succeq 0 \\ & W \in \mathbb{R}^{n \times n} \end{array}$$

and proved that if (D) has a finite optimal objective value then so deos the $(ELSD)$ and their optimal values are the same, and $(ELSD)$ attains its optimal.

Theorem 3.3. *Let $\mathcal{A} : \Sigma^n \mapsto \mathbb{R}^m$, $C \in \Sigma^n$ be given, then exactly one of the following systems has a solution.*
(I) $\mathcal{A}^\star(y) \preceq C$,
(II) $\mathcal{A}(U + W) = 0$, $\mathcal{A}(V) = 0$,$V \succeq WW^T$, $U \succeq 0$ and $\langle C, U + W + W^T \rangle = -1$

We will talk something about when the Slater condition holds in SDP relaxations. We may encounter problem like in a linear form inf $\left\{c^T x : x \in F\right\}$ or a quadratic form inf $\left\{c^T x + x^T C x : x \in F\right\}$ where $F \subset \mathbb{R}^n$ and F is nonconvex. We need to do convexations on the feasible region and get Slater condition held.

Definition 3.4. Assume F and $\mathcal{A}$ have the property that $F = \left\{x \in \mathbb{R}^n : \mathcal{A}\begin{pmatrix} 1 & x^T \\ x & xx^T \end{pmatrix} = 0\right\}$ where $\mathcal{A} : \Sigma^{n+1} \mapsto \mathbb{R}^m$ is a linear transformation. We call this form the Homogeneous Equality Form (HEF).

We can put every system of finitely many quadratic inequalities into the Homogeneous Equality Form. The following Homogeneous Equality Form

$$\left\langle \begin{pmatrix} \gamma & q^T & 0 \\ q & Q & 0 \\ 0 & 0^T & 1 \end{pmatrix}, \begin{pmatrix} 1 & x^T & \tilde{s} \\ x & xx^T & \tilde{s}x \\ \tilde{s} & \tilde{s}x^T & \tilde{s}^2 \end{pmatrix} \right\rangle$$

is equivalent to the quadratic inequality $\gamma + 2q^T x + x^T Q x \leq 0$ where $Q \in \Sigma^n$,$q \in \mathbb{R}^n$, $\gamma \in \mathbb{R}$ and $\tilde{s} \in \mathbb{R}$. Actually every finite system of polynomial inequalities can be put into Homogeneous Equality Form.

Example 3.5. Put $x_1 x_2^3 x_4 + x_3^2 + x_5^3 \leq 0$ into HEF.
Let $y_1 = x_2^2$, $y_2 = y_1 x_2$, $y_3 = x_1 x_4$, $y_4 = x_5^2$, we arrive at $y_2 y_3 + x_5 y_4 + x_3^2 \leq 0$ ....

SDP relaxation on $\mathcal{F} \triangleq conv\left\{\begin{pmatrix} 1 & x^T \\ x & X \end{pmatrix} \in \Sigma^{n+1} : x \in F\right\}$ gives

$$\hat{P} \triangleq \left\{\begin{pmatrix} 1 & x^T \\ x & X \end{pmatrix} \in \Sigma^{n+1} : \mathcal{A}\begin{pmatrix} 1 & x^T \\ x & xx^T \end{pmatrix} = 0, \begin{pmatrix} 1 & x^T \\ x & X \end{pmatrix} \succeq 0\right\}$$

since $X = xx^T \succeq 0$ is also a feasible point in $\mathcal{F}$. It is obvious $\hat{P} \supseteq \mathcal{F}$ and F is the projection of the intersection of $\hat{P}$ with rank-1 matrices. Keep in mind

$$\inf\left\{c^T x + x^T C x : x \in F\right\} = \inf\left\{\left\langle \begin{pmatrix} 0 & \frac{1}{2}c^T \\ \frac{1}{2}c & C \end{pmatrix}, \begin{pmatrix} 1 & x^T \\ x & X \end{pmatrix} \right\rangle : \begin{pmatrix} 1 & x^T \\ x & X \end{pmatrix} \in \mathcal{F}\right\}$$

We have the following nice theorem about the Slater condition for $\hat{P}$.

Theorem 3.6. *Suppose F and $\hat{P}$ are given as above, then the Slater condition holds for $\hat{P}$ if $conv(F)$ is full dimensional.*

Proof. $v^{(1)}, v^{(2)}, \cdots v^{(k)} \in \mathbb{R}^n$ are affinely independent iff $\left(v^{(2)} - v^{(1)}\right)$, $\left(v^{(3)} - v^{(1)}\right)$,$\cdots$, $\left(v^{(k)} - v^{(1)}\right)$ are linearly independent or if $\begin{pmatrix} v^{(1)} \\ 1 \end{pmatrix}, \cdots \begin{pmatrix} v^{(k)} \\ 1 \end{pmatrix} \in \mathbb{R}^{n+1}$ in the lifted space are linearly independent. Since $conv(F)$ is full dimensional, there exists an affinely independent vectors in F. We define a new matrix $V_\lambda \triangleq \sum_{i=1}^{n+1} \lambda_i \begin{pmatrix} 1 \\ v^{(i)} \end{pmatrix} \begin{pmatrix} 1 & \left(v^{(i)}\right)^T \end{pmatrix}$ where $\lambda \geq 0$ and $\bar{e}^T \lambda = 1$. It is obvious $V_\lambda \in \mathcal{F} \subseteq \hat{P}$ and V_λ is nonsingular if $\lambda > 0$ and $\bar{e}^T \lambda = 1$. We can further get $V_\lambda \in \left(\hat{P} \cap \Sigma^n_{++}\right)$ as desired. □

What if the dimension of the convex hull of F is not full? We can define a lower dimensional convex hulls which still satisfy the Slater condition in a lower dimension problem. Suppose $\dim(conv(F)) = d < n$, there exist $L \in \mathbb{R}^{d \times n}$, $l \in \mathbb{R}^n$ such that L has full row rank and $x \in F \Rightarrow x = l + L^T y$ for some $y \in \mathbb{R}^d$. In the lifted space we define $\mathcal{L}(Z) : \Sigma^{n+1} \mapsto \Sigma^{d+1}$

$$\mathcal{L}(Z) \triangleq \begin{pmatrix} 1 & l^T \\ 0 & L \end{pmatrix} Z \begin{pmatrix} 1 & 0^T \\ l & L^T \end{pmatrix}$$

and define $\bar{\mathcal{A}} : \Sigma^{d+1} \mapsto \mathbb{R}^m$ as

$$\bar{\mathcal{L}}(W) \triangleq \mathcal{A}\left(\mathcal{L}^*(W)\right) = \mathcal{A}\left(\left(\begin{array}{cc} 1 & 0^T \\ l & L^T \end{array}\right) W \left(\begin{array}{cc} 1 & l^T \\ 0 & L \end{array}\right)\right)$$

