Lecture 12 - Duality

\[ f^* = \min \ f(x) \]
\[ \text{s.t. } g_i(x) \leq 0, \ i = 1, 2, \ldots, m \]
\[ h_j(x) = 0, \ j = 1, 2, \ldots, p, \]
\[ x \in X, \]  

(1)

- \( f, g_i, h_j(i = 1, 2, \ldots, m, j = 1, 2, \ldots, p) \) are functions defined on the set \( X \subseteq \mathbb{R}^n \).

- Problem (1) will be referred to as the **primal problem**.

- The Lagrangian is

\[ L(x, \lambda, \mu) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{p} \mu_j h_j(x) \quad (x \in X, \lambda \in \mathbb{R}_+^m, \mu \in \mathbb{R}^p) \]

- The **dual objective function** \( q : \mathbb{R}_+^m \times \mathbb{R}^p \rightarrow \mathbb{R} \cup \{-\infty\} \) is defined to be

\[ q(\lambda, \mu) = \min_{x \in X} L(x, \lambda, \mu). \]  

(2)
The Dual Problem

- The domain of the dual objective function is

\[ \text{dom}(q) = \{(\lambda, \mu) \in \mathbb{R}_+^m \times \mathbb{R}^p : q(\lambda, \mu) > -\infty\} \].

- The dual problem is given by

\[ q^* = \max_{(\lambda, \mu) \in \text{dom}(q)} q(\lambda, \mu) \quad \text{s.t.} \quad (\lambda, \mu) \in \text{dom}(q) \] (3)
Convexity of the Dual Problem

Theorem. Consider problem (1) with \( f, g_i, h_j(i = 1, 2, \ldots, m, j = 1, 2, \ldots, p) \) being functions defined on the set \( X \subseteq \mathbb{R}^n \), and let \( q \) be the dual function defined in (2). Then

(a) \( \text{dom}(q) \) is a convex set.

(b) \( q \) is a concave function over \( \text{dom}(q) \).

Proof.

\( \triangleright \) (a) Take \((\lambda_1, \mu_1), (\lambda_2, \mu_2) \in \text{dom}(q) \) and \( \alpha \in [0, 1] \). Then

\[
\min_{x \in X} L(x, \lambda_1, \mu_1) > -\infty, \quad (4)
\]

\[
\min_{x \in X} L(x, \lambda_2, \mu_2) > -\infty. \quad (5)
\]
Proof Contd.

▶ Therefore, since the Lagrangian $L(x, \lambda, \mu)$ is affine w.r.t. $\lambda, \mu$,

\[
q(\alpha \lambda_1 + (1 - \alpha) \lambda_2, \alpha \mu_1 + (1 - \alpha) \mu_2) \\
= \min_{x \in X} L(x, \alpha \lambda_1 + (1 - \alpha) \lambda_2, \alpha \mu_1 + (1 - \alpha) \mu_2) \\
= \min_{x \in X} \{ \alpha L(x, \lambda_1, \mu_1) + (1 - \alpha) L(x, \lambda_2, \mu_2) \} \\
\geq \alpha \min_{x \in X} L(x, \lambda, \mu_1) + (1 - \alpha) \min_{x \in X} L(x, \lambda_2, \mu_2) \\
= \alpha q(\lambda_1, \mu_1) + (1 - \alpha) q(\lambda_2, \mu_2) \\
> -\infty.
\]

▶ Hence, $\alpha(\lambda_1, \mu_1) + (1 - \alpha)(\lambda_2, \mu_2) \in \text{dom}(q)$, and the convexity of $\text{dom}(q)$ is established.

▶ (b) $L(x, \lambda, \mu)$ is an affine function w.r.t. $(\lambda, \mu)$.

▶ In particular, it is a concave function w.r.t. $(\lambda, \mu)$.

▶ Hence, since $q$ is the minimum of concave functions, it must be concave.
The Weak Duality Theorem

Theorem. Consider the primal problem (1) and its dual problem (3). Then

\[ q^* \leq f^*, \]

where \( f^*, q^* \) are the primal and dual optimal values respectively.

Proof.

- The feasible set of the primal problem is

\[ S = \{ x \in X : g_i(x) \leq 0, h_j(x) = 0, i = 1, 2, \ldots, m, j = 1, 2, \ldots, p \}. \]

- Then for any \((\lambda, \mu) \in \text{dom}(q)\) we have

\[
q(\lambda, \mu) = \min_{x \in X} L(x, \lambda, \mu) \leq \min_{x \in S} L(x, \lambda, \mu)
\]

\[
= \min_{x \in S} \left\{ f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{p} \mu_j h_j(x) \right\}
\]

\[
\leq \min_{x \in S} f(x) = f^*.
\]

- Taking the maximum over \((\lambda, \mu) \in \text{dom}(q)\), the result follows.
Example

\[\begin{align*}
\text{min} & \quad x_1^2 - 3x_2^2 \\
\text{s.t.} & \quad x_1 = x_2^3.
\end{align*}\]

