A Brief Introduction to Duality Theory

Paul J. Atzberger

Comments should be sent to:
atzberg@math.ucsb.edu

Abstract

These notes give an introduction to duality theory in the context of linear and positive semidefinite programming. These notes are based on material from *Convex Analysis and Nonlinear Optimization* by Borwein and Lewis and *Numerical Optimization* by Nocedal and Wright. Two examples are given to show how duality can be used. The first optimization application is to find the matrix in an affine family that has the least 2-norm. This problem can be formulated as the dual to a positive semi-definite program. The second problem is to find for an empirical correlation matrix $\tilde{Q}$ the closest positive semidefinite matrix Q with respect to the Frobenius norm. This problem can be formulated as a positive semidefinite program. We also derive it's dual problem. Lastly, we briefly discuss some numerical methods useful in solving these constrained optimization problems.

Cones

An important construct that arises in optimization problems is a cone K. Cones are often used to formulate constraints in optimization problems and have some important properties with respect to the dual. We review here some basic definitions and results concerning cones.

Definition 1 *A cone K is a set invariant under multiplication by non-negative scalars. If $x \in K$ and $\lambda \geq 0$ then $\lambda x \in K$.*

Definition 2 *The dual of a cone is*

$$Dual(K) = \{z \in R^n | z^T x \geq 0\}$$

Three important cones are:

Definition 3 *Quadratic Cone:*

$$K_q = \{z \in R^m | \ \|(z_2, \ldots, z_m)\|_2 \leq z_1\}$$

Definition 4 *Positive Orthant Cone:*

$$K_+ = \{z \in R^m | \ z_1 \geq 0, z_2 \geq 0, \ldots, z_m \geq 0\}$$

Definition 5 *Positive Semidefinite Cone:*

$$K_{S+} = \{X \in S^{n \times n} | \ X \succeq 0\}$$

Each of these cones is self-dual:

Lemma 1 *The positive orthant cone K_+ in R^m is equal to its dual cone:*
Proof:
We will show that $z^T x \geq 0 \ \forall x \in K_+ \iff z_i \geq 0 \ \forall i$.

If $z^T x \geq 0 \ \forall x \in K_+$ then since $e_i \in K$ we have that $z_i \geq 0 \ \forall i$. If $z_i \geq 0 \ \forall i$ then since any $x \in K_+$ has $x_i \geq 0 \ \forall i$ we have that $z^T x \geq 0$. Therefore $z \in \text{Dual}(K_+) \iff z \in K_+$. ■

Lemma 2 *The positive semidefinite cone K_{S+} in $S^{n\times n}$ is equal to its dual cone:*
Proof:
We will show that $Z \cdot X \geq 0 \; \forall X \in K_{S+} \iff Z \in K_{S+}$.

If $Z \cdot X \geq 0 \; \forall X \in K_{S+}$ then since $Z \in S^{n\times n}$ we can choose a basis so that Z is diagonal. Let $X = e_k e_k^T$ then $X \in K_{S+}$ and $Z \cdot X = Z_{k,k} \geq 0$. This requires that Z be positive semidefinite and $Z \in K_{S+}$.

Let $Z \in K_{S+}$ then by definition of positive semidefiniteness $v^T Z v \geq 0, \forall v$. In particular $e_k^T Z e_k = Z_{k,k} \geq 0$. If for any $X \in K_{S+}$ we represent matrices in the eigenbasis of X we obtain ($\mu_k \geq 0$ eigenvalues of X):

$$\begin{aligned} Z \cdot X &= \mu_1 Z_{1,1} + \mu_2 Z_{2,2} + \cdots + \mu_n Z_{n,n} \\ &\geq \min\{\mu_k\} \cdot \mathrm{tr}(Z) \\ &\geq 0 \end{aligned}$$

Therefore $\mathrm{dual}(K_{S+}) = K_{S+}$.
■

Lemma 3 *The quadratic cone K_q in R^{m+1} is equal to its dual cone:*
Proof:

We first will show that the quadratic code is contained in the dual cone. Since all vectors in the quadratic cone have first component non-negative we have:

$$\begin{aligned} & z^T x \geq 0 \\ \iff & z_1 x_1 \geq -z_2 x_2 - \cdots - z_{n+1} x_{n+1} \end{aligned}$$

$$\begin{aligned} -z_2 x_2 - \cdots - z_{n+1} x_{n+1} &\leq |(z_2, \ldots, z_{n+1})^T \cdot (x_2, \ldots, x_{n+1})| \\ &= \|(z_2, \ldots, z_{n+1})\| \cdot \|(x_2, \ldots, x_{n+1})\| \cdot |cos(\theta)| \\ &\leq z_1 x_1 \end{aligned}$$

Thus $K \subseteq \mathrm{Dual}(K)$.

