Optimization Methods and Applications

Lecture 3. Constrained Optimization Problems, Convex Programming and Lagrangian Duality

Dr. Huiqi Guan

Fudan University

October 14, 2025

Roadmap

Last lecture

- separation theorems
  - Farkas' Lemma
  - Gordan's Lemma
- unconstrained optimization problems

This lecture

- constrained optimization problems
  - first-order conditions
- convex programming
- Lagrangian duality

Nonlinear Programming Problems

Problem P: Let X_0 be an open convex set in $\mathbb{R}^n$. Let $f, \mathbf{g}$, and $\mathbf{h}$ be $C^{(1)}$ functions with domain X_0 and ranges in $\mathbb{R}, \mathbb{R}^m$, and $\mathbb{R}^k$, respectively. Let

$$X = \{\mathbf{x} : \mathbf{x} \in X_0, \mathbf{g}(\mathbf{x}) \leq \mathbf{0}, \mathbf{h}(\mathbf{x}) = \mathbf{0}\}.$$

Minimize f over X.
Or

$$\begin{aligned} &\text{minimize } f(\mathbf{x}) \\ &\text{subject to } \mathbf{g}(\mathbf{x}) \leq \mathbf{0}, \qquad \mathbf{h}(\mathbf{x}) = \mathbf{0}, \qquad \mathbf{x} \in X_0. \end{aligned}$$

Descent Direction

Definition 2.1 (Descent Direction)

Let f be differentiable at $\mathbf{x}$. If $\nabla f(\mathbf{x})\mathbf{d} < 0$, $\mathbf{d}$ is denoted as a descent direction of f at $\mathbf{x}$.

Lemma 2.2

Let X be an open set in $\mathbb{R}^n$ and let f be differentiable at $\mathbf{x}$. If there is a vector $\mathbf{d}$ such that $\nabla f(\mathbf{x})\mathbf{d} < 0$, then there exists a $\delta > 0$ such that $f(\mathbf{x} + \lambda\mathbf{d}) < f(\mathbf{x})$ for all $\lambda \in (0, \delta)$, so that $\mathbf{d}$ is a descent direction of f at $\mathbf{x}$.

Proof.

By differentiability of f at $\mathbf{x}$, we must have

$$f(\mathbf{x} + \lambda\mathbf{d}) = f(\mathbf{x}) + \lambda\nabla f(\mathbf{x})\mathbf{d} + \lambda\|\mathbf{d}\|\alpha(\mathbf{x}; \lambda\mathbf{d})$$

where $\alpha(\mathbf{x}; \lambda\mathbf{d}) \to 0$ as $\lambda \to 0$. You can readily derive the lemma. $\square$

If $\nabla f(\mathbf{x}) \neq 0$, then $-\nabla f(\mathbf{x})$ is the direction of steepest descent at $\mathbf{x}$. Denote by $\mathcal{D}_{\mathbf{x}}$ the cone of descent directions at $\mathbf{x}$ or

$$\mathcal{D}_{\mathbf{x}} = \{\mathbf{d} \in \mathbb{R}^n : \nabla f(\mathbf{x})\mathbf{d} < 0\}.$$

Feasible Direction

Definition 2.3 (Feasible Direction)

Let X be a set in $\mathbb{R}^n$ and $\mathbf{x} \in X$. Nonzero vector $\mathbf{d}$ is denoted as a feasible direction of X at $\mathbf{x}$ if there is a δ and a differentiable curve $\boldsymbol{\xi}(t)$ such that

$$\boldsymbol{\xi}(0) = \mathbf{x}, \boldsymbol{\xi}'(0) = \mathbf{d}, \boldsymbol{\xi}(t) \in X, t \in (0, \delta).$$

At feasible point $\mathbf{x}$, denote the cone of feasible directions at $\mathbf{x}$ as $\mathcal{F}_{\mathbf{x}}$. If there is a $\gamma > 0$ such that $\mathbf{x} + t\mathbf{d} \in X$ for all $t \in (0, \gamma)$, then $\mathbf{d}$ is a feasible direction of X at $\mathbf{x}$. Therefore, $\{\mathbf{d} \in \mathbb{R}^n : \mathbf{d} \neq \mathbf{0}, \mathbf{x} + t\mathbf{d} \in X$ for all $t \in (0, \gamma)$ for some $\gamma > 0\} \subseteq \mathcal{F}_{\mathbf{x}}$.

Example 2.4

Let $X = \{\mathbf{x} : A\mathbf{x} = \mathbf{0}\}$. Then $\mathcal{F}_{\mathbf{x}} = \{\mathbf{d} : \mathbf{d} \neq \mathbf{0}, A\mathbf{d} = \mathbf{0}.\}$.

Example 2.5

Let $X = \{\mathbf{x} : \mathbf{x} \in X_0, \mathbf{g}(\mathbf{x}) \leq \mathbf{0}\}$. Let $\mathbf{x}^* \in X$ and $E = \{i : g_i(\mathbf{x}^*) = 0\}$. If $\nabla \mathbf{g}_E(\mathbf{x}^*)\mathbf{d} < \mathbf{0}$, then $\mathbf{d}$ is a feasible direction of X at $\mathbf{x}^*$.

Hint: $\mathbf{x}^* + t\mathbf{d} \in X$ with small enough t.

Geometric Optimality Conditions

A fundamental question in Optimization is: what are the necessary conditions in order to have $\bar{\mathbf{x}}$ as a local optimizer?

A general answer is: the intersection of the descent and feasible direction sets at $\bar{\mathbf{x}}$ must be empty. That is, $\mathcal{D}_{\bar{\mathbf{x}}} \cap \mathcal{F}_{\bar{\mathbf{x}}} = \phi$ can be regarded as a geometric condition for $\bar{\mathbf{x}}$ to be a local minimizer. It is a necessary condition.