In class
Supporting Hyperplane Theorem Let $C \subseteq \mathbb{R}^n$ be a convex set and let $y \not\in C$. Then there exists $0 \neq p \in \mathbb{R}^n$ such that

$$p^T x \leq p^T y \text{ for any } x \in C.$$

Proof.
- Although the theorem holds for any convex set $C$, we will prove it only for sets with a nonempty interior.
- Since $y \not\in \text{int}(C)$, it follows that $y \not\in \text{int}(\text{cl}(C))$.
- Therefore, there exists a sequence $\{y_k\}_{k \geq 1}$ such that $y_k \not\in \text{cl}(C)$ and $y_k \to y$.
- By the separation theorem of a point from a closed and convex set, there exists $0 \neq p_k \in \mathbb{R}^n$ such that

$$p_k^T x < p_k^T y_k \quad \forall x \in \text{cl}(C)$$

- Thus,

$$\frac{p_k^T}{\|p_k\|}(x - y_k) < 0 \text{ for any } x \in \text{cl}(C). \quad (6)$$
Since the sequence \( \left\{ \frac{p_k}{\|p_k\|} \right\} \) is bounded, it follows that there exists a subsequence \( \left\{ \frac{p_k}{\|p_k\|} \right\}_{k \in T} \) such that \( \frac{p_k}{\|p_k\|} \to p \) as \( k \to T \to \infty \) for some \( p \in \mathbb{R}^n \).

- Obviously, \( \|p\| = 1 \) and hence in particular \( p \neq 0 \).

- Taking the limit as \( k \to T \to \infty \) in inequality (6) we obtain that

\[
p^T(x - y) \leq 0 \text{ for any } x \in \text{cl}(C),
\]

which readily implies the result since \( C \subseteq \text{cl}(C) \).
Separation of Two Convex Sets

**Theorem.** Let $C_1, C_2 \subseteq \mathbb{R}^n$ be two nonempty convex sets such that $C_1 \cap C_2 = \emptyset$. Then there exists $0 \neq p \in \mathbb{R}^n$ for which

$$p^T x \leq p^T y$$

for any $x \in C_1, y \in C_2$.

**Proof.**

- The set $C_1 - C_2$ is a convex set.
- $C_1 \cap C_2 = \emptyset \Rightarrow 0 \notin C_1 - C_2$.
- By the supporting hyperplane theorem, there exists $0 \neq p \in \mathbb{R}^n$ such that

$$p^T (x - y) \leq p^T 0$$

for any $x \in C_1, y \in C_2$. 
The Nonlinear Farkas Lemma

**Theorem.** Let $X \subseteq \mathbb{R}^n$ be a convex set and let $f, g_1, g_2, \ldots, g_m$ be convex functions over $X$. Assume that there exists $\hat{x} \in X$ such that

$$g_1(\hat{x}) < 0, g_2(\hat{x}) < 0, \ldots, g_m(\hat{x}) < 0.$$ 

Let $c \in \mathbb{R}$. Then the following two claims are equivalent:

(a) the following implication holds:

$$x \in X, g_i(x) \leq 0, i = 1, 2, \ldots, m \Rightarrow f(x) \geq c.$$ 

(b) there exist $\lambda_1, \lambda_2, \ldots, \lambda_m \geq 0$ such that

$$\min_{x \in X} \left\{ f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) \right\} \geq c. \quad (7)$$
Proof of (b) \implies (a)

- Suppose that there exist $\lambda_1, \lambda_2, \ldots, \lambda_m \geq 0$ such that (7) holds, and let $x \in X$ satisfy $g_i(x) \leq 0, i = 1, 2, \ldots, m$.

- By (7) we have

$$f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) \geq c,$$

- Hence,

$$f(x) \geq c - \sum_{i=1}^{m} \lambda_i g_i(x) \geq c.$$
Proof of \((a) \Rightarrow (b)\)

- Assume that the implication \((a)\) holds.
- Consider the following two sets:
  \[
  S = \{ u = (u_0, u_1, \ldots, u_m) : \exists x \in X, f(x) \leq u_0, g_i(x) \leq u_i, i = 1, 2, \ldots, m \},
  \]
  \[
  T = \{ (u_0, u_1, \ldots, u_m) : u_0 < c, u_1 \leq 0, u_2 \leq 0, \ldots, u_m \leq 0 \}.
  \]
- \(S, T\) are nonempty and convex and in addition \(S \cap T = \emptyset\).
- By the supporting hyperplane theorem, there exists a vector \(a = (a_0, a_1, \ldots, a_m) \neq 0\), such that
  \[
  \min_{(u_0, u_1, \ldots, u_m) \in S} \sum_{j=0}^{m} a_j u_j \geq \max_{(u_0, u_1, \ldots, u_m) \in T} \sum_{j=0}^{m} a_j u_j. \tag{8}
  \]
- \(a \geq 0\).
- Since \(a \geq 0\), it follows that the right-hand side is \(a_0 c\), and we thus obtained
  \[
  \min_{(u_0, u_1, \ldots, u_m) \in S} \sum_{j=0}^{m} a_j u_j \geq a_0 c. \tag{9}
  \]
Proof of (a) ⇒ (b) Contd.

We will show that $a_0 > 0$. Suppose in contradiction that $a_0 = 0$. Then

$$\min_{(u_0, u_1, \ldots, u_m) \in S} \sum_{j=1}^{m} a_j u_j \geq 0.$$ 

Since we can take $u_i = g_i(\hat{x})$, we can deduce that $\sum_{j=1}^{m} a_j g_j(\hat{x}) \geq 0$, which is impossible since $g_j(\hat{x}) < 0$ and $a \neq 0$.

Since $a_0 > 0$, we can divide (9) by $a_0$ to obtain

$$\min_{(u_0, u_1, \ldots, u_m) \in S} \left\{ u_0 + \sum_{j=1}^{m} \tilde{a}_j u_j \right\} \geq c, \quad (10)$$

where $\tilde{a}_j = \frac{a_j}{a_0}$.