Next we show that the dual cone is contained in the quadratic cone. If $z \in \mathrm{Dual}(K)$ then for all $x \in K$:

$$\begin{aligned} z_1 x_1 &\geq -z_2 x_2 - \cdots - z_{n+1} x_{n+1} \\ &= -(z_2, \ldots, z_{n+1})^T \cdot (x_2, \ldots, x_{n+1}) cos(\theta) \\ &= -\|(z_2, \ldots, z_{n+1})\| \cdot \|(x_2, \ldots, x_{n+1})\| cos(\theta) \end{aligned}$$

Now choose the member x^* of K so that $|x_1^*| = 1$, $\mathrm{sign}(x_1^*) = \mathrm{sign}(z_1)$, $\|(x_2^*, \ldots, x_{n+1}^*)\| = 1$ and $cos(\theta) = -1$ then $z_1 x_1 = |z_1|$ and we obtain:

$$|z_1| \geq \|(z_2, \ldots, z_{n+1})\|$$

Thus $z \in K$ and $K \supseteq \text{Dual}(K)$.

Therefore $K = \text{Dual}(K)$.

■

Definition 6 *The indicator function for a set U denoted δ_U is defined as*

$$\delta_U(x) = \left\{ \begin{array}{ll} 0 & \text{if } x \in U \\ \infty & \text{if } x \notin U \end{array} \right\}$$

Definition 7 *The Fenchel Conjugate of a function $f(x)$ is*

$$f^*(\phi) = \sup_{x \in R^n} \{< \phi, x > - f(x)\}$$

Lemma 4 *If K is a cone and $\delta_K(x)$ is the indicator function the Fenchel Conjugate of the indicator function is $\delta_K^*(\phi) = \delta_{-Dual(K)}(\phi)$.*

Proof:

$$\begin{aligned} \delta_K^*(\phi) &= \sup_{x \in R^n} \{< \phi, x > - \delta_K(x)\} \\ &= \sup_{x \in K} \{< \phi, x >\} \\ &= \left\{ \begin{array}{ll} \infty & \text{if } \exists x \in K \text{ s.t. } < \phi, x >> 0 \\ 0 & \text{if } < \phi, x > \leq 0, \, \forall x \in K \end{array} \right. \\ &= \delta_{-Dual(K)}(\phi) \end{aligned}$$

■

Linear and Semidefinite Programs

Many optimization problems can be arranged into a standard form. Below we discuss some important standard forms that can arise.

Optimization problems involving vectors with linear inequality and equality constraints and a linear objective function can be formulated as:

LP Program:

$$\inf_{x \in R^n} \{< c, x > | x \geq 0, Ax = b\}$$

This optimization problem is referred to as a *linear program.*

To formulate inequalities a *slack variable* z can be introduced and the linear inequality can be formulated as an equality. For example the inequality

$$ax_1 \geq 2$$

can be formulated as

$$\begin{aligned} z &\geq 0 \\ ax_1 - z &= 2 \end{aligned}$$

Optimization problems involving matrices with linear equality constraints, linear objective function, and inequality constraints involving the definiteness of the matrix can often be formulated as:

SDP Program:

$$\inf_{X \in S^{n \times n}} \{C \cdot X | X \succeq 0, A_i \cdot X = b_i\}$$

This optimization problem is referred to as a *semidefinite program*.

The dot product refers to vector dot product regarding the entries of each matrix as a vector $A \cdot B = \text{tr}(AB)$.

Duality

The central result we will use in the derivation of the dual problem is the Fenchel-Young Inequality.

Theorem 1 (Fenchel-Young Inequality) *Let* $f^*(\phi) = \sup_{x \in R}\{< \phi, x > -f(x)\}$. *This is called the Fenchel Conjugate and the following inequality holds:*

$$f(x) + f^*(\phi) \geq < \phi, x >$$

Equality holds if ϕ *is a subgradient of* f *at* x, $\phi \in \partial f(x)$.