We can rewrite the feasible region F as

$$F = \left\{ l + L^T y : y \in \mathbb{R}^d, \bar{\mathcal{A}}\left(\begin{array}{cc} 1 & y^T \\ y & Y \end{array}\right) = 0 \right\}.$$

We define

$$F_{\mathcal{L}} \triangleq \left\{ y \in \mathbb{R}^d : \bar{\mathcal{A}}\left(\begin{array}{cc} 1 & y^T \\ y & yy^T \end{array}\right) = 0 \right\}$$

and

$$\mathcal{F}_{\mathcal{L}} \triangleq conv\left\{ \left(\begin{array}{cc} 1 & y^T \\ y & yy^T \end{array}\right) \in \Sigma^{d+1} : y \in F_{\mathcal{L}} \right\}$$

and

$$\hat{P}_{\mathcal{L}} \triangleq \left\{ \left(\begin{array}{cc} 1 & y^T \\ y & Y \end{array}\right) \in \Sigma^{d+1} : \bar{\mathcal{A}}\left(\begin{array}{cc} 1 & y^T \\ y & Y \end{array}\right) = 0, \left(\begin{array}{cc} 1 & y^T \\ y & Y \end{array}\right) \succeq 0 \right\}.$$

Notice that $conv(F_{\mathcal{L}})$ is full dimensional thus Slater condition holds for $\hat{P}_{\mathcal{L}}$. In addtion, we have the following SDP relaxation:

$$\begin{array}{ll} & \inf \left\langle \mathcal{L}\left(\begin{array}{cc} 0 & \frac{1}{2}c^T \\ \frac{1}{2}c & C \end{array}\right), \left(\begin{array}{cc} 1 & y^T \\ y & yy^T \end{array}\right) \right\rangle \\ st. & \bar{\mathcal{A}}\left(\begin{array}{cc} 1 & y^T \\ y & Y \end{array}\right) = 0 \\ & \left(\begin{array}{cc} 1 & y^T \\ y & Y \end{array}\right) \succeq 0 \end{array}$$

Note: $\mathcal{A}\left(\begin{array}{cc} 1 & x^T \\ x & xx^T \end{array}\right) = \mathcal{A}\left(\mathcal{L}^*\left(\begin{array}{cc} 1 & y^T \\ y & yy^T \end{array}\right)\right)$ where $x = l + L^T y$.

4. Ellipsoid Method and Primal-Dual Interior-Point Method

An $E \in \mathbb{R}^d$ is called an ellipsoid if $\exists c \in \mathbb{R}^d$ and $A \in \Sigma^d_{++}$ such that $E \triangleq E(c, A) \triangleq \left\{ x \in \mathbb{R}^d : (x-c)^T A^{-1}(x-c) \le 1 \right\}$. Let $B_d(0,1)$ denote an unit ball in $\mathbb{R}^d$ at the origin. The volumn of $E(c,A)$ is given by $vol(E(c,A)) = \sqrt{\det(A)} vol(B_d(0,1))$ where $vol(B_d(0,1)) = \frac{\pi^{d/2}}{\Gamma(\frac{d}{2}+1)}$. Suppose $G \in \mathbb{R}^d$ is the convex set of interest. Define δ-relaxation of G as $relax(G,\delta) \triangleq \left\{ u \in \mathbb{R}^d : \|u - x\|_2 \le \delta, some\, x \in G \right\}$. We will resort to a weak separation oracle for G which takes input as $\bar{x} \in \mathbb{R}^d$ and correctly returns either (i) $x \in relax(G,\delta)$ or (ii) there eixsts $a \in \mathbb{R}^d$ such that $\|a\|_\infty = 1$ and $\langle a, \bar{x} \rangle \ge \langle a, x \rangle - \delta, \forall x \in relax(G,\delta)$. The oracle tells us either the given point is in the relaxed region of G or we can find a separation hyperplane such that the given point and G lies in the opposite sides of this hyperplane.

Theorem 4.1. *Every compact, convex set in $\mathbb{R}^d$ with nonempty interior, there exists a unique minimal volumn ellipsoid (Lowner-John ellipsoid) containing that set. Moreover, shrinking that ellipsoid around that center by a factor of at most d gives an ellipsoid contained in the convex set.*

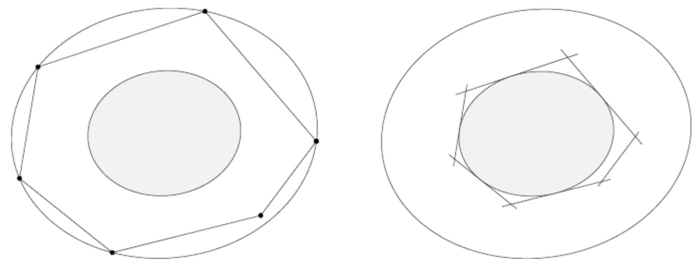

FIGURE 4.1. Lowner-John Ellipsoid- Minimum Volumn Ellipsoid

Proof. Suppose $C \subset \mathbb{R}^d$ is the compact, convex set with nonempty interior. Let $\bar{A} \in \Sigma_{++}^d, c \in \mathbb{R}^d$ such that ellipsoid $\bar{E} \triangleq \left\{x \in \mathbb{R}^d : (x-c)^T \bar{A}(x-c) \leq 1\right\}$ containing C. In order to find the minimal volumn ellipsoid, we want to solve the following optimization problem as given by

$$\begin{array}{rc} (P_{\bar{A}}) & \min -\ln(\det(\bar{A})) \\ st. & (x-c)^T \bar{A}(x-c) \leq 1, \forall x \in C \\ & \bar{A} \in \Sigma_{++}^d, c \in \mathbb{R}^d \end{array}$$

