Lecture 8: Constrained Optimization: Duality Theory

- Convex Cost/Linear Constraints
- Duality Theorem
- Linear Programming Duality
- Quadratic Programming Duality

A General Sufficiency Condition

Consider the problem

$$\begin{array}{ll} \text{minimize} & f(\boldsymbol{x}) \\ \text{subject to} & \boldsymbol{x} \in X,\ g_j(\boldsymbol{x}) \leq 0,\ j = 1, ..., r. \end{array}$$

Let $\boldsymbol{x}^*$ be feasible and $\boldsymbol{\mu}^*$ satisfy

$$\mu_j^* \geq 0,\ j = 1, ..., r,\ \mu_j^* = 0,\ \forall\, j \notin A(\boldsymbol{x}^*),\ \boldsymbol{x}^* = \arg\min_{\boldsymbol{x} \in X} L(\boldsymbol{x}, \boldsymbol{\mu}^*).$$

Then $\boldsymbol{x}^*$ is a global minimum of the problem.

Proof: We have

$$\begin{aligned} f(\boldsymbol{x}^*) &= f(\boldsymbol{x}^*) + (\boldsymbol{\mu}^*)'\boldsymbol{g}(\boldsymbol{x}^*) = \min_{\boldsymbol{x} \in X} \{f(\boldsymbol{x}) + (\boldsymbol{\mu}^*)'\boldsymbol{g}(\boldsymbol{x})\} \\ &\leq \min_{\boldsymbol{x} \in X, \boldsymbol{g}(\boldsymbol{x}) \leq \boldsymbol{0}} \{f(\boldsymbol{x}) + (\boldsymbol{\mu}^*)'\boldsymbol{g}(\boldsymbol{x})\} \\ &\leq \min_{\boldsymbol{x} \in X, \boldsymbol{g}(\boldsymbol{x}) \leq \boldsymbol{0}} f(\boldsymbol{x}), \end{aligned}$$

where the first equality follows from the hypothesis, which implies that $(\boldsymbol{\mu}^*)'\boldsymbol{g}(\boldsymbol{x}^*) = 0$, and the last inequality follows from the nonnegativity of $\boldsymbol{\mu}^*$.

Lagrangian Multiplier Result

Let $f : \mathbb{R}^n \mapsto \mathbb{R}$ be convex, continuously differentiable. Consider a linear inequality constrained problem

$$\boxed{\begin{array}{ll} \text{minimize} & f(\boldsymbol{x}) \\ \text{subject to} & \boldsymbol{a}_j'\boldsymbol{x} \le b_j, \quad j = 1, ..., r. \end{array}}$$

- Let $J \subseteq 1, ..., r$. Then $\boldsymbol{x}^*$ is a global min if and only if $\boldsymbol{x}^*$ is feasible and there exist $\mu_j^* \ge 0,\ j \in J$, such that $\mu_j^* = 0$ for all $j \in J \notin A(\boldsymbol{x}^*)$, and

$$\boldsymbol{x}^* = \underset{\boldsymbol{a}_j'\boldsymbol{x} \le b_j,\ j \notin J}{\arg\min} \left\{ f(\boldsymbol{x}) + \sum_{j \in J} \mu_j^*(\boldsymbol{a}_j'\boldsymbol{x} - b_j) \right\}.$$

- Proof: Assume $\boldsymbol{x}^*$ is global min. Then there exist $\mu_j^* \ge 0$, such that $\mu_j^*(\boldsymbol{a}_j'\boldsymbol{x}^* - b_j) = 0$ for all j and $\nabla f(\boldsymbol{x}^*) + \sum_{j=1}^r \mu_j^* \boldsymbol{a}_j = 0$, implying

$$\boldsymbol{x}^* = \arg\min_{\boldsymbol{x} \in \mathbb{R}^n} \left\{ f(\boldsymbol{x}) + \sum_{j=1}^r \mu_j^*(\boldsymbol{a}_j'\boldsymbol{x} - b_j) \right\}.$$