Proof:

$$\begin{aligned} < \phi, x > -f(x) &\leq f^*(\phi) \\ \Rightarrow f(x) + f^*(\phi) &\geq < \phi, x > \end{aligned}$$

If ϕ is a subgradient of f at x then:

$$\begin{aligned} f(x) + < \phi, (y - x) > &\leq f(y) \\ \Rightarrow f(x) - < \phi, x > &\leq f(y) - < \phi, y > \\ \Rightarrow f(x) - < \phi, x > &\leq -f^*(\phi) \\ \Rightarrow f(x) + f^*(\phi) &\leq < \phi, x > \end{aligned}$$

Therefore:

$$f(x) + f^*(\phi) = < \phi, x >$$

■

The Fenchel-Young Inequality has a geometric interpretation if we view the graph of $f(x)$ as a surface over the x hyperplane in the z-direction. We can rearrange the inequality to become:

$$< \phi, x > -f^*(\phi) \leq f(x)$$

The inequality then states that out of all hyperplanes with slope vector ϕ the greatest z-intercept attainable while still remaining below the surface is $-f^*(\phi)$.

Theorem 2

$$\inf_{x\in R}\{f(x)\} = \inf_{x\in R}\sup_{\phi\in R}\{<\phi, x> - f^*(\phi)\}$$

Proof:
By the Fenchel-Young Inequality:

$$\begin{aligned} f(x) &\geq \sup_{\phi\in R}\{<\phi, x> - f^*(\phi)\} \\ \Rightarrow \inf_{x\in R}\{f(x)\} &\geq \inf_{x\in R}\sup_{\phi\in R}\{<\phi, x> - f^*(\phi)\} \end{aligned}$$

For the other inequality we use:

$$f^*(0) = \sup_{x\in R}\{-f(x)\} = -\inf_{x\in R}\{f(x)\}$$

Then

$$\sup_{\phi\in R}\{<\phi, x> - f^*(\phi)\} \geq \inf_{x\in R}\{f(x)\}$$

■

This consequence of the Fenchel-Young Inequality offers an alternative formulation an optimization problem.

Geometrically this can be thought of as the following procedure to find the minimum:
1. At a given x, find for all the hyperplanes that lie below the surface the one that is closest to attaining f(x) at x, denote the value at x by g(x). (If no one hyperplane exists then use the monotonic sequence of approximations to the value of f(x) at x to define g(x))
2. Minimize the function g(x).

Below are some examples using the alternative approach to find the minimizer. For the first example we demonstrate that the alternative problem amounts to the same thing as the primal problem.

Example 1

$$\inf_{x\in R^2}\{x^2+1\}$$

The Fenchel Conjugate is

$$\begin{aligned} f^*(\phi) &= \sup_{x\in R}\{\phi x - x^2 - 1\} \\ &= \sup_{x\in R}\{-(x-\frac{\phi}{2})^2 + \frac{\phi^2}{4} - 1\} \\ &= \frac{\phi^2}{4} - 1 \end{aligned}$$

This gives for the dual formulation of the problem:

$$\begin{aligned}
\inf_{x\in R}\sup_{\phi\in R}\{\phi x - \frac{\phi^2}{4} + 1\} &= \inf_{x\in R}\sup_{\phi\in R}\{-\frac{1}{4}(\phi - 2x)^2 + x^2 + 1\} \\
&= \inf_{x\in R}\{x^2 + 1\} \\
&= 1
\end{aligned}$$

■

Example 2 The alternative formulation of the problem can sometimes lead to a problem with poor properties. One example were this occurs is the following:

$$f(x) = \left\{ \begin{array}{ll} \ln(1+x) & \text{if } x \geq 0 \\ \ln(1-x) & \text{if } x \leq 0 \end{array} \right\}$$

The minimizer is at $x = 0$. Since the tangent lines to $\ln(1 + x)$ have slopes $\frac{1}{1+x} \to 0$ as $x \to \pm\infty$ we see no line other than the one with slope 0 lies below the graph of $f(x)$. This means $f^*(\phi) = \infty$ for $\phi \neq 0$ and $f^*(0) = 0$. The new objective function inside the infimum becomes discontinuous and infinity everywhere except at the solution point. In fact even formulating the problem which involves finding the Fenchel Conjugate (z-intercept) is the same as having found the minimizer of the problem.