Since C is bounded, we are able to define an ellipsoid centered at origin that contains C in it. I.e., let $M \triangleq \max\{\|x\|_2 : x \in C\} + 1$ and $\bar{A} \triangleq \frac{1}{M^2} I, c \triangleq 0$, then $\bar{E} \triangleq \left\{x \in \mathbb{R}^d : (x-c)^T \bar{A}(x-c) \leq 1\right\} \supseteq C$. This minimization problem is also bounded from below since the required minimal volumn ellipsoid has to be at least as large as the maximal ellipsoid inscribed in C. Notice that $\bar{A}$ in $(P_{\bar{A}})$ is also a function of x thus the inequality constraint is nonlinear (at least third order polynomial and not convex) and we want to formulate in a way such that SDP algorithms can be applied. This motivates us to lift the problem to $\mathbb{R}^{d+1}$ with C in $\mathbb{R}^d$ as contained in an embedded hyperplane $\left\{x \in \mathbb{R}^{d+1} : x_0 = 1\right\}$ in the lifted space. We will focus on the revised problem given by

$$\begin{array}{rc} P_A & -\ln(\det(A)) \\ st. & (1, x^T) \begin{pmatrix} \alpha & a^T \\ a & A \end{pmatrix} \begin{pmatrix} 1 \\ x \end{pmatrix} \leq 1 \quad \forall x \in C \\ & \begin{pmatrix} \alpha & a^T \\ a & A \end{pmatrix} \succeq 0 \\ & A \in \Sigma_{++}^d, a \in \mathbb{R}^d, \alpha \in \mathbb{R} \end{array}$$

Above P_A still has the desired properties as $P_{\bar{A}}$and the solutions to each problem are highly correlated. Let $(\bar{A}, c)$ be the feasible solution of $(P_{\bar{A}})$, then $(A \triangleq \bar{A}, a \triangleq -\bar{A}c, \alpha \triangleq c^T \bar{A} c)$ is a feasible solution of (P_A) with the same objective value. This is because we have the following equality

$$(x-c)^T \bar{A}(x-c) = x^T \bar{A} x - 2(\bar{A}c)^T x + c^T \bar{A} c = (1, x^T) \begin{pmatrix} \alpha & a^T \\ a & A \end{pmatrix} \begin{pmatrix} 1 \\ x \end{pmatrix}$$

In the other direction, let (A, a, α) be an optimal solution of (P_A), we can claim $\alpha = a^T A^{-1} a$. Otherwise we can manipulate the inequality constraint $\alpha \searrow + x^T A x \nearrow + 2a^T x$ in a way such that it still stays feasible for $\forall x \in C$ and $A \leftarrow A + \epsilon a a^T \succeq 0$ for small enough $\epsilon > 0$, which will contradict the optimality assumption that (A, a, α) is optimal. Therefore, setting $\bar{A} \triangleq A, c \triangleq= A^{-1} a$ will also make $(P_{\bar{A}})$ feasible with the same objective value. We know the objective function is strictly convex in Σ_{++}^d thus the local minimimizer of (P_A)

is also a global minimizer. Let $E \triangleq \left\{x \in \mathbb{R}^d : (x-c)^T A^{-1}(x-c) \le 1\right\}$ and $\tilde{E} \triangleq \left\{x \in E : a^T x \le a^T c\right\}$(half ellipsoid whose points are derived from one side of the hyperplane cutting through the center) for some $a \in \mathbb{R}^d \backslash \{0\}$.We would like to construct the minimal volumn ellipsoid E_+ that contains the half ellipsoid $\tilde{E}$. Let c_+ and A_+ denote the center and the positive definite matrix determining E_+. We have the following relationships established:

$$c_+ = c - \frac{1}{(d+1)\sqrt{a^T A a}} A a$$

and

$$A_+ = \frac{d^2}{d^2-1}\left[A - \frac{2}{(d+1)a^T A a} A a a^T A\right].$$

This problem can be also converted to finding a minimal volume ball containing the half ball after an affine transformation

$$B_d(0,1) = A^{-1/2} E(c, A) - A^{-1/2} c$$

and the results can be reverted to the original problem domain by applying the inverse affine transformation $A^{1/2}(E_+) + c$. First, we should observe that $A_+ \succ 0$ if $A \succ 0$. This can be proved by noticing that for arbitrary $h \in \mathbb{R}^d \backslash \{0\}$, the quadratic form

$$\begin{aligned} h^T A_+ h = & \quad \tfrac{d^2}{d^2-1} h^T A h - \tfrac{2d^2}{(d-1)(d+1)^2 a^T A a}(a^T A h)^2 \\ \underbrace{\ge}_{Cauchy-Schwarz} & \quad \tfrac{d^2}{d^2-1}\|A^{1/2}\|_2^2 - \tfrac{2d^2\|A^{1/2}a\|_2^2\|A^{1/2}h\|_2^2}{(d-1)(d+1)^2\|A^{1/2}h\|_2^2} \\ = & \quad \left(\tfrac{d}{d+1}\right)^2 \|A^{1/2}h\|_2^2 \qquad > 0 \end{aligned}$$

Suppose we take the image of $\mathbb{R}^d$ under the mapping $A^{-1/2}$ to make our ellipsoid E be a unit ball $B_d(0,1)$. Then A_+ corresponds to

$$I_+ = \frac{d^2}{d^2-1}\left[I - \frac{2}{(d+1)\bar{a}^T \bar{a}} \bar{a}\bar{a}^T\right]$$

where $\bar{a} = A^{1/2}a$. This matrix in the brackets has two eigenvalues $\frac{d-1}{d+1}$ and 1 (with multiplicity $(d-1)$). Therefore $\det(I_+) = \left(\frac{d^2}{d^2-1}\right)^d \cdot \frac{d-1}{d+1} \cdot 1$, then we get the volume E_+ 's formula as $vol(E_+) = \left(\frac{d}{d-1}\right)^{d-1} \frac{d}{d+1} vol(E)$. $\square$

Sherman-Morrison-Woodbury formula is very useful when computing the inverse in form of $(A+uv^T)^{-1}$. Let $A \in \Sigma^d$ be nonsingular and $u, v \in \mathbb{R}^d$ be given. Then $(A+uv^T)$ is nonsingular if and only if $(1+u^T A^{-1} v) \neq 0$. If the latter condition holds, then its inverse can be written as

$$(A+uv^T)^{-1} = A^{-1} - \frac{1}{1+u^T A^{-1} v} A^{-1} u v^T A^{-1}.$$

This can be easily verified by taking multiplication of $(A+uv^T)$ on both sides.