First-order Conditions

Theorem 2.6

If $\mathbf{x}^*$ *is a solution of the problem* $\min\{f(\mathbf{x}) : \mathbf{x} \in X_0, \mathbf{g}(\mathbf{x}) \leq \mathbf{0}\}$*, then there exists a real number* $\lambda_0 \geq 0$*, a vector* $\boldsymbol{\lambda} \geq \mathbf{0}$ *in* $\mathbb{R}^m$*, such that*

(i) $(\lambda_0, \boldsymbol{\lambda}) \neq \mathbf{0}$,

(ii) $\langle \boldsymbol{\lambda}, \mathbf{g}(\mathbf{x}^*) \rangle = 0$, *and*

(iii) $\lambda_0 \nabla f(\mathbf{x}^*) + \boldsymbol{\lambda}^t \nabla \mathbf{g}(\mathbf{x}^*) = \mathbf{0}$.

Proof.

Let $E = \{i : g_i(\mathbf{x}^*) = 0\}$. Since x^* is a solution of the minimization problem, the following system has no solution:

$$\nabla \mathbf{g}_E(\mathbf{x}^*)\mathbf{d} < \mathbf{0}, \qquad \nabla \mathbf{f}(\mathbf{x}^*)\mathbf{d} < \mathbf{0}.$$

By setting $\lambda_i = 0$ if $g_i(\mathbf{x}^*) < 0$, the results we need can be derived from Gordan's lemma. $\square$

Theorem 2.7 (Lagrange multiplier rule)

If $\mathbf{x}^*$ *is a solution of the problem* $\min\{f(\mathbf{x}) : \mathbf{h}(\mathbf{x}) = \mathbf{0}\}$ *and if the* k *gradient vectors* $\nabla h_j(\mathbf{x}^*)$ *are linearly independent, then there exist unique scalars* $\mu_j, j = 1, \ldots, k$*, such that for each* $i = 1, \ldots, n$

$$\frac{\partial f}{\partial x_i}(\mathbf{x}^*) + \sum_{j=1}^{k} \mu_j \frac{\partial h_j(\mathbf{x}^*)}{\partial x_i} = 0$$

and

$$h_j(\mathbf{x}^*) = 0, j = 1, \ldots, k.$$

If $\mathbf{h}(\mathbf{x}) = A\mathbf{x} + \mathbf{b}$, since $\mathbf{x}^*$ is a minimizer, then there is no $\mathbf{d}$ such that $A\mathbf{d} = \mathbf{0}$ and $\nabla f(\mathbf{x}^*)\mathbf{d} < 0$. The result is derived from Farkas' Lemma and we don't need the property "$\nabla \mathbf{h}(\mathbf{x}^*)$ has full rank" at $\mathbf{x}^*$.
For the general case, note that
$\mathbf{0} = \mathbf{h}(\mathbf{x}) \approx \mathbf{h}(\mathbf{x}^*) + \nabla \mathbf{h}(\mathbf{x}^*)(\mathbf{x} - \mathbf{x}^*) = \nabla_{\mathbf{x}_1} \mathbf{h}(\mathbf{x}^*)(\mathbf{x}_1 - \mathbf{x}_1^*) + \nabla_{\mathbf{x}_2} \mathbf{h}(\mathbf{x}^*)(\mathbf{x}_2 - \mathbf{x}_2^*)$. If $\nabla_{\mathbf{x}_2} \mathbf{h}(\mathbf{x}^*)$ is nonsingular, then $\mathbf{x}_2 \approx \mathbf{x}_2^* - (\nabla_{\mathbf{x}_2} \mathbf{h}(\mathbf{x}^*))^{-1} \nabla_{\mathbf{x}_1} \mathbf{h}(\mathbf{x}^*)(\mathbf{x}_1 - \mathbf{x}_1^*)$.

Implicit Function Theorem

Theorem 2.8 (Implicit Function Theorem)

Let D_1 be an open set in $\mathbb{R}^n$ and let D_2 be an open set in $\mathbb{R}^m$. Let $D = D_1 \times D_2$. Let

$$\mathbf{u} : (\mathbf{x}, \mathbf{y}) \to \mathbf{u}(\mathbf{x}, \mathbf{y}), \quad \mathbf{x} \in D_1, \mathbf{y} \in D_2,$$

be a mapping that is of class $C^{(p)}$ on D and with range in $\mathbb{R}^m$. Let $(\mathbf{x}_0, \mathbf{y}_0), \mathbf{x}_0 \in D_1, \mathbf{y}_0 \in D_2$, be a solution of

$$\mathbf{u}(\mathbf{x}, \mathbf{y}) = \mathbf{0}$$

and suppose that the $m \times m$ matrix

$$J_{\mathbf{y}}(\mathbf{x}_0, \mathbf{y}_0) = (\frac{\partial u_i}{\partial y_j}(\mathbf{x}_0, \mathbf{y}_0)), \quad i = 1, \ldots, m, \quad j = 1, \ldots, m$$

is nonsingular. Then there exists an $\alpha > 0$, a $\beta > 0$ and a function $\mathbf{y}(\cdot)$ defined on $B(\mathbf{x}_0, \alpha) \in D_1$ with range in $B(\mathbf{y}_0, \beta) \in D_2$ such that $\mathbf{y}(\cdot)$ is of class $C^{(p)}$, $\mathbf{u}(\mathbf{x}, \mathbf{y}(\mathbf{x})) = \mathbf{0}$ for all $\mathbf{x} \in B(\mathbf{x}_0, \alpha)$. Moreover, for $\mathbf{x}$ in $B(\mathbf{x}_0, \alpha)$, $\mathbf{y}(\mathbf{x})$ is the only solution of $\mathbf{u}(\mathbf{x}, \mathbf{y}) = \mathbf{0}$ such that $\mathbf{y}(\mathbf{x}) \in B(\mathbf{y}_0, \beta)$.