■

While the alternative formulation in the examples above might not offer much of an advantage. We will see below that there are cases in which the alternative form does offer some advantages after an appropriate change of variable in ϕ. One important advantage is that it is often possible to turn equality constraints into inequality constraints under an appropriate change of variable. In addition many practical problems can be formulated as the alternative problem associated with one of the standard forms above.

We refer to the alternative formulation of the minimization problem as the *dual problem*. The original minimization problem is referred to as the *primal problem*.

We now derive the dual problem for linear programming.

LP Primal:

$$\inf_{x\in R^n}\{< c, x > | x \geq 0, Ax = b\}$$

LP Dual:

$$\sup_{y\in R^m}\{< y, b > | A^T y - c \leq 0\}$$

Derivation:

Let

$$\begin{aligned}
V &= \{x \in R^n | Ax = b, A \in R^{m\times n}\} \\
K_+ &= \{x \in R^n | x \geq 0\}
\end{aligned}$$

Then

$$\begin{aligned} f^*(\phi) &= \sup_{x\in R^n}\{<\phi - c, x> - \delta_{K_+}(x) - \delta_V(x)\} \\ &\leq \sup_{x\in R^n}\{<\phi - c, x> - \delta_{K_+}(x)\} + \sup_{x\in R^n}\{-\delta_V(x)\} \\ &= \delta_{-K_+}(\phi - c) - \inf_{x\in R^n}\{\delta_V(x)\} \end{aligned}$$

If we make a change of variable:

$$\phi = A^T y$$

The alternative formulation becomes:

$$\begin{aligned} \inf_{x\in R^n}\{f(x)\} &= \inf_{x\in R^n}\sup_{\phi\in R^m}\{<\phi, x> - f^*(\phi)\} \\ &\geq \inf_{x\in R^n}\sup_{y\in R^m}\{<y, b> - f^*(A^T y)\} \\ &\geq \sup_{y\in R^m}\{<y, b> - \delta_{-K_+}(A^T y - c) + \inf_{x\in R^n}\{\delta_V(x)\}\} \\ &\geq \sup_{y\in R^m, A^T y - c\leq 0}\{<y, b>\} \end{aligned}$$

■

Next we find the dual problem for semidefinite programming.

SDP Primal:

$$\inf_{X\in S^{n\times n}}\{C\cdot X | X\succeq 0, A_i\cdot X = b_i\}$$

SDP Dual:

$$\sup_{y\in R^{m+1}, C-\sum_{i=1}^{m+1} y_i A_i\succeq 0}\{y\cdot b\}$$

Derivation of dual:
Let

$$\begin{aligned} K_{S+} &= \{X\in S^{n\times n} | X\succeq 0\} \\ V &= \{X\in S^{n\times n} | A_i\cdot X = b_i\} \end{aligned}$$

By the theorems above:

$$\begin{aligned} &\inf_{X\in S^{n\times n}}\{C\cdot X | X\succeq 0, A_i\cdot X = b_i\} \\ = &\inf_{X\in S^{n\times n}}\{C\cdot X + \delta_K(X) + \delta_V(X)\} \\ = &\inf_{X\in S^{n\times n}}\sup_{\Phi\in R^{n\times n}}\{\Phi\cdot X - f^*(\Phi)\} \end{aligned}$$

The Fenchel Conjugate of $f(X) = C \cdot X + \delta_K(X) + \delta_V(X)$ is by the lemmas above:

$$
\begin{aligned}
f^*(\Phi) &= \sup_{X \in S^{n\times n}} \{(\Phi - C)\cdot X - \delta_{K_{S+}}(X) - \delta_V(X)\} \\
&\le \delta_{-\mathrm{Dual}(K)}(\Phi - C) - \inf_{X \in S^{n\times n}} \{\delta_V(X)\} \\
&= \delta_{-K}(\Phi - C) - \inf_{X \in S^{n\times n}} \{\delta_V(X)\}
\end{aligned}
$$

Under the change of variable:

$$\Phi = \sum_{i=1}^{m+1} y_i A_i$$

We obtain:

$$
\begin{aligned}
& \inf_{X\in S^{n\times n}, A_i\cdot X = b_i} \sup_{\Phi \in R^{n\times n}} \{\Phi \cdot X - f^*(\Phi)\} \\
\ge & \sup_{y\in R^{m+1}} \{\sum_{i=1}^{m+1} y_i b_i - f^*(\sum_{i=1}^{m+1} y_i A_i)\} \\
\ge & \sup_{y\in R^{m+1}} \{y\cdot b - \delta_{-K}(\sum_{i=1}^{m+1} y_i A_i - C) + \inf_{X\in S^{n\times n}} \{\delta_V(X)\}\} \\
\ge & \sup_{y\in R^{m+1}, C - \sum_{i=1}^{m+1} y_i A_i \succeq 0} \{y \cdot b\}
\end{aligned}
$$

■

By examining the derivations above we obtain the following theorem.