By the definition of $S$ we have

$$\min_{(u_0, u_1, \ldots, u_m) \in S} \left\{ u_0 + \sum_{j=1}^{m} \tilde{a}_j u_j \right\} \leq \min_{x \in X} \left\{ f(x) + \sum_{j=1}^{m} \tilde{a}_j g_j(x) \right\},$$

which combined with (10) yields the desired result

$$\min_{x \in X} \left\{ f(x) + \sum_{j=1}^{m} \tilde{a}_j g_j(x) \right\} \geq c.$$
Theorem. Consider the optimization problem

\[
    f^* = \min_{x \in X} f(x) \quad \text{s.t.} \quad g_i(x) \leq 0, \quad i = 1, 2, \ldots, m, \quad (11)
\]

where \( X \) is a convex set and \( f, g_i, i = 1, 2, \ldots, m \) are convex functions over \( X \). Suppose that there exists \( \hat{x} \in X \) for which \( g_i(\hat{x}) < 0, \quad i = 1, 2, \ldots, m \). If problem (11) has a finite optimal value, then

(a) the optimal value of the dual problem is attained.
(b) \( f^* = q^* \).
Proof of Strong Duality Theorem

- Since $f^* > -\infty$ is the optimal value of (11), it follows that the following implication holds:

$$x \in X, g_i(x) \leq 0, i = 1, 2, \ldots, m \Rightarrow f(x) \geq f^*,$$

- By the nonlinear Farkas Lemma there exists $\tilde{\lambda}_1, \tilde{\lambda}_2, \ldots, \tilde{\lambda}_m \geq 0$ such that

$$q(\tilde{\lambda}) = \min_{x \in X} \left\{ f(x) + \sum_{j=1}^{m} \tilde{\lambda}_j g_j(x) \right\} \geq f^*.$$

- By the weak duality theorem,

$$q^* \geq q(\tilde{\lambda}) \geq f^* \geq q^*,$$

- Hence $f^* = q^*$ and $\tilde{\lambda}$ is an optimal solution of the dual problem.
Example

\[
\begin{align*}
\text{min} & \quad x_1^2 - x_2 \\
\text{s.t.} & \quad x_2^2 \leq 0.
\end{align*}
\]
Duffin’s Duality Gap

\[
\min \left\{ e^{-x_2} : \sqrt{x_1^2 + x_2^2} - x_1 \leq 0 \right\}.
\]

- The feasible set is in fact \( F = \{(x_1, x_2) : x_1 \geq 0, x_2 = 0\} \Rightarrow f^* = 1 \)
- Slater condition is not satisfied.
- Lagrangian: \( L(x_1, x_2, \lambda) = e^{-x_2} + \lambda(\sqrt{x_1^2 + x_2^2} - x_1) \ (\lambda > 0) \).
- \( q(\lambda) = \min_{x_1, x_2} L(x_1, x_2, \lambda) \geq 0 \)
- For any \( \varepsilon > 0 \), take \( x_2 = -\log \varepsilon, \ x_1 = \frac{x_2^2 - \varepsilon^2}{2\varepsilon} \).

\[
\sqrt{x_1^2 + x_2^2} - x_1 = \sqrt{\frac{(x_2^2 - \varepsilon^2)}{4\varepsilon^2}} + x_2 - \frac{x_2^2 - \varepsilon^2}{2\varepsilon} = \sqrt{\frac{(x_2^2 + \varepsilon^2)^2}{4\varepsilon^2}} - \frac{x_2^2 - \varepsilon^2}{2\varepsilon} = \frac{x_2^2 + \varepsilon^2}{2\varepsilon} - \frac{x_2^2 - \varepsilon^2}{2\varepsilon} = \varepsilon.
\]

- Hence, \( L(x_1, x_2, \lambda) = e^{-x_2} + \lambda(\sqrt{x_1^2 + x_2^2} - x_1) = \varepsilon + \lambda \varepsilon = (1 + \lambda)\varepsilon \),
- \( q(\lambda) = 0 \) for all \( \lambda \geq 0 \).
- \( q^* = 0 \Rightarrow f^* - q^* = 1 \Rightarrow \) duality gap of 1.
Complementary Slackness Conditions

**Theorem.** Consider the optimization problem

\[ f^* = \min \{ f(x) : g_i(x) \leq 0, i = 1, 2, \ldots, m, x \in X \}, \quad (12) \]

and assume that \( f^* = q^* \) where \( q^* \) is the optimal value of the dual problem. Let \( x^*, \lambda^* \) be feasible solutions of the primal and dual problems. Then \( x^*, \lambda^* \) are **optimal** solutions of the primal and dual problems iff

\[ x^* \in \arg \min_{x \in X} L(x, \lambda^*), \quad (13) \]

\[ \lambda^*_i g_i(x^*) = 0, i = 1, 2, \ldots, m. \quad (14) \]

**Proof.**

- \( q(\lambda^*) = \min_{x \in X} L(x, \lambda^*) \leq L(x^*, \lambda^*) = f(x^*) + \sum_{i=1}^{m} \lambda^*_i g_i(x^*) \leq f(x^*) \)
- By strong duality, \( x^*, \lambda^* \) are optimal iff \( f(x^*) = q(\lambda^*) \)
- iff \( \min_{x \in X} L(x, \lambda^*) = L(x^*, \lambda^*), \sum_{i=1}^{m} \lambda^*_i g_i(x^*) = 0. \)
- iff (13), (14) hold.
A More General Strong Duality Theorem