Theorem 4.2. *We have* $\tilde{E} \subseteq E_+$ *and* $\ln\left(\frac{vol(E_+)}{vol(E)}\right) \le -\frac{1}{2d}$

Suppose $vol(E_0) = R$ and we want $E_k \le \epsilon$, from above theorem, by $O(d \ln(\frac{R}{\epsilon}))$ iterations it suffices. The exact arithmetic algorithm for ellipsoid method is presented below:

(1) Given an ellipsoid $E_0 \supseteq G$ and a weak separation oracle for G, let $E_0 \triangleq E(c, A)$
(2) Ask oracle "is c (center of the ellipsoid) in G". If "yes", STOP. Otherwise, we have $a \in \mathbb{R}^d$ such that $G \subseteq \tilde{E} \triangleq \{x \in E(c,A) : \langle a, x\rangle \le \langle a, c\rangle\}$
(3) Construct the smallest volume ellipsoid E_+containing $\tilde{E}$

(4) If $vol(E_+) < \epsilon$ STOP. Report $vol(G) < \epsilon$
(5) $E(c, A) \triangleq E_+$and repeat step 2.

Ellipsoid method is indeed a bisection method over higher dimensional problems. Separating and supporting hyperplanes for a convex set G have fundamental connections to the oracle. Even if the problem is convex but non-differentiable, we can use sub-gradient oracle for problem like inf $\{f(x) : x \in G\}$. The ellipsoid method's elegance in theoretical proof of complexities is far reaching.

Ellipsoid method can be applied to SDP problems in a very elegant way. Suppose (P) and (D) both have Slater points $\bar{X}$ and $(\bar{y}, \bar{S})$,respectively. We can replace our primal problem with a variant which can use ellipsoid method to solve.

$$\begin{array}{rc} (\tilde{P}) & \inf \langle C, X \rangle \\ st. & \mathcal{A}(X) = b \\ & \langle \bar{S}, X \rangle \leq 2 \langle \bar{X}, \bar{S} \rangle \\ & X \succeq 0 \end{array}$$

We have the following theorem with respect to the newly created dual problem $(\tilde{P})$.

Theorem 4.3. *(a) Both (P) and $(\tilde{P})$ have optimal solutions, and their optimal values coincide and are attained.*

(b) Let G denote the feasible region of $(\tilde{P})$. Then G is convex and bounded. Moreover,

$$B_G(\bar{X}, \lambda_n(\bar{X})) \subseteq G \subseteq B_G\left(0, \frac{2\langle \bar{X}, \bar{S} \rangle}{\lambda_n(\bar{S})}\right)$$

where B_G denotes the unit ball in the affined space spanned by G.

(c) We have $\sup\{\langle C, X \rangle : X \in G\} - \inf\{\langle C, X \rangle : X \in G\} \leq \frac{4n\|C\|_2 \langle \bar{X}, \bar{S} \rangle}{\lambda_n(\bar{S})}$.

Primal-dual interior point method generates a sequence $\{X^{(k)}, y^{(k)}, S^{(k)}\} \in \Sigma^n_{++} \oplus \mathbb{R}^m \oplus \Sigma^n_{++}$ such that $\mathcal{A}(X^{(k)}) = b$,$\mathcal{A}^*(y^{(k)}) + S^{(k)} = C$ and $\langle X^{(k)}, S^{(k)} \rangle \to 0$ as $k \to \infty$.

We define a strictly convex function $f : \Sigma^n_{++} \mapsto \mathbb{R}$

$$f(X) := -\ln(\det(X))$$

Notice that for any sequence $\{X^{(k)}\} \subset \Sigma^n_{++}$,converging to the boundary of Σ^n_+ we will have $f(X^{(k)}) \to +\infty$. For every $X \in \Sigma^n_{++}$and $H \in \Sigma^n$, we have $\left\langle f^{'}(X), H \right\rangle = -tr(X^{-1}H)$ and $\left\langle f^{''}(X)H, H \right\rangle = tr(X^{-1}HX^{-1}H) = tr\left[\left(X^{-1/2}HX^{-1/2}\right)^2\right] \geq 0$.

Let's give a proof of above. $\det(X) = X_{11}\det(X\{1,1\}) - X_{12}\det(X\{1,2\}) + \cdots$ where $X\{i,j\}$ denotes the matrix deleting row i and column j. Let $adj(X)$ be the matrix of cofactors so $\frac{d(\det(X))}{dX} = adj(X)$. By chain rule we get $\nabla_X \ln(\det(X)) = \frac{adj(X)}{\det(X)} = X^{-1}$. By Taylor's theorem,

$$f(X+H) = f(X) + \underbrace{\langle \nabla f(X), H \rangle}_{=-tr(X^{-1}H)} + O(\|H\|)$$

and the Hessian of $f(X)$ is $\nabla^2 f(X) = -\frac{\partial}{\partial X}(X^{-1})$. We use the fact that $0 = G(X) = XX^{-1} - I$. Taking first derivative on both sides we arrive at

$$\frac{\partial}{\partial X} 0 \cdot \Delta X = \frac{\partial}{\partial X} G(X) \cdot \Delta X = \frac{\partial X}{\partial X} \cdot X^{-1} \cdot \Delta X + X \frac{\partial X^{-1}}{\partial X} \cdot \Delta X = X^{-1} \cdot \Delta X + X \frac{\partial X^{-1}}{\partial X} \cdot \Delta X.$$

Multiplying X^{-1}on both sides and we get $\nabla^2 f(X)H = X^{-1}HX^{-1}$ as expected. As μ is given, we can define a SDP problem

$$(P_\mu) \quad \inf \tfrac{1}{\mu}\langle C, X\rangle + f(X)$$
$$\mathcal{A}(X) = b$$

Be aware of the fact that (P_μ) is strictly convex and tends to infinity for every sequence tending to the boundary of Σ^n_+. By applying the optimality condition for this CONVEX problem, we get

$$\mathcal{A}(X) = b, X \succ 0$$

and

$$-\mathcal{A}^*(y) - X^{-1} + \frac{1}{\mu}C = 0.$$

If we force $S = \mu X^{-1}$ we will get a new system

$$\begin{aligned} \mathcal{A}(X) = b \qquad X \succ 0 \\ \mathcal{A}^*(y) + S = C \\ S = \mu X^{-1} \end{aligned}$$

Above system has a unique solution and the solution defines the primal-dual central path. We should pay attention to that $\langle X(\mu), S(\mu)\rangle = n\mu = \langle C, X\rangle - b^T y$. Thus we will decrease the duality gap by decreasing μ. A measure of centrality is given by

$$\psi(X, S) \triangleq n \ln\left(\frac{\langle X, S\rangle}{n}\right) - \ln\det(X) - \ln\det(S)$$

We have the following theorem about the centrality measure function $\psi(X, S)$.