Theorem 3 *(Weak Duality) For the linear and positive semidefinite program if we let p denote the solution value obtained in the primal problem and d denote the solution value obtained in the dual problem. Then*

$$d \le p$$

Proof:(see the derivations of the dual problem) ■

Under addition conditions on the linear and semi-definite programs it is possible to show that the primal and dual problems obtain the same values $p = d$. See reference [1] for the Strong Duality Theorem.

Two Applications of SDP Programming and Duality

Before discussing the applications we first prove a theorem that will be useful.

Theorem 4 *Let M be any $m \times n$ matrix. Let*

$$C = \begin{bmatrix} 0 & M \\ M^T & 0 \end{bmatrix}$$

The non-negative eigenvalues of C are the singular values of M.

Proof:
Every matrix has a SVD decomposition. For M we have:

$$WMV^* = \Sigma$$

Let

$$U = \frac{1}{\sqrt{2}} \begin{bmatrix} W^* & -W^* \\ V^* & V^* \end{bmatrix}$$

then

$$\begin{aligned} UU^* &= \frac{1}{2} \begin{bmatrix} W & V \\ -W & V \end{bmatrix} \cdot \begin{bmatrix} W^* & -W^* \\ V^* & V^* \end{bmatrix} \\ &= \frac{1}{2} \begin{bmatrix} WW^* + VV^* & -WW^* + VV^* \\ -WW^* + VV^* & WW^* + VV^* \end{bmatrix} \\ &= \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix} \end{aligned}$$

and

$$\begin{aligned} UCU^* &= \frac{1}{2} \begin{bmatrix} W & V \\ -W & V \end{bmatrix} \cdot \begin{bmatrix} MV^* & MV^* \\ M^TW^* & -M^TW^* \end{bmatrix} \\ &= \frac{1}{2} \begin{bmatrix} WMV^* + VM^TW^* & WMV^* - VM^TW^* \\ -WMV^* + VM^TW^* & -WMV^* - VM^TW^* \end{bmatrix} \\ &= \begin{bmatrix} \Sigma & 0 \\ 0 & -\Sigma \end{bmatrix} \end{aligned}$$

Since C is symmetric and U is unitary the eigenvalues are σ_i and $-\sigma_i$ where

$$\Sigma = [\delta_{ij}\sigma_i]$$

■

Problem 1

Find the member of the matrix family:

$$B(y) = B_0 + \sum_{k=1}^{m} y_k B_k$$

that has the least 2-norm.

This problem can be formulated as the dual problem of a semidefinite program.

<u>SDP Dual:</u>

$$\sup_{y \in R^{m+1}, C - \sum_{i=1}^{m+1} y_i A_i \succeq 0} \{y \cdot b\}$$

Let

$$
\begin{aligned}
t &:= y_{m+1} \\
b_i &:= -1 \text{ for } i = m+1 \\
b_i &:= 0 \text{ for } 1 \le i \le m \\
A_i &:= -\begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix} \text{ for } i = m+1 \\
A_i &:= -\begin{bmatrix} 0 & B_i \\ B_i^T & 0 \end{bmatrix} \text{ for } 1 \le i \le m \\
C &:= \begin{bmatrix} 0 & B_0 \\ B_0^T & 0 \end{bmatrix} \\
Q(y,t) &:= \begin{bmatrix} tI & B(y) \\ B(y)^T & tI \end{bmatrix} \\
&= U^* \left(\begin{bmatrix} \Sigma & 0 \\ 0 & -\Sigma \end{bmatrix} + \begin{bmatrix} tI & 0 \\ 0 & tI \end{bmatrix} \right) U
\end{aligned}
$$

The dual SDP problem becomes

$$\sup_{(y,t)\in R^m \times R, Q(y,t)\succeq 0} \{-t\}$$

Now we show that this formulation is equivalent to the problem of finding the matrix with least 2-norm in the affine family.