Proof of Lagrange Multiplier Rule

Proof of Lemma 2.7.

Suppose the last k columns of $\nabla \mathbf{h}(\mathbf{x}^*)$ is linearly independent. Let $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2)$, where $\mathbf{x}_2$ is the last k elements in $\mathbf{x}$. By Implicit Function Theorem, there exists an $\alpha > 0$, a $\beta > 0$ and a function $\mathbf{y}(\cdot)$ such that $\mathbf{h}(\mathbf{x}_1, \mathbf{x}_2) = \mathbf{0}$ if and only if $\mathbf{x}_2 = \mathbf{y}(\mathbf{x}_1)$ when $\mathbf{x}_1 \in B(\mathbf{x}_1^*, \alpha)$ and $\mathbf{x}_2 \in B(\mathbf{x}_2^*, \beta)$. Consider $\tilde{f}(\mathbf{x}_1) = f(\mathbf{x}_1, \mathbf{y}(\mathbf{x}_1))$. Then $\mathbf{x}_1^*$ is a local minimizer of $\tilde{f}(\mathbf{x}_1)$ in $B(\mathbf{x}_1^*, \alpha)$, which yields $\nabla \tilde{f}(\mathbf{x}_1^*) = \mathbf{0}$, i.e. $\nabla_{\mathbf{x}_1} f(\mathbf{x}^*) + \nabla_{\mathbf{x}_2} f(\mathbf{x}^*) \nabla_{\mathbf{x}_1} \mathbf{y}(\mathbf{x}_1^*) = \mathbf{0}$. Since $\mathbf{h}(\mathbf{x}_1, \mathbf{y}(\mathbf{x}_1)) = \mathbf{0}$, we have $\nabla_{\mathbf{x}_1} \mathbf{h}(\mathbf{x}^*) + \nabla_{\mathbf{x}_2} \mathbf{h}(\mathbf{x}^*) \nabla_{\mathbf{x}_1} \mathbf{y}(\mathbf{x}_1^*) = \mathbf{0}$. Hence, we have

$$\begin{bmatrix} \nabla f(\mathbf{x}^*) \\ \nabla \mathbf{h}(\mathbf{x}^*) \end{bmatrix} \begin{bmatrix} I_{n-k} \\ \nabla_{\mathbf{x}_1} \mathbf{y}(\mathbf{x}_1^*) \end{bmatrix} = \mathbf{0}.$$

Since rank($\begin{bmatrix} I_{n-k} \\ \nabla \mathbf{y}(\mathbf{x}_1^*) \end{bmatrix}$) $\geq n-k$, then the rows of $\begin{bmatrix} \nabla f(\mathbf{x}^*) \\ \nabla \mathbf{h}(\mathbf{x}^*) \end{bmatrix}$ are linearly dependent. Note that the rows of $\nabla(\mathbf{h}(\mathbf{x}^*))$ are linearly independent and we complete the proof. $\square$

Example

The property "$\nabla \mathbf{h}(\mathbf{x}^*)$ has full row rank" is called a regularity condition or constraint qualification. Lagrange's theorem is not valid without a regularity condition when constraints are nonlinear.

Consider the problem

$$\begin{aligned} \text{minimize } & x_1 \\ \text{subject to } & x_1^2 + (x_2 - 1)^2 - 1 = 0 \\ & x_1^2 + (x_2 + 1)^2 - 1 = 0 \end{aligned}$$

This problem has just one feasible point: $\bar{\mathbf{x}} = (0, 0)$. Note that $\nabla f(\bar{\mathbf{x}}) = (1, 0), \nabla h_1(\bar{\mathbf{x}}) = (0, -2), \nabla h_2(\bar{\mathbf{x}}) = (0, 2)$. One can see that the Lagrange Theorem does not hold.

Feasible Direction with Equality Constraints

Lemma 2.9

Let $\mathbf{g}$ *and* $\mathbf{h}$ *be of class* $C^{(p)}$ *on* $X_0, p \geq 1$. *Let* $\mathbf{x}^*$ *be feasible,* $I = \{i : g_i(\mathbf{x}^*) < 0\}$, $E = \{i : g_i(\mathbf{x}^*) = 0\}$, *and let* $\mathbf{z}$ *satisfy* $\nabla \mathbf{g}_E(\mathbf{x}^*)\mathbf{z} \leq \mathbf{0}, \nabla \mathbf{h}(\mathbf{x}^*)\mathbf{z} = \mathbf{0}$. *Suppose that the vectors*

$$\nabla g_1(\mathbf{x}^*), \ldots, \nabla g_q(\mathbf{x}^*), \nabla h_1(\mathbf{x}^*), \ldots, \nabla h_k(\mathbf{x}^*)$$

are linearly independent, where $E_1 = \{1, 2, \ldots, q\} = \{i : i \in E, \langle \nabla g_i(\mathbf{x}^*), \mathbf{z} \rangle = 0\}$. *Then there exists a* $\tau > 0$ *and a* $C^{(p)}$ *mapping* $\boldsymbol{\xi}(\cdot)$ *from* $(-\tau, \tau)$ *to* $\mathbb{R}^n$ *such that*

$$\begin{aligned} \boldsymbol{\xi}(0) &= \mathbf{x}^*, & \boldsymbol{\xi}'(0) &= \mathbf{z} & & \\ \mathbf{g}_{E_1}(\boldsymbol{\xi}(t)) &= \mathbf{0}, & \mathbf{h}(\boldsymbol{\xi}(t)) &= \mathbf{0}, & \mathbf{g}_I(\boldsymbol{\xi}(t)) &< \mathbf{0}, \\ \mathbf{g}_{E \setminus E_1}(\boldsymbol{\xi}(t)) &< \mathbf{0}, & \textit{for} \quad 0 < |t| &< \tau. & & \end{aligned}$$

Proof.