Theorem 4.4. *For every $(X, S) \in \Sigma^n_{++} \oplus \Sigma^n_{++}$, $\psi(X, S) \geq 0$ and the equality holds iff. $S = \mu X^{-1}$ for some $\mu > 0$*

Proof. For every strictly feasible (X, S) we will have

$$\langle X, S\rangle - \ln\det(X) + \sup_{X \in \Sigma^n_{++}} \{-\langle S, X\rangle + \ln\det(X)\} \geq 0$$

The equality holds iff $X = S^{-1}$since logarithm determinant function is strictly convex and $\sup\{-\langle S, X\rangle + \ln\det(X)\} = -\langle S, X\rangle + \ln\det(X)$. Since X is chosen to be the optimal value of $\sup\{-\langle S, X\rangle + \ln\det(X)\}$, the first order necessary and sufficient condition for a minimizer implies that $-S = -X^{-1}$. Therefore we can say for every $(X, S) \in \Sigma^n_{++} \oplus \Sigma^n_{++}$,

$$\langle X, S\rangle - \ln\det(X) - \ln\det(S) - n \geq 0$$

Notice that $\Sigma^n_{++}$is a convex cone, thus we can replace X by $\alpha_1 X, \alpha_1 > 0$ and S by $\alpha_2 S, \alpha_2 > 0$ while still satisfying the inequality.Thus

$$\alpha_1\alpha_2\langle X, S\rangle - n\ln(\alpha_1\alpha_2) - \ln\det(X) - \ln\det(S) - n \geq 0$$

and above equality holds iff. $S = \frac{1}{\alpha_1\alpha_2}X^{-1}$. Let $t \triangleq \alpha_1\alpha_2 > 0$, the above inequality can be reformulated as

$$t\langle X, S\rangle - n\ln t - \ln\det(X) - \ln\det(S) - n \geq 0$$

which is a strictly convex function of t. By the first order necessary and sufficient condition for a minimizer, we have $\langle X, S\rangle - \frac{n}{t} = 0$ which implies

$$t = \frac{n}{< X, S >}.$$

Thus we have

$$\psi(X,S)\geq 0$$

as required. Moreover, the equality holds above iff $S=\mu X^{-1}$ for $\mu=\frac{\langle X,S\rangle}{n}$. $\square$

We will define the primal-dual potential function $\phi_q(X,S)\triangleq q\ln(\langle X,S\rangle)+\psi(X,S)$, which serves the purpose to guide us to find the next appropriate primal-dual pair. Suppose we are in the k-th iteration and we want to update $\left(X^{(k)},S^{(k)}\right)$ with $\left(X^{(k+1)},S^{(k+1)}\right)$ such that $\mathcal{A}(X)=b,\mathcal{A}^*(y)+S=C,X\succ 0,S\succ 0$ are still satisfied. In addition, we want $\langle X,S\rangle$ is decreased and $\psi(X,S)$ is not increased too much. We will find the search directions $D_X,D_S\in\Sigma^n$ and a step size $\alpha\in\mathbb{R}_{++}$. Thus the next primal-dual pair can be expressed as

$$X(\alpha)\triangleq X+\alpha D_X,S(\alpha)\triangleq S+\alpha D_S.$$

It is obvious we should have D_X in the null space of $\mathcal{A}(\cdot)$ and D_S in the range of $\mathcal{A}^*(\cdot)$.

Theorem 4.5. *Suppose $X^{(0)},S^{(0)}\in\Sigma^n_{++}$ feasible in (P) and (D) respectively are given such that for some $\epsilon>0$, $\psi(X^{(0)},S^{(0)})\leq\sqrt{n}\ln(\frac{1}{\epsilon})$. We also assume that we can guarantee $\phi_q(X^{(k+1)},S^{(k+1)})-\phi_q(X^{(k)},S^{(k)})\leq-\delta$ for all $k\in\{0,1,\cdots\}$ where $\delta>0$ is an absolute constant. Then after order of $\bar{k}=O\left(\sqrt{n}\ln(\frac{1}{\epsilon})\right)$ iterations we have the duality gap $\left\langle X^{(k)},S^{(k)}\right\rangle\leq\epsilon\left\langle X^{(0)},S^{(0)}\right\rangle$ for $\forall k\geq\bar{k}$ and $X^{(k)},S^{(k)}$ feasible in (P) and (D).*

Proof. The proof simply follows from the definition of the potential function and the initial assumptions. The details will be omitted here $\square$

We wanna find an automorphism linear, self-adjoint, positive definite transformation $T:\Sigma^n\mapsto\Sigma^n$ such that

(i) $T\in Aut(\Sigma^n_+)$
(ii) $T(S)=T^{-1}(X)=:V$
(iii) $T(X^{-1})=T^{-1}(S^{-1})=:V^{-1}$

Above linear transform gives $\mu=\frac{\langle X,S\rangle}{n}=\frac{\left\langle T^{-1}(X),T(S)\right\rangle}{n}=\frac{\langle V,V\rangle}{n}$. We will define $\bar{D}_X:=T^{-1}(D_X)$ and $\bar{D}_S:=T(D_S)$ respectively. Then

$$\langle X(\alpha),S(\alpha)\rangle=\langle X,S\rangle+\alpha\langle D_X,S\rangle+\alpha\langle X,D_S\rangle+\alpha^2\langle D_X,D_S\rangle=\langle X,S\rangle+\alpha\left\langle V,\bar{D}_X+\bar{D}_S\right\rangle$$

and we utilize the fact that $\langle D_X,S\rangle=\left\langle T^{-1}(D_X),T(S)\right\rangle=\left\langle\bar{D}_X,V\right\rangle$, $\langle X,D_S\rangle=\left\langle T^{-1}(X),T(D_S)\right\rangle=\left\langle V,\bar{D}_S\right\rangle$ and $\langle D_X,D_S\rangle=0$. Since we want to keep the duality gap as small as possible, so $\bar{D}_X+\bar{D}_S=-V$ is the best choice. How about the barrier part of the potential function? Similarly, we have the first order approximation of the difference between two iterations