The 2-norm of a matrix is the maximum singular value. For $Q(y,t)$ we have:

$$\|Q(y,t)\|_2 = t + \sigma_{\max}(y)$$

The positive semidefinite constraint on the matrix Q requires that the minimum eigenvalue is non-negative:

$$t - \sigma_{\max}(y) \ge 0$$

Since the positive semidefinite program takes $\sup\{-t\}$ we see that it is maximized when t is smallest. For a given y the smallest value of t is obtained when:

$$t = \sigma_{\max}(y) = \|B(y)\|_2$$

Therefore the 2-norm problem is equivalent to the dual positive semi-definite program.

Now we derive the primal version of the positive semidefinite program.

SDP Primal:

$$\inf_{X\in S^{n\times n}} \{C \cdot X | X \succeq 0, A_i \cdot X = b_i\}$$

Which can be simplified a little.

For a general matrix P of the type above:

$$P \cdot X = \begin{bmatrix} 0 & M \\ M^T & 0 \end{bmatrix} \cdot X$$

$$
\begin{aligned}
&= \sum_{(1\le i\le n),(1\le j\le n)} M_{i,j} x_{i,(j+n)} + \sum_{(1\le i\le n),(1\le j\le n)} M^T_{i,j} x_{(i+n),j} \\
&= \sum_{(1\le i\le n),(1\le j\le n)} 2M_{i,j} x_{i,(j+n)}
\end{aligned}
$$

This gives:

$$
C \cdot X \quad = \sum_{(1\le i\le n),(1\le j\le n)} B_{0i,j} x_{i,(j+n)}
$$

$$
\begin{aligned}
A_k \cdot X &= 0 \\
\Longleftrightarrow \sum_{(1\le i\le n),(1\le j\le n)} B_{ki,j} x_{i,(j+n)} &= 0
\end{aligned}
$$

$$
\begin{aligned}
\begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix} \cdot X &= \sum_{i=1}^{2n} x_{i,i} \\
&= 1
\end{aligned}
$$

Now we show that the primal and dual problems both have strictly feasible points:

The dual problem requires:

$$
\begin{aligned}
&(y,t) \in R^m \times R \\
&\begin{bmatrix} 0 & B(y) \\ B(y)^T & 0 \end{bmatrix} + \begin{bmatrix} tI & 0 \\ 0 & tI \end{bmatrix} \succeq 0
\end{aligned}
$$

Given any y there is a minimum eigenvalue of the first matrix, so if t is chosen sufficiently large all eigenvalues will be positive. Therefore there is a strictly feasible point for the dual problem.

The primal problem requires:

$$
\begin{aligned}
X \in S^{n\times n}& \\
X \succeq 0& \\
\sum_{(1\le i\le n),(1\le j\le n)} B_{ki,j} x_{i,(j+n)} &= 0 \\
\sum_{i=1}^{2n} x_{i,i} &= 1
\end{aligned}
$$

If all $x_{i,j} = 0$ with $|i-j| > 0$ and $x_{i,i} = \frac{1}{2n}$ then the above equality constraints are satisfied and the matrix X is positive definite. Therefore there is a strictly feasible point for the primal problem.

Problem 2

Given a symmetric matrix $\tilde{Q}$ find the best approximating positive semidefinite matrix Q with respect to the Frobenius norm.

This problem arises naturally when computing a correlation matrix from empirical data. The empirical correlation matrix may be symmetric but lack positive definiteness.

The problem can be formulated as a mixed semidefinite-quadratic cone program.

$$\inf_{X \in S^{n\times n}, x \in R^m} \{C \cdot X + c \cdot x | X \succeq 0, x \in K_{quadratic}, A_i \cdot X + a_i \cdot x = b_i\}$$

$$K_{quadratic} := \{x \in R^m | \; \|(x_2, x_3, \ldots, x_m)\|_2 \leq x_1\}$$

Let

$$D = Q - \tilde{Q}$$

then since D is symmetric we can represent the matrix as a vector d of length $\frac{n(n+1)}{2} = m$.

$$d_k = \tilde{Q}_{\alpha(k),\beta(k)} - Q_{\alpha(k),\beta(k)}$$

where $\alpha(\cdot)$,$\beta(\cdot)$ are the maps that give the corresponding i, j entry of the matrix.

Since the 2-norm of d is proportional the the Frobenius norm of D finding the best Frobenius approximation is the same as minimizing the 2-norm of d.