Theorem. Consider the optimization problem

\[ f^* = \min_{x \in X} f(x) \]
\[ \text{s.t. } g_i(x) \leq 0, \quad i = 1, 2, \ldots, m, \]
\[ h_j(x) \leq 0, \quad j = 1, 2, \ldots, p, \]
\[ s_k(x) = 0, \quad k = 1, 2, \ldots, q, \]

(15)

where \( X \) is a convex set and \( f, g_i, i = 1, 2, \ldots, m \) are convex functions over \( X \). The functions \( h_j, s_k \) are affine functions. Suppose that there exists \( \hat{x} \in \text{int}(X) \) for which \( g_i(\hat{x}) < 0, h_j(\hat{x}) \leq 0, s_k(\hat{x}) = 0 \). Then if problem (15) has a finite optimal value, then the optimal value of the dual problem

\[ q^* = \max \{ q(\lambda, \eta, \mu) : (\lambda, \eta, \mu) \in \text{dom}(q) \}, \]

where

\[ q(\lambda, \eta, \mu) = \min_{x \in X} \left[ f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{p} \eta_j h_j(x) + \sum_{k=1}^{q} \mu_k s_k(x) \right] \]

is attained, and \( f^* = q^* \).
Importance of the Underlying Set

\[
\begin{align*}
\text{(P)} \quad & \min \ x_1^3 + x_2^3 \\
\text{s.t.} \quad & x_1 + x_2 \geq 1, \\
& x_1, x_2 \geq 0.
\end{align*}
\]

\begin{itemize}

\item \((\frac{1}{2}, \frac{1}{2})\) is the optimal solution of (P) with an optimal value \(f^* = \frac{1}{4}\).

\item First dual problem is constructed by taking \(X = \{(x_1, x_2) : x_1, x_2 \geq 0\}\).

\item The primal problem is \(\min \{x_1^3 + x_2^3 : x_1 + x_2 \geq 1, (x_1, x_2) \in X\}\).

\item Strong duality holds for the problem and hence in particular \(q^* = \frac{1}{4}\).

\item Second dual is constructed by taking \(X = \mathbb{R}^2\).

\item Objective function is not convex \(\Rightarrow\) strong duality is not necessarily satisfied.

\item \(L(x_1, x_2, \lambda, \eta_1, \eta_2) = x_1^3 + x_2^3 - \lambda(x_1 + x_2 - 1) - \eta_1 x_1 - \eta_2 x_2\).

\item \(q(\lambda, \eta_1, \eta_2) = -\infty\) for all \((\lambda, \mu_1, \mu_2) \Rightarrow q^* = -\infty\).

\end{itemize}
Linear Programming

Consider the linear programming problem

\[ \begin{align*}
\min & \quad c^T x \\
\text{s.t.} & \quad Ax \leq b,
\end{align*} \]

- \( c \in \mathbb{R}^n, A \in \mathbb{R}^{m \times n} \) and \( b \in \mathbb{R}^m \).
- We assume that the problem is feasible \( \Rightarrow \) strong duality holds.
- \( L(x, \lambda) = c^T x + \lambda^T (Ax - b) = (c + A^T \lambda)^T x - b^T \lambda \).
- Dual objective function:

\[
q(\lambda) = \min_{x \in \mathbb{R}^n} L(x, \lambda) = \min_{x \in \mathbb{R}^n} (c + A^T \lambda)^T x - b^T \lambda = \begin{cases} 
- b^T \lambda & c + A^T \lambda = 0, \\
- \infty & \text{else}.
\end{cases}
\]

- Dual problem:

\[
\begin{align*}
\max & \quad -b^T \lambda \\
\text{s.t.} & \quad A^T \lambda = -c, \\
& \quad \lambda \geq 0.
\end{align*}
\]
Strictly Convex Quadratic Programming

Consider the strictly convex quadratic programming problem

\[
\begin{align*}
\min & \quad x^T Q x + 2 f^T x \\
\text{s.t.} & \quad A x \leq b,
\end{align*}
\]  

(16)

\[Q \in \mathbb{R}^{n \times n} \text{ positive definite, } f \in \mathbb{R}^n, A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m.\]

\[\text{Lagrangian: } (\lambda \in \mathbb{R}^m) \quad L(x, \lambda) = x^T Q x + 2 f^T x + 2 \lambda^T (A x - b) = x^T Q x + 2 (A^T \lambda + f)^T x - 2 b^T \lambda.\]

\[\text{The minimizer of the Lagrangian is attained at } x^* = -Q^{-1}(f + A^T \lambda).\]

\[q(\lambda) = L(x^*, \lambda) = (f + A^T \lambda)^T Q^{-1} QQ^{-1}(f + A^T \lambda) - 2(f + A^T \lambda)^T Q^{-1}(f + A^T \lambda) - 2 b^T \lambda \]

\[= -(f + A^T \lambda)^T Q^{-1}(f + A^T \lambda) - 2 b^T \lambda \]

\[= -\lambda^T AQ^{-1} A^T \lambda - 2 f^T Q^{-1} A^T \lambda - f^T Q^{-1} f - 2 b^T \lambda \]

\[= -\lambda^T AQ^{-1} A^T \lambda - 2(AQ^{-1} f + b)^T \lambda - f^T Q^{-1} f.\]

\[\text{The dual problem is } \max \{q(\lambda) : \lambda \geq 0\}.\]
Dual of Convex QCQP with strictly convex objective

Consider the QCQP problem

\[
\begin{align*}
\min & \quad x^T A_0 x + 2 b_0^T x + c_0 \\
\text{s.t.} & \quad x^T A_i x + 2 b_i^T x + c_i \leq 0, \quad i = 1, 2, \ldots, m,
\end{align*}
\]

where \( A_i \succeq 0 \) is an \( n \times n \) matrix, \( b_i \in \mathbb{R}^n, c_i \in \mathbb{R}, i = 0, 1, \ldots, m \).

