Lecture 12 - Duality

$$f^* = \begin{array}{ll} \min & f(\mathbf{x}) \\ \text{s.t.} & g_i(\mathbf{x}) \leq 0, i = 1, 2, \ldots, m \\ & h_j(\mathbf{x}) = 0, j = 1, 2, \ldots, p, \\ & \mathbf{x} \in X, \end{array} \tag{1}$$

- $f, g_i, h_j (i = 1, 2, \ldots, m, j = 1, 2, \ldots, p)$ are functions defined on the set $X \subseteq \mathbb{R}^n$.
- Problem (1) will be referred to as the primal problem.
- The Lagrangian is

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

- The dual objective function $q : \mathbb{R}^m_+ \times \mathbb{R}^p \to \mathbb{R} \cup \{-\infty\}$ is defined to be

$$q(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \min_{\mathbf{x} \in X} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}). \tag{2}$$

The Dual Problem

- The domain of the dual objective function is

$$\text{dom}(q) = \{(\boldsymbol{\lambda}, \boldsymbol{\mu}) \in \mathbb{R}^m_+ \times \mathbb{R}^p : q(\boldsymbol{\lambda}, \boldsymbol{\mu}) > -\infty\}.$$

- The dual problem is given by

$$\begin{aligned} q^* = \quad & \max \quad q(\boldsymbol{\lambda}, \boldsymbol{\mu}) \\ & \text{s.t.} \quad (\boldsymbol{\lambda}, \boldsymbol{\mu}) \in \text{dom}(q) \end{aligned} \tag{3}$$

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.

- (a) Take $(\boldsymbol{\lambda}_1, \boldsymbol{\mu}_1), (\boldsymbol{\lambda}_2, \boldsymbol{\mu}_2) \in \text{dom}(q)$ and $\alpha \in [0, 1]$. Then

$$\min_{\mathbf{x} \in X} L(\mathbf{x}, \boldsymbol{\lambda}_1, \boldsymbol{\mu}_1) > -\infty, \quad (4)$$

$$\min_{\mathbf{x} \in X} L(\mathbf{x}, \boldsymbol{\lambda}_2, \boldsymbol{\mu}_2) > -\infty. \quad (5)$$

Proof Contd.

- Therefore, since the Lagrangian $L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu})$ is affine w.r.t. $\boldsymbol{\lambda}, \boldsymbol{\mu}$,

$$\begin{aligned}
& q(\alpha\boldsymbol{\lambda}_1 + (1-\alpha)\boldsymbol{\lambda}_2, \alpha\boldsymbol{\mu}_1 + (1-\alpha)\boldsymbol{\mu}_2) \\
&= \min_{\mathbf{x}\in X} L(\mathbf{x}, \alpha\boldsymbol{\lambda}_1 + (1-\alpha)\boldsymbol{\lambda}_2, \alpha\boldsymbol{\mu}_1 + (1-\alpha)\boldsymbol{\mu}_2) \\
&= \min_{\mathbf{x}\in X} \{\alpha L(\mathbf{x}, \boldsymbol{\lambda}_1, \boldsymbol{\mu}_1) + (1-\alpha) L(\mathbf{x}, \boldsymbol{\lambda}_2, \boldsymbol{\mu}_2)\} \\
&\geq \alpha \min_{\mathbf{x}\in X} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}_1) + (1-\alpha) \min_{\mathbf{x}\in X} L(\mathbf{x}, \boldsymbol{\lambda}_2, \boldsymbol{\mu}_2) \\
&= \alpha q(\boldsymbol{\lambda}_1, \boldsymbol{\mu}_1) + (1-\alpha) q(\boldsymbol{\lambda}_2, \boldsymbol{\mu}_2) \\
&> -\infty.
\end{aligned}$$

- Hence, $\alpha(\boldsymbol{\lambda}_1, \boldsymbol{\mu}_1) + (1-\alpha)(\boldsymbol{\lambda}_2, \boldsymbol{\mu}_2) \in \text{dom}(q)$, and the convexity of dom(q) is established.
- (b) $L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu})$ is an affine function w.r.t. $(\boldsymbol{\lambda}, \boldsymbol{\mu})$.
- In particular, it is a concave function w.r.t. $(\boldsymbol{\lambda}, \boldsymbol{\mu})$.
- 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

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

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

$$\begin{aligned} q(\boldsymbol{\lambda}, \boldsymbol{\mu}) &= \min_{\mathbf{x} \in X} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) \leq \min_{\mathbf{x} \in S} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) \\ &= \min_{\mathbf{x} \in S} \left\{ f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x}) + \sum_{j=1}^{p} \mu_j h_j(\mathbf{x}) \right\} \\ &\leq \min_{\mathbf{x} \in S} f(\mathbf{x}) = f^*. \end{aligned}$$

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

Example

$$\begin{array}{ll} \min & x_1^2 - 3x_2^2 \\ \text{s.t.} & x_1 = x_2^3. \end{array}$$

In class

Strong Duality in the Convex Case - Back to Separation

Supporting Hyperplane Theorem Let $C \subseteq \mathbb{R}^n$ be a convex set and let $\mathbf{y} \notin C$. Then there exists $\mathbf{0} \neq \mathbf{p} \in \mathbb{R}^n$ such that

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

Proof.

- Although the theorem holds for any convex set C, we will prove it only for sets with a nonempty interior.
- Since $\mathbf{y} \notin \mathrm{int}(C)$, it follows that $\mathbf{y} \notin \mathrm{int}(\mathrm{cl}(C))$.
- Therefore, there exists a sequence $\{\mathbf{y}_k\}_{k\geq 1}$ such that $\mathbf{y}_k \notin \mathrm{cl}(C)$ and $\mathbf{y}_k \to \mathbf{y}$.
- By the separation theorem of a point from a closed and convex set, there exists $\mathbf{0} \neq \mathbf{p}_k \in \mathbb{R}^n$ such that

$$\mathbf{p}_k^T\mathbf{x} < \mathbf{p}_k^T\mathbf{y}_k \quad \forall \mathbf{x} \in \mathrm{cl}(C)$$

- Thus,

$$\frac{\mathbf{p}_k^T}{\|\mathbf{p}_k\|}(\mathbf{x} - \mathbf{y}_k) < 0 \text{ for any } \mathbf{x} \in \mathrm{cl}(C). \tag{6}$$

Proof Contd.

