CHAPTER 2

Duality Theory

Given an optimization problem which is to minimize a function subject to some constraints, how do we evaluate the quality of a feasible solution? One way would be to have a tight lower bound on the optimal objective function value if well-defined. If the optimum value is not well-defined then we would like to have easy to evaluate certificates of this fact. These questions can be answered by a suitable notion of *the dual problem*. In LP theory, the properties of the dual are wonderful. Only some of these wonderful properties remain valid when we move to the settings of SDP and more general convex optimization. As a result, the duality theory for SDP is much richer than that of LP, more interesting, more powerful and in some way perhaps even more elegant.

2.1 Dual Cones

Let $K \subseteq \mathbb{R}^d$ be a convex cone. Recall that its dual is

$$K^* := \left\{ s \in \mathbb{R}^d : \langle x, s \rangle \geq 0, \ \forall x \in K \right\}.$$

Note that K^* is always closed (intersection of closed sets) and convex (intersections of convex sets). Using the results of this chapter (for instance using Theorem 2.8 which will be proven from the first principles, independent of the results of this section), we can prove the following three fundamental results.

Theorem 2.1 *For every convex cone K in $\mathbb{R}^d$, we have*

$$K^{**} = \text{cl}(K).$$

Moreover, a convex cone K is closed iff $K = K^{**}$.

In fact, put slightly differently, K is a closed convex cone iff $K = K^{**}$. Now, recall our formulation of primal and dual convex optimization problems with cone constraints involving K and K^* in the previous chapter. LP and SDP special cases are based on symmetric cones ($\mathbb{R}^n_+$ and Σ^n_+ respectively); so, in these cases $K = K^* = K^{**}$. In general, we have the following property.

Theorem 2.2 *Let $K \subset \mathbb{R}^d$ be a pointed, closed convex cone with nonempty interior. Then so is K^*.*

In many instances, we intersect convex sets (or cones) and consider their duals in our arguments. Another common algebraic operation is the *Minkowski sum* of two sets:

$$K_1 + K_2 := \{u + v : u \in K_1, v \in K_2\}.$$

For pairs of convex cones, these two operations are "dual to each other" in the sense of the next theorem.

Theorem 2.3 *Let $K_1, K_2 \subseteq \mathbb{R}^d$ be nonempty convex cones. Then*

- $(K_1 + K_2)^* = K_1^* \cap K_2^*$;
- $(\mathrm{cl}\,(K_1) \cap \mathrm{cl}\,(K_2))^* = \mathrm{cl}\,(K_1^* + K_2^*)$.

If K_1 *and* K_2 *are closed and* $\mathrm{relint}(K_1) \cap \mathrm{relint}(K_2) \neq \emptyset$ *in the above then*

$$(K_1 \cap K_2)^* = K_1^* + K_2^*.$$

2.2 Polars of (Compact) Sets

Let $K \subset \mathbb{R}^d$ be a pointed closed convex cone with nonempty interior. Let $\bar{s} \in \mathrm{int}\,(K^*)$. Define

$$G := \{x \in K : \langle \bar{s}, x \rangle \leq 1\}.$$

Then G is a compact convex set in $\mathbb{R}^d$ which possesses all the information about K in a very nice way. Sometimes it is slightly more convenient to deal with G and sometimes it is the other way around. So, we will use both.

Now, let $G \subseteq \mathbb{R}^d$. Duality for sets is also very nice. We define the *polar of* G as

$$G^o := \left\{s \in \mathbb{R}^d : \langle x, s \rangle \leq 1, \forall x \in G\right\}.$$

Theorem 2.4 *Let* $G \subset \mathbb{R}^d$ *be a compact convex set such that* $0 \in \mathrm{int}(G)$. *Then so is* G^o. *Moreover,* $G^{oo} = G$.

We also have the analogous result to Theorem 2.3 in the current setting.

Theorem 2.5 *Let* $G_1, G_2 \subset \mathbb{R}^d$ *be compact convex sets such that* $0 \in \mathrm{int}(G_1) \cap \mathrm{int}(G_2)$. *Then*

- $[\mathrm{conv}\,(G_1 \cup G_2)]^o = G_1^o \cap G_2^o$;
- $(G_1 \cap G_2)^o = \mathrm{conv}\,(G_1^o \cup G_2^o)$.

2.3 Conjugates of (Convex) Functions

Another very closely related domain to study duality is that of (convex) functions. For a given function $f : \mathbb{R}^d \to \mathbb{R} \cup \{\infty\}$, we define its Legendre-Fenchel conjugate as

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

This notation and definition are slightly different than the usual ones in convex analysis. (There f^* is used and $f^*(s) = f_*(-s)$; of course, the definition of dual cone is also slightly different in most references in convex analysis.)

Given f as above, the *epigraph of* f is defined as

$$\mathrm{epi}(f) := \left\{ \begin{pmatrix} t \\ x \end{pmatrix} \in \mathbb{R} \oplus \mathbb{R}^d : f(x) \leq t \right\}.$$

It is easy to prove that

Theorem 2.6 *Let* f *be defined as above. Then,* f *is convex iff* $\mathrm{epi}(f)$ *is a convex set.*

Therefore, studying convex sets is essentially equivalent to studying convex functions, as the results for one can be translated via the above correspondence to the results about the other. In particular, we can express the theorems of Sections 2.1 and 2.2 in terms of convex functions and their conjugates (the reader not familiar with these correspondences should try this exercise).

2.4 A Strong Duality Theorem via the Hahn-Banach Theorem

The Hahn-Banach Theorem is one of the most important tools of functional analysis. The most common form of the Hahn-Banach Theorem allows extensions of bounded linear operators from a subspace of a vector space to the whole vector space. Perhaps more importantly, it proves that every normed vector space has sufficiently many linear functionals (dual elements) defined on it so that it makes the space of linear functionals (the dual space) very useful and interesting. This very general issue of *having enough dual elements for a specific purpose* will arise in many disguises throughout the rest of the chapters. Shortly, we will discuss the Strong Duality Theorem for SDP and then try to extend the theorem to the general case where we do not require any constraint qualification (in that case, not having *enough* dual variables will be the initial hurdle). Towards the end of the monograph, when we discuss Successive Convex Relaxation Methods, Polynomial Optimization Problems and Sums-of-Squares proofs of nonnegativity, such issues will again be paramount.

Perhaps the second most common form of the Hahn-Banach theorem is stated as the *Hahn-Banach Separation Theorem* (see for example Rudin [309]), and this separation theorem is usually referred to as *Mazur's geometric form of the Hahn-Banach Theorem.*

Theorem 2.7 *(Hahn-Banach Separation Theorem) Let $\mathbb{E}$ be a topological vector space ($\mathbb{E}^*$ is its dual) and $G_1, G_2 \subset \mathbb{E}$. Suppose G_1 and G_2 are nonempty, convex and disjoint. Then*

(a) *if G_1 is open, then there exist a continuous linear functional $\ell \in \mathbb{E}^*$ and $\alpha \in \mathbb{R}$ such that*

$$\sup\{\ell(u) : u \in G_1\} < \alpha \le \inf\{\ell(v) : v \in G_2\};$$

(b) *if G_1 is compact, G_2 is closed and $\mathbb{E}$ is locally convex, then there exist a continuous linear functional $\ell \in \mathbb{E}^*$ and $\alpha \in \mathbb{R}$ such that*

$$\sup\{\ell(u) : u \in G_1\} < \alpha < \inf\{\ell(v) : v \in G_2\}.$$

In the above, we stated the theorem with the specification that the underlying scalar field is $\mathbb{R}$. (When the underlying scalar field is $\mathbb{C}$, we simply replace $\ell(u)$ and $\ell(v)$ by their real parts.)

Below, we derive finite-dimensional counterparts of the above theorem from the first principles. These theorems serve a similarly central role in finite-dimensional optimization.

Theorem 2.8 *(Separation Theorem) Let $G \subset \mathbb{R}^d$ be a nonempty, closed convex set. Suppose $0 \notin G$. Then there exist $a \in \mathbb{R}^d \backslash \{0\}$ and $\alpha \in \mathbb{R}_{++}$ such that*

$$G \subseteq \left\{x \in \mathbb{R}^d : a^T x \ge \alpha\right\}.$$

Proof Let $\bar{x} \in G$. Define

$$\bar{G} := \{x \in G : \|x\|_2 \le \|\bar{x}\|_2\}.$$

Let $a \in \bar{G}$ be the point in $\bar{G}$ with the minimum norm. (Such a point $a \in \mathbb{R}^d$ exists, since $\bar{G} \neq \emptyset$, $\bar{G}$ is compact, $\|x\|_2^2$ is continuous on $\bar{G}$; in fact a is unique since $\|x\|_2^2$ is strictly convex on $\bar{G}$.) Define $\alpha := a^T a > 0$. Note that for every $x \in G$, the line segment $[a, x]$ lies completely in G. Thus, for every $x \in G$ and for every $\lambda \in (0, 1]$,

$$\begin{aligned} [\lambda x + (1-\lambda)a] \in G \quad &\Rightarrow \quad \|\lambda x + (1-\lambda)a\|_2^2 \geq \|a\|_2^2 = \alpha \\ &\Rightarrow \quad \|\lambda(x-a) + a\|_2^2 \geq \alpha \\ &\Rightarrow \quad \lambda^2\|x-a\|_2^2 + 2\lambda a^T(x-a) \geq 0 \\ &\Rightarrow \quad a^T(x-a) \geq -\frac{\lambda}{2}\|x-a\|_2^2. \end{aligned}$$

Now, we take limits as $\lambda \to 0^+$ and conclude $a^T(x-a) \geq 0$. Therefore, for every $x \in G$, we have $a^T x \geq \alpha$. □