Assume that \( A_0 \succ 0 \).

▶ Lagrangian \((\lambda \in \mathbb{R}_+^m)\):

\[
L(x, \lambda) = x^T A_0 x + 2 b_0^T x + c_0 + \sum_{i=1}^m \lambda_i (x^T A_i x + 2 b_i^T x + c_i)
\]

\[
= x^T (A_0 + \sum_{i=1}^m \lambda_i A_i) x + 2 (b_0 + \sum_{i=1}^m \lambda_i b_i)^T x + c_0 + \sum_{i=1}^m \lambda_i c_i.
\]

▶ The minimizer of the Lagrangian w.r.t. \( x \) is attained at \( \bar{x} \) satisfying

\[
2 (A_0 + \sum_{i=1}^m \lambda_i A_i) \bar{x} = -2 (b_0 + \sum_{i=1}^m \lambda_i b_i).
\]

▶ Thus, \( \bar{x} = - (A_0 + \sum_{i=1}^m \lambda_i A_i)^{-1} (b_0 + \sum_{i=1}^m \lambda_i b_i) \).
Plugging this expression back into the Lagrangian, we obtain the following expression for the dual objective function

\[
q(\lambda) = \min_x L(x, \lambda) = L(\tilde{x}, \lambda) = \tilde{x}^T (A_0 + \sum_{i=1}^m \lambda_i A_i) \tilde{x} + 2 (b_0 + \sum_{i=1}^m \lambda_i b_i)^T \tilde{x} + c_0 + \sum_{i=1}^m \lambda_i c_i
\]

\[
= - (b_0 + \sum_{i=1}^m \lambda_i b_i)^T (A_0 + \sum_{i=1}^m \lambda_i A_i)^{-1} (b_0 + \sum_{i=1}^m \lambda_i b_i) + c_0 + \sum_{i=1}^m \lambda_i c_i.
\]

The dual problem is thus

\[
\max \quad - (b_0 + \sum_{i=1}^m \lambda_i b_i)^T (A_0 + \sum_{i=1}^m \lambda_i A_i)^{-1} (b_0 + \sum_{i=1}^m \lambda_i b_i) + c_0 + \sum_{i=1}^m \lambda_i c_i \\
\text{s.t.} \quad \lambda_i \geq 0, \quad i = 1, 2, \ldots, m.
\]
Dual of Convex QCQPs

\(A_0\) is only assumed to be positive semidefinite.

- The previous dual is not well defined since the matrix \(A_0 + \sum_{i=1}^{m} \lambda_i A_i\) is not necessarily PD.

- Decompose \(A_i\) as \(A_i = D_i^T D_i\) \((D_i \in \mathbb{R}^{n \times n})\) and rewrite the problem as

\[
\begin{align*}
\min & \quad x^T D_0^T D_0 x + 2b_0^T x + c_0 \\
\text{s.t.} & \quad x^T D_i^T D_i x + 2b_i^T x + c_i \leq 0, \quad i = 1, 2, \ldots, m,
\end{align*}
\]

- Define additional variables \(z_i = D_i x\), giving rise to the formulation

\[
\begin{align*}
\min & \quad \|z_0\|^2 + 2b_0^T x + c_0 \\
\text{s.t.} & \quad \|z_i\|^2 + 2b_i^T x + c_i \leq 0, \quad i = 1, 2, \ldots, m, \\
& \quad z_i = D_i x, \quad i = 0, 1, \ldots, m.
\end{align*}
\]
Dual of Convex QCQPs

The Lagrangian is \((\lambda \in \mathbb{R}^m_+, \mu_i \in \mathbb{R}^n, i = 0, 1, \ldots, m)\):

\[
L(x, z_0, \ldots, z_m, \lambda, \mu_0, \ldots, \mu_m) = \|z_0\|^2 + 2b_0^T x + c_0 + \sum_{i=1}^{m} \lambda_i (\|z_i\|^2 + 2b_i^T x + c_i) + 2 \sum_{i=0}^{m} \mu_i^T (z_i - D_i x)
\]

\[
= \|z_0\|^2 + 2\mu_0^T z_0 + \sum_{i=1}^{m} (\lambda_i \|z_i\|^2 + 2\mu_i^T z_i) + 2 \left( b_0 + \sum_{i=1}^{m} \lambda_i b_i - \sum_{i=0}^{m} D_i^T \mu_i \right)^T x + c_0 + \sum_{i=1}^{m} c_i \lambda_i.
\]
Dual of Convex QCQPs

For any \(\lambda \in \mathbb{R}_+\), \(\mu \in \mathbb{R}^n\),

\[
g(\lambda, \mu) \equiv \min_z \{ \lambda \|z\|^2 + 2\mu^T z \} = \begin{cases} -\frac{\|\mu\|^2}{\lambda} & \lambda > 0, \\ 0 & \lambda = 0, \mu = 0, \\ -\infty & \lambda = 0, \mu \neq 0. \end{cases}
\]