We can use the quadratic cone constraint to formulate this by introducing equality constraints:

$$x_i = d_{i-1} \text{ for } 2 \leq i \leq m+1$$

By the quadratic cone constraint

$$\|(x_2, x_3, \ldots, x_{m+1})\|_2 \leq x_1$$

The objective is then to minimizer x_1 subject to the constraints.

Let

$$\begin{aligned}
C &= 0 \\
c_1 &= e_1 \\
c_i &= 0 \text{ for } (2 \leq i \leq m+1) \\
A_1 &= 0 \\
A_i &= \frac{1}{2} e_{\alpha(i-1)} e^T_{\beta(i-1)} + \frac{1}{2} e_{\beta(i-1)} e^T_{\alpha(i-1)} \text{ for } (2 \leq i \leq m+1) \\
a_1 &= 0 \\
a_i &= e_i \text{ for } (2 \leq i \leq m+1) \\
b_1 &= 0 \\
b_i &= \tilde{Q}_{\alpha(i-1),\beta(i-1)} \text{ for } (2 \leq i \leq m+1)
\end{aligned}$$

The primal problem becomes:

$$\inf_{X \in S^{n\times n}, x \in R^{m+1}} \{x_1 | X \succeq 0, x \in K_{quadratic}, x_i + X_{\alpha(i-1),\beta(i-1)} = \tilde{Q}_{\alpha(i-1),\beta(i-1)}\}$$

Now we derive the dual problem. Using the derivation above for the semidefinite problem and the fact that the dual cone of the quadratic cone K is itself we obtain by a similar computation:

$$\sup_{z\in R^{m+1}}\{z\cdot b|\sum_{i=1}^{m+1} z_i a_i - c \in -K_{quadratic}, \sum_{i=1}^{m+1} z_i A_i - C \preceq 0\}$$

For the given problem this becomes:

$$\sup_{z\in R^{m+1}}\{\sum_{i=2}^{m+1} z_i \tilde{Q}_{\alpha(i-1),\beta(i-1)}|e_1 - \sum_{i=2}^{m+1} z_i e_i \in K_{quadratic}, \sum_{i=2}^{m+1} z_i \left(e_{\alpha(i-1)}e^T_{\beta(i-1)} + e_{\beta(i-1)}e^T_{\alpha(i-1)}\right) \preceq 0\}$$

If we let

$$Z \quad := \quad \frac{1}{2}\sum_{i=2}^{m+1} z_i \left(e_{\alpha(i-1)}e^T_{\beta(i-1)} + e_{\beta(i-1)}e^T_{\alpha(i-1)}\right)$$

then the problem can be written more simply as:

$$\sup_{z\in R^{m+1}}\{Z{\cdot}\tilde{Q}|(1, -z_2, \ldots, -z_{m+1}) \in K_{quadratic}, Z \preceq 0\}$$

We now show each method has a strictly feasible point.
For the primal problem the constraints require:

$$\begin{aligned}
&X \in S^{n\times n}\\
&x \in R^{m+1}\\
&X \succeq 0\\
&x \in K_{quadratic}\\
&x_i + X_{\alpha(i-1),\beta(i-1)} = \tilde{Q}_{\alpha(i-1),\beta(i-1)} \text{ for } (2 \le i \le m+1)
\end{aligned}$$

If we set $X_{\alpha(i-1),\beta(i-1)} = \tilde{Q}_{\alpha(i-1),\beta(i-1)} + M_{\alpha(i-1),\beta(i-1)}$ where $\lambda_{\min}$ is the minimum eigenvalue of $\tilde{Q}$ and $M = 2\lambda_{\min}I$ then $X \succ 0$. We set corresponding entries of X by symmetry. We then must set $x_i = -M_{\alpha(i-1),\beta(i-1)}$ for $(2 \le i \le m+1)$ to satisfy the equality constraint and since x_1 is free we can set it to $x_1 = 2\|(x_2, \ldots, x_{m+1}\|_2$. This is a strictly feasible point for the primal problem.
For the dual problem the constraints require:

$$\begin{aligned}
&z \in R^{m+1}\\
&e_1 - \sum_{i=2}^{m+1} z_i e_i \in K_{quadratic}\\
&\sum_{i=2}^{m+1} z_i \left(e_{\alpha(i-1)}e^T_{\beta(i-1)} + e_{\beta(i-1)}e^T_{\alpha(i-1)}\right) \preceq 0
\end{aligned}$$

Let $z_{\gamma(k)} = -\frac{1}{\sqrt{2n}}$ and all other $z_i = 0$ where $\gamma(k)$ is the map which gives the corresponding index of the k^{th} diagonal entry of Z. Then the negative semidefinite constraint will be satisfied strictly. The first constraint will be satisfied since there are only n diagonal entries of Z. This gives a feasible point for the dual problem.