- Since the sequence $\left\{\frac{\mathbf{p}_k}{\|\mathbf{p}_k\|}\right\}$ is bounded, it follows that there exists a subsequence $\left\{\frac{\mathbf{p}_k}{\|\mathbf{p}_k\|}\right\}_{k\in T}$ such that $\frac{\mathbf{p}_k}{\|\mathbf{p}_k\|} \to \mathbf{p}$ as $k \xrightarrow{T} \infty$ for some $\mathbf{p} \in \mathbb{R}^n$.
- Obviously, $\|\mathbf{p}\| = 1$ and hence in particular $\mathbf{p} \neq 0$.
- Taking the limit as $k \xrightarrow{T} \infty$ in inequality (6) we obtain that

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

which readily implies the result since $C \subseteq \mathrm{cl}(C)$.

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 $\mathbf{0} \neq \mathbf{p} \in \mathbb{R}^n$ for which

$$\mathbf{p}^T\mathbf{x} \leq \mathbf{p}^T\mathbf{y} \text{ for any } \mathbf{x} \in C_1, \mathbf{y} \in C_2.$$

Proof.

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

$$\mathbf{p}^T(\mathbf{x} - \mathbf{y}) \leq \mathbf{p}^T\mathbf{0} \text{ for any } \mathbf{x} \in C_1, \mathbf{y} \in C_2,$$

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 $\hat{\mathbf{x}} \in X$ such that

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

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

(a) the following implication holds:

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

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

$$\min_{\mathbf{x} \in X} \left\{ f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x}) \right\} \geq c. \tag{7}$$

Proof of (b)$\Rightarrow$ (a)

- Suppose that there exist $\lambda_1, \lambda_2, \ldots, \lambda_m \geq 0$ such that (7) holds, and let $\mathbf{x} \in X$ satisfy $g_i(\mathbf{x}) \leq 0, i = 1, 2, \ldots, m$.
- By (7) we have
$$f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x}) \geq c,$$
- Hence,
$$f(\mathbf{x}) \geq c - \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x}) \geq c.$$

Proof of (a) $\Rightarrow$ (b)

- Assume that the implication (a) holds.
- Consider the following two sets:

$$\begin{aligned} S &= \{\mathbf{u} = (u_0, u_1, \ldots, u_m) : \exists \mathbf{x} \in X, f(\mathbf{x}) \le u_0, g_i(\mathbf{x}) \le u_i, i = 1, 2, \ldots, m\}, \\ T &= \{(u_0, u_1, \ldots, u_m) : u_0 < c, u_1 \le 0, u_2 \le 0, \ldots, u_m \le 0\}. \end{aligned}$$

- S, T are nonempty and convex and in addition $S \cap T = \emptyset$.
- By the supporting hyperplane theorem, there exists a vector $\mathbf{a} = (a_0, a_1, \ldots, a_m) \neq \mathbf{0}$, such that

$$\min_{(u_0, u_1, \ldots, u_m) \in S} \sum_{j=0}^{m} a_j u_j \ge \max_{(u_0, u_1, \ldots, u_m) \in T} \sum_{j=0}^{m} a_j u_j. \tag{8}$$

- $\mathbf{a} \ge \mathbf{0}$.
- Since $\mathbf{a} \ge 0$, it follows that the right-hand side is $a_0 c$, and we thus obtained

$$\min_{(u_0, u_1, \ldots, u_m) \in S} \sum_{j=0}^{m} a_j u_j \ge a_0 c. \tag{9}$$

Proof of (a) $\Rightarrow$ (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 $u_i = g_i(\hat{\mathbf{x}})$, we can deduce that $\sum_{j=1}^m a_j g_j(\hat{\mathbf{x}}) \geq 0$, which is impossible since $g_j(\hat{\mathbf{x}}) < 0$ and $\mathbf{a} \neq \mathbf{0}$.
- Since $a_0 > 0$, we can divide (9) by a_0 to obtain

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

  where $\tilde{a}_j = \frac{a_j}{a_0}$.
- By the definition of S we have

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

  which combined with (10) yields the desired result

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

Strong Duality of Convex Problems with Inequality Constraints

Theorem. Consider the optimization problem

$$\begin{array}{rll} f^* = & \min & f(\mathbf{x}) \\ & \text{s.t.} & g_i(\mathbf{x}) \leq 0, \quad i = 1, 2, \ldots, m, \ , \\ & & \mathbf{x} \in X, \end{array} \tag{11}$$

where X is a convex set and $f, g_i, i = 1, 2, \ldots, m$ are convex functions over X. Suppose that there exists $\hat{\mathbf{x}} \in X$ for which $g_i(\hat{\mathbf{x}}) < 0, i = 1, 2, \ldots, m$. 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:

$$\mathbf{x} \in X, g_i(\mathbf{x}) \le 0, i = 1, 2, \ldots, m \Rightarrow f(\mathbf{x}) \ge f^*,$$

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

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

- By the weak duality theorem,

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

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

Example

$$\begin{array}{ll} \min & x_1^2 - x_2 \\ \text{s.t.} & x_2^2 \leq 0. \end{array}$$

In class

Duffin's Duality Gap

$$\min\left\{e^{-x_2} : \sqrt{x_1^2 + x_2^2} - x_1 \le 0\right\}.$$

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

$$\begin{aligned}
\sqrt{x_1^2 + x_2^2} - x_1 &= \sqrt{\frac{(x_2^2 - \varepsilon^2)}{4\varepsilon^2} + x_2^2} - \frac{x_2^2 - \varepsilon^2}{2\varepsilon} = \sqrt{\frac{(x_2^2 + \varepsilon^2)^2}{4\varepsilon^2}} - \frac{x_2^2 - \varepsilon^2}{2\varepsilon} \\
&= \frac{x_2^2 + \varepsilon^2}{2\varepsilon} - \frac{x_2^2 - \varepsilon^2}{2\varepsilon} = \varepsilon.
\end{aligned}$$

- 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 $\lambda \ge 0$.
- $q^* = 0 \Rightarrow f^* - q^* = 1 \Rightarrow$ duality gap of 1.