Since the Lagrangian is separable with respect to \(z_i\) and \(x\), we can perform the minimization with respect to each of the variables vectors:

\[
\min_{z_0} \left[ \|z_0\|^2 + 2\mu_0^T z_0 \right] = g(1, \mu_0) = -\|\mu_0\|^2,
\]

\[
\min_{z_i} \left[ \lambda_i \|z_i\|^2 + 2\mu_i^T z_i \right] = g(\lambda_i, \mu_i),
\]

\[
\min_x \left( b_0 + \sum_{i=1}^m \lambda_i b_i - \sum_{i=0}^m D_i^T \mu_i \right)^T x = \begin{cases} 0 & b_0 + \sum_{i=1}^m \lambda_i b_i - \sum_{i=0}^m D_i^T \mu_i = 0, \\ -\infty & \text{else}, \end{cases}
\]

Hence,

\[
q(\lambda, \mu_0, \ldots, \mu_m) = \min_{x,z_0,\ldots,z_m} L(x, z_0, \ldots, z_m, \lambda, \mu_0, \ldots, \mu_m)
\]

\[
= \begin{cases} g(1, \mu_0) + \sum_{i=1}^m g(\lambda_i, \mu_i) + c_0 + c^T \lambda & b_0 + \sum_{i=1}^m \lambda_i b_i - \sum_{i=0}^m D_i^T \mu_i = 0, \\ -\infty & \text{else}. \end{cases}
\]
The dual problem is therefore

$$\max \quad g(1, \mu_0) + \sum_{i=1}^m g(\lambda_i, \mu_i) + c_0 + \sum_{i=1}^m c_i \lambda_i$$

s.t. $$b_0 + \sum_{i=1}^m \lambda_i b_i - \sum_{i=0}^m D_i^T \mu_i = 0,$$

$$\lambda \in \mathbb{R}^m_+, \mu_0, \ldots, \mu_m \in \mathbb{R}^n.$$
Dual of Nonconvex QCQPs

Consider the problem

\[
\begin{align*}
\min & \quad x^T A_0 x + 2 b_0^T x + c_0 \\
\text{s.t.} & \quad x^T A_i x + 2 b_i^T x + c_i \leq 0, \quad i = 1, 2, \ldots, m,
\end{align*}
\]

- \( A_i = A_i^T \in \mathbb{R}^{n \times n}, b_i \in \mathbb{R}^n, c_i \in \mathbb{R}, i = 0, 1, \ldots, m. \)
- We do not assume that \( A_i \) are positive semidefinite, and hence the problem is in general nonconvex.
- Lagrangian \((\lambda \in \mathbb{R}^m_+)\):

\[
L(x, \lambda) = x^T A_0 x + 2 b_0^T x + c_0 + \sum_{i=1}^{m} \lambda_i \left( x^T A_i x + 2 b_i^T x + c_i \right)
\]

\[
= x^T \left( A_0 + \sum_{i=1}^{m} \lambda_i A_i \right) x + 2 \left( b_0 + \sum_{i=1}^{m} \lambda_i b_i \right)^T x + c_0 + \sum_{i=1}^{m} c_i \lambda_i.
\]

- Note that

\[
q(\lambda) = \min_{x} L(x, \lambda) = \max_{t} \{ t : L(x, \lambda) \geq t \text{ for any } x \in \mathbb{R}^n \}.
\]
Dual of Nonconvex QCQPs

The following holds:

\[ L(x, \lambda) \geq t \text{ for all } x \in \mathbb{R}^n \]

is equivalent to

\[
\begin{pmatrix}
A_0 + \sum_{i=1}^{m} \lambda_i A_i \\
(b_0 + \sum_{i=1}^{m} \lambda_i b_i)^T
\end{pmatrix}
\begin{pmatrix}
b_0 + \sum_{i=1}^{m} \lambda_i b_i \\
c_0 + \sum_{i=1}^{m} \lambda_i c_i - t
\end{pmatrix} \succeq 0,
\]

Therefore, the dual problem is

\[
\max_{t, \lambda_i} t \\
\text{s.t.} \quad \begin{pmatrix}
A_0 + \sum_{i=1}^{m} \lambda_i A_i \\
(b_0 + \sum_{i=1}^{m} \lambda_i b_i)^T
\end{pmatrix}
\begin{pmatrix}
b_0 + \sum_{i=1}^{m} \lambda_i b_i \\
c_0 + \sum_{i=1}^{m} \lambda_i c_i - t
\end{pmatrix} \succeq 0,
\]

\[ \lambda_i \geq 0, \quad i = 1, 2, \ldots, m. \]
Orthogonal Projection onto the Unit Simplex

- Given a vector \( y \in \mathbb{R}^n \), the orthogonal projection of \( y \) onto \( \Delta_n \) is the solution to

\[
\begin{align*}
\min & \quad \| x - y \|^2 \\
\text{s.t.} & \quad e^T x = 1, \\
& \quad x \geq 0.
\end{align*}
\]

- Lagrangian:

\[
L(x, \lambda) = \| x - y \|^2 + 2\lambda(e^T x - 1) = \| x \|^2 - 2(y - \lambda e)^T x + \| y \|^2 - 2\lambda
\]

\[
= \sum_{j=1}^{n} (x_j^2 - 2(y_j - \lambda)x_j) + \| y \|^2 - 2\lambda.
\]

- The optimal \( x_j \) is the solution to the 1D problem \( \min_{x_j \geq 0} [x_j^2 - 2(y_j - \lambda)x_j] \).