Corollary 2.9 *Let G_1, $G_2 \subset \mathbb{R}^d$ be disjoint, nonempty closed convex sets. If G_1 or G_2 is bounded then there exists $a \in \mathbb{R}^d \setminus \{0\}$ such that*

$$\inf\left\{a^T x : x \in G_1\right\} > \sup\left\{a^T x : x \in G_2\right\}.$$

Proof Let $G_0 := G_1 - G_2$. G_0 is nonempty, convex. Since G_1 and G_2 are disjoint, $0 \notin G_0$. We will prove that G_0 is closed. Then applying the previous theorem will conclude the proof. Let $\left\{g_0^{(k)}\right\}$ be a sequence in G_0 such that $g_0^{(k)} \to \overline{g}_0$. Then $g_0^{(k)} = g_1^{(k)} - g_2^{(k)}$ for some $g_1^{(k)} \in G_1$ and $g_2^{(k)} \in G_2$. Without loss of generality, we can assume that G_2 is bounded. Hence G_2 is compact. By refining to a subsequence if necessary, we can assume $g_2^{(k)} \to \overline{g}_2 \in G_2$. Then the corresponding refined subsequence $g_1^{(k)} \to (\overline{g}_0 - \overline{g}_2) =: \overline{g}_1$. Since G_1 is closed, $\overline{g}_1 \in G_1$. Therefore, $\overline{g}_0 = (\overline{g}_1 - \overline{g}_2) \in G_0$. Hence, G_0 is closed, as desired. Applying the previous theorem, we conclude that there exist $a \in \mathbb{R}^d \setminus \{0\}$ and $\alpha \in \mathbb{R}_{++}$ such that

$$\begin{aligned} & a^T g_0 \geq \alpha > 0, \ \forall g_0 \in (G_1 - G_2) \\ \Rightarrow \quad & a^T g_1 \geq \alpha + a^T g_2, \ \forall g_1 \in G_1, g_2 \in G_2 \\ \Rightarrow \quad & \inf\left\{a^T x : x \in G_1\right\} > \sup\left\{a^T x : x \in G_2\right\} \end{aligned}$$

as desired. □

Remark 2.10 We need at least one of G_1, G_2 bounded. Consider

$$G_1 := \left\{x \in \mathbb{R}^2_{++} : x_2 \geq \frac{1}{x_1}\right\}$$

and

$$G_2 := \left\{x \in \mathbb{R}^2 : x_1 > 0, x_2 \leq -\frac{1}{x_1}\right\}.$$

Then the only viable candidate for $a \in \mathbb{R}^2 \setminus \{0\}$ is $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ which leads to

$$\inf\left\{a^T x : x \in G_1\right\} = \sup\left\{a^T x : x \in G_2\right\}.$$

Theorem 2.11 *Let $G \subset \mathbb{R}^d$ be a nonempty convex set. Suppose $0 \notin G$. Then there exists $a \in \mathbb{R}^d \setminus \{0\}$ such that*

$$G \subseteq \left\{x \in \mathbb{R}^d : a^T x \geq 0\right\}.$$

Proof For every $x \in G$, we define the unit hemisphere in $\mathbb{R}^d$ intersecting the halfspace $\{x\}^*$:

$$HS(x^*) := \left\{s \in \mathbb{R}^d : x^T s \geq 0, \ \|s\|_2 = 1\right\}.$$

We would like to find $a \in \bigcap_{x \in G} HS(x^*)$. By definition, $HS(x^*)$ is a hemisphere in $\mathbb{R}^d$ and therefore is compact. Hence, if every finite collection of $HS(x^*)$'s (for $x \in G$) has a nonempty intersection, then so do all $HS(x^*)$ (where the intersection runs over all elements of G). Let $x^{(1)}, x^{(2)}, \ldots, x^{(k)}$ be an arbitrary collection of points in G. Consider $\text{conv}\left(x^{(1)}, x^{(2)}, \ldots, x^{(k)}\right) \subseteq G$. The set $\text{conv}\left(x^{(1)}, x^{(2)}, \ldots, x^{(k)}\right)$ is a nonempty, compact, convex set not containing 0. Therefore, by the separation theorem, there exists $a \in \mathbb{R}^d \setminus \{0\}$, such that

$$\text{conv}\left(x^{(1)}, x^{(2)}, \ldots, x^{(k)}\right) \subseteq \left\{x \in \mathbb{R}^d : a^T x \geq 0\right\}.$$

We set $a := \frac{a}{\|a\|_2}$. Therefore, every finite collection of $HS(x^*)$'s (for $x \in G$) has a nonempty intersection implying, there exists $a \in \mathbb{R}^d \setminus \{0\}$ such that

$$G \subseteq \left\{x \in \mathbb{R}^d : a^T x \geq 0\right\}.$$

□

Aside: We used a standard fact from elementary real analysis in the above proof. Let $H_t, t \in T$ be an arbitrary collection of closed sets and let $\hat{H}$ be a compact set. If the intersection of all H_t has no intersection with $\hat{H}$, then the union of the complements of H_t (open sets) cover $\hat{H}$. Since $\hat{H}$ is compact, any open cover of it must have a finite subcover (Heine-Borel characterization of compactness). Upon complementing this subcover, we see that the intersection of these finitely many H_t has no intersection with $\hat{H}$ either. In the above proof, we took $H_t := \{x\}^*$, $T := G$ and $\hat{H}$ as the unit hypersphere in $\mathbb{R}^d$.

Corollary 2.12 *Let $G_1, G_2 \subset \mathbb{R}^d$ be nonempty convex sets such that $G_1 \cap G_2 = \emptyset$. Then there exists $a \in \mathbb{R}^d \setminus \{0\}$ such that*

$$\inf\left\{a^T x : x \in G_1\right\} \geq \sup\left\{a^T x : x \in G_2\right\}.$$

Proof We simply apply the previous theorem to $G := G_1 - G_2$. □

Now, we are ready to begin discussing a *Strong Duality Theorem* for SDP. In the statement of our theorem, a *constraint qualification* is necessary. This is what we define next. Recall,

$$\begin{array}{llrcl} (P) & \inf & \langle C, X\rangle & & \\ & \text{subject to:} & \mathcal{A}(X) & = & b, \\ & & X & \succeq & 0, \end{array}$$

$$\begin{array}{llrcrcl} (D) & \sup & b^T y & & & & \\ & \text{subject to:} & \mathcal{A}^*(y) & + & S & = & C, \\ & & & & S & \succeq & 0. \end{array}$$

Definition 2.13 We say that (P) *satisfies the Slater condition*, or (P) *has a Slater point* if there exists $\bar{X} \in \Sigma^n$ such that $\mathcal{A}(\bar{X}) = b$ and $\bar{X} \succ 0$. Similarly, we say that (D) *satisfies the Slater condition*, or (D) *has a Slater point* if there exist $\bar{S} \in \Sigma^n$ and $\bar{y} \in \mathbb{R}^m$ such that $\mathcal{A}^*(\bar{y}) + \bar{S} = C$ and $\bar{S} \succ 0$.

The next theorem has different assumptions and conclusions about (P) and (D). We are assuming that we are interested in the optimal solutions of the primal problem; hence, the attainment of the optimal objective value of (P) is more critical to us. (If it is the attainment of the optimal objective value of (D) that is more important, then we switch the roles of (P) and (D).)

Theorem 2.14 *(A Strong Duality Theorem) 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 Suppose there exist $\bar{S} \in \Sigma^n_{++}$, $\bar{y} \in \mathbb{R}^m$ such that $\mathcal{A}^*(\bar{y}) + \bar{S} = C$. Let $z^* := \sup\left\{b^T y : \mathcal{A}^*(y) \preceq C\right\}$. We can assume without loss of generality that $b \neq 0$. If $b = 0$, then $\bar{X} := 0$ is an optimal solution of (P), because $(\bar{y}, \bar{S})$ is feasible in (D) with objective value zero, which matches the objective value of $\bar{X}$ which is feasible in (P). Consider

$$G_1 := \left\{S \in \Sigma^n : S = C - \mathcal{A}^*(y) \text{ for some } y \in \mathbb{R}^m, b^T y \geq z^*\right\}.$$

Note that $G_1 \neq \emptyset$. There exists a sequence $\left\{\left(y^{(k)}, S^{(k)}\right)\right\} \subset \mathbb{R}^m \oplus \Sigma^n$ such that $\mathcal{A}^*\left(y^{(k)}\right) + S^{(k)} = C$, $S^{(k)} \succeq 0$, and $b^T y^{(k)} \to z^*$. This last statement implies that there exists $\left(\hat{y}, \hat{S}\right) \in \mathbb{R}^m \oplus \Sigma^n$ such that $\mathcal{A}^*(\hat{y}) + \hat{S} = C$ and $b^T \hat{y} = z^*$ (a linear function over an affine subspace attains its limit; note that $\hat{S} \notin \Sigma^n_+$ is allowed; there are many elementary ways of proving this, an unusual proof is to apply Theorem 1.4 to the LP:

$$\max\left\{b^T y : b^T y \leq z^*, \mathcal{A}^*(y) + S = C, y \in \mathbb{R}^m, S \in \Sigma^n\right\}).$$

Now, let us define $G_2 := \Sigma^n_{++}$. We claim that $G_1 \cap G_2 = \emptyset$. We prove this claim by contradiction. Suppose there exists $\bar{y} \in \mathbb{R}^m$ such that $\mathcal{A}^*(\bar{y}) \prec C$, $b^T \bar{y} \geq z^*$; then $\hat{y} := \bar{y} + \epsilon b$ for some $\epsilon > 0$ satisfies $\mathcal{A}^*(\hat{y}) \prec C$ and $b^T \hat{y} > z^*$ (here we used the fact that $b \neq 0$). Now we arrived at a contradiction, since z^* is the optimal value of (D). We apply the last corollary to G_1 and G_2. Then, there exists $\tilde{X} \in \Sigma^n \setminus \{0\}$ such that

$$\sup\left\{\langle \tilde{X}, S\rangle : S \in G_1\right\} \ \leq \ \inf\left\{\langle \tilde{X}, S\rangle : S \in \Sigma^n_{++}\right\}.$$

Since $G_1 \neq \emptyset$, we conclude that $\inf_{S \in \Sigma^n_{++}} \langle \tilde{X}, S\rangle$ is bounded from below. Since Σ^n_{++} is a cone, this yields $\langle \tilde{X}, S\rangle \geq 0$, for every $S \in \Sigma^n_{++}$. So, $\langle \tilde{X}, S\rangle \geq 0$, for every $S \in \text{cl}\left(\Sigma^n_{++}\right) = \Sigma^n_+$. Hence, $\tilde{X} \in \Sigma^n_+$, by Proposition 1.10 part (f). We proved $\inf_{S \in \Sigma^n_{++}} \langle \tilde{X}, S\rangle \geq 0$. Sending $\left\{S^{(k)}\right\} \subset \Sigma^n_{++}$ to the zero matrix, we see that the infimum is equal to zero. Therefore, $\sup_{S \in G_1} \langle \tilde{X}, S\rangle \leq 0$. This is equivalent to

$$\langle \tilde{X}, C\rangle - \langle \tilde{X}, \mathcal{A}^*(y)\rangle \leq 0, \text{ for every } y \in \mathbb{R}^m \text{ such that } b^T y \geq z^*.$$

Equivalently,

$$\mathcal{A}(\tilde{X})^T y \geq \langle C, \tilde{X}\rangle, \text{ for every } y \in \mathbb{R}^m \text{ such that } b^T y \geq z^*.$$