Complementary Slackness Conditions

Theorem. Consider the optimization problem

$$f^* = \min\{f(\mathbf{x}) : g_i(\mathbf{x}) \le 0, i = 1, 2, \ldots, m, \mathbf{x} \in X\}, \qquad (12)$$

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

$$\mathbf{x}^* \in \operatorname{argmin} L_{\mathbf{x} \in X}(\mathbf{x}, \boldsymbol{\lambda}^*), \qquad (13)$$

$$\lambda_i^* g_i(\mathbf{x}^*) = 0, i = 1, 2, \ldots, m. \qquad (14)$$

Proof.

- $q(\boldsymbol{\lambda}^*) = \min_{\mathbf{x} \in X} L(\mathbf{x}, \boldsymbol{\lambda}^*) \le L(\mathbf{x}^*, \boldsymbol{\lambda}^*) = f(\mathbf{x}^*) + \sum_{i=1}^m \lambda_i^* g_i(\mathbf{x}^*) \le f(\mathbf{x}^*)$
- By strong duality, $\mathbf{x}^*, \boldsymbol{\lambda}^*$ are optimal iff $f(\mathbf{x}^*) = q(\boldsymbol{\lambda}^*)$
- iff $\min_{\mathbf{x} \in X} L(\mathbf{x}, \boldsymbol{\lambda}^*) = L(\mathbf{x}^*, \boldsymbol{\lambda}^*)$, $\sum_{i=1}^m \lambda_i^* g_i(\mathbf{x}^*) = 0$.
- iff (13), (14) hold.

A More General Strong Duality Theorem

Theorem. Consider the optimization problem

$$f^* = \begin{array}{ll} \min & f(\mathbf{x}) \\ \text{s.t.} & g_i(\mathbf{x}) \le 0, \quad i = 1, 2, \ldots, m, \\ & h_j(\mathbf{x}) \le 0, \quad j = 1, 2, \ldots, p, \\ & s_k(\mathbf{x}) = 0, \quad k = 1, 2, \ldots, q, \\ & \mathbf{x} \in X, \end{array} \tag{15}$$

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 $\hat{\mathbf{x}} \in \mathrm{int}(X)$ for which $g_i(\hat{\mathbf{x}}) < 0, h_j(\hat{\mathbf{x}}) \le 0, s_k(\hat{\mathbf{x}}) = 0$. Then if problem (15) has a finite optimal value, then the optimal value of the dual problem

$$q^* = \max\{q(\boldsymbol{\lambda}, \boldsymbol{\eta}, \boldsymbol{\mu}) : (\boldsymbol{\lambda}, \boldsymbol{\eta}, \boldsymbol{\mu}) \in \mathrm{dom}(q)\},$$

where

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

is attained, and $f^* = q^*$.

Importance of the Underlying Set

$$
\begin{array}{lll}
 & \min & x_1^3 + x_2^3 \\
(\text{P}) & \text{s.t.} & x_1 + x_2 \geq 1, \\
 & & x_1, x_2 \geq 0.
\end{array}
$$

- $(\frac{1}{2}, \frac{1}{2})$ is the optimal solution of (P) with an optimal value $f^* = \frac{1}{4}$.
- First dual problem is constructed by taking $X = \{(x_1, x_2) : x_1, x_2 \geq 0\}$.
- The primal problem is $\min\{x_1^3 + x_2^3 : x_1 + x_2 \geq 1, (x_1, x_2) \in X\}$.
- Strong duality holds for the problem and hence in particular $q^* = \frac{1}{4}$.
- Second dual is constructed by taking $X = \mathbb{R}^2$.
- Objective function is not convex $\Rightarrow$ strong duality is not necessarily satisfied.
- $L(x_1, x_2, \lambda, \eta_1, \eta_2) = x_1^3 + x_2^3 - \lambda(x_1 + x_2 - 1) - \eta_1 x_1 - \eta_2 x_2$.
- $q(\lambda, \eta_1, \eta_2) = -\infty$ for all $(\lambda, \mu_1, \mu_2) \Rightarrow q^* = -\infty$.

Linear Programming

Consider the linear programming problem

$$\begin{array}{ll} \min & \mathbf{c}^T\mathbf{x} \\ \text{s.t.} & \mathbf{A}\mathbf{x} \le \mathbf{b}, \end{array}$$

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

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

- Dual problem:

$$\begin{array}{ll} \max & -\mathbf{b}^T\boldsymbol{\lambda} \\ \text{s.t.} & \mathbf{A}^T\boldsymbol{\lambda} = -\mathbf{c}, \\ & \boldsymbol{\lambda} \ge \mathbf{0}. \end{array}$$

Strictly Convex Quadratic Programming

Consider the strictly convex quadratic programming problem

$$\begin{array}{ll} \min & \mathbf{x}^T\mathbf{Q}\mathbf{x} + 2\mathbf{f}^T\mathbf{x} \\ \text{s.t.} & \mathbf{A}\mathbf{x} \le \mathbf{b}, \end{array} \tag{16}$$