Suppose the first $q + k$ columns of $\left[\nabla \mathbf{g}_1(\mathbf{x}^*); \ldots; \nabla \mathbf{g}_q(\mathbf{x}^*); \nabla \mathbf{h}_1(\mathbf{x}^*); \ldots; \nabla \mathbf{h}_k(\mathbf{x}^*)\right]$ are linearly independent. Consider $\mathbf{u}(\mathbf{x}, t) = \left[\mathbf{g}_{E_1}(\mathbf{x}); \mathbf{h}(\mathbf{x}); x_{q+k+1} - x^*_{q+k+1} - tz_{q+k+1}; \ldots; x_n - x^*_n - tz_n\right]$. The result is derived from Implicit Function Theorem. □

Fritz John Theorem

Theorem 2.10 (Fritz John Theorem)

Let $\mathbf{x}^*$ *be a solution of the problem P. Then there exists a real number* $\lambda_0 \geq 0$, *a vector* $\boldsymbol{\lambda} \geq \mathbf{0}$ *in* $\mathbb{R}^m$, *and a vector* $\boldsymbol{\mu}$ *in* $\mathbb{R}^k$ *such that*

(i) $(\lambda_0, \boldsymbol{\lambda}, \boldsymbol{\mu}) \neq \mathbf{0}$,

(ii) $\langle \boldsymbol{\lambda}, \mathbf{g}(\mathbf{x}^*) \rangle = 0$, *and*

(iii) $\lambda_0 \nabla f(\mathbf{x}^*) + \boldsymbol{\lambda}^t \nabla \mathbf{g}(\mathbf{x}^*) + \boldsymbol{\mu}^t \nabla \mathbf{h}(\mathbf{x}^*) = \mathbf{0}$.

Definition 2.11

Critical Point: A (feasible) point $\mathbf{x}^*$ at which the conclusion of Theorem 2.10 holds will be called a critical point for problem **P**.

Note: Theorem 2.10, with additional hypotheses guaranteeing that $\lambda_0 > 0$, is often referred to as the Karush–Kuhn–Tucker Theorem.

Proof of Fritz John Theorem

Proof of Theorem 2.10.

If the vectors $\nabla h_1(\mathbf{x}^*), \ldots, \nabla h_k(\mathbf{x}^*)$ are linearly dependent, the result is trivial. Now suppose the vectors $\nabla h_1(\mathbf{x}^*), \ldots, \nabla h_k(\mathbf{x}^*)$ are linearly independent.

Let $E = \{i : g_i(\mathbf{x}^*) = 0\}$. Then the following system has no solution:

$$\nabla \mathbf{g}_E(\mathbf{x}^*)\mathbf{d} < \mathbf{0}, \qquad \nabla \mathbf{h}(\mathbf{x}^*)\mathbf{d} = \mathbf{0}, \qquad \nabla \mathbf{f}(\mathbf{x}^*)\mathbf{d} < 0.$$

By setting $\lambda_i = 0$ if $g_i(\mathbf{x}^*) < 0$, the results we need can be derived from separation theorem. □

Example

Example 2.12

Let $X_0 = \mathbb{R}^2$, let

$$f(\mathbf{x}) = (x_1 - 1)^2 + x_2^2,$$

and let

$$g(\mathbf{x}) = 2kx_1 - x_2^2 \leq 0, \quad k > 0.$$

Find all the critical points for problem **PI**.

Solution: $\nabla f = (2(x_1 - 1), 2x_2), \nabla g = (2k, -2x_2)$. Suppose $\lambda_0 \nabla f + \lambda \nabla g = \mathbf{0}, \lambda g(\mathbf{x}) = 0, \lambda_0, \lambda \geq 0$ and $(\lambda_0, \lambda) \neq (0, 0)$.
If $\lambda_0 = 0$, then $\lambda \neq 0$. Since $k > 0$, a contradiction.
If $\lambda_0 \neq 0$, we may take $\lambda_0 = 1$. If $\lambda = 0$, then $(x_1, x_2) = (1, 0)$, which is not feasible since $g((1, 0)) > 0$. If $\lambda \neq 0$, then $2kx_1 - x_2^2 = 0$. We have $(x_1 - 1, x_2) + \lambda(k, -x_2) = (0, 0)$ and $x_2^2 = 2kx_1$. If $x_2 \neq 0$, then $\lambda = 1$ and so $x_1 = 1 - k, x_2 = \pm\sqrt{2k(1-k)}$ if $0 < k < 1$. If $x_2 = 0$, then $x_1 = 0, \lambda = \frac{1}{k}$.

Constraint Qualifications

Definition 2.13

The functions $\mathbf{g}$ and $\mathbf{h}$ satisfy the *constraint qualification* CQ at a feasible point $\mathbf{x}_0$ if

(i) the vectors $\nabla h_1(\mathbf{x}_0), \ldots, \nabla h_k(\mathbf{x}_0)$ are linearly independent and

(ii) the system

$$\nabla \mathbf{g}_E(\mathbf{x}_0)\mathbf{z} < \mathbf{0}, \qquad \nabla \mathbf{h}(\mathbf{x}_0)\mathbf{z} = \mathbf{0}, \tag{1}$$

has a solution $\mathbf{z}$ in $\mathbb{R}^n$. Here, $E = \{i : g_i(\mathbf{x}_0) = 0\}$.