- The optimal \( x_j \) is

\[
x_j = \begin{cases} 
  y_j - \lambda & y_j \geq \lambda \\
  0 & \text{else} 
\end{cases} = [y_j - \lambda]_+ , \text{ with optimal value } -[y_j - \lambda]^2_+.
\]

- The dual problem is

\[
\max_{\lambda \in \mathbb{R}} \left\{ g(\lambda) \equiv - \sum_{j=1}^{n} [y_j - \lambda]^2_+ - 2\lambda + \| y \|^2 \right\}.
\]
Orthogonal Projection onto the Unit Simplex

- $g$ is concave, differentiable, $\lim_{\lambda \to \infty} g(\lambda) = \lim_{\lambda \to -\infty} g(\lambda) = -\infty$.
- Therefore, there exists an optimal solution to the dual problem attained at a point $\lambda^*$ in which $g'(\lambda^*) = 0$.
- $\sum_{j=1}^{n} [y_j - \lambda^*]_+ = 1$.
- $h(\lambda) = \sum_{j=1}^{n} [y_j - \lambda]_+ - 1$ is nonincreasing over $\mathbb{R}$ and is in fact strictly decreasing over $(-\infty, \max_j y_j]$.

$$h(y_{\max}) = -1,$$
$$h\left(y_{\min} - \frac{2}{n}\right) = \sum_{j=1}^{n} y_j - ny_{\min} + 2 - 1 > 0,$$

where $y_{\max} = \max_{j=1,2,...,n} y_j$, $y_{\min} = \min_{j=1,2,...,n} y_j$.

- We can therefore invoke a bisection procedure to find the unique root $\lambda^*$ of the function $h$ over the interval $[y_{\min} - \frac{2}{n}, y_{\max}]$, and then define $P_\Delta_n(y) = [y - \lambda^* e]_+$. 
Orthogonal Projection Onto the Unit Simplex

The MATLAB function `proj_unit_simplex`:

```matlab
function xp=proj_unit_simplex(y)
    f=@(lam)sum(max(y-lam,0))-1;
    n=length(y);
    lb=min(y)-2/n;
    ub=max(y);
    lam=bisection(f,lb,ub,1e-10);
    xp=max(y-lam,0);
```
Dual of the Chebyshev Center Problem

- Formulation:
  \[
  \begin{align*}
  \min_{x,r} & \quad r \\
  \text{s.t.} & \quad \|x - a_i\| \leq r, \quad i = 1, 2, \ldots, m.
  \end{align*}
  \]

- Reformulation:
  \[
  \begin{align*}
  \min_{x,\gamma} & \quad \gamma \\
  \text{s.t.} & \quad \|x - a_i\|^2 \leq \gamma, \quad i = 1, 2, \ldots, m.
  \end{align*}
  \]

\[
L(x, \gamma, \lambda) = \gamma + \sum_{i=1}^{m} \lambda_i (\|x - a_i\|^2 - \gamma)
\]

\[
= \gamma (1 - \sum_{i=1}^{m} \lambda_i) + \sum_{i=1}^{m} \lambda_i \|x - a_i\|^2.
\]

- The minimization of the above expression must be \(-\infty\) unless \(\sum_{i=1}^{m} \lambda_i = 1\), and in this case we have
  \[
  \min_{\gamma} \gamma \left(1 - \sum_{i=1}^{m} \lambda_i\right) = 0.
  \]
Dual of Chebyshev Center Contd.

- Need to solve $\min_x \sum_{i=1}^m \lambda_i \|x - a_i\|^2$.
- We have

$$\sum_{i=1}^m \lambda_i \|x - a_i\|^2 = \|x\|^2 - 2 (\sum_{i=1}^m \lambda_i a_i)^T x + \sum_{i=1}^m \lambda_i \|a_i\|^2,$$

(17)

- The minimum is attained at the point in which the gradient vanishes:

$$x^* = \sum_{i=1}^m \lambda_i a_i = A\lambda,$$

$A$ is the $n \times m$ matrix whose columns are $a_1, a_2, \ldots, a_m$.

- Substituting this expression back into (17),

$$q(\lambda) = \|A\lambda\|^2 - 2(A\lambda)^T (A\lambda) + \sum_{i=1}^m \lambda_i \|a_i\|^2 = -\|A\lambda\|^2 + \sum_{i=1}^m \lambda_i \|a_i\|^2.$$

- The dual problem is therefore

$$\max \quad -\|A\lambda\|^2 + \sum_{i=1}^m \lambda_i \|a_i\|^2$$

s.t. $\lambda \in \Delta_m$. 
function [xp,r]=chebyshev_center(A)

d=size(A);

m=d(2);

Q=A'*A;

L=2*max(eig(Q));

b=sum(A.^2)';

%initialization with the uniform vector

lam=1/m*ones(m,1);

old_lam=zeros(m,1);

while (norm(lam-old_lam)>1e-5)
    old_lam=lam;
    lam=proj_unit_simplex(lam+1/L*(-2*Q*lam+b));
end

xp=A*lam;

r=0;

for i=1:m
    r=max(r,norm(xp-A(:,i)));
end
Denoising

Suppose that we are given a signal contaminated with noise.

\[ y = x + w, \]

\( x \) - unknown “true” signal, \( w \) - unknown noise, \( y \) - known observed signal.