$$\left[\left\langle f^{'}(X),D_X\right\rangle+\left\langle f^{'}(S),D_S\right\rangle\right]=\left\langle V^{-1},\bar{D}_X+\bar{D}_S\right\rangle$$

so we need to choose $\bar{D}_X+\bar{D}_S$ to be V^{-1} or a positive multiple. Combining both duality gap and barrier function, we choose a joint $\tilde{U}:=-\frac{(n+\sqrt{n})}{\langle X,S\rangle}V+V^{-1}$. The Frobenius norm of $\|\tilde{U}\|_F>0$ otherwise suppose $\|\tilde{U}\|=0$ will imply $V^{-1}=\frac{(n+\sqrt{n})}{\langle X,S\rangle}V$. If we take inner product with V on both sides of above, we arrive at a contradiction $n=\left\langle V^{-1},V\right\rangle=\left\langle\frac{(n+\sqrt{n})}{\langle X,S\rangle}V,V\right\rangle=\frac{(n+\sqrt{n})}{\langle T^{-1}(X),T(S)\rangle}\langle V,V\rangle=(n+\sqrt{n})$. Therefore we can scale $U:=\frac{\tilde{U}}{\|\tilde{U}\|_F}$ and it has a deep connection to the measure of the centrality. Define $\tilde{\mu}:=\frac{\left\langle X^{-1},S^{-1}\right\rangle}{n}=\frac{\left\langle T(X^{-1}),T^{-1}(S^{-1})\right\rangle}{n}=\frac{\left\langle V^{-1},V^{-1}\right\rangle}{n}$ and we have the former definition of $\mu=\frac{\langle X,S\rangle}{n}=\frac{\left\langle T^{-1}(X),T(S)\right\rangle}{n}=\frac{\langle V,V\rangle}{n}$. We can derive $\|\tilde{U}\|_F^2=\frac{(n+\sqrt{n})^2}{n\mu}-2\frac{(n+\sqrt{n})}{\mu}+n\tilde{\mu}=\frac{1}{\mu}\left[n(\mu\tilde{\mu}-1)+1\right]$. We have the following theorem about $\mu\tilde{\mu}$.

Theorem 4.6. *Let $X, S \in \Sigma^n_{++}$, then $\mu\tilde{\mu} \geq 1$ and the equality holds if and only if $S = \mu X^{-1}$.*

Proof. We can construct $\frac{\langle T(S-\mu X^{-1}), T(S-\mu X^{-1})\rangle}{n\mu} = \frac{1}{n\mu}\left\langle V - \mu V^{-1}, V - \mu V^{-1}\right\rangle = \frac{1}{n\mu}\left(n\mu - 2n\mu + n\mu^2\tilde{\mu}\right) = \mu\tilde{\mu} - 1 \geq 0$ and the equality holds iff $S = \mu X^{-1}$. And we can immediately say that $\|\tilde{U}\|_F^2 \geq \frac{1}{\mu}$ and equality holds iff $S = \mu X^{-1}$. □

We come back to the search direction problem, how can we choose the appropriate $\bar{D}_X, d_y, \bar{D}_S$ such that duality gap is decreased and we don't deviate too much from the center path while maintaining the strict feasibility? We need to solve the following system to give the answers:

$$\bar{\mathcal{A}}(\bar{D}_X) = 0, \bar{\mathcal{A}}^*(d_y) + \bar{D}_S = 0, \bar{D}_X + \bar{D}_S = U$$

where $\bar{\mathcal{A}} := \mathcal{A}(T(\cdot))$, $\bar{D}_X := T^{-1}(D_X)$ and $\bar{D}_S := T(D_S)$. By the definition of Frobenius form and orthogonality of $\bar{D}_X$ and $\bar{D}_S$, we get $\|\bar{D}_X\|_F^2 + \|\bar{D}_S\|_F^2 = \|U\|_F^2 = 1$. Thus we get the desired property $\|\bar{D}_X\|_F \leq 1$ and $\|\bar{D}_S\|_F \leq 1$. By the following theorem, we can guarantee that $X(\alpha)$ and $S(\alpha)$ is positive (semi)definite.

Theorem 4.7. *Suppose $X \in \Sigma^n_{++}, D \in \Sigma^n$ satisfies $\|D\|_X := \left\langle D, X^{-1}DX^{-1}\right\rangle^{1/2} \leq 1$ then*

$$f(X) + \left\langle f^{'}(X), D\right\rangle \leq f(X+D) \leq f(X) + \left\langle f^{'}(X), D\right\rangle + \frac{\|D\|_X^2}{2(1-\|D\|_X)^2}$$

Proof. The LHS can be easily proved by the convexity of f. The right hand side inequality follows from the Taylor's theorem. □

The condition $\|D_X\|_X \leq 1(<1)$ implies $(X + D_X) \succeq 0(\succ 0)$ which follows by

$$\begin{aligned} 1 \quad \geq \quad & \|D_X\|_X := \left\langle D_X, X^{-1}D_XX^{-1}\right\rangle^{1/2} = tr\left(D_XX^{-1}D_XX^{-1}\right)^{1/2} \\ & = tr\left(X^{-1/2}D_XX^{-1/2}X^{-1/2}D_XX^{-1/2}\right)^{1/2} = \|X^{-1/2}D_XX^{-1/2}\|_F \end{aligned}$$

which is equivalent to $I \pm X^{-1/2}D_XX^{-1/2} \succeq 0$. Applying automorphism $X^{1/2} \cdot X^{1/2}$ to both sides as above we can get $X \pm D_X \succeq 0$.