- Since $\mu_j^*(\boldsymbol{a}_j'\boldsymbol{x}^* - b_j) = 0$ for all j, we have

$$f(\boldsymbol{x}^*) = \min_{\boldsymbol{x}\in\mathbb{R}^n} \left\{ f(\boldsymbol{x}) + \sum_{j=1}^{r} \mu_j^*(\boldsymbol{a}_j'x - b_j) \right\}.$$

Since $\mu_j^*(\boldsymbol{a}_j'\boldsymbol{x} - b_j) \leq 0$ if $\boldsymbol{a}_j'\boldsymbol{x} - b_j \leq 0$,

$$\begin{aligned} f(\boldsymbol{x}^*) &\leq \min_{\boldsymbol{a}_j'\boldsymbol{x}\leq b_j,\ j\notin J} \left\{ f(\boldsymbol{x}) + \sum_{j=1}^{r} \mu_j^*(\boldsymbol{a}_j'\boldsymbol{x} - b_j) \right\} \\ &\leq \min_{\boldsymbol{a}_j'\boldsymbol{x}\leq b_j,\ j\notin J} \left\{ f(\boldsymbol{x}) + \sum_{j\in J} \mu_j^*(\boldsymbol{a}_j'\boldsymbol{x} - b_j) \right\}. \end{aligned}$$

- Conversely, if $\boldsymbol{x}^*$ is feasible and there exist scalars $\mu_j^*,\ j \in J$ with the stated properties, then $\boldsymbol{x}^*$ is a global min by the general sufficiency condition of the preceding slide (where X is taken to be the set of $\boldsymbol{x}$ such that $\boldsymbol{a}_j'\boldsymbol{x} \leq b_j$ for all $j \notin J$). Q.E.D.

- Interesting observation: The same set of μ_j^* works for all index sets J.

- The flexibility to split the set of constraints into those that are handled by Lagrange multipliers (set J) and those that are handled explicitly comes handy in many analytical and computational contexts.

The Dual Problem

Consider the problem

$$\min_{\boldsymbol{x}\in X,\ \boldsymbol{a}_j'\boldsymbol{x}\le b_j,\ j=1,\ldots,r} f(\boldsymbol{x})$$

where f is convex and continuously differentiable over $\mathbb{R}^n$ and X is polyhedral.

- Define the dual function $q : \mathbb{R}^r \mapsto [-\infty, \infty)$

  $$q(\boldsymbol{\mu}) = \inf_{\boldsymbol{x}\in X} L(\boldsymbol{x}, \boldsymbol{\mu}) = \inf_{\boldsymbol{x}\in X} \{f(\boldsymbol{x}) + \sum_{j=1}^{r} \mu_j(\boldsymbol{a}_j' x - b_j)\}$$

  and the dual problem

  $$\max_{\boldsymbol{\mu}\ge 0} q(\boldsymbol{\mu}).$$

- If X is bounded, the dual function takes real values. In general, $q(\boldsymbol{\mu})$ can take the value $-\infty$. The effective constraint set of the dual is

  $$Q = \{\boldsymbol{\mu} \mid \boldsymbol{\mu} \ge \mathbf{0},\ q(\boldsymbol{\mu}) > -\infty\}$$

Duality Theorem

(a) If the primal problem has an optimal solution, the dual problem also has an optimal solution and the optimal values are equal.