Note: For $i \in E$, if $\mathbf{g}_i$ is convex and $g_i(\bar{\mathbf{x}}) < 0$, then let $z = \bar{\mathbf{x}} - \mathbf{x}_0$, and we have $\nabla g_i(\mathbf{x}_0)\mathbf{z} < 0$. Since $\mathbf{g}_i$ is convex, then
$\nabla g_i(\mathbf{x}_0)\mathbf{z} = \nabla g_i(\mathbf{x}_0)\mathbf{z} + g_i(\mathbf{x}_0) \leq g_i(\bar{\mathbf{x}}) < 0.$

Karush–Kuhn–Tucker Theorem

Theorem 2.14 (Karush–Kuhn–Tucker Theorem)

Let $\mathbf{x}^*$ *be a solution of problem* ***P*** *and let CQ hold at* $\mathbf{x}^*$. *Then* $\lambda_0 > 0$ *and there exists a vector* $\boldsymbol{\lambda} \geq \mathbf{0}$ *in* $\mathbb{R}^m$ *and a vector* $\boldsymbol{\mu}$ *in* $\mathbb{R}^k$ *such that*

(i) $\langle \boldsymbol{\lambda}, \mathbf{g}(\mathbf{x}^*) \rangle = 0$, *and*

(ii) $\nabla f(\mathbf{x}^*) + \boldsymbol{\lambda}^t \nabla \mathbf{g}(\mathbf{x}^*) + \boldsymbol{\mu}^t \nabla \mathbf{h}(\mathbf{x}^*) = \mathbf{0}$.

Definition 2.15

KKT (stationary) Point: A point $\mathbf{x}^*$ at which the conclusion of Theorem 2.14 holds will be called a KKT point for problem **PI**.
KKT Pair : A vector $(\mathbf{x}^*; \boldsymbol{\lambda}; \boldsymbol{\mu})$ at which the conclusion of Theorem 2.14 holds will be called a KKT pair for problem **P**.

To say that $\mathbf{x}^*$ is a KKT stationary point means that there exists a vector $(\boldsymbol{\lambda}, \boldsymbol{\mu})$ such that $(\mathbf{x}^*; \boldsymbol{\lambda}; \boldsymbol{\mu})$ is a KKT pair, i.e., satisfies the KKT first-order necessary conditions of local optimality.

Karush–Kuhn–Tucker Theorem continued

Corollary 2.16 (LICQ)

Let $\mathbf{x}^*$ *be a solution of problem* ***P*** *such that* $E = \{i : g_i(\mathbf{x}^*) = 0\} = \{1, \ldots, r\}$ *and such that the vectors*

$$\nabla g_1(\mathbf{x}^*), \ldots, \nabla g_r(\mathbf{x}^*), \nabla h_1(\mathbf{x}^*), \ldots, \nabla h_k(\mathbf{x}^*)$$

are linearly independent. Then the conclusion of Theorem 2.14 holds.

The KKT Conditions are not Sufficient

Consider the optimization problem

$$\text{minimize } x_2$$
$$\text{subject to } x_1^2 + x_2 \geq 0$$

Its KKT points satisfy

$$\begin{cases} \begin{pmatrix} 0 \\ 1 \end{pmatrix} - \lambda \begin{pmatrix} 2x_1 \\ 1 \end{pmatrix} = 0 \\ \lambda(x_1^2 + x_2) = 0, x_1^2 + x_2 \geq 0, \lambda \geq 0. \end{cases}$$

This shows that the problem has the unique KKT point $\mathbf{x}^* = (0; 0)$ with $\lambda = 1$, and the LICQ is satisfied at $\mathbf{x}^*$. But $\mathbf{x}^*$ is not optimal.

The need for a CQ

Consider the optimization problem

$$\begin{aligned} \text{minimize } & (x_1 - 1)^2 + (x_2 - 1)^2 \\ \text{subject to } & (1 - x_1 - x_2)^3 \geq 0 \\ & x_1 \geq 0 \end{aligned}$$

The feasible region of this example is the same as $x_1 + x_2 \leq 1, x_1 \geq 0$. This problem has a unique optimal solution: $\mathbf{x}^* = (\frac{1}{2}, \frac{1}{2})$. Clearly, the (KKT) constraint qualification, $(\nabla g_i(\mathbf{x}), i \in E)$, is linearly independent, is not satisfied at x^*, and x^* is not a KKT point.

The need for a CQ continued

Let $E = \{i : g_i(\mathbf{x}^*) = 0\}$. If $\mathbf{x}^*$ is an optimal solution, then $\mathcal{D}^0_{\mathbf{x}^*} \cap \mathcal{F}^0_{\mathbf{x}^*} = \phi$, where $\mathcal{D}^0_{\mathbf{x}^*} = \{\mathbf{d} \in \mathbb{R}^n : \nabla \mathbf{f}(\mathbf{x}^*)\mathbf{d} < 0\}, \mathcal{F}^0_{\mathbf{x}^*} = \{\mathbf{d} \in \mathbb{R}^n : \nabla \mathbf{g}_E(\mathbf{x}^*)\mathbf{d} < \mathbf{0}, \nabla \mathbf{h}(\mathbf{x}^*)\mathbf{d} = \mathbf{0}\}$. This condition is not sufficient.

Let $\tilde{x} = (\tilde{x}_1; \tilde{x}_2)$ be any point satisfying $\tilde{x}_1 + \tilde{x}_2 = 1$ with $\tilde{x}_1 \neq 0$ and $\tilde{x}_2 \neq 0$. With $g_1(\mathbf{x}) = -(1 - x_1 - x_2)^3$, we have $\nabla g_1(\tilde{\mathbf{x}}) = 3(1 - \tilde{x}_1 - \tilde{x}_2)^2(1, 1) = (0, 0)$. Clearly, $\mathcal{D}^0_{\tilde{\mathbf{x}}} \cap \mathcal{F}^0_{\tilde{\mathbf{x}}} = \phi$. This illustrates that the condition $\mathcal{D}^0_{\tilde{\mathbf{x}}} \cap \mathcal{F}^0_{\tilde{\mathbf{x}}} = \phi$ can be satisfied by infinitely many nonoptimal points in the feasible region as well as by the optimal solution, and hence it is not sufficient.

Is the KKT constraint qualification indispensable?