- $\mathbf{Q} \in \mathbb{R}^{n\times n}$ positive definite, $\mathbf{f} \in \mathbb{R}^n, \mathbf{A} \in \mathbb{R}^{m\times n}$, $\mathbf{b} \in \mathbb{R}^m$.
- Lagrangian: $(\boldsymbol{\lambda} \in \mathbb{R}^m_+)$ $\quad L(\mathbf{x}, \boldsymbol{\lambda}) = \mathbf{x}^T\mathbf{Q}\mathbf{x} + 2\mathbf{f}^T\mathbf{x} + 2\boldsymbol{\lambda}^T(\mathbf{A}\mathbf{x} - \mathbf{b}) = \mathbf{x}^T\mathbf{Q}\mathbf{x} + 2(\mathbf{A}^T\boldsymbol{\lambda} + \mathbf{f})^T\mathbf{x} - 2\mathbf{b}^T\boldsymbol{\lambda}$.
- The minimizer of the Lagrangian is attained at $\mathbf{x}^* = -\mathbf{Q}^{-1}(\mathbf{f} + \mathbf{A}^T\boldsymbol{\lambda})$.
- $$\begin{aligned} q(\boldsymbol{\lambda}) &= L(\mathbf{x}^*, \boldsymbol{\lambda}) \\ &= (\mathbf{f} + \mathbf{A}^T\boldsymbol{\lambda})^T\mathbf{Q}^{-1}\mathbf{Q}\mathbf{Q}^{-1}(\mathbf{f} + \mathbf{A}^T\boldsymbol{\lambda}) - 2(\mathbf{f} + \mathbf{A}^T\boldsymbol{\lambda})^T\mathbf{Q}^{-1}(\mathbf{f} + \mathbf{A}^T\boldsymbol{\lambda}) - 2\mathbf{b}^T\boldsymbol{\lambda} \\ &= -(\mathbf{f} + \mathbf{A}^T\boldsymbol{\lambda})^T\mathbf{Q}^{-1}(\mathbf{f} + \mathbf{A}^T\boldsymbol{\lambda}) - 2\mathbf{b}^T\boldsymbol{\lambda} \\ &= -\boldsymbol{\lambda}^T\mathbf{A}\mathbf{Q}^{-1}\mathbf{A}^T\boldsymbol{\lambda} - 2\mathbf{f}^T\mathbf{Q}^{-1}\mathbf{A}^T\boldsymbol{\lambda} - \mathbf{f}^T\mathbf{Q}^{-1}\mathbf{f} - 2\mathbf{b}^T\boldsymbol{\lambda} \\ &= -\boldsymbol{\lambda}^T\mathbf{A}\mathbf{Q}^{-1}\mathbf{A}^T\boldsymbol{\lambda} - 2(\mathbf{A}\mathbf{Q}^{-1}\mathbf{f} + \mathbf{b})^T\boldsymbol{\lambda} - \mathbf{f}^T\mathbf{Q}^{-1}\mathbf{f}. \end{aligned}$$
- The dual problem is $\max\{q(\boldsymbol{\lambda}) : \boldsymbol{\lambda} \ge \mathbf{0}\}$.

Dual of Convex QCQP with strictly convex objective

Consider the QCQP problem

$$\begin{array}{ll} \min & \mathbf{x}^T\mathbf{A}_0\mathbf{x} + 2\mathbf{b}_0^T\mathbf{x} + c_0 \\ \text{s.t.} & \mathbf{x}^T\mathbf{A}_i\mathbf{x} + 2\mathbf{b}_i^T\mathbf{x} + c_i \leq 0, \quad i = 1, 2, \ldots, m, \end{array}$$

where $\mathbf{A}_i \succeq \mathbf{0}$ is an $n \times n$ matrix, $\mathbf{b}_i \in \mathbb{R}^n, c_i \in \mathbb{R}, i = 0, 1, \ldots, m$.

Assume that $\mathbf{A}_0 \succ \mathbf{0}$.

- Lagrangian ($\boldsymbol{\lambda} \in \mathbb{R}^m_+$):

$$\begin{aligned} L(\mathbf{x}, \boldsymbol{\lambda}) &= \mathbf{x}^T\mathbf{A}_0\mathbf{x} + 2\mathbf{b}_0^T\mathbf{x} + c_0 + \sum_{i=1}^m \lambda_i(\mathbf{x}^T\mathbf{A}_i\mathbf{x} + 2\mathbf{b}_i^T\mathbf{x} + c_i) \\ &= \mathbf{x}^T\left(\mathbf{A}_0 + \textstyle\sum_{i=1}^m \lambda_i\mathbf{A}_i\right)\mathbf{x} + 2\left(\mathbf{b}_0 + \textstyle\sum_{i=1}^m \lambda_i\mathbf{b}_i\right)^T\mathbf{x} + c_0 + \textstyle\sum_{i=1}^m \lambda_i c_i. \end{aligned}$$

- The minimizer of the Lagrangian w.r.t. $\mathbf{x}$ is attained at $\tilde{\mathbf{x}}$ satisfying

$$2\left(\mathbf{A}_0 + \textstyle\sum_{i=1}^m \lambda_i\mathbf{A}_i\right)\tilde{\mathbf{x}} = -2\left(\mathbf{b}_0 + \textstyle\sum_{i=1}^m \lambda_i\mathbf{b}_i\right).$$

- Thus, $\tilde{\mathbf{x}} = -\left(\mathbf{A}_0 + \sum_{i=1}^m \lambda_i\mathbf{A}_i\right)^{-1}\left(\mathbf{b}_0 + \sum_{i=1}^m \lambda_i\mathbf{b}_i\right).$

QCQP contd.

- Plugging this expression back into the Lagrangian, we obtain the following expression for the dual objective function

$$
\begin{aligned}
q(\boldsymbol{\lambda}) &= \min_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}) = L(\tilde{\mathbf{x}}, \boldsymbol{\lambda}) \\
&= \tilde{\mathbf{x}}^T \left(\mathbf{A}_0 + \sum_{i=1}^m \lambda_i \mathbf{A}_i\right) \tilde{\mathbf{x}} + 2\left(\mathbf{b}_0 + \sum_{i=1}^m \lambda_i \mathbf{b}_i\right)^T \tilde{\mathbf{x}} + c_0 + \sum_{i=1}^m \lambda_i c_i \\
&= -\left(\mathbf{b}_0 + \sum_{i=1}^m \lambda_i \mathbf{b}_i\right)^T \left(\mathbf{A}_0 + \sum_{i=1}^m \lambda_i \mathbf{A}_i\right)^{-1} \left(\mathbf{b}_0 + \sum_{i=1}^m \lambda_i \mathbf{b}_i\right) + \\
&\quad\; c_0 + \sum_{i=1}^m \lambda_i c_i.
\end{aligned}
$$

- The dual problem is thus

$$
\begin{array}{ll}
\max & -\left(\mathbf{b}_0 + \sum_{i=1}^m \lambda_i \mathbf{b}_i\right)^T \left(\mathbf{A}_0 + \sum_{i=1}^m \lambda_i \mathbf{A}_i\right)^{-1} \left(\mathbf{b}_0 + \sum_{i=1}^m \lambda_i \mathbf{b}_i\right) + \\
& c_0 + \sum_{i=1}^m \lambda_i c_i \\
\text{s.t.} & \lambda_i \geq 0, \quad i = 1, 2, \ldots, m.
\end{array}
$$

Dual of Convex QCQPs

$\mathbf{A}_0$ is only assumed to be positive semidefinite.