(b) $\boldsymbol{x}^*$ is primal-optimal and $\boldsymbol{\mu}^*$ is dual-optimal if and only if $\boldsymbol{x}^*$ is primal-feasible, $\boldsymbol{\mu}^* \geq \mathbf{0}$, and

$$f(\boldsymbol{x}^*) = L(\boldsymbol{x}^*, \boldsymbol{\mu}^*) = \min_{\boldsymbol{x} \in X} L(\boldsymbol{x}, \boldsymbol{\mu}^*).$$

- Proof: (a) Let $\boldsymbol{x}^*$ be a primal optimal solution. For all primal feasible $\boldsymbol{x}$, and all $\boldsymbol{\mu} \geq \mathbf{0}$, we have $\boldsymbol{\mu}_j'(\boldsymbol{a}_j'\boldsymbol{x} - b_j) \leq 0$ for all j, so

$$\begin{aligned} q(\boldsymbol{\mu}) \quad &\leq \quad \inf_{\boldsymbol{x} \in X,\ \boldsymbol{a}_j'\boldsymbol{x} \leq b_j,\ j=1,\ldots,r} \left\{ f(\boldsymbol{x}) + \sum_{j=1}^{r} \mu_j(\boldsymbol{a}_j'\boldsymbol{x} - b_j) \right\} \\ &\leq \quad \inf_{\boldsymbol{x} \in X,\ \boldsymbol{a}_j'\boldsymbol{x} \leq b_j,\ j=1,\ldots,r} f(\boldsymbol{x}) = f(\boldsymbol{x}^*). \end{aligned}$$

- By Lagrangian Multiplier Theorem, there exists $\boldsymbol{\mu}^* \geq \mathbf{0}$ such that $\mu_j^*(\boldsymbol{a}_j'\boldsymbol{x}^* - b_j) = 0$ for all j, and $\boldsymbol{x}^* = \arg\min_{\boldsymbol{x}\in X} L(\boldsymbol{x}, \boldsymbol{\mu}^*)$, so

$$q(\boldsymbol{\mu}^*) = L(\boldsymbol{x}^*, \boldsymbol{\mu}^*) = f(\boldsymbol{x}^*) + \sum_{j=1}^{r} \mu_j^*(\boldsymbol{a}_j'\boldsymbol{x}^* - b_j) = f(\boldsymbol{x}^*).$$

- To prove (b), suppose $\boldsymbol{x}^*$ is primal-optimal and $\boldsymbol{\mu}^*$ is dual-optimal, by part (a) $f(\boldsymbol{x}^*) = q(\boldsymbol{\mu}^*)$, further implying

$$f(\boldsymbol{x}^*) = L(\boldsymbol{x}^*, \boldsymbol{\mu}^*) = q(\boldsymbol{\mu}^*) = \min_{\boldsymbol{x}\in X} L(\boldsymbol{x}, \boldsymbol{\mu}^*).$$

Conversely, the relation $f(\boldsymbol{x}^*) = \min_{\boldsymbol{x}\in X} L(\boldsymbol{x}, \boldsymbol{\mu}^*)$ is written as $f(\boldsymbol{x}^*) = q(\boldsymbol{\mu}^*)$, and since $\boldsymbol{x}^*$ is primal-feasible and $\boldsymbol{\mu}^* \geq \mathbf{0}$, it follows that $\boldsymbol{x}^*$ is primal optimal and $\boldsymbol{\mu}^*$ is dual-optimal. Q.E.D.

- Linear equality constraints are treated similar to inequality constraints, except that the sign of the Lagrange multipliers is unrestricted:

$$\boxed{\text{Primal:} \quad \min_{\boldsymbol{x}\in X,\ \boldsymbol{e}_i'\boldsymbol{x}=d_i,\ i=1,\ldots,m,\ \boldsymbol{a}_j'\boldsymbol{x}\le b_j,\ j=1,\ldots,r} f(\boldsymbol{x})}$$

$$\boxed{\text{Dual:} \quad \max_{\boldsymbol{\lambda}\in\mathbb{R}^m,\ \boldsymbol{\mu}\ge\mathbf{0}} q(\boldsymbol{\lambda},\boldsymbol{\mu}) = \max_{\boldsymbol{\lambda}\in\mathbb{R}^m,\ \boldsymbol{\mu}\ge\mathbf{0}} \inf_{\boldsymbol{x}\in X} L(\boldsymbol{x},\boldsymbol{\lambda},\boldsymbol{\mu}).}$$