If $\bar{\mathbf{x}}$ is a local minimizer and the KKT conditions hold at $\bar{\mathbf{x}}$, does the KKT constraint qualification have to hold there as well? It does not.
Consider the following primal problem

$$\begin{aligned} \text{minimize } & f(\mathbf{x}) = x_2 \\ \text{subject to } & c_1(\mathbf{x}) = r(x_1) - x_2 + x_1^2 \geq 0, \\ & c_2(\mathbf{x}) = x_2 - x_1^2 - s(x_1) \geq 0 \\ & c_3(\mathbf{x}) = 1 - x_1^2 \geq 0 \end{aligned}$$

where the functions $r(t) = \begin{cases} t^4 \sin\frac{1}{t} & \text{if } t \neq 0 \\ 0 & \text{if } t = 0 \end{cases}$ and $s(t) = \begin{cases} t^4 \cos\frac{1}{t} & \text{if } t \neq 0 \\ 0 & \text{if } t = 0 \end{cases}$ and they are continuously differentiable. The feasible region lies between the curves $x_2 = x_1^2 + x_1^4$ and $x_2 = x_1^2 - x_1^4$. Indeed, $x_2 - x_1^4 \leq x_1^2 + s(x_1) \leq x_2 \leq x_1^2 + r(x_1) \leq x_1^2 + x_1^4$. The feasible region also lies between the lines $x_1 = -1$ and $x_1 = 1$. It is easy to see that the unique optimal solution to the problem is $\bar{\mathbf{x}} = \mathbf{0}$.

Is the KKT constraint qualification indispensable?

In this instance, we have $E = \{i : c_i(\bar{\mathbf{x}}) = 0\} = \{1, 2\}$. Moreover, $\nabla c_1(\mathbf{0}) = (0, -1), \nabla c_2(\mathbf{0}) = (0, 1)$. It can be shown that $(\nabla c_i(0), i \in E)$ is linearly dependent. This means that the KKT constraint qualification does not hold at $\bar{\mathbf{x}} = \mathbf{0}$.

On the other hand, let $X_0 = \{\mathbf{x} : c_3(\mathbf{x}) \geq 0\}$ and the KKT conditions for this problem give:

$$-\mu_1 r'(x_1) - 2\mu_1 x_1 + 2\mu_2 x_1 + \mu_2 s'(x_1) = 0,$$

$$1 + \mu_1 - \mu_2 = 0,$$

$$\mu_1 \geq 0, \mu_2 \geq 0,$$

$$\mu_1 c_1(\mathbf{x}) = 0, \mu_2 c_2(\mathbf{x}) = 0,$$

$$c_1(\mathbf{x}) \geq 0, c_2(\mathbf{x}) \geq 0, \mathbf{x} \in X_0.$$

Thus, $\bar{\mathbf{x}} = \mathbf{0}$ and any $\bar{\boldsymbol{\mu}} = (\bar{\mu}_1, \bar{\mu}_2) \geq \mathbf{0}$ such that $1 + \bar{\mu}_1 - \bar{\mu}_2 = 0$ (e.g.$\bar{\mu}_1 = 1$ and $\bar{\mu}_2 = 2$) will satisfy the KKT conditions.

Sufficient Condition

Theorem 2.17 (Sufficient Condition)

*Let f and $\mathbf{g}$ be as in the statement of problem **P** and let X_0, f, and $\mathbf{g}$ be convex and $\mathbf{h}$ be affine, i.e. $\mathbf{h} = \mathbf{A}\mathbf{x} + \mathbf{b}$. Let $\mathbf{x}^* \in X_0$ be such that*

(i) $\mathbf{g}(\mathbf{x}^*) \leq \mathbf{0}, \mathbf{h}(\mathbf{x}^*) = \mathbf{0}$

(ii) $\boldsymbol{\lambda} \geq \mathbf{0}$,

(iii) $\langle \boldsymbol{\lambda}, \mathbf{g}(\mathbf{x}^*) \rangle = 0$, *and*

(iv) $\nabla f(\mathbf{x}^*) + \boldsymbol{\lambda}^t \nabla \mathbf{g}(\mathbf{x}^*) + \boldsymbol{\mu}^t \nabla \mathbf{h}(\mathbf{x}^*) = \mathbf{0}$.

*Then $\mathbf{x}^*$ is a solution of problem **P**.*

Proof.

Since f and $\mathbf{g}$ are convex and $\boldsymbol{\lambda} \geq \mathbf{0}$, then $L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\mathbf{x}) + \boldsymbol{\lambda}^t \mathbf{g}(\mathbf{x}) + \boldsymbol{\mu}^t \mathbf{h}(\mathbf{x})$ is a convex function on X_0. Note $\nabla f(\mathbf{x}^*) + \boldsymbol{\lambda}^t \nabla \mathbf{g}(\mathbf{x}^*) + \boldsymbol{\mu}^t \nabla \mathbf{h}(\mathbf{x}^*) = \mathbf{0}$. Then $\mathbf{x}^*$ is a minimizer of $L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu})$ on X_0. For each $\mathbf{x} \in X_0$ such that $g(\mathbf{x}) \leq \mathbf{0}, \mathbf{h}(\mathbf{x}) = \mathbf{0}$, since $\langle \boldsymbol{\lambda}, \mathbf{g}(\mathbf{x}^*) \rangle = 0$, then $f(\mathbf{x}) \geq f(\mathbf{x}) + \boldsymbol{\lambda}^t \mathbf{g}(\mathbf{x}) + \boldsymbol{\mu}^t \mathbf{h}(\mathbf{x}) = L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) \geq L(\mathbf{x}^*, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\mathbf{x}^*) + \boldsymbol{\lambda}^t \mathbf{g}(\mathbf{x}^*) + \boldsymbol{\mu}^t \mathbf{h}(\mathbf{x}^*) = f(\mathbf{x}^*)$. □