- The previous dual is not well defined since the matrix $\mathbf{A}_0 + \sum_{i=1}^m \lambda_i \mathbf{A}_i$ is not necessarily PD.
- Decompose $\mathbf{A}_i$ as $\mathbf{A}_i = \mathbf{D}_i^T \mathbf{D}_i$ ($\mathbf{D}_i \in \mathbb{R}^{n \times n}$) and rewrite the problem as

$$\begin{array}{ll} \min & \mathbf{x}^T \mathbf{D}_0^T \mathbf{D}_0 \mathbf{x} + 2\mathbf{b}_0^T \mathbf{x} + c_0 \\ \text{s.t.} & \mathbf{x}^T \mathbf{D}_i^T \mathbf{D}_i \mathbf{x} + 2\mathbf{b}_i^T \mathbf{x} + c_i \leq 0, i = 1, 2, \ldots, m, \end{array}$$

- Define additional variables $\mathbf{z}_i = \mathbf{D}_i \mathbf{x}$, giving rise to the formulation

$$\begin{array}{ll} \min & \|\mathbf{z}_0\|^2 + 2\mathbf{b}_0^T \mathbf{x} + c_0 \\ \text{s.t.} & \|\mathbf{z}_i\|^2 + 2\mathbf{b}_i^T \mathbf{x} + c_i \leq 0, i = 1, 2, \ldots, m, \\ & \mathbf{z}_i = \mathbf{D}_i \mathbf{x}, \quad i = 0, 1, \ldots, m. \end{array}$$

Dual of Convex QCQPs

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

$$\begin{aligned}
& L(\mathbf{x}, \mathbf{z}_0, \ldots, \mathbf{z}_m, \boldsymbol{\lambda}, \boldsymbol{\mu}_0, \ldots, \boldsymbol{\mu}_m) \\
= \quad & \|\mathbf{z}_0\|^2 + 2\mathbf{b}_0^T\mathbf{x} + c_0 + \sum_{i=1}^m \lambda_i(\|\mathbf{z}_i\|^2 + 2\mathbf{b}_i^T\mathbf{x} + c_i) + \\
& 2\sum_{i=0}^m \boldsymbol{\mu}_i^T(\mathbf{z}_i - \mathbf{D}_i\mathbf{x}) \\
= \quad & \|\mathbf{z}_0\|^2 + 2\boldsymbol{\mu}_0^T\mathbf{z}_0 + \sum_{i=1}^m(\lambda_i\|\mathbf{z}_i\|^2 + 2\boldsymbol{\mu}_i^T\mathbf{z}_i) + \\
& 2\left(\mathbf{b}_0 + \sum_{i=1}^m \lambda_i\mathbf{b}_i - \sum_{i=0}^m \mathbf{D}_i^T\boldsymbol{\mu}_i\right)^T \mathbf{x} \\
& + c_0 + \sum_{i=1}^m c_i\lambda_i.
\end{aligned}$$

Dual of Convex QCQPs

- For any $\lambda \in \mathbb{R}_+, \boldsymbol{\mu} \in \mathbb{R}^n$,

$$g(\lambda, \boldsymbol{\mu}) \equiv \min_{\mathbf{z}} \left\{ \lambda \|\mathbf{z}\|^2 + 2\boldsymbol{\mu}^T \mathbf{z} \right\} = \begin{cases} -\frac{\|\boldsymbol{\mu}\|^2}{\lambda} & \lambda > 0, \\ 0 & \lambda = 0, \boldsymbol{\mu} = \mathbf{0}, \\ -\infty & \lambda = 0, \boldsymbol{\mu} \neq \mathbf{0}. \end{cases}$$

- Since the Lagrangian is separable with respect to $\mathbf{z}_i$ and $\mathbf{x}$, we can perform the minimization with respect to each of the variables vectors:

$$\begin{aligned} \min_{\mathbf{z}_0} \left[\|\mathbf{z}_0\|^2 + 2\boldsymbol{\mu}_0^T \mathbf{z}_0 \right] &= g(1, \boldsymbol{\mu}_0) = -\|\boldsymbol{\mu}_0\|^2, \\ \min_{\mathbf{z}_i} \left[\lambda_i \|\mathbf{z}_i\|^2 + 2\boldsymbol{\mu}_i^T \mathbf{z}_i \right] &= g(\lambda_i, \boldsymbol{\mu}_i), \end{aligned}$$

$$\min_{\mathbf{x}} \left(\mathbf{b}_0 + \sum_{i=1}^m \lambda_i \mathbf{b}_i - \sum_{i=0}^m \mathbf{D}_i^T \boldsymbol{\mu}_i \right)^T \mathbf{x} = \begin{cases} 0 & \mathbf{b}_0 + \sum_{i=1}^m \lambda_i \mathbf{b}_i - \sum_{i=0}^m \mathbf{D}_i^T \boldsymbol{\mu}_i = \mathbf{0}, \\ -\infty & \text{else}, \end{cases}$$

- Hence,

$$\begin{aligned} q(\boldsymbol{\lambda}, \boldsymbol{\mu}_0, \ldots, \boldsymbol{\mu}_m) &= \min_{\mathbf{x}, \mathbf{z}_0, \ldots, \mathbf{z}_m} L(\mathbf{x}, \mathbf{z}_0, \ldots, \mathbf{z}_m, \boldsymbol{\lambda}, \boldsymbol{\mu}_0, \ldots, \boldsymbol{\mu}_m) \\ &= \begin{cases} g(1, \boldsymbol{\mu}_0) + \sum_{i=1}^m g(\lambda_i, \boldsymbol{\mu}_i) + c_0 + \mathbf{c}^T \boldsymbol{\lambda} & \mathbf{b}_0 + \sum_{i=1}^m \lambda_i \mathbf{b}_i - \sum_{i=0}^m \mathbf{D}_i^T \boldsymbol{\mu}_i = \mathbf{0}, \\ -\infty & \text{else}. \end{cases} \end{aligned}$$

Dual of Convex QCQPs

The dual problem is therefore

$$\begin{array}{ll} \max & g(1,\boldsymbol{\mu}_0) + \sum_{i=1}^m g(\lambda_i,\boldsymbol{\mu}_i) + c_0 + \sum_{i=1}^m c_i\lambda_i \\ \text{s.t.} & \mathbf{b}_0 + \sum_{i=1}^m \lambda_i \mathbf{b}_i - \sum_{i=0}^m \mathbf{D}_i^T \boldsymbol{\mu}_i = \mathbf{0}, \\ & \boldsymbol{\lambda} \in \mathbb{R}^m_+, \boldsymbol{\mu}_0, \ldots, \boldsymbol{\mu}_m \in \mathbb{R}^n. \end{array}$$