So, $\mathcal{A}(\tilde{X})^T y$ is bounded below on $\left\{y \in \mathbb{R}^m : b^T y \geq z^*\right\}$. For instance, LP duality implies $\mathcal{A}(\tilde{X}) = \alpha b$, for some $\alpha \geq 0$. If $\alpha = 0$ then $\mathcal{A}(\tilde{X}) = 0$ implies $\langle C, \tilde{X}\rangle \leq 0$. Since $\tilde{X} \in \Sigma^n_+$, $\tilde{X} \neq 0$, and there exists $\left(\bar{y}, \bar{S}\right) \in \mathbb{R}^m \oplus \Sigma^n$ such that $\mathcal{A}^*(\bar{y}) + \bar{S} = C$, $\bar{S} \succ 0$, we have

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

(we used Proposition 1.11 (f)) a contradiction. So, we can assume $\alpha > 0$. We define $\bar{X} := \frac{1}{\alpha}\tilde{X} \in \Sigma^n_+$. Also, $\mathcal{A}(\bar{X}) = b$. Hence, $\bar{X}$ is feasible in (P). Finally,

$$\begin{aligned} & \mathcal{A}(\tilde{X})^T y \geq \langle C, \tilde{X}\rangle, \quad \text{for all } y \in \mathbb{R}^m \text{ such that } b^T y \geq z^* \\ \Rightarrow \quad & \langle C, \bar{X}\rangle \leq z^*. \end{aligned}$$

Therefore, by the weak duality relation, $\bar{X}$ is an optimal solution of (P). □

Remark 2.15 Note that

(i) under the assumption that (D) has a Slater point, the above theorem implies that the set of optimal solutions of (P) is compact;
(ii) under the assumptions of the above theorem, the optimum value of (D) may not be attained even though it is guaranteed to exist;
(iii) this theorem and its proof generalize to the conic convex optimization setting.

Noting that the dual of (D) is equivalent to (P), we have the following corollary.

Corollary 2.16 *If (P) has a feasible solution and (D) has a Slater point then (P) attains its optimal objective value and the optimal objective values of (P) and (D) are the same.*

Corollary 2.17 *If (P) and (D) both have Slater points, then they both attain their optimal objective values and the optimal objective values of (P) and (D) coincide.*

Consider

$$\begin{array}{llrcl} (P) & \inf & \left\langle \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, X \right\rangle & & \\ & \text{subject to:} & \left\langle \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, X \right\rangle & = & 2, \\ & & X & \succeq & 0. \end{array}$$

Its dual is

$$\begin{array}{llrcl} (D) & \sup & 2y & & \\ & \text{subject to:} & \begin{pmatrix} 1 & -y \\ -y & 0 \end{pmatrix} & \succeq & 0. \end{array}$$

Then $X(\epsilon) := \begin{pmatrix} \epsilon & 1 \\ 1 & \frac{1}{\epsilon} \end{pmatrix}$ is feasible in (P) for every $\epsilon > 0$ and $\langle C, X(\epsilon)\rangle \to 0$ as $\epsilon \to 0$. However, $\langle C, X\rangle = 0$ implies $X_{11} = 0$ which in turn implies $X_{12} = X_{21} = 0$ (since X is positive semidefinite). This contradicts the equality constraint in the primal $X_{12} + X_{21} = 2$. Therefore, the optimal value of (P) is zero but it is not attained. For (D), we see that $\bar{y} := 0$ is the only feasible solution (hence is also optimal). For this example, $\bar{X} := I + \bar{e}\bar{e}^T$ is a Slater point for (P) and the optimal value of (D) is attained. Nevertheless, (D) does not have a Slater point and even though (P) has a well-defined optimal value, it is not attained.

Problem (P) is equivalent to $\inf\{X_{11} : X_{11}X_{22} \geq 1, X_{11} \geq 0, X_{22} \geq 0\}$. Even though the feasible region of this problem is a closed set, its projection onto $\begin{pmatrix} \mathbb{R} \\ 0 \end{pmatrix}$ is not.

Another example (for $\gamma > 0$):

$$C := \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, A_1 := \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, A_2 := \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix},$$

$$A_3 := \begin{pmatrix} 0 & -1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 2 \end{pmatrix}, b := \begin{pmatrix} 0 \\ 0 \\ 2\gamma \end{pmatrix}.$$

Note that $X^* := \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \gamma \end{pmatrix}$ is an optimal solution in (P) with $\langle C, X^* \rangle = 0$ and $y^* := \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix}$ is an optimal solution in (D) with $b^T y^* = -2\gamma$. Both optimal values are attained but there is a duality gap of 2γ. Of course, neither problem has Slater points.

We can also give examples of SDP problems where neither (P) nor (D) attains the common optimal value:

$$n := 4, m := 3, C := e_1 e_2^T + e_2 e_1^T + e_4 e_4^T,$$
$$A_1 := -e_1 e_1^T, A_2 := -e_2 e_2^T, A_3 := e_3 e_4^T + e_4 e_3^T,$$
$$b := [0, -1, 2]^T.$$

By modifying this example, we can also give a primal-dual pair such that there is a finite positive duality gap and neither problem attains its (finite) optimal value:

$$n := 5, m := 4, C := e_1 e_2^T + e_2 e_1^T + e_4 e_4^T + e_5 e_5^T,$$
$$A_1 := -e_1 e_1^T, A_2 := -e_2 e_2^T, A_3 := e_3 e_4^T + e_4 e_3^T, A_4 := e_1 e_3^T + e_3 e_1^T + e_5 e_5^T,$$
$$b := [0, -1, 2, 1]^T.$$

Another case that is impossible in the linear optimization setting but can happen in SDP is that (P) has a finite optimal value that is unattained and (D) is infeasible:

$$n := 4, m := 4, C := e_1 e_2^T + e_2 e_1^T,$$
$$A_1 := -e_1 e_1^T, A_2 := -e_2 e_2^T, A_3 := e_1 e_3^T + e_3 e_1^T, A_4 := e_1 e_4^T + e_4 e_1^T + e_3 e_3^T,$$
$$b := [0, -1, 1, 0]^T.$$

A useful and intriguing property of the LP problems is the following (Goldman-Tucker Theorem of strict complementarity):

Theorem 2.18 *Suppose (LP) and (LD) have feasible solutions. Then (LP) and (LD) have optimal solutions x^*, s^* respectively, such that*

$$x^* + s^* > 0.$$

This theorem does not generalize to SDP in its full power. There are examples of SDP problems which do not have the *strict complementarity property*. There are even examples with (P) and (D) both having Slater points as well as unique optimal solutions such that the strict complementarity property fails for (P) and (D).

Example 2.19 Consider the SDP problem with the following data:

$$C := \begin{pmatrix} -1 & -1 & -1 \\ -1 & 0 & 1 \\ -1 & 1 & 0 \end{pmatrix}, A_1 := \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, A_2 := \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix},$$
$$A_3 := \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, b := \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}.$$

This SDP problem has a unique primal optimal solution X^* and a unique dual optimal solution (y^*, S^*), there is no duality gap; however, $X^* + S^* \not\succ 0$.

We say that *strict complementarity fails* for a given SDP, if there does not exist a pair of optimal solutions $(\bar{X}, \bar{S})$ (for (P) and (D) respectively) satisfying $\bar{X} + \bar{S} \succ 0$.

It is worth noting that for the above SDP,

$$\bar{X} := \begin{bmatrix} 1 & -\frac{1}{4} & -\frac{1}{4} \\ -\frac{1}{4} & \frac{1}{2} & 0 \\ -\frac{1}{4} & 0 & \frac{1}{2} \end{bmatrix} \succ 0$$

is a Slater point for (P) and $\bar{y}_1 := -4$, $\bar{y}_2 := -3$, $\bar{y}_3 := -2$ yields

$$\bar{S} := C - \mathcal{A}^*(\bar{y}) = \begin{bmatrix} 3 & 1 & 2 \\ 1 & 3 & 1 \\ 2 & 1 & 2 \end{bmatrix} \succ 0$$

a Slater point for (D). Therefore, in the sense of the Strong Duality Theorem and Corollary 2.17, the example is well-behaved. In the domain of such SDP problems (when both problems (P) and (D) have Slater points), we can ask "*how common are the instances failing strict complementarity, like the one in the last example?*" While working on this question, it is convenient to parameterize the data in the space $\mathcal{GM}_m \oplus \Sigma^n_+ \oplus \Sigma^n_+$ representing the data space for $(\mathcal{A}, b, c)$. Here, $\mathcal{GM}_m$ denotes the (Grassmannian) manifold of all linear subspaces of dimension m in Σ^n (representing the range of the linear transformation $\mathcal{A}^*$) and we represent b by those $\tilde{X} \in \Sigma^n$ which lie in the affine space

$$\left\{ \tilde{X} \in \Sigma^n : \mathcal{A}(\tilde{X}) = b \right\};$$

we represent C by those $\tilde{S}$ which lie in the affine space

$$\left\{ \tilde{S} \in \Sigma^n : \tilde{S} = C - \mathcal{A}^*(y), \text{ for some } y \in \mathbb{R}^m \right\}.$$

Alizadeh, Haeberly and Overton [7] showed that the SDPs failing strict complementarity are not common at all. Their notion of the failure of strict complementarity is slightly different than ours; also they assume that the data satisfy the Slater condition for the primal and the dual (as we stated above); however, their results can be used to prove the following version of the theorem as well.

Theorem 2.20 *For every $m, n \in \mathbb{Z}_{++}$, the set of data points for SDP, in the space $\mathcal{GM}_m \oplus \Sigma^n_+ \oplus \Sigma^n_+$, for which strict complementarity fails has measure zero relative to the set of data points $(\mathcal{A}, b, c)$ for which there is no duality gap. In particular, for every $\epsilon > 0$, there exists a countable family of sets in $\mathcal{GM}_m \oplus \Sigma^n_+ \oplus \Sigma^n_+$, of total measure less than ϵ, which cover all the data points failing strict complementarity; moreover, the set of data points satisfying strict complementarity has a positive measure.*

This theorem generalizes to all convex optimization problems in conic form.