Numerical Methods

Two important classes of numerical methods used for solving constrained optimization problems are Log Barrier Methods and Primal-Dual Interior Point Methods. We will only briefly sketch the ideas behind the numerical methods here and refer the reader to reference [2] for more details.

The log barrier methods approximately solve constrained optimization problems by augmenting the objective function by a penalty term:

$$h(x) = f(x) + \mu \sum_{i=1}^{m} \log(c_i(x))$$

The constraints are formulated as $c_i(x) \geq 0$.

A sequence of optimization problems is then solve for successively smaller values of μ. Typically the initial values of μ are set to be large so that the constraints can be easily satisfied. Next μ is gradually reduced to zero which has the effect of increasing the importance of the objective function and results in a better approximation to the true solution. The advantage of this approach is that if a solution lies on the boundary of the feasible set the approximate solutions will be guided along a path in the interior of the set of feasible points as $\mu \to 0$. It can be shown that if at each stage the optimization problem corresponding to μ is solved to sufficient accuracy and the parameter μ is taken sufficiently close to zero then the solutions become close to the true constrained minimizer of the objective function. For more details see Chapter 17 of Nocedal and Wright [2].

A second class of numerical methods useful in solving linear and semidefinite programs with constraints are the Primal-Dual Methods. The basic idea behind these methods is to search for a point that satisfies the Karush-Kuhn-Tucker necessary conditions for the constrained minimization problem. The Karush-Kuhn-Tucker conditions are:

$$\begin{aligned} \nabla f(x^*) &= \sum_{i \in A} \lambda_i \nabla c_i(x^*) \\ \lambda_i &\geq 0 \\ \lambda_i c_i(x^*) &= 0 \end{aligned}$$

where $A = \{i | c_i(x) = 0\}$ is the active set.

In addition it is usually required that the constraints be such that the gradients of $\{\nabla c_i(x) | i \in A(x)\}$ are linearly independent. This is referred to as the *Linear Independent Constraint Qualification (LICQ)*.

Geometrically the KKT conditions with LICQ can be interpreted as stating that at a local minimizer of the constrained optimization problem the objective function $f(x)$ must have either all derivatives zero or any non-zero components of the steepest descent vector must point outside the feasible set. In other words the steepest descent vector must be in the span of the negative gradients of the constraint functions $c_i(x)$. Thus any step in a descent direction leaves the feasible set.

The Primal-Dual Interior Methods search for a minimizing feasible point by satisfying approximate KKT conditions and the constraints of the primal and dual problems. In particular they search for a point $(x_\tau, \lambda_\tau, s_\tau)$ solving:

$$\begin{aligned} A^T \lambda + s &= c \\ Ax &= b \\ x_i s_i &= \tau \\ (x, s) \geq 0 & \end{aligned}$$

where s is the slack variables used to represent the dual inequality constraint.

If the RHS is subtracted from each equation and the difference is denoted by the vector valued function $F(x, \lambda, s)$ then finding a root of F amounts to finding a point satisfying the above conditions. This can be achieved numerically using Newton's method. As τ is reduced to zero the sequence of points $(x_\tau, \lambda_\tau, s_\tau)$ approaches a point satisfying the constraints and KKT conditions. It turns out the path followed by the points in primal space x is similar to the path obtained from the log barrier method. For a more thorough discussion of the methods and implementation see Chapter 14 of Nocedal and Wright [2].

Two implementations of Primal-Dual Methods available freely and worth experimenting with are SeDuMi and SDPT3. As websites quickly become outdated we do not give the URL's here. The distributions are easy to locate using any of the popular search engines.

References

[1] Convex Analysis and Nonlinear Optimization, Borwein and Lewis, Canadian Mathematical Society, 2000.
[2] Numerical Optimization, Nocedal and Wright, Springer, 1999.

Any comments should be send to:
Paul J. Atzberger
atzberg@math.ucsb.edu