Dual of Nonconvex QCQPs

Consider the problem

$$\begin{array}{ll} \min & \mathbf{x}^T\mathbf{A}_0\mathbf{x} + 2\mathbf{b}_0^T\mathbf{x} + c_0 \\ \text{s.t.} & \mathbf{x}^T\mathbf{A}_i\mathbf{x} + 2\mathbf{b}_i^T\mathbf{x} + c_i \leq 0, \quad i = 1, 2, \ldots, m, \end{array}$$

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

$$\begin{aligned} L(\mathbf{x}, \boldsymbol{\lambda}) &= \mathbf{x}^T\mathbf{A}_0\mathbf{x} + 2\mathbf{b}_0^T\mathbf{x} + c_0 + \sum_{i=1}^m \lambda_i \left(\mathbf{x}^T\mathbf{A}_i\mathbf{x} + 2\mathbf{b}_i^T\mathbf{x} + c_i\right) \\ &= \mathbf{x}^T\left(\mathbf{A}_0 + \sum_{i=1}^m \lambda_i\mathbf{A}_i\right)\mathbf{x} + 2\left(\mathbf{b}_0 + \sum_{i=1}^m \lambda_i\mathbf{b}_i\right)^T\mathbf{x} + c_0 + \sum_{i=1}^m c_i\lambda_i. \end{aligned}$$

- Note that

$$q(\boldsymbol{\lambda}) = \min_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}) = \max_t\{t : L(\mathbf{x}, \boldsymbol{\lambda}) \geq t \text{ for any } \mathbf{x} \in \mathbb{R}^n\}.$$

Dual of Nonconvex QCQPs

- The following holds:

$$L(\mathbf{x}, \boldsymbol{\lambda}) \geq t \text{ for all } \mathbf{x} \in \mathbb{R}^n$$

is equivalent to

$$\begin{pmatrix} \mathbf{A}_0 + \sum_{i=1}^m \lambda_i \mathbf{A}_i & \mathbf{b}_0 + \sum_{i=1}^m \lambda_i \mathbf{b}_i \\ (\mathbf{b}_0 + \sum_{i=1}^m \lambda_i \mathbf{b}_i)^T & c_0 + \sum_{i=1}^m \lambda_i c_i - t \end{pmatrix} \succeq \mathbf{0},$$

- Therefore, the dual problem is

$$\begin{array}{ll} \max_{t,\lambda_i} & t \\ \text{s.t.} & \begin{pmatrix} \mathbf{A}_0 + \sum_{i=1}^m \lambda_i \mathbf{A}_i & \mathbf{b}_0 + \sum_{i=1}^m \lambda_i \mathbf{b}_i \\ (\mathbf{b}_0 + \sum_{i=1}^m \lambda_i \mathbf{b}_i)^T & c_0 + \sum_{i=1}^m \lambda_i c_i - t \end{pmatrix} \succeq \mathbf{0}, \\ & \lambda_i \geq 0, \quad i = 1, 2, \ldots, m. \end{array}$$

Orthogonal Projection onto the Unit Simplex

- Given a vector $\mathbf{y} \in \mathbb{R}^n$, the orthogonal projection of $\mathbf{y}$ onto Δ_n is the solution to

$$\begin{array}{ll} \min & \|\mathbf{x} - \mathbf{y}\|^2 \\ \text{s.t.} & \mathbf{e}^T\mathbf{x} = 1, \\ & \mathbf{x} \geq \mathbf{0}. \end{array}$$

- Lagrangian:

$$\begin{aligned} L(\mathbf{x}, \lambda) &= \|\mathbf{x} - \mathbf{y}\|^2 + 2\lambda(\mathbf{e}^T\mathbf{x} - 1) = \|\mathbf{x}\|^2 - 2(\mathbf{y} - \lambda\mathbf{e})^T\mathbf{x} + \|\mathbf{y}\|^2 - 2\lambda \\ &= \sum_{j=1}^n (x_j^2 - 2(y_j - \lambda)x_j) + \|\mathbf{y}\|^2 - 2\lambda. \end{aligned}$$

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

$$\max_{\lambda \in \mathbb{R}} \left\{ g(\lambda) \equiv -\sum_{j=1}^n [y_j - \lambda]_+^2 - 2\lambda + \|\mathbf{y}\|^2 \right\}.$$

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 λ^* 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 $(-\infty, \max_j y_j]$.
- $$\begin{aligned} h(y_{\max}) &= -1, \\ h\left(y_{\min} - \frac{2}{n}\right) &= \sum_{j=1}^n y_j - ny_{\min} + 2 - 1 > 0, \end{aligned}$$

  where $y_{\max} = \max_{j=1,2,\ldots,n} y_j$, $y_{\min} = \min_{j=1,2,\ldots,n} y_j$.
- We can therefore invoke a bisection procedure to find the unique root λ^* of the function h over the interval $[y_{\min} - \frac{2}{n}, y_{\max}]$, and then define $P_{\Delta_n}(\mathbf{y}) = [\mathbf{y} - \lambda^*\mathbf{e}]_+$.