2.5 Linear Consequences, Proving Unboundedness, Strong Infeasibility Certificates

Consider

$$\begin{array}{llrcl} (P) & \inf & \left\langle \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, X \right\rangle & & \\ & \text{subject to:} & \left\langle \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, X \right\rangle & = & 1, \\ & & X & \succeq & 0. \end{array}$$

Its dual is

$$\begin{array}{llrcl} (D) & \sup & y & & \\ & \text{subject to:} & \begin{pmatrix} y & 0 \\ 0 & 0 \end{pmatrix} & \preceq & \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}. \end{array}$$

Let $X(\alpha) := \begin{pmatrix} 1 & \alpha \\ \alpha & \alpha^2 \end{pmatrix}$. $X(\alpha)$ is feasible for every $\alpha \in \mathbb{R}$. Therefore (P) is unbounded. For an unbounded LP problem in an analogous form, we have a nice proof of unboundedness (a feasible point and a direction d such that $Ad = 0$, $d \geq 0$, $c^T d < 0$). However, there is no such *linear unboundedness proof* for this example. That is, there does not exist $D \in \Sigma^n_+$ such that $\mathcal{A}(D) = 0$ and $\text{Tr}(CD) < 0$. Note that for this example, $\text{Tr}(A_1 D) = 0$ and $D \succeq 0$ imply $D_{11} = D_{12} = D_{21} = 0$ and $D_{22} \geq 0$. But for every such D, we must have $\text{Tr}(CD) = 0$. Therefore, no *LP-like* unboundedness proof can exist for unbounded SDPs in general.

Indeed, such an issue also arises in infeasibility detection. Consider the convex set in $\mathbb{R}^2$ defined by the constraint

$$\begin{pmatrix} 1 & X_{21} \\ X_{21} & X_{22} \end{pmatrix} \succeq 0.$$

This is the region:

$$\left\{ \begin{pmatrix} X_{21} \\ X_{22} \end{pmatrix} \in \mathbb{R}^2 : X_{22} \geq X_{21}^2 \right\}.$$

Let the objective function be $\inf\{2X_{21}\}$. Any ray of the form

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} + \begin{pmatrix} -\alpha \\ \beta \end{pmatrix}$$

for $\alpha > 0, \beta > 0$, must leave the parabolic region. Then (D) is infeasible, since the dual constraint

$$\begin{pmatrix} -y & 1 \\ 1 & 0 \end{pmatrix} \succeq 0$$

is impossible to satisfy. However, (D) is *almost feasible*. That is, for every $\epsilon > 0$, there exists $C' \in \Sigma^n$ such that $\|C - C'\| < \epsilon$ and $\mathcal{A}^*(y) \preceq C'$ is feasible.

Theorem 2.21 *Suppose $\mathcal{A} : \Sigma^n \to \mathbb{R}^m$ and $C \in \Sigma^n$ are given. Then*

(a) *if there exists $D \in \Sigma^n$ such that $D \succeq 0$, $\mathcal{A}(D) = 0$, $\text{Tr}(CD) < 0$ then there does not exist $y \in \mathbb{R}^m$ such that $\mathcal{A}^*(y) \preceq C$;*

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

Proof We can prove part (a) in a similar way to the proof of the weak duality relation. Suppose D is as promised and we have $\bar{y} \in \mathbb{R}^m$ such that $\mathcal{A}^*(\bar{y}) \preceq C$. Then

$$0 \leq \langle C, D\rangle - \langle \mathcal{A}^*(\bar{y}), D\rangle = \langle C, D\rangle - \bar{y}^T \mathcal{A}(D) = \langle C, D\rangle < 0.$$

We arrived at a contradiction. To prove part (b), we set up a primal-dual pair of SDPs and apply the Strong Duality Theorem. Consider

$$\begin{array}{llrcl} (D_1) & \sup & & \eta & \\ & \text{subject to:} & \mathcal{A}^*(y) + \eta I & \preceq & C, \\ & & \eta & \leq & 0. \end{array}$$

Its dual is

$$\begin{array}{llrcl} (P_1) & \inf & \langle C, X\rangle & & \\ & \text{subject to:} & \mathcal{A}(X) & = & 0, \\ & & \mathrm{Tr}(X) & \leq & 1, \\ & & X & \succeq & 0. \end{array}$$

Let $\bar{y} := 0$, $\bar{\eta} := -\|C\|_2 - 1$. This proves that (D_1) has a Slater point. $\bar{X} := 0$ is feasible in (P_1). Therefore the Strong Duality Theorem for SDP applies. (P_1) must attain its optimal value and the objective values are the same. Suppose there does not exist $D \in \Sigma^n$ such that $D \succeq 0$, $\mathcal{A}(D) = 0$, $\mathrm{Tr}(CD) < 0$. Then the optimal objective value of (P_1) is zero. By the Strong Duality Theorem, the optimal objective value of (D_1) is also zero. Either (D_1) attains this value (i.e., $\mathcal{A}^*(y) \preceq C$ has a solution) or it doesn't (but then there must exist a sequence $\left\{y^{(k)}, \eta_k\right\}$ such that

$$\mathcal{A}^*\left(y^{(k)}\right) + \eta_k I \preceq C, \quad \text{and } \eta_k \to 0^-).$$

Therefore, $\mathcal{A}^*(y) \preceq C$ is almost feasible. □

As a consequence, we have

Theorem 2.22 *There exists $D \in \Sigma^n$ such that $D \succeq 0$, $\mathcal{A}(D) = 0$, $\mathrm{Tr}(CD) < 0$ iff $\mathcal{A}^*(y) \preceq C$ is not almost feasible.*

Proof Part (b) of Theorem 2.21 establishes one direction. For the other direction, suppose there exists $D \in \Sigma^n$ such that $D \succeq 0$, $\mathcal{A}(D) = 0$, $\mathrm{Tr}(CD) < 0$. We may choose (in addition) D such that $\mathrm{Tr}(CD) = -1$. Then, for every $C' \in \Sigma^n$ such that $\|C - C'\| < \frac{1}{\|D\|}$, $\mathcal{A}^*(y) \preceq C'$ is infeasible (its infeasibility is certified by the same D above). Therefore, $\mathcal{A}^*(y) \preceq C$ is not almost feasible. □

An interpretation of the above theorem is that it characterized when

$$\mathcal{A}(D) = 0, D \succeq 0 \Rightarrow \langle C, D\rangle \geq 0.$$

Alternatively, it characterized when the linear inequality $\langle C, D\rangle \geq 0$ is a *consequence of the system* $\{\mathcal{A}(D) = 0, D \succeq 0\}$. The next fundamental question along these lines, which has ties to duality theory is:

When is a quadratic inequality a consequence of other quadratic inequalities?

Let us start with the simplest version:

When is a quadratic inequality a consequence of another?

The answer to this question is the so-called $\mathcal{S}$-Lemma:

Theorem 2.23 *Let $A, C \in \Sigma^n$. Assume $\lambda_1(A) > 0$. Then, the following are equivalent:*

(I) *for every* $h \in \mathbb{R}^n$, $h^T Ah \geq 0 \implies h^T Ch \geq 0$;
(II) *there exists* $y \geq 0$ *such that* $yA \preceq C$.

We can prove the above theorem, utilizing the duality theory developed so far.

Proof Suppose $A, C \in \Sigma^n$ with $\lambda_1(A) > 0$ such that (II) holds for $\bar{y} \in \mathbb{R}_+$. Then $\bar{y}A \preceq C$ and $\bar{y} \geq 0$ imply that for every $h \in \mathbb{R}^n$ with $h^T Ah \geq 0$, we have

$$0 \leq \bar{y} h^T Ah \leq h^T Ch \text{ and (I) follows.}$$

Suppose $A, C \in \Sigma^n$ with $\lambda_1(A) > 0$ such that (I) holds. Then

$$\min \left\{ h^T Ch : h^T Ah \geq 0 \right\} = 0. \tag{2.1}$$

The above minimum value is also equal to

$$\min \left\{ \langle C, X \rangle : \langle A, X \rangle \geq 0, X \succeq 0, \operatorname{rank}(X) \leq 1 \right\}.$$

If we can prove that the SDP relaxation of this problem (obtained by removing the rank restriction constraint) also has the optimal objective value equal to zero with dual attainment, then we are done. We have the SDP relaxation

$$(P) \min \left\{ \langle C, X \rangle : \langle A, X \rangle \geq 0, X \succeq 0 \right\}$$

ands its dual

$$(D) \max \left\{ 0 : yA \preceq C, y \geq 0 \right\}.$$

Since $\lambda_1(A) > 0$, (P) has a Slater point. So, if we prove that the objective function value of (P) is bounded from below, then by the Strong Duality Theorem, (D) has an optimal solution and the optimal objective values of (P) and (D) are the same. In particular, there exists $\bar{y} \geq 0$ such that $\bar{y}A \preceq C$. We will prove that the optimal objective value of (P) is zero, which will finish the proof. For a contradiction, suppose not, i.e., there exists $\bar{X} \succeq 0$ such that $\langle A, \bar{X} \rangle \geq 0$ and $\langle C, \bar{X} \rangle < 0$. Apply Theorem 1.8 to $\bar{X}$, let $r := \operatorname{rank}(\bar{X})$, and rewrite the current problem in the space Σ^r. So, we have $\hat{X} \in \Sigma^r_{++}$ such that $\langle \bar{A}, \hat{X} \rangle \geq 0$ and $\langle \bar{C}, \hat{X} \rangle < 0$. Note that $r \geq 2$ (if $r = 0$, then $\bar{X} = 0$, a contradiction to $\langle C, \bar{X} \rangle < 0$; if $r = 1$, then $\bar{X} = hh^T$ for some $h \in \mathbb{R}^n$ such that $h^T Ah \geq 0$ and $h^T Ch < 0$, a contradiction to (2.1)). Then, there exists $D \in \Sigma^r \setminus \{0\}$ such that $\langle \bar{A}, D \rangle = 0$ and $\langle \hat{X}, D \rangle = 0$. If $\langle C, D \rangle > 0$, replace D by $(-D)$ to make sure that $\langle C, D \rangle \leq 0$. Now, $\hat{X} + \alpha D$ satisfies

$$\langle \bar{A}, \hat{X} + \alpha D \rangle \geq 0, \text{ and } \langle \bar{C}, \hat{X} + \alpha D \rangle < 0, \text{ for every } \alpha \geq 0.$$

We choose the largest $\alpha > 0$ such that $\hat{X} + \alpha D \succeq 0$. Such α exists, since $D \neq 0$, $\hat{X} \succ 0$ and $\langle \hat{X}, D \rangle = 0$. Moreover, $\left(\hat{X} + \alpha D\right)$ has at most $(r-1)$ positive eigenvalues. So, we repeat this process until $r = 0$ or $r = 1$ at which time we arrive at a contradiction. Therefore, the result follows. □

Indeed, as we mentioned before the last theorem, a much more interesting question is:

When is the quadratic inequality $h^T Ch \geq 0$ a consequence of the system $h^T A_1 h \geq 0, h^T A_2 h \geq 0, \cdots, h^T A_m h \geq 0$?

The above question led to very interesting research results. See for instance, Ben-Tal and Nemirovski [34] and Pólik and Terlaky [291].