The only remaining unsolved problem is what is the automorphism T? Indeed, T can be chosen as $W \cdot W$ where

$$W^2 := S^{-1/2}\left(S^{1/2}XS^{1/2}\right)^{1/2}S^{-1/2}.$$

We need to verify $WSW = W^{-1}XW^{-1} := V$ and $WX^{-1}W = W^{-1}S^{-1}W^{-1} := V^{-1}$. I.e.

$$WSW = W^{-1}XW^{-1} \Leftrightarrow W^2SW^2 = X \Leftrightarrow S^{1/2}W^2SW^2S^{1/2} = S^{1/2}XS^{1/2} \Leftrightarrow W^2 = S^{-1/2}\left(S^{1/2}XS^{1/2}\right)^{1/2}S^{-1/2}.$$

The primal-dual interior point method is given as follows:

Algorithm 4.8. *Given $X^{(0)}, S^{(0)} \in \Sigma^n_{++}$ such that $X^{(0)}$,$S^{(0)}$ are feasible in (P) and (D) respectively, $\psi(X^{(0)}, S^{(0)}) \leq \sqrt{n}\ln(\frac{1}{\epsilon})$ for some given $\epsilon > 0$.*

$k = 0$

while $\left(\left\langle X^{(k)}, S^{(k)}\right\rangle > \epsilon\left\langle X^{(0)}, S^{(0)}\right\rangle\right)$

$W^2 := \left(S^{(k)}\right)^{-1/2}\left[\left(S^{(k)}\right)^{1/2}X^{(k)}\left(S^{(k)}\right)^{1/2}\right]^{1/2}\left(S^{(k)}\right)^{-1/2}$

$\bar{\mathcal{A}}(\cdot) := \mathcal{A}\left(W \cdot W\right)\left[\bar{\mathcal{A}}_i := WA_iW\right]$

$V := WS^{(k)}W$

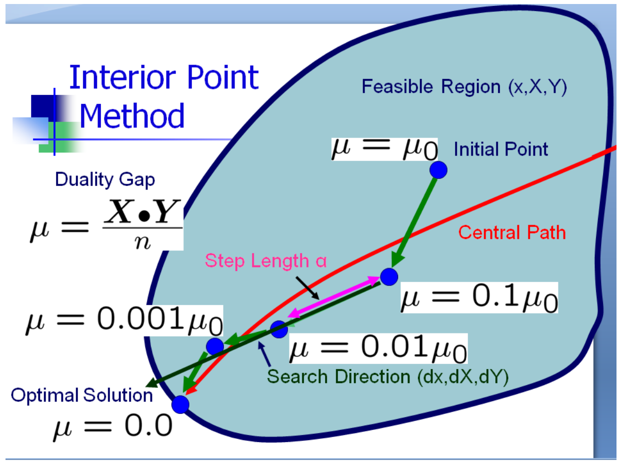

FIGURE 4.2. Primal-Dual Interior Point Method

$\tilde{U} := -\frac{n+\sqrt{n}}{\langle X^{(k)}, S^{(k)} \rangle} V + V^{-1}$, $\bar{U} := \frac{\tilde{U}}{\|\tilde{U}\|_F}$
Solve system

$$\begin{aligned} \bar{\mathcal{A}}(\bar{D}_X) &= 0 \\ \bar{\mathcal{A}}^*(d_y) + \bar{D}_S &= 0 \\ \bar{D}_X + \bar{D}_S &= U \end{aligned}$$

Compute $\bar{\alpha} := argmin \left\{ \phi_{\sqrt{n}} \left(X^{(k)} + \alpha W \bar{D}_X W, S^{(k)} + \alpha W^{-1} \bar{D}_S W^{-1} \right) : \alpha > 0 \right\}$
Let $X^{(k+1)} := X^{(k)} + \bar{\alpha} W \bar{D}_X W$ *and* $S^{(k+1)} := S^{(k)} + \bar{\alpha} W^{-1} \bar{D}_S W^{-1}$
$k := k + 1$
end {while}

The above algorithm terminates at most $24\sqrt{n}\ln(\frac{1}{\epsilon})$ iterations with feasible $X^{(k)}, S^{(k)}$ such that $\langle X^{(k)}, S^{(k)} \rangle \leq \epsilon \langle X^{(0)}, S^{(0)} \rangle$.

5. Application on Approximation Algorithms

SDP is very useful in many applications. In this section, I will give three examples which are based on SDP problem formulations. We can see the beauty of SDP programming to give a better performance in contrast to its counterpart LP programming.

A max cut problem is defined as below: Given a simple undirected graph $G = (V, E)$ a cut is $(S, V \backslash S)$ where $S \subseteq V$. Given a cut S in G, we are interested in the set edges that cross the cut. $\delta(S) := \{\{i, j\} \in E : i \in S, j \in V \backslash S\}$. The max cut problem indicates given $W_{ij} \geq 0$ for $\forall \{i, j\} \in E$, find a cut S in G such that $\sum_{\{i,j\} \in \delta(S)} W_{ij}$ is maximized. This problem actually is an $\mathcal{NP}$-hard problem. In SDP formulation, we will encode vertices instead of edges in LP. The cut will be encodes by $\{-1, 1\}$vector $u \in \{-1, 1\}^{|V|}$ where $u_i := 1$ if $i \in S$ and $u_i = -1$ otherwise. Therefore the objective function can be written as $\max \frac{1}{4} \sum_{i,j} W_{ij} (1 - u_i u_j)$. We can further convert it to the inner product form $\max \frac{1}{4} \langle W, \bar{e}\bar{e}^T \rangle - \frac{1}{4} \langle W, X \rangle$

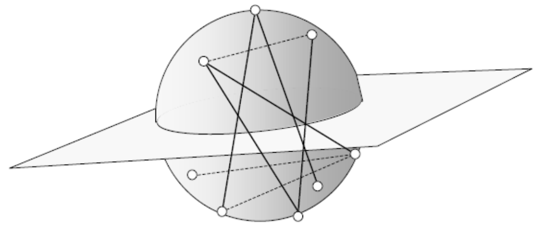

FIGURE 5.1. Random Hyperplane Technique

where $X := uu^T \in \Sigma_+^{|V|}$. The complete SDP formulation is given by

$$\begin{aligned} (P) \quad & \max -\tfrac{1}{4}\langle W, X\rangle \quad \left(+\frac{1}{4}\langle W, \bar{e}\bar{e}^T\rangle\right) \\ st. \quad & diag(X) = \bar{e} \\ & X \succeq 0 \\ & rank(X) = 1 \end{aligned}$$

where the first equality constraint forces $X_{ii} \in \{-1, 1\}, \forall i \in \{1, \cdots |V| = n\}$ and the second and the third constraint force X to be the outer product of a vector. We will delete the $rank$-1 constraint to relax the problem in a tractable form. the dual problem is given by

$$\begin{aligned} (D) \quad & \min \bar{e}^T y \quad \left(+\frac{1}{4}\langle W, \bar{e}\bar{e}^T\rangle\right) \\ st. \quad & Diag(y) - S = -\tfrac{1}{4}W \\ & S \succeq 0 \end{aligned}$$