Orthogonal Projection Onto the Unit Simplex

The MATLAB function proj_unit_simplex:

```
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:

$$\begin{array}{ll} \min_{\mathbf{x},r} & r \\ \text{s.t.} & \|\mathbf{x}-\mathbf{a}_i\| \le r, \quad i=1,2,\ldots,m. \end{array}$$

- Reformulation:

$$\begin{array}{ll} \min_{\mathbf{x},\gamma} & \gamma \\ \text{s.t.} & \|\mathbf{x}-\mathbf{a}_i\|^2 \le \gamma, \quad i=1,2,\ldots,m. \end{array}$$

-

$$\begin{aligned} L(\mathbf{x},\gamma,\boldsymbol{\lambda}) &= \gamma + \sum_{i=1}^m \lambda_i(\|\mathbf{x}-\mathbf{a}_i\|^2 - \gamma) \\ &= \gamma\left(1-\sum_{i=1}^m \lambda_i\right) + \sum_{i=1}^m \lambda_i \|\mathbf{x}-\mathbf{a}_i\|^2. \end{aligned}$$

- 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 $\min_{\mathbf{x}} \sum_{i=1}^m \lambda_i \|\mathbf{x} - \mathbf{a}_i\|^2$.
- We have

$$\sum_{i=1}^m \lambda_i \|\mathbf{x} - \mathbf{a}_i\|^2 = \|\mathbf{x}\|^2 - 2\left(\sum_{i=1}^m \lambda_i \mathbf{a}_i\right)^T \mathbf{x} + \sum_{i=1}^m \lambda_i \|\mathbf{a}_i\|^2, \qquad (17)$$

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

$$\mathbf{x}^* = \sum_{i=1}^m \lambda_i \mathbf{a}_i = \mathbf{A}\boldsymbol{\lambda},$$

  $\mathbf{A}$ is the $n \times m$ matrix whose columns are $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m$.
- Substituting this expression back into (17),

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

- The dual problem is therefore

$$\begin{array}{ll} \max & -\|\mathbf{A}\boldsymbol{\lambda}\|^2 + \sum_{i=1}^m \lambda_i \|\mathbf{a}_i\|^2 \\ \text{s.t.} & \boldsymbol{\lambda} \in \Delta_m. \end{array}$$

MATLAB code

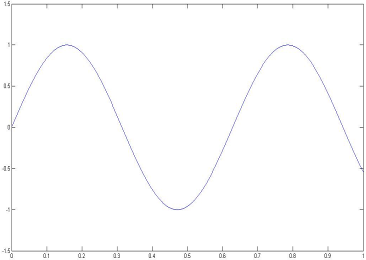
```
function [xp,r]=chebyshev_center(A)

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
```

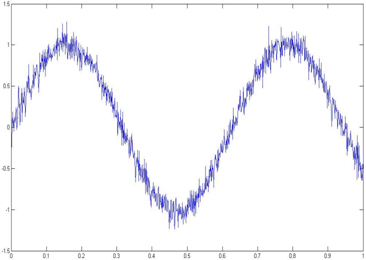
Denoising

Suppose that we are given a signal contaminated with noise.

$$\mathbf{y} = \mathbf{x} + \mathbf{w},$$

x - unknown "true" signal, **w** - unknown noise, **y** - known observed signal.





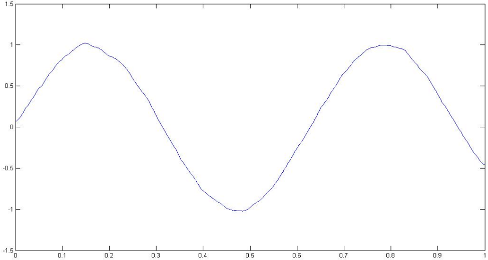
The denoising problem: find a "good" estimate for **x** given **y**.

A Tikhonov Regularization Approach

Quadratic Penalty:

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

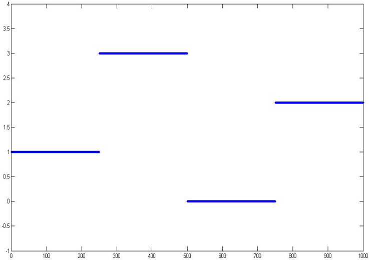
The solution with $\lambda = 1$:

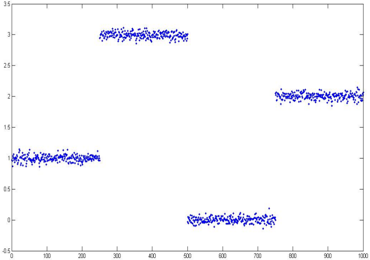


Pretty good!

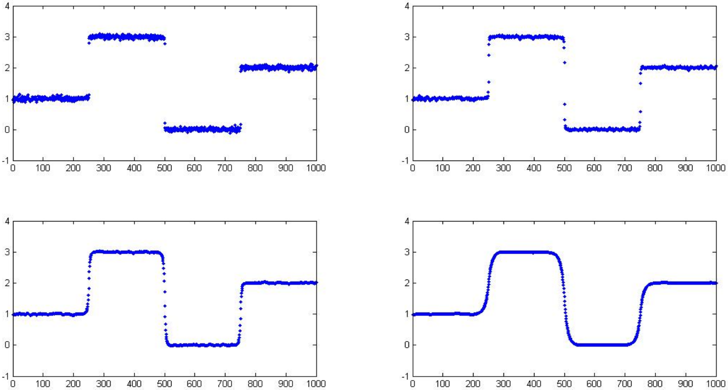
Weakness of Quadratic Regularization

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





Failure of Quadratic Regularization



l_1 regularization

$$\min \|\mathbf{x} - \mathbf{y}\|^2 + \lambda \|\mathbf{L}\mathbf{x}\|_1. \tag{18}$$

- The problem is equivalent to the optimization problem

$$\begin{array}{ll} \min_{\mathbf{x},\mathbf{z}} & \|\mathbf{x} - \mathbf{y}\|^2 + \lambda \|\mathbf{z}\|_1 \\ \text{s.t.} & \mathbf{z} = \mathbf{L}\mathbf{x}. \end{array}$$

  $\mathbf{L}$ 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

$$\begin{aligned} L(\mathbf{x}, \mathbf{z}, \boldsymbol{\mu}) &= \|\mathbf{x} - \mathbf{y}\|^2 + \lambda \|\mathbf{z}\|_1 + \boldsymbol{\mu}^T (\mathbf{L}\mathbf{x} - \mathbf{z}) \\ &= \|\mathbf{x} - \mathbf{y}\|^2 + (\mathbf{L}^T \boldsymbol{\mu})^T \mathbf{x} + \lambda \|\mathbf{z}\|_1 - \boldsymbol{\mu}^T \mathbf{z}. \end{aligned}$$

- The dual problem is

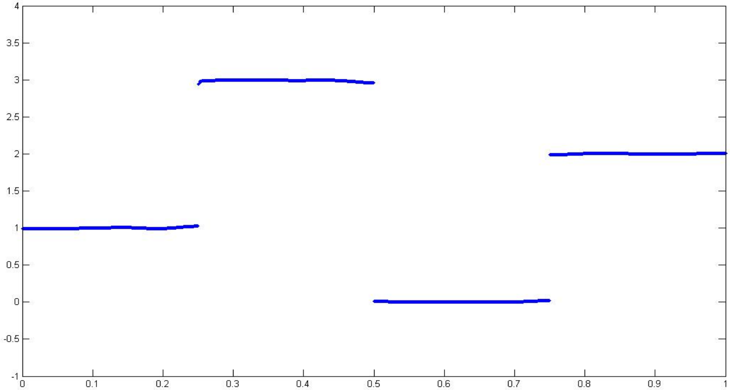
$$\begin{array}{ll} \max & -\frac{1}{4} \boldsymbol{\mu}^T \mathbf{L}\mathbf{L}^T \boldsymbol{\mu} + \boldsymbol{\mu}^T \mathbf{L}\mathbf{y} \\ \text{s.t.} & \|\boldsymbol{\mu}\|_\infty \le \lambda. \end{array} \tag{19}$$

A MATLAB code

Employing the gradient projection method on the dual:

```
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

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

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

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

where $C > 0$ is a penalty parameter.

Dual SVM

- The same as

$$\begin{array}{ll} \min & \frac{1}{2}\|\mathbf{w}\|^2 + C(\mathbf{e}^T\boldsymbol{\xi}) \\ \text{s.t.} & \mathbf{Y}(\mathbf{X}\mathbf{w} + \beta\mathbf{e}) \geq \mathbf{e} - \boldsymbol{\xi}, \\ & \boldsymbol{\xi} \geq \mathbf{0}, \end{array}$$

 where $\mathbf{Y} = \text{diag}(y_1, y_2, \ldots, y_m)$ and $\mathbf{X}$ is the $m \times n$ matrix whose rows are $\mathbf{x}_1^T, \mathbf{x}_2^T, \ldots, \mathbf{x}_m^T$.
- Lagrangian ($\boldsymbol{\alpha} \in \mathbb{R}^m_+$):

$$\begin{aligned} L(\mathbf{w}, \beta, \boldsymbol{\xi}, \boldsymbol{\alpha}) &= \frac{1}{2}\|\mathbf{w}\|^2 + C(\mathbf{e}^T\boldsymbol{\xi}) - \boldsymbol{\alpha}^T[\mathbf{Y}\mathbf{X}\mathbf{w} + \beta\mathbf{Y}\mathbf{e} - \mathbf{e} + \boldsymbol{\xi}] \\ &= \frac{1}{2}\|\mathbf{w}\|^2 - \mathbf{w}^T[\mathbf{X}^T\mathbf{Y}\boldsymbol{\alpha}] - \beta(\boldsymbol{\alpha}^T\mathbf{Y}\mathbf{e}) + \boldsymbol{\xi}^T(C\mathbf{e} - \boldsymbol{\alpha}) + \boldsymbol{\alpha}^T\mathbf{e}. \end{aligned}$$

-

$$q(\boldsymbol{\alpha}) = \left[\min_{\mathbf{w}} \frac{1}{2}\|\mathbf{w}\|^2 - \mathbf{w}^T[\mathbf{X}^T\mathbf{Y}\boldsymbol{\alpha}]\right] + \left[\min_{\beta}(-\beta(\boldsymbol{\alpha}^T\mathbf{Y}\mathbf{e}))\right] + \left[\min_{\boldsymbol{\xi}\geq\mathbf{0}} \boldsymbol{\xi}^T(C\mathbf{e} - \boldsymbol{\alpha})\right] + \boldsymbol{\alpha}^T\mathbf{e}.$$

Dual SVM

- $$\begin{aligned} \min_{\mathbf{w}} \frac{1}{2}\|\mathbf{w}\|^2 - \mathbf{w}^T[\mathbf{X}^T\mathbf{Y}\boldsymbol{\alpha}] &= -\frac{1}{2}\boldsymbol{\alpha}^T\mathbf{Y}\mathbf{X}\mathbf{X}^T\mathbf{Y}\boldsymbol{\alpha}, \\ \min_{\beta}(-\beta(\boldsymbol{\alpha}^T\mathbf{Y}\mathbf{e})) &= \begin{cases} 0 & \boldsymbol{\alpha}^T\mathbf{Y}\mathbf{e} = 0, \\ -\infty & \text{else}, \end{cases} \\ \min_{\boldsymbol{\xi}\geq\mathbf{0}} \boldsymbol{\xi}^T(C\mathbf{e} - \boldsymbol{\alpha}) &= \begin{cases} 0 & \boldsymbol{\alpha} \leq C\mathbf{e}, \\ -\infty & \text{else}, \end{cases} \end{aligned}$$
- Therefore, the dual objective function is given by
$$q(\boldsymbol{\alpha}) = \begin{cases} \boldsymbol{\alpha}^T\mathbf{e} - \frac{1}{2}\boldsymbol{\alpha}^T\mathbf{Y}\mathbf{X}\mathbf{X}^T\mathbf{Y}\boldsymbol{\alpha} & \boldsymbol{\alpha}^T\mathbf{Y}\mathbf{e} = 0, \mathbf{0} \leq \boldsymbol{\alpha} \leq C\mathbf{e} \\ -\infty & \text{else.} \end{cases}$$
- The dual problem is $\begin{array}{ll} \max & \boldsymbol{\alpha}^T\mathbf{e} - \frac{1}{2}\boldsymbol{\alpha}^T\mathbf{Y}\mathbf{X}\mathbf{X}^T\mathbf{Y}\boldsymbol{\alpha} \\ \text{s.t.} & \boldsymbol{\alpha}^T\mathbf{Y}\mathbf{e} = 0, \\ & \mathbf{0} \leq \boldsymbol{\alpha} \leq C\mathbf{e}. \end{array}$
- or
$$\begin{array}{ll} \max & \sum_{i=1}^m \alpha_i - \frac{1}{2}\sum_{i=1}^m \sum_{j=1}^m \alpha_i\alpha_j y_i y_j (\mathbf{x}_i^T\mathbf{x}_j) \\ \text{s.t.} & \sum_{i=1}^m y_i\alpha_i = 0, \\ & 0 \leq \alpha_i \leq C, \quad i = 1, 2, \ldots, m. \end{array}$$