Convex Programming Problems

Convex Programming Problems (CP): Let X_0 be a convex set in $\mathbb{R}^n$. Let f and $\mathbf{g}$ be convex functions with domain X_0 and ranges in $\mathbb{R}$ and $\mathbb{R}^m$, respectively. Let $\mathbf{h}$ be affine. Let

$$X = \{\mathbf{x} : \mathbf{x} \in X_0, \mathbf{g}(\mathbf{x}) \le \mathbf{0}, \mathbf{h}(\mathbf{x}) = \mathbf{0}\}.$$

Minimize f over X.
Or

$$\begin{aligned} &\text{minimize } f(\mathbf{x}) \\ &\text{subject to } \mathbf{g}(\mathbf{x}) \le \mathbf{0}, \qquad \mathbf{h}(\mathbf{x}) = \mathbf{0}, \qquad \mathbf{x} \in X_0. \end{aligned}$$

Lemma 3.1

Let X be a nonempty convex set in $\mathbb{R}^n$. Let $\alpha : \mathbb{R}^n \to \mathbb{R}$ and $\mathbf{g} : \mathbb{R}^n \to \mathbb{R}^m$ be convex and let $\mathbf{h}$ be affine, i.e. $\mathbf{h}(\mathbf{x}) = \mathbf{A}\mathbf{x} - \mathbf{b}$. If System 1 below has no solution $\mathbf{x}$, then System 2 has a solution $(\lambda_0, \boldsymbol{\lambda}, \boldsymbol{\mu})$. The converse holds if $\lambda_0 > 0$.

System 1 $\alpha(\mathbf{x}) < 0, \quad \mathbf{g}(\mathbf{x}) \leq \mathbf{0}, \quad \mathbf{h}(\mathbf{x}) = \mathbf{0}$ *for some* $\mathbf{x} \in X$

System 2 $\lambda_0 \alpha(\mathbf{x}) + \boldsymbol{\lambda}^t \mathbf{g}(\mathbf{x}) + \boldsymbol{\mu}^t \mathbf{h}(\mathbf{x}) \geq 0$ *for all* $\mathbf{x} \in X$

$$(\lambda_0, \boldsymbol{\lambda}) \geq \mathbf{0}, \quad (\lambda_0, \boldsymbol{\lambda}, \boldsymbol{\mu}) \neq \mathbf{0}.$$

Convex Programming continued

Definition 3.2

Generalized Lagrangian:
$\wedge(\mathbf{x}, \lambda_0, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \lambda_0 f(\mathbf{x}) + \langle \boldsymbol{\lambda}, g(\mathbf{x}) \rangle + \langle \boldsymbol{\mu}, \mathbf{h}(\mathbf{x}) \rangle.$

Theorem 3.3

Let f and $\mathbf{g}$ be convex on $\mathbb{R}^n$ and let $\mathbf{h}$ be affine. Let $\mathbf{x}^*$ be a solution of the problem CP. Then there exists a real number $\lambda_0^* \geq 0$, a vector $\boldsymbol{\lambda}^* \geq \mathbf{0}$ in $\mathbb{R}^m$, and a vector $\boldsymbol{\mu}^*$ in $\mathbb{R}^m$ such that

(i) $(\lambda_0^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) \neq \mathbf{0}$,

(ii) $\langle \boldsymbol{\lambda}^*, \mathbf{g}(\mathbf{x}^*) \rangle = 0$, *and*

(iii) $\wedge(\mathbf{x}^*, \lambda_0^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = \lambda_0^* f(\mathbf{x}^*)$, *and*

(iv) $\wedge(\mathbf{x}^*, \lambda_0^*, \boldsymbol{\lambda}, \boldsymbol{\mu}) \leq \wedge(\mathbf{x}^*, \lambda_0^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) \leq \wedge(\mathbf{x}, \lambda_0^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ *for all* $\mathbf{x}$ *in* $\mathbb{R}^n$, $\boldsymbol{\lambda} \geq \mathbf{0}$ *in* $\mathbb{R}^n$ *and* $\boldsymbol{\mu}$ *in* $\mathbb{R}^k$.

Hint: Since $\mathbf{x}^*$ is a solution of **CP**, then
$f(\mathbf{x}) - f(\mathbf{x}^*) < 0, \mathbf{g}(\mathbf{x}) \leq \mathbf{0}, \mathbf{h}(\mathbf{x}) = \mathbf{0}$, has no solution in $X_0 \subseteq \mathbb{R}^n$.

Convex Programming continued

Definition 3.4

Strongly Consistent: The problem CP is said to be strongly consistent if there exists an $\mathbf{x}_0$ such that $\mathbf{g}(\mathbf{x}_0) < \mathbf{0}$ and $\mathbf{h}(\mathbf{x}_0) = \mathbf{0}$.

Theorem 3.5

Let $X_0 = \mathbb{R}^n$, f and $\mathbf{g}$ be convex on $\mathbb{R}^n$ and $\mathbf{h} = \mathbf{A}\mathbf{x} - \mathbf{b}$. Let $\mathbf{A}$ have full rank. Let $\mathbf{x}^*$ be a solution of the problem CP and the problem be strongly consistent. Then there exists a vector $\boldsymbol{\lambda}^* \geq \mathbf{0}$ in $\mathbb{R}^m$, and a vector $\boldsymbol{\mu}^*$ in $\mathbb{R}^k$ such that

(i) $\langle \boldsymbol{\lambda}^*, \mathbf{g}(\mathbf{x}^*) \rangle = 0$, *and*

(ii) $L(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = f(\mathbf{x}^*)$, *and*

(iii) $L(\mathbf{x}^*, \boldsymbol{\lambda}, \boldsymbol{\mu}) \leq L(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) \leq L(\mathbf{x}, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ *for all* $\mathbf{x} \in X_0$ *and* $\boldsymbol{\lambda} \geq \mathbf{0}$ *in* $\mathbb{R}^n$,

where $L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \wedge(\mathbf{x}, 1, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\mathbf{x}) + \langle \boldsymbol{\lambda}, \mathbf{g}(\mathbf{x}) \rangle + \langle \boldsymbol{\mu}, \mathbf{h}(\mathbf{x}) \rangle$.