The Dual of a Linear Program

- Consider the linear program

$$\begin{array}{ll} \text{minimize} & \boldsymbol{c}'\boldsymbol{x} \\ \text{subject to} & \boldsymbol{e}_i'\boldsymbol{x} = d_i, \quad i = 1, ..., m, \;\; \boldsymbol{x} \geq \boldsymbol{0}. \end{array}$$

- Dual function

$$q(\boldsymbol{\lambda}) = \inf_{\boldsymbol{x} \geq \boldsymbol{0}} \left\{ \sum_{j=1}^{n} \left(c_j - \sum_{i=1}^{m} \lambda_i e_{ij} \right) x_j + \sum_{i=1}^{m} \lambda_i d_i \right\}.$$

- It can be checked that

$$q(\boldsymbol{\lambda}) = \begin{cases} \sum_{i=1}^{m} \lambda_i d_i, & \text{if } c_j - \sum_{i=1}^{m} \lambda_i e_{ij} \geq 0 \text{ for all } j \\ -\infty, & \text{if } c_j - \sum_{i=1}^{m} \lambda_i e_{ij} < 0 \text{ for some } j \end{cases}$$

- Thus, the dual problem is

$$\begin{array}{ll} \text{maximize} & \sum_{i=1}^{m} \lambda_i d_i \\ \text{subject to} & \sum_{i=1}^{m} \lambda_i e_{ij} \leq c_j, \ j = 1, ..., n. \end{array}$$

The Dual of a Quadratic Program

- Consider the quadratic program

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\boldsymbol{x}'\boldsymbol{Q}\boldsymbol{x} + \boldsymbol{c}'\boldsymbol{x} \\ \text{subject to} & \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}, \end{array}$$

where $\boldsymbol{Q}$ is a given $n \times n$ positive definite symmetric matrix, $\boldsymbol{A}$ is a given $r \times n$ matrix, and $\boldsymbol{b} \in \mathbb{R}^r$ and $\boldsymbol{c} \in \mathbb{R}^n$ are given vectors.

- Dual function:

$$q(\boldsymbol{\mu}) = \inf_{\boldsymbol{x} \in \mathbb{R}^n} \frac{1}{2}\boldsymbol{x}'\boldsymbol{Q}\boldsymbol{x} + \boldsymbol{c}'\boldsymbol{x} + \boldsymbol{\mu}'(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}).$$

The infimum is attained for $\boldsymbol{x} = -\boldsymbol{Q}^{-1}(\boldsymbol{c} + \boldsymbol{A}'\boldsymbol{\mu})$, and, after substitution and calculation,

$$q(\boldsymbol{\mu}) = -\frac{1}{2}\boldsymbol{\mu}'\boldsymbol{A}\boldsymbol{Q}^{-1}\boldsymbol{A}'\boldsymbol{\mu} - \boldsymbol{\mu}'(\boldsymbol{b} + \boldsymbol{A}\boldsymbol{Q}^{-1}\boldsymbol{c}) - \frac{1}{2}\boldsymbol{c}'\boldsymbol{Q}^{-1}\boldsymbol{c}.$$

- The dual problem, after a sign change, is

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\boldsymbol{\mu}'\boldsymbol{P}\boldsymbol{\mu} + \boldsymbol{t}'\boldsymbol{\mu} \\ \text{subject to} & \boldsymbol{\mu} \geq \mathbf{0}, \end{array}$$

where $\boldsymbol{P} = \boldsymbol{A}\boldsymbol{Q}^{-1}\boldsymbol{A}'$ and $\boldsymbol{t} = \boldsymbol{b} + \boldsymbol{A}\boldsymbol{Q}^{-1}\boldsymbol{c}$.