2.6 Slater Condition, Borwein-Wolkowicz approach

Suppose we know that there exists $\bar{X} \in \Sigma^n$ feasible in (P) and that (P) is not unbounded. Is it possible to define a dual that in addition to the usual properties we expect from primal-dual pairs, this pair will have the property that

$$\text{infimum} = \text{supremum?}$$

Of course, in addition we would like to guarantee that at least one of the problems (P) or (D) (especially the one for which we want an optimal solution), attains its optimal objective value. So, we want *strong duality* to hold. Let us introduce some geometric notions that will be useful in this theory.

Definition 2.24 Let $K \subseteq \mathbb{R}^d$ be a closed convex cone. A convex cone $G \subseteq K$ is a *face of* K if for every $u, v \in K$ such that $(u + v) \in G$, we have $u \in G$, $v \in G$.

A face G of K is *exposed* if there exists $a \in \mathbb{R}^d \setminus \{0\}$ such that

$$G = \{x \in K : \langle a, x\rangle = 0\} \text{ and } K \subseteq \left\{x \in \mathbb{R}^d : \langle a, x\rangle \le 0\right\},$$

i.e., G is the intersection of K with one of its supporting hyperplanes.

A face G of K is a *proper face of* K if

$$\{0\} \subset G \subset K.$$

Notice that our earlier definition of a face of a polyhedron is precisely the above definition of exposed faces. Indeed, if we were to define a face of a polyhedral cone using the above definition for a general convex cone, we could easily prove that every proper face of every polyhedral cone is exposed.

Theorem 2.25 (a) *Every nonempty face G of Σ^n_+ is characterized by a unique subspace $L \subset \mathbb{R}^n$ such that*

$$G = \left\{X \in \Sigma^n_+ : \text{Null}(X) \supseteq L\right\},$$

$$\text{relint}(G) = \left\{X \in \Sigma^n_+ : \text{Null}(X) = L\right\}.$$

(b) *Every proper face G of Σ^n_+ is exposed.*

(c) *The cone Σ^n_+ is* projectionally exposed. *That is, every nonempty face G of Σ^n_+ can be expressed as*

$$G = (I - Q)\Sigma^n_+(I - Q),$$

where $Q \in \Sigma^n$ is the projection onto the unique subspace L defining G.

Note that every proper face of Σ^n_+ is isomorphic to Σ^k_+ for some $k < n$. For instance, once we "find" the minimal face $\bar{G}$ containing the feasible region, we can apply a linear isomorphism (actually an automorphism of Σ^n_+ works) so that the image of $\bar{G}$ is

$$\left\{ \begin{pmatrix} \hat{X} & 0 \\ 0 & 0 \end{pmatrix} \in \Sigma^n : \hat{X} \in \Sigma^k_+ \right\}.$$

Now, we can rewrite (P) as

$$\begin{array}{lccc} (\tilde{P}) & \inf & \langle C, X\rangle & \\ & \text{subject to:} & \mathcal{A}(X) & = & b, \\ & & X & \in & \bar{G}. \end{array}$$

Clearly the feasible region and the set of optimal solutions of (P) and $(\tilde{P})$ are the same. But since the algebraic description is different, the (algebraic) dual is also different.

A *poor* way of seeing this is to find $\tilde{\mathcal{A}} : \Sigma^n \to \mathbb{R}^p$ such that the primal problem is equivalent to

$$\begin{array}{llrcl} (\tilde{P}) & \inf & \langle C, X \rangle & & \\ & \text{subject to:} & \mathcal{A}(X) & = & b, \\ & & \tilde{\mathcal{A}}(X) & = & 0, \\ & & X & \in & \Sigma^n_+. \end{array}$$

Then the dual is

$$\begin{array}{llrcl} (\tilde{D}) & \sup & b^T y & & \\ & \text{subject to:} & \mathcal{A}^*(y) + \tilde{\mathcal{A}}^*(\tilde{y}) & \preceq & C, \end{array}$$

whose feasible region (when projected onto the y variables) is potentially larger than that of (D). Therefore, adding redundant equations to the primal can only make the dual feasible region larger. While this makes no difference in the special case of LP, in nonlinear optimization (including SDP) there is a significant difference.

Let us consider another example:

$$C := \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, A_1 := \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, A_2 := \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, b := \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Then for every feasible solution of (D), we have $y_2 = 0$ and $y_1 \leq 0$. So, the set of feasible solutions of (D) (in the S-space) is

$$\left\{ \begin{pmatrix} 1 & 0 & 0 \\ 0 & S_{22} & 0 \\ 0 & 0 & 0 \end{pmatrix} : S_{22} \geq 0 \right\}.$$

Moreover, every feasible solution of (D) is optimal in (D). For the primal, it is also true that the set of feasible solutions and the set of optimal solutions coincide, which is (in the X-space):

$$\left\{ \begin{pmatrix} 1 & 0 & X_{31} \\ 0 & 0 & 0 \\ X_{31} & 0 & X_{33} \end{pmatrix} : X_{33} \geq X_{31}^2 \right\}.$$

Even though both (P) and (D) have finite optimum objective values, both of these values are attained, there is a duality gap of 1.

Now, suppose we add to (P) the following redundant constraint:

$$\langle A_3, X \rangle = b_3, \text{ where } A_3 := \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \text{ and } b_3 := 0.$$

The primal SDP problem (P) still has the same set of feasible solutions, the same set of optimal solutions and the same optimal value of 1. However, the dual constraints become:

$$\begin{pmatrix} (1 - y_2) & 0 & 0 \\ 0 & -y_1 & -(y_2 + y_3) \\ 0 & -(y_2 + y_3) & 0 \end{pmatrix} \succeq 0.$$

The latter constraint is equivalent to:

$$y_1 \leq 0, \quad y_2 \leq 1, \quad y_2 = -y_3.$$

Since we are maximizing y_2, the new dual has $y_2 = 1$ and $y_3 = -1$ at every optimal solution. The set of optimal solutions of the new dual is (in the S-space):

$$\left\{ \begin{pmatrix} 0 & 0 & 0 \\ 0 & S_{22} & 0 \\ 0 & 0 & 0 \end{pmatrix} : S_{22} \geq 0 \right\}.$$

Clearly, there is no duality gap between the primal and the new dual.

Borwein and Wolkowicz [59] showed how to arrive from (P) to $(\tilde{P})$ in finitely many (at most n) steps, in each step strictly reducing the dimension of the current cone. An insightful, geometric result in this theory is the following.

Lemma 2.26 *Let G be a proper face of Σ_+^n. Then $\left(\Sigma_+^n + G^\perp\right)$ is closed. However, $\left(\Sigma_+^n + \operatorname{span}(G)\right)$ is not closed.*

In the above result,

$$G^\perp := \{S \in \Sigma^n : \langle X, S\rangle = 0, \forall X \in G\},$$

and $\operatorname{span}(G)$ is the smallest linear subspace in Σ^n containing G.

To arrive at the minimal face $\bar{G}$, Borwein and Wolkowicz repeatedly utilize the following key lemma (which is a theorem of the alternative):

Lemma 2.27 *Let $\mathcal{A} : \Sigma^n \to \mathbb{R}^m$ be a linear transformation and let $b \in \mathbb{R}^m$ be given. Then, exactly one of the following two systems has a solution:*

(I) $\mathcal{A}(X) = b$, $X \in \Sigma_{++}^n$;

(II) $\mathcal{A}^*(y) \in \Sigma_{++}^n \setminus \{0\}$, $b^T y = 0$.

It follows from the above lemma that either (P) has a Slater point or we can find a nonzero, symmetric positive semidefinite matrix $\bar{S}$ which is in the range of $\mathcal{A}^*$ such that $\langle \bar{S}, X\rangle = 0$ for every feasible solution of (P). Suppose (P) does not have a Slater point. Then by Lemma 2.27, there exists $\bar{u} \in \mathbb{R}^m$ such that

$$\mathcal{A}^*(\bar{u}) \in \Sigma_{++}^n \setminus \{0\}, \text{ and } b^T \bar{u} = 0.$$

Let $\bar{S} := \mathcal{A}^*(\bar{u})$. Now, for every feasible X in (P), we have

$$\langle \bar{S}, X\rangle = \langle \mathcal{A}^*(\bar{u}), X\rangle = \bar{u}^T \mathcal{A}(X) = b^T \bar{u} = 0.$$

Moreover, since $\bar{S} \succeq 0$ and $\bar{S} \neq 0$, the set

$$\{X \in \Sigma_+^n : \langle \bar{S}, X\rangle = 0\}$$

is a proper face of Σ_+^n. We can replace (P) by an equivalent primal problem which explicitly includes the linear equation $\langle \bar{S}, X\rangle = 0$. If the new primal problem (now over Σ_+^k, for some $k < n$) does not have a Slater point then we apply Lemma 2.27 to the new problem. Since every proper face of Σ_+^n is isomorphic to Σ_+^k for $k < n$ and in each step we strictly reduce k, we are guaranteed to terminate with the minimal face in at most n steps.

Let us reconsider

$$\begin{array}{llrcl} (\tilde{P}) & \inf & \langle C, X\rangle & & \\ & \text{subject to:} & \mathcal{A}(X) & = & b, \\ & & X & \in & \bar{G}. \end{array}$$

Since $\bar{G}$ is the minimal face of Σ_+^n containing the feasible region, the Slater Condition holds for $(\tilde{P})$. So, the Strong Duality Theorem can be applied.

As SDP became re-popularized in 1990's, Ramana asked (and answered) the question of how to arrive at $(\tilde{P})$ in a way that is efficient from a computational complexity viewpoint. Given the problem

$$\begin{array}{llll} (D) & \sup & b^T y & \\ & \text{subject to:} & \mathcal{A}^*(y) & \preceq \; C, \end{array}$$

Ramana proposed the following dual (named Extended Lagrangian-Slater Dual):

$$\begin{array}{lllll} (ELSD) & \inf & \langle C, U+W \rangle & & \\ & \text{subject to:} & \mathcal{A}(U+W) & = & b, \\ & & W & \in & \mathcal{W}_n, \\ & & U & \succeq & 0. \end{array}$$

The set $\mathcal{W}_n$ admits an SDP representation as follows:

$$\begin{aligned} \mathcal{C}_k \quad := \quad & \big\{(U_1, W_1; U_2, W_2; \ldots; U_k, W_k) \in \Sigma^n \oplus \mathbb{R}^{n\times n} \oplus \cdots \oplus \Sigma^n \oplus \mathbb{R}^{n\times n} : W_0 := 0, \\ & \mathcal{A}(U_i + W_{i-1} + W_{i-1}^T) = 0, \langle C, U_i + W_{i-1} + W_{i-1}^T \rangle = 0, \\ & U_i \succeq W_i W_i^T, \forall i \in \{1, 2, \ldots, k\}\big\}, \end{aligned}$$

$$\mathcal{W}_k := \big\{W_k + W_k^T : (U_1, W_1; U_2, W_2; \ldots; U_k, W_k) \in \mathcal{C}_k, \text{ for some } U_1, W_1, \ldots, U_k\big\}.$$

Noticing the following Schur complement structure, it is easy to see that $\mathcal{C}_k$ is a convex set:

$$U_i \succeq W_i W_i^T \iff \begin{pmatrix} I & W_i^T \\ W_i & U_i \end{pmatrix} \succeq 0.$$

Now, suppose that $h \in \text{Null}(U_i)$. Then $U_i \succeq W_i W_i^T$ implies $h^T W_i W_i^T h \leq 0$, whence $h \in \text{Null}(W_i^T)$. We proved $\text{Null}(W_i^T) \supseteq \text{Null}(U_i)$. Note the connection to Theorem 2.25. The definitions and the above properties imply that $\mathcal{W}_k$ is a subspace.