Convex Programming continued

Proof of Theorem 3.5.

If $\lambda_0 \neq 0$ in Theorem 3.3, the result is derived. If $\lambda_0 = 0$, since A has full rank, then $\boldsymbol{\lambda}^* \neq \mathbf{0}$. Note that $\langle \boldsymbol{\lambda}^*, \mathbf{g}(\mathbf{x}) \rangle + \langle \boldsymbol{\mu}^*, \mathbf{h}(\mathbf{x}) \rangle \geq 0$ for all $\mathbf{x} \in X_0$. Since the problem CP is strongly consistent, then there exists an $\mathbf{x}_0$ such that $\mathbf{g}(\mathbf{x}_0) < \mathbf{0}$ and $\mathbf{h}(\mathbf{x}_0) = \mathbf{0}$. Therefore, $\langle \boldsymbol{\lambda}^*, \mathbf{g}(\mathbf{x}_0) \rangle + \langle \boldsymbol{\mu}^*, \mathbf{h}(\mathbf{x}_0) \rangle < 0$, which contradicts $\langle \boldsymbol{\lambda}^*, \mathbf{g}(\mathbf{x}) \rangle + \langle \boldsymbol{\mu}^*, \mathbf{h}(\mathbf{x}) \rangle \geq 0$ for all $\mathbf{x} \in X_0$. $\square$

Theorem 3.6

*Let **CP** be strongly consistent, let $\mathbf{x}^*$ be a solution, let ν denote the value of the minimum, and let $\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*$ be the multipliers associated with $\mathbf{x}^*$ as in Theorem 3.5. Then*

$$\begin{aligned}\nu = L(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) &= \min_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) \\ &= \sup_{(\boldsymbol{\lambda},\boldsymbol{\mu})} \inf_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \inf_{\mathbf{x}} \sup_{(\boldsymbol{\lambda},\boldsymbol{\mu})} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}),\end{aligned}$$

where $\boldsymbol{\lambda} \in \mathbb{R}^m, \boldsymbol{\lambda} \geq \mathbf{0}, \boldsymbol{\mu} \in \mathbb{R}^k, \boldsymbol{x} \in \mathbb{R}^n$.

Hint: It's obvious that $\sup_{(\boldsymbol{\lambda},\boldsymbol{\mu})} \inf_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) \leq \inf_{\mathbf{x}} \sup_{(\boldsymbol{\lambda},\boldsymbol{\mu})} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu})$. $\sup_{(\boldsymbol{\lambda},\boldsymbol{\mu})} L(\mathbf{x}^*, \boldsymbol{\lambda}, \boldsymbol{\mu}) = L(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = \inf_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$.

Convex Programming with Differentiable Functions

Theorem 3.7

Let f and $\mathbf{g}$ be convex and differentiable on $X_0 = \mathbb{R}^n$ and let $\mathbf{h} = \mathbf{A}\mathbf{x} - \mathbf{b}$. Let $\mathbf{A}$ have full rank and the problem be strongly consistent. A necessary and sufficient condition that a feasible $\mathbf{x}^*$ be a solution of the problem CP is that there exists a vector $\boldsymbol{\lambda} \geq \mathbf{0}$ in $\mathbb{R}^m$ and a vector $\boldsymbol{\mu}$ in $\mathbb{R}^k$ such that

(i) $\langle \boldsymbol{\lambda}, \mathbf{g}(\mathbf{x}^*) \rangle = 0$, *and*

(ii) $\nabla f(\mathbf{x}^*) + \boldsymbol{\lambda}^t \nabla \mathbf{g}(\mathbf{x}^*) + \boldsymbol{\mu}^t \nabla \mathbf{h}(\mathbf{x}^*) = \mathbf{0}$.

Proof.

If $\mathbf{x}^*$ is the optimal solution of the problem CP, then $L(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) \leq L(\mathbf{x}, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ for all $\mathbf{x} \in X_0$ where $L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\mathbf{x}) + \langle \boldsymbol{\lambda}, \mathbf{g}(\mathbf{x}) \rangle + \langle \boldsymbol{\mu}, \mathbf{h}(\mathbf{x}) \rangle$. Therefore, $\mathbf{x}^*$ is also the minimizer of $L(\mathbf{x}, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ over X_0. $\square$

Exercise

Consider the following problem

$$\begin{aligned} \text{minimize } & \mathbf{c^t x} \\ \text{subject to } & A\mathbf{x} = \mathbf{0}, \\ & \mathbf{x^t x} \leq 1, \end{aligned}$$

where $A \in \mathbb{R}^{m\times n}$ with rank$(A) = m$, and $\|A^t(AA^t)^{-1}A\mathbf{c} - \mathbf{c}\| \neq 0$. Please write the analytic expression of its optimal solution.
Hint: $\mathbf{x}^* = \frac{A^t(AA^t)^{-1}A\mathbf{c}-\mathbf{c}}{\|A^t(AA^t)^{-1}A\mathbf{c}-\mathbf{c}\|}$.

Intuition

- $Y_0 = \{\boldsymbol{\eta} : \boldsymbol{\eta} = (\boldsymbol{\lambda}, \boldsymbol{\mu}) : \boldsymbol{\lambda} \geq \mathbf{0}, \boldsymbol{\mu} \in \mathbb{R}^k\}$
- $\theta(\boldsymbol{\eta}) = \inf\{f(\mathbf{x}) + \langle \boldsymbol{\lambda}, \mathbf{g}(\mathbf{x}) \rangle + \langle \boldsymbol{\mu}, \mathbf{h}(\mathbf{x}) \rangle : \mathbf{x} \in \mathbb{R}^n\}$
- $\delta = \sup\{\theta(\boldsymbol{\