We will notice that $\bar{X} := I$ is a Slater point in (P) and $\bar{y} = \eta\bar{e}$ where $\eta := \frac{1}{4}\left(\sum_{i,j}|W_{ij}|\right) + 1$ can make $\bar{S} := Diag(\bar{y}) + \frac{1}{4}W \succ 0$.We can apply aforementioned primal-dual interior method to find the optimal solution X^* of (P) and factorize $X^* = VV^T$where $V^T := \left[v^{(1)}, \cdots v^{(n)}\right]$ and $v^{(i)} \in \mathbb{R}^d$ $d \leq n$ if $rank(X^*) = d$. We can then pick a random point r on the hypersphere (since every $v^{(i)}$ is on the hypersphere in $\mathbb{R}^d$) in $\mathbb{R}^d$ and define $S := \left\{i \in V : r^T v^{(i)} \geq 0\right\}$. I will omit the proof of the performance in the mean sense, which states that the expected value of the cut generated by this random hyperplane technique (RHT) yields at least $\rho := 0.87856$ approximation of the optimal objective value of the SDP problem. All these random algorithm can be de-randomized to give a deterministic algorithm. The best known worst case for SDP relaxation on Max -Cut problem is given by an odd cycle with length 5, the max cut is therefore 4 (assume edge weight is equal to 1), but the optimal value of the SDP problem is $2\sqrt{5}$. Max cut based on SDP is superior to its counterpart based on LP because in SDP formulations, we use a vector to represent each node and more underlying information will be utilized in classification.

Satisfiability problem is defined as: Find binary variables $x_1, x_2, \cdots x_n \in \{true = 1, false = 0\}$ such that formula $C_1 \cap C_2 \cdots \cap C_m$ is true where clause $C_i := l_1 \cup \cdots \cup l_k$. Literal l_i is either x_j or $\bar{x}_j$. Maximum satisfiability problem is defined as: Given weights w_i for each clause, find $x \in \{0, 1\}^n$ such that total weights of satisfied clauses is maximized. Max k-Sat problem forces each clause has at most k-literals. We can have a ρ-approximation polynomial time algorithm for max-2 SAT problem in contrast to its 0.75 approximation in LP formulation.

The SDP formulation for Max 2-SAT problem will use some tricks and we can see Max 2-SAT problem is indeed equivalent to Max Cut problem. We will use an additional variable $x_0 \in \{-1,1\}$ and $\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \in \{-1,1\}^n$ such that $x_0$determines what is "true". Therefore x_j is true if and only if $x_j = x_0$. Then we can easily verify that literal x_i can be represented by $\frac{(x_0 x_1 + 1)}{2}$ and $\bar{x}_i$ can be denoted as $\frac{1 - x_0 x_i}{2}$. Clause $(x_i \cup x_j)$ by De-Morgan's law can be written as

$$(x_i \cup x_j) = 1 - (\bar{x}_i)(\bar{x}_j) = \frac{1}{4}(1 + x_0 x_i) + \frac{1}{4}(1 + x_0 x_j) + \frac{1}{4}(1 - x_i x_j)$$

which implies the max 2-SAT problem can be formulated very similar to max cut problem and can be applied the same RHT to assign true or false to the literals. To be more specific. the objective value is given by

$$\max \frac{1}{4}\left[\sum_{\{i,j\}\in E^+} W_{ij}(1 + x_i x_j) + \sum_{\{i,j\}\in E^-} W_{ij}(1 - x_i x_j)\right]$$

where $x \in \{-1,1\}^{n+1}$, $W_{ij} \geq 0$ for $\forall i, j$. Let's do another example to see what SDP can do for us.

Example 5.1. Consider the optimization problem $\bar{f}(W) := \max_{x\in\{-1,1\}^n} x^T W x$ and $\underline{f}(W) := \min_{x\in\{-1,1\}^n} x^T W x$. They can be rewritten in their original and relaxed SDP forms, respectively, i.e:

$$\begin{aligned} & \bar{f}(W) = \max \langle W, X\rangle \\ st. \qquad & diag(X) = \bar{e} \\ & X \succeq 0 \\ & rank(X) = 1 \end{aligned}$$

and

$$\begin{aligned} & \bar{F}(W) = \max \langle W, X\rangle \\ st. \qquad & diag(X) = \bar{e} \\ & X \succeq 0 \end{aligned}$$

We have

$$\underline{\mathrm{F}}(W) \leq \underline{\mathrm{f}}(W) \leq \frac{2}{\pi}\underline{\mathrm{F}}(W) + (1 - \frac{2}{\pi})\bar{F}(W) \leq (1 - \frac{2}{\pi})\underline{\mathrm{F}}(W) + \frac{2}{\pi}\bar{F}(W) \leq \bar{f}(W) \leq \bar{F}(W).$$

6. Summary

Above materials are based on Prof. Levent Tuncel's monograph and lectures given on Semidefinite Optimization in the summer term, 2007. To this far, we only give the basic ideas of SDP optimization (What is a SDP problem? Duality theory of SDP, What is Slater condition and its implications? How to solve a general SDP problem?) and some applications in combinatorial problems (Max-Cut, Max-2 SAT and General Quadratic Programming). Geometric representation of graphs, lift-and-project method and successive regression will be covered in a separate survey along with graph theory on coloring and algebraic method to solve complicated graph problems.

References

[1] CO671/471 Lecture notes on Semidefinite Optimization by Prof. Levent Tuncel, University of Waterloo, August 2007
[2] Lecture notes on Semidefinite Programs and Combinatorial Optimization by L.Lovasz, Microsoft Research,
[3] CO350 Lecture Notes on Linear Programming, University of Waterloo, April 2005
[4] CO754 Lecture notes on Approximation Algorithms by Joseph Cheriyan, University of Waterloo, March 2008
[5] CO663/463 Lecture notes on Convex Optimization and Analysis by Henry Wolkowicz, University of Waterloo, December 2007
[6] Textbook on Convex Optimization by Stephen Boyd and Lieven Vandenberghe, Stanford University, Cambridge University Press, 2004