Theorem 2.28 *If (D) has a finite optimal value then $(ELSD)$ is feasible and the optimal objective values of (D) and $(ELSD)$ coincide, $(ELSD)$ attains its optimal value.*

Recall Theorem 2.22. An alternative way of stating that theorem is

Let $\mathcal{A} : \Sigma^n \to \mathbb{R}^m$, $C \in \Sigma^n$ be given. Then exactly one of the following systems has a solution:

(I) *there exists $D \in \Sigma^n$ such that $D \succeq 0$, $\mathcal{A}(D) = 0$, $\text{Tr}(CD) < 0$;*

(II) *$\mathcal{A}^*(y) \preceq C$ is almost feasible.*

So, Theorem 2.22 generalized a theorem of the alternative to the SDP setting by keeping the infeasibility certificate "the same" and modifying (weakening) the requirements on the "solution" of the system $\mathcal{A}^*(y) \preceq C$. The next theorem is a complement of the Theorem 2.22 in the sense that it keeps the requirements on the "solution" y the same as it was in the polyhedral setting but modifies (weakens) the requirements on the infeasibility certificate D.

Theorem 2.29 *Let $\mathcal{A} : \Sigma^n \to \mathbb{R}^m$, $C \in \Sigma^n$ be given. Then exactly one of the following systems has a solution:*

(I) $\mathcal{A}^*(y) \preceq C$,

(II) $\mathcal{A}(U + W) = 0$, $W \in \mathcal{W}_n$, $U \succeq 0$, *and* $\langle C, U + W \rangle = -1$.

Now, we turn to the consequences of such results for the computational complexity of Semidefinite Optimization. When system (I) has a feasible solution, we can certify such fact by exhibiting a $\bar{y} \in \mathbb{R}^m$ such that $\mathcal{A}^*(\bar{y}) \preceq C$. If (I) does not have any solutions then by Theorem 2.29, we can certify such a fact by exhibiting a $\bar{U} \in \Sigma_+^n$ and a $\bar{W} \in \Sigma^n$ such that $\mathcal{A}(\bar{U}+\bar{W}) = 0$, $\bar{W} \in \mathcal{W}_n$, and $\langle C, \bar{U}+\bar{W}\rangle = -1$.

This implies the next two theorems:

Theorem 2.30 *In the* real number computational model*, the problem of deciding SDP feasibility is in* $\mathcal{NP} \cap \text{co-}\mathcal{NP}$*.*

Theorem 2.31 *In the Turing machine model, if SDP feasibility is in* $\mathcal{NP}$ *then it is in* $\mathcal{NP} \cap \text{co-}\mathcal{NP}$*.*

It turns out that the approach of Ramana is closely related to the work of Borwein and Wolkowicz. (The duals obtained are essentially the same, including those obtained for smaller k in the intermediate steps.)

The next example proves that in general, there are at least two steps required in the above ELSD. Consider

$$n := m := 3, \ C := 0,$$

$$A_1 := \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \ A_2 := \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \ A_3 := \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \ b := \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

Then (D) has optimal objective value zero and it is attained. However, (P) is infeasible. In ELSD with $k = 1$,
$\text{Tr}(A_1(U + W + W^T)) = 0$ implies $U_{22} = 0$. $\text{Tr}(A_3(U + W + W^T)) = 1$ implies $U_{12} = U_{21} = 1/2$. Since $U \succeq 0$, we have a contradiction. Therefore, ELSD is infeasible and at least two steps of the above procedure is required.

We can further generalize this example to generate a family of instances with $n = m$ which requires at least $n - 1 = m - 1$ steps (layers in ELSD): For every $n \geq 3$, let $m := n$, $C := 0$, $b := e_2$, $A_1 := e_1 e_1^T$, $A_2 := e_1 e_2^T + e_2 e_1^T$,

$$A_{2+i} := e_{i+1}e_{i+1}^T + e_1 e_{2+i}^T + e_{2+i}e_1^T, \forall i \in \{1, 2, \ldots, n-2\}.$$

The minimum number of layers required in ELSD (or equivalently the number of steps required by Borwein-Wolkowicz construction) is interesting elsewhere too. Sturm [338] calls this number *degree of singularity* and proves error bounds which are better for the smaller degrees of singularity.

2.7 When does the Slater Condition Hold in SDP Relaxations?

Suppose $c \in \mathbb{R}^n$ is given and we are interested in solving (and/or obtaining good lower bounds for) the following linear optimization problem over a **nonconvex** set of feasible solutions $F \subset \mathbb{R}^n$:

$$\inf \left\{ c^T x : \ x \in F \right\}. \tag{2.2}$$

First, we will focus on the above problem (2.2). However, given $c \in \mathbb{R}^n$, $C \in \Sigma^n$, we can also handle the nonconvex quadratic objective functions in our same framework below:

$$\inf \left\{ c^T x + x^T C x : \ x \in F \right\}. \tag{2.3}$$

2.7.1 Homogeneous Equality Form. For this subsection, we assume that the set of feasible solutions F and the linear transformation $\mathcal{A}$ 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 $\mathcal{A} : \Sigma^{n+1} \to \mathbb{R}^m$ is a linear transformation. We call this form of expressing the set of feasible solutions F, *the Homogeneous Equality Form.* Note that the name we gave to this form is based on the superficial fact that the right hand side is zero. Indeed, in the matrix variable we forced one of the entries to be 1 which allows us to express a wide range of problems in this form.

A very wide range of optimization problems (including some very difficult ones in finite dimensional spaces) can be put into this form. For instance, given $Q^{(i)} \in \Sigma^n$, $q^{(i)} \in \mathbb{R}^n$, $\gamma_i \in \mathbb{R}$, for $i \in \{1, 2, \ldots, m\}$,

$$\left\langle \begin{pmatrix} \gamma_i & \left(q^{(i)}\right)^T \\ q^{(i)} & Q^{(i)} \end{pmatrix}, \begin{pmatrix} 1 & x^T \\ x & xx^T \end{pmatrix} \right\rangle = 0, \quad i \in \{1, 2, \ldots, m\}$$

models a system of quadratic equations. We can also put every system of finitely many quadratic *inequalities* into the Homogeneous Equality Form:

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

Since any finite system of polynomial inequalities and equations can be expressed as a system of quadratic inequalities (with the use of new variables), we have

Proposition 2.32 *Every finite system of polynomial inequalities and equations can be put into Homogeneous Equality Form.*

Indeed, since the set of quadratic equations $x_j^2 - x_j = 0$, for every $j \in \{1, 2, \ldots, n\}$ models the constraint $x \in \{0, 1\}^n$, any combinatorial optimization problem as well as any mixed integer 0,1 programming problem can be put into Homogeneous Equality Form.

We define a relaxation in the lifted space:

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

That is, F is a projection onto the x-space of the intersection of $\hat{\mathcal{P}}$ with rank-1 matrices. Let's focus on another set in this lifted space:

$$\mathcal{F} := \operatorname{conv} \left\{ \begin{pmatrix} 1 & x^T \\ x & xx^T \end{pmatrix} \in \Sigma^{n+1} : x \in F \right\}.$$

Ideally, we would like to have a description of $\mathcal{F}$, since its projection onto the x-space yields

$$\left\{ x \in \mathbb{R}^n : \ \begin{pmatrix} 1 & x^T \\ x & X \end{pmatrix} \in \mathcal{F} \right\} = \operatorname{conv}(F). \tag{2.4}$$

To see this let $\bar{x} \in \operatorname{conv}(F)$. Then, by Carathédory's theorem, there exists $v^{(1)}, v^{(2)}, \ldots, v^{(n+1)} \in F$ and $\lambda \in \mathbb{R}_+^{n+1}$ such that

$$\bar{x} = \sum_{i=1}^{n+1} \lambda_i v^{(i)} \text{ and } \bar{e}^T \lambda = 1.$$

By the definition of $\mathcal{F}$, $\begin{pmatrix} 1 & \bar{x}^T \\ \bar{x} & \sum_{i=1}^{n+1} \lambda_i v^{(i)} \left(v^{(i)}\right)^T \end{pmatrix} \in \mathcal{F}$. Hence, $\bar{x}$ is in the set described in the right hand side of the above claimed equation (2.4). For the converse, let $\bar{x}$ be an element of the set in the right hand side of the above claimed equation (2.4). Then, by the definition of $\mathcal{F}$ and Carathédory's theorem, there exist $k = O\left(n^2\right)$, $v^{(1)}, v^{(2)}, \ldots, v^{(k)} \in F$, and $\lambda \in \mathbb{R}_+^k$ such that

$$\bar{x} = \sum_{i=1}^{k} \lambda_i v^{(i)} \text{ and } \bar{e}^T \lambda = 1.$$

Whence, $\bar{x} \in F$ as desired.

Clearly, we have $\hat{\mathcal{P}} \supseteq \mathcal{F}$. Therefore, $\hat{\mathcal{P}}$ is a relaxation of $\mathcal{F}$. Finally, note that

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

Hence, $\hat{\mathcal{P}}$ together with the above-mentioned objective function, gives a relaxation of the original nonconvex optimization problem (2.2). For the nonconvex optimization problems with a quadratic objective function as in form (2.3), we note

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

Next, we show that our SDP relaxation is guaranteed to behave well under the duality theory.