The denoising problem: find a “good” estimate for \( x \) given \( y \).
A Tikhonov Regularization Approach

Quadratic Penalty:

$$\min \|x - y\|^2 + \lambda \sum_{i=1}^{n-1} (x_i - x_{i+1})^2,$$

The solution with $\lambda = 1$:

Pretty good!
Weakness of Quadratic Regularization

The quadratic regularization method does not work so well for all types of signals. True and noisy step functions:

![Graphs showing true and noisy step functions]
Failure of Quadratic Regularization
\( l_1 \) regularization

\[
\min \|x - y\|^2 + \lambda \|Lx\|_1. 
\] (18)

- The problem is equivalent to the optimization problem

\[
\begin{align*}
\min_{x,z} & \quad \|x - y\|^2 + \lambda \|z\|_1 \\
\text{s.t.} & \quad z = Lx.
\end{align*}
\]

\( L \) is the \((n-1) \times n\) matrix whose components are \( L_{i,i} = 1, \ L_{i,i+1} = -1 \) and 0 otherwise.

- The Lagrangian of the problem is

\[
L(x, z, \mu) = \|x - y\|^2 + \lambda \|z\|_1 + \mu^T(Lx - z) = \|x - y\|^2 + (L^T \mu)^T x + \lambda \|z\|_1 - \mu^T z.
\]

- The dual problem is

\[
\max \quad -\frac{1}{4} \mu^T LL^T \mu + \mu^T Ly \\
\text{s.t.} \quad \|\mu\|_\infty \leq \lambda. 
\] (19)
A MATLAB code

Employing the gradient projection method on the dual:

```matlab
lambda=1;
mu=zeros(n-1,1);
for i=1:1000
    mu=mu-0.25*L*(L'*mu)+0.5*(L*y);
    mu=lambda*mu./max(abs(mu),lambda);
    xde=y-0.5*L'*mu;
end
figure(5)
plot(t,xde,’.’);
axis([0,1,-1,4])
```
$l_1$-regularized solution
Dual of the Linear Separation Problem (Dual SVM)

- \( \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m \in \mathbb{R}^n. \)
- For each \( i \), we are given a scalar \( y_i \) which is equal to 1 if \( \mathbf{x}_i \) is in class A or \(-1\) if it is in class B.
- The problem of finding a maximal margin hyperplane that separates the two sets of points is

\[
\begin{align*}
\min & \quad \frac{1}{2} \| \mathbf{w} \|^2 \\
\text{s.t.} & \quad y_i(\mathbf{w}^T \mathbf{x}_i + \beta) \geq 1, \quad i = 1, 2, \ldots, m.
\end{align*}
\]

- The above assumes that the two classes are linearly separable.
- A formulation that allows violation of the constraints (with an appropriate penalty):

\[
\begin{align*}
\min & \quad \frac{1}{2} \| \mathbf{w} \|^2 + C \sum_{i=1}^{m} \xi_i \\
\text{s.t.} & \quad y_i(\mathbf{w}^T \mathbf{x}_i + \beta) \geq 1 - \xi_i, \quad i = 1, 2, \ldots, m, \\
& \quad \xi_i \geq 0, \quad i = 1, 2, \ldots, m,
\end{align*}
\]

where \( C > 0 \) is a penalty parameter.
Dual SVM

- The same as

\[
\begin{align*}
\min & \quad \frac{1}{2} \| w \|^2 + C (e^T \xi) \\
\text{s.t.} & \quad Y (Xw + \beta e) \geq e - \xi, \\
& \quad \xi \geq 0,
\end{align*}
\]

where \( Y = \text{diag}(y_1, y_2, \ldots, y_m) \) and \( X \) is the \( m \times n \) matrix whose rows are \( x_1^T, x_2^T, \ldots, x_m^T \).

- Lagrangian (\( \alpha \in \mathbb{R}_+^m \)):

\[
L(w, \beta, \xi, \alpha) = \frac{1}{2} \| w \|^2 + C (e^T \xi) - \alpha^T [YXw + \beta Ye - e + \xi]
\]

\[
= \frac{1}{2} \| w \|^2 - w^T [X^T Y \alpha] - \beta (\alpha^T Ye) + \xi^T (Ce - \alpha) + \alpha^T e.
\]

- 

\[
q(\alpha) = \left[ \min_w \frac{1}{2} \| w \|^2 - w^T [X^T Y \alpha] \right] + \left[ \min_{\beta} (-\beta (\alpha^T Ye)) \right] + \left[ \min_{\xi \geq 0} \xi^T (Ce - \alpha) \right] + \alpha^T e.
\]
Dual SVM

\[
\min_w \frac{1}{2} \|w\|^2 - w^T [X^T Y \alpha] = -\frac{1}{2} \alpha^T YXX^T Y \alpha,
\]

\[
\min_{\beta} (-\beta (\alpha^T Ye)) = \begin{cases} 
0 & \alpha^T Ye = 0, \\ 
-\infty & \text{else},
\end{cases}
\]

\[
\min_{\xi \geq 0} \xi^T (Ce - \alpha) = \begin{cases} 
0 & \alpha \leq Ce, \\ 
-\infty & \text{else},
\end{cases}
\]

Therefore, the dual objective function is given by

\[
q(\alpha) = \begin{cases} 
\alpha^T e - \frac{1}{2} \alpha^T YXX^T Y \alpha & \alpha^T Ye = 0, 0 \leq \alpha \leq Ce \\
-\infty & \text{else}.
\end{cases}
\]

\[
\max \quad \alpha^T e - \frac{1}{2} \alpha^T YXX^T Y \alpha \\
\text{s.t.} \\
\alpha^T Ye = 0, \\
0 \leq \alpha \leq Ce.
\]

or

\[
\max \quad \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i \alpha_j y_i y_j (x_i^T x_j) \\
\text{s.t.} \\
\sum_{i=1}^{m} y_i \alpha_i = 0, \\
0 \leq \alpha_i \leq C, \quad i = 1, 2, \ldots, m.
\]