Theorem 2.33 *Suppose F and $\hat{\mathcal{P}}$ are given as in the Homogeneous Equality Form above and that* $\operatorname{conv}(F)$ *is full dimensional. Then, the Slater condition holds for $\hat{\mathcal{P}}$.*

Proof Suppose the feasible solution set F above is such that $\operatorname{conv}(F)$ is full dimensional. Then there must exist an affinely independent set of vectors $\{v^{(1)}, v^{(2)}, \ldots, v^{(n+1)}\} \subseteq F$. We then let

$$V_\lambda := \sum_{i=1}^{n+1} \lambda_i \begin{pmatrix} 1 \\ v^{(i)} \end{pmatrix} \begin{pmatrix} 1, & \left(v^{(i)}\right)^T \end{pmatrix},$$

for $\lambda > 0$ such that $\bar{e}^T \lambda = 1$. Then $V_\lambda \in \mathcal{F} \subseteq \hat{\mathcal{P}}$. Moreover, by the definition of affine independence,

$$\left\{\begin{pmatrix} 1 \\ v^{(1)} \end{pmatrix}, \begin{pmatrix} 1 \\ v^{(2)} \end{pmatrix}, \ldots, \begin{pmatrix} 1 \\ v^{(n+1)} \end{pmatrix}\right\} \text{ is linearly independent.}$$

Thus, by Proposition 1.11 part (c), we have $V_\lambda \in \left(\hat{\mathcal{P}} \cap \Sigma_{++}^n\right)$. Therefore, V_λ is a Slater point for $\hat{\mathcal{P}}$. □

In many applications, in natural domains for formulations of F, $\operatorname{conv}(F)$ may not be full-dimensional. Next, we address the more general situation by removing the assumption on the full dimensionality of $\operatorname{conv}(F)$.

2.7.1.1 *What if* $\operatorname{conv}(F)$ *is not full dimensional?* Suppose $\dim(\operatorname{conv}(F)) = d < n$. Then our last theorem no longer applies, also the proof technique of the theorem always generates singular positive semidefinite matrices V_λ. To remedy the situation, we do the obvious thing and focus on the smallest affine space containing $\operatorname{conv}(F)$. We find $L \in \mathbb{R}^{d\times n}$, $\ell \in \mathbb{R}^n$ such that $\operatorname{rank}(L) = d$ and

$$x \in F \Rightarrow x = \ell + L^T y \quad \text{for some } y \in \mathbb{R}^d.$$

Is it reasonable to assume that we know the $\dim(\operatorname{conv}(F))$ and L as well as ℓ? The answer from a theoretical worst-case viewpoint would be "no!" Of course, we can embed many hard optimization problems as feasibility problems and even just deciding whether $F = \emptyset$ could be hard enough (equivalently, deciding whether $\dim(F) \geq 0$ —note that $\dim(\emptyset) = -1$).

However, the answer from a practical viewpoint is "usually yes!" For example, in combinatorial optimization, one of the first things we do when we start studying a polytope is to determine its dimension. Next, we usually study its facets and for this purpose, if $\operatorname{conv}(F)$ is not full-dimensional, then we compute L and ℓ. Therefore, in many applications, $\dim(F)$, L and ℓ are readily available in the combinatorial optimization literature. In many other applications, these are easy to derive.

Let's define a linear transformation $\mathcal{L} : \Sigma^{n+1} \to \Sigma^{d+1}$,

$$\mathcal{L}(Z) := \begin{pmatrix} 1 & \ell^T \\ 0 & L \end{pmatrix} Z \begin{pmatrix} 1 & 0^T \\ \ell & L^T \end{pmatrix}.$$

This map $\mathcal{L}$ and its adjoint:

$$\mathcal{L}^*(W) = \begin{pmatrix} 1 & 0^T \\ \ell & L^T \end{pmatrix} W \begin{pmatrix} 1 & \ell^T \\ 0 & L \end{pmatrix}$$

are all we need to define new variables, to reduce dimension and most importantly to ensure the existence of Slater points for the transformed problem. Now, we can define $\bar{\mathcal{A}} : \Sigma^{d+1} \to \mathbb{R}^m$ which will take the role of the linear map $\mathcal{A}$ in the smaller dimensional space:

$$\bar{\mathcal{A}}(W) := \mathcal{A}\left(\mathcal{L}^*(W)\right).$$

With the above definitions, we can express the set of feasible solutions in an alternative form:

$$F = \left\{ \ell + L^T y : \ y \in \mathbb{R}^d, \ \bar{\mathcal{A}} \begin{pmatrix} 1 & y^T \\ y & yy^T \end{pmatrix} = 0 \right\}.$$

Based on the above construction, we can define a counterpart of F in the d-dimensional space:

$$F_{\mathcal{L}} := \left\{ y \in \mathbb{R}^d : \bar{\mathcal{A}} \begin{pmatrix} 1 & y^T \\ y & yy^T \end{pmatrix} = 0 \right\}.$$

Then in the lifted space, ideally we would like to get our hands on:

$$\mathcal{F}_{\mathcal{L}} := \operatorname{conv} \left\{ \begin{pmatrix} 1 & y^T \\ y & yy^T \end{pmatrix} \in \Sigma^{d+1} : \ y \in F_{\mathcal{L}} \right\}.$$

The latter leads to the relaxation

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

Theorem 2.34 *The Slater condition holds for* $\hat{\mathcal{P}}_{\mathcal{L}}$.

Proof By construction, $\dim(\operatorname{conv}(F_{\mathcal{L}})) = d$, i.e., $\operatorname{conv}(F_{\mathcal{L}})$ is full dimensional. Therefore, by Theorem 2.33, Slater condition holds for $\hat{\mathcal{P}}_{\mathcal{L}}$. □

Now that we obtained an SDP relaxation of the feasible solution set in the lifted space, we need to construct a linear objective function in this space which will lead to an SDP relaxation. First, we observe that $\inf\left\{c^T x + x^T C x : \; x \in F\right\}$ is equal to

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

By our construction of $\bar{\mathcal{A}}, L$, and ℓ the latter is also equal to

$$\inf\left\{c^T\left(\ell + L^T y\right) + \left(\ell + L^T y\right)^T C\left(\ell + L^T y\right) : y \in \mathbb{R}^d, \; \bar{\mathcal{A}}\begin{pmatrix} 1 & y^T \\ y & yy^T \end{pmatrix} = 0\right\}.$$

Using the definition of $F_{\mathcal{L}}$ we can rewrite the last infimum as

$$\inf\left\{\left\langle\mathcal{L}\begin{pmatrix} 0 & \frac{1}{2}c^T \\ \frac{1}{2}c & C \end{pmatrix}, \begin{pmatrix} 1 & y^T \\ y & yy^T \end{pmatrix}\right\rangle : \; y \in F_{\mathcal{L}}\right\}.$$

Now, recall the definition of $\mathcal{F}_{\mathcal{L}}$; this is the object we would ideally like to have our hands on, in the lifted space. Therefore, the last infimum is also equal to

$$\inf\left\{\left\langle\mathcal{L}\begin{pmatrix} 0 & \frac{1}{2}c^T \\ \frac{1}{2}c & C \end{pmatrix}, \begin{pmatrix} 1 & y^T \\ y & Y \end{pmatrix}\right\rangle : \; \begin{pmatrix} 1 & y^T \\ y & Y \end{pmatrix} \in \mathcal{F}_{\mathcal{L}}\right\}.$$

So, we have the following SDP relaxation:

$$\begin{array}{lrcl} \inf & \left\langle\mathcal{L}\begin{pmatrix} 0 & \frac{1}{2}c^T \\ \frac{1}{2}c & C \end{pmatrix}, \begin{pmatrix} 1 & y^T \\ y & Y \end{pmatrix}\right\rangle & & \\ \text{subject to:} & \bar{\mathcal{A}}\begin{pmatrix} 1 & y^T \\ y & Y \end{pmatrix} & = & 0, \\ & \begin{pmatrix} 1 & y^T \\ y & Y \end{pmatrix} & \succeq & 0. \end{array}$$

Moreover, it is guaranteed that the above problem has a Slater point.

Now, let us talk about L and ℓ. Note that in general, there may be infinitely many L and ℓ describing the (same) affine hull of F. So, we should use this flexibility to our advantage and choose those L and ℓ which help us in our applications. Usually, in many large-scale applications, the most important property to preserve is sparsity. I.e., if $\mathcal{A}$ has a very sparse representation and, the L and ℓ that we pick lead to a completely dense representation for $\bar{\mathcal{A}}$, we may be unhappy with the approach.

To make the discussion more concrete, let's move from the abstract representation of the linear map $\mathcal{A}$ to an explicit representation with respect to a chosen basis. Let $A_i \in \Sigma^n$ be such that $[\mathcal{A}(X)]_i = \langle A_i, X\rangle$ for $i \in \{1, 2, \ldots, m\}$. Then

$$\mathcal{L}(A_i) = \begin{pmatrix} 1 & \ell^T \\ 0 & L \end{pmatrix} A_i \begin{pmatrix} 1 & 0^T \\ \ell & L^T \end{pmatrix}, \quad i \in \{1, 2, \ldots, m\}.$$

So, now, assume that each A_i is very sparse. Then among all possible choices of L and ℓ, we want to pick the one that makes $\mathcal{L}(A_i)$ as sparse as possible for every i. For an application of this kind of idea, see [189].

2.7.2 Nonhomogeneous Equality Form. Let us consider a form more closely related to the form of popular SDP relaxations of the maximum cut and related combinatorial optimization problems. For this section, we assume that we have F given in *Nonhomogeneous Equality Form*; namely,

$$F = \left\{x \in \mathbb{R}^n : \mathcal{A}(xx^T) = b\right\},$$

where $\mathcal{A} : \Sigma^n \to \mathbb{R}^m$ is a given linear transformation as before and $b \in \mathbb{R}^m$ is also given. There is an obvious choice for an SDP relaxation:

$$\hat{\mathcal{P}} := \{X \in \Sigma^n : \ \mathcal{A}(X) = b, \ X \succeq 0\}.$$

Theorem 2.35 *Suppose there exists a linearly independent set of vectors* $\{v^{(1)}, v^{(2)}, \ldots, v^{(n)}\} \subseteq F$*, and* F *is given in the Nonhomogeneous Equality Form. Then* $\hat{\mathcal{P}}$ *has Slater points.*

Proof Suppose $\{v^{(1)}, v^{(2)}, \ldots, v^{(n)}\} \subseteq F$ is linearly independent. For every $\lambda \in \mathbb{R}^n_{++}$ such that $\bar{e}^T\lambda = 1$, define

$$V_\lambda := \sum_{i=1}^{n} \lambda_i v^{(i)} \left(v^{(i)}\right)^T.$$

Then

$$\mathcal{A}\left(V_\lambda\right) = \sum_{i=1}^{n} \lambda_i \mathcal{A}\left(v^{(i)} \left(v^{(i)}\right)^T\right) = \sum_{i=1}^{n} \lambda_i b = b.$$

Moreover, by Proposition 1.11 part (c), $V_\lambda \succ 0$. Therefore, $\hat{\mathcal{P}}$ has Slater points. □

What kind of objective functions are easily handled in this case? Here are some examples: If we have a combinatorial optimization problem which asks for a maximum (or minimum) cardinality or maximum (or minimum) weight set of certain type and the feasible solutions are encoded by 0,1 characteristic vectors, then the objective function $c^Tx \geq 0$ for every $x \in F$. Then, we can equivalently minimize (or maximize) $(c^Tx)^2$ over F, yielding the objective function $\langle cc^T, X\rangle$ for the SDP relaxation. Handling the pure quadratic objective functions is also straightforward. In this case, in problem (2.3), we have $c := 0$. The objective function x^TCx in (2.3) would be modeled by the objective function $\langle C, X\rangle$ for the SDP relaxation. The typical example is a way of obtaining the most popular SDP relaxation for Max Cut which can be obtained by using $\operatorname{diag}(xx^T) = \bar{e}$ for the general form $\mathcal{A}(xx^T) = b$ above.

This Nonhomogeneous Equality Form is also very general. We already noticed $\operatorname{diag}(xx^T) = \bar{e}$ could be a part of the constraints $\mathcal{A}(xx^T) = b$. Clearly, we can formulate systems of quadratic equations in which every monomial is degree two except the constant.

We can also formulate systems of quadratic inequalities in which every monomial is degree two except the constant by using additional "slack" variables:

$$\begin{aligned} \left\langle \begin{pmatrix} Q & 0 \\ 0^T & 1 \end{pmatrix}, \begin{pmatrix} xx^T & \tilde{s}x \\ \tilde{s}x^T & \tilde{s}^2 \end{pmatrix} \right\rangle &= \beta \\ \iff x^TQx + \tilde{s}^2 &= \beta \\ \iff x^TQx &\leq \beta. \end{aligned}$$

2.7.3 Homogeneous Convex Inequality Form. To generalize our setting to nonlinear constraints, we consider once more the representation of the set of feasible solutions in the lifted space:

$$\mathcal{F} := \operatorname{conv}\left\{ \begin{pmatrix} 1 & x^T \\ x & xx^T \end{pmatrix} \in \Sigma^{n+1} : x \in F \right\}.$$

Now, we want a map $\mathcal{A} : \Sigma^{n+1} \to \mathbb{R}^m$, not necessarily linear, such that

$$F = \left\{ x \in \mathbb{R}^n : \mathcal{A}\begin{pmatrix} 1 & x^T \\ x & xx^T \end{pmatrix} \leq 0 \right\}.$$

We also want $\mathcal{A}$ to have some form of convexity, namely

$$\mathcal{A}(\lambda U + (1-\lambda)V) \;\leq\; \lambda \mathcal{A}(U) + (1-\lambda)\mathcal{A}(V),$$

for every $U, V \in \Sigma_+^{n+1}$ and for every $\lambda \in [0,1]$. Finally, we want $\mathcal{A}$ to be good enough of a formulation of F so that it can identify the interior of convex hull of F:

$$\operatorname{int}(\operatorname{conv}(F)) = \operatorname{conv}\left\{ x \in \mathbb{R}^n : \; \mathcal{A}\begin{pmatrix} 1 & x^T \\ x & xx^T \end{pmatrix} < 0 \right\} \neq \emptyset.$$

Let's call such a form the *Homogeneous Convex Inequality Form.* Our construction above is only for F whose convex hull is full dimensional. If this assumption fails (i.e., $\dim(\operatorname{conv}(F)) = d < n$), then we first apply our technique from our earlier discussion, find the smallest affine space containing F, identify the best choices for L and ℓ and then continue as above in the transformed space.

Now, it is easy to propose a convex relaxation of the feasible solution set in the lifted space:

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

The corresponding convex relaxation problem is

$$\begin{array}{lrcl} \inf & \left\langle \begin{pmatrix} 0 & \frac{1}{2}c^T \\ \frac{1}{2}c & C \end{pmatrix}, \begin{pmatrix} 1 & x^T \\ x & X^T \end{pmatrix} \right\rangle & & \\ \text{subject to:} & \mathcal{A}\begin{pmatrix} 1 & x^T \\ x & X \end{pmatrix} & \leq & 0, \\ & \begin{pmatrix} 1 & x^T \\ x & X \end{pmatrix} & \succeq & 0. \end{array}$$

The above problem is not necessarily an SDP since $\mathcal{A}$ is not necessarily a linear map. However, this convex relaxation and/or its dual may still be useful in some applications.

In this general setting, we can construct Slater points for the above convex optimization problem as follows: Since we may assume that $\operatorname{conv}(F)$ is full dimensional, there exists an affinely independent set $\left\{v^{(1)}, v^{(2)}, \ldots, v^{(n+1)}\right\} \subseteq F$. We define V_λ as in the proof of Theorem 2.33. Then, it is easy to check that $\mathcal{A}(V_\lambda) \leq 0$ and by construction, $V_\lambda \succ 0$.

Also, for every $\lambda \in \mathbb{R}^n_{++}$ such that $\bar{e}^T\lambda = 1$, we have

$$v_\lambda := \sum_{i=1}^{n+1} \lambda_i v^{(i)} \in \operatorname{int}(\operatorname{conv}(F)).$$

Moreover, by definition,

$$\mathcal{A}\left(\begin{pmatrix} 1 \\ v_\lambda \end{pmatrix}\begin{pmatrix} 1, & v_\lambda^T \end{pmatrix}\right) < 0.$$

Now, we can define a feasible point for the convex relaxation which will be easily shown to be a Slater point:

$$\bar{V}_\lambda := \frac{1}{2}V_\lambda + \frac{1}{2}\begin{pmatrix} 1 \\ v_\lambda \end{pmatrix}\begin{pmatrix} 1, & v_\lambda^T \end{pmatrix}.$$

Clearly, $\bar{V}_\lambda \succ 0$ and $\mathcal{A}\left(\bar{V}_\lambda\right) < 0$. Therefore, $\bar{V}_\lambda$ is a Slater point for the above given convex relaxation. We showed:

Theorem 2.36 *Suppose F, $\mathcal{A}$ are such that F can be expressed in the Homogeneous Convex Inequality Form based on $\mathcal{A}$ and conv(F) is full dimensional. Then, the Slater condition holds for $\hat{\mathcal{P}}$.*

2.7.4 Adding Inequalities to a Convex Relaxation. In many applications, we may want to utilize SDP relaxations in a branch-and-bound or branch-and-cut scheme. In some other applications, we may want to optimize over the x-space projection of the SDP relaxation intersected with a polyhedral relaxation (which itself may have exponentially many facets but admit an efficient separation oracle). Such situations usually require adding valid linear inequalities for F in the space of the SDP relaxations. In some settings, we may have quadratic inequalities.

Suppose $Q \in \Sigma^n$, $q \in \mathbb{R}^n$, $\gamma \in \mathbb{R}$ are given so that the inequality

$$x^TQx + 2q^Tx + \gamma \le 0$$

is a valid inequality for F (perhaps violated by some points in the current convex relaxation that we have). Then the constraint

$$\left\langle \begin{pmatrix} \gamma & q^T \\ q & Q \end{pmatrix}, \begin{pmatrix} 1 & x^T \\ x & X \end{pmatrix} \right\rangle \le 0$$

can be added to the SDP relaxation or the current convex relaxation (since it is valid for $\mathcal{F}$). The new relaxation, after the addition of the above valid inequality, still has V_λ (in the previous subsection, $\bar{V}_\lambda$) as a Slater point.

2.8 Bibliographical Notes

Duality theory is extremely important and useful in optimization. Most books on optimization contain significant amount of duality theory or its applications. See for instance, Rockafellar [305], Rockafellar and Wets [306], Barvinok [24], Ben-Tal and Nemirovski [34], Borwein and Lewis [55], Nesterov and Nemirovski [275], the Handbook of SDP [366]. Also see Borwein and Wolkowicz [57], Duffin [109], Kretschmer [215], Ramana [299], Ramana, Tunçel, Wolkowicz [300], Shapiro [323], Tunçel [347], and Wolkowicz [365]. For applications of our approach in Section 2.7, see Benson and Ye [33], Wolkowicz and Zhao [367] and Zhao, Karisch, Rendl and Wolkowicz [376].

Many of the examples we used are well-known and can be found in the Handbook of SDP [366], Alizadeh, Haeberly and Overton [7], Freund [128], Luo, Sturm and Zhang [248], Ramana [299] and [349].

For various properties of Σ_+^n from a linear algebraic viewpoint, see Barker [21], Barker and Carlson [22], Hill and Waters [173], Lim [229], Loewy and Schneider [236], Saunders and Schneider [312] and the references therein.

As we pointed out in Lemma 2.26, there is a very fundamental connection between strong duality and the closure of the Minkowski sum of a linear subspace and a convex cone. Such connections go back at least to Guignard [156]. Also see some more recent related work in [58, 59, 56, 40, 41, 335, 126, 290, 54].

Convexity and duality are covered in the above-mentioned references as well as in many others in many different contexts; see, Berman [38], Berman and Plemons [39], Marshall and Olkin [249], Ben-Israel, Ben-Tal and Zlobec [37], Brøndsted [67], Anderson and Nash [9], Hiriart-Urruty and Lemaréchal [174, 175], Webster [363], Schneider [315], Nesterov [272].

Separation theorems (either separating two disjoint convex sets, or a point from a convex set) play a central role in duality theory as well as in convex analysis and optimization. The origins of such theorems go back at least to Minkowski (the late 1800's and the early 1900's). There are many versions of these theorems in many different settings. See for instance, Klee [199, 198, 197, 196], Gale and Klee [136], Klee, Maluta and Zanco [200].

The work on S-Lemma goes back at least to Yakubovich [370]; for more recent work, see the survey [291] and Jeyakumar, Lee and Li [182].

Theorem 2.33 (Theorem 3.1 in [347]) is related to Lemma 3 of Mitchell [256]. The strict complementarity result we quoted for SDP (Theorem 2.20) is due to Alizadeh, Haeberly and Overton [7]. Their paper also contains results pertaining to degeneracy/nondegeneracy as well as the uniqueness of optimal solutions in SDPs. Shortly after that, these results were generalized to all convex optimization problems in conic form [288]. For some related results in the setting of *nonlinear* SDP problems, see Shapiro [323]. For connections between the positive duality gaps for the pair (P), (D) and the failure of strict complementarity in the homogenizations of (P) and (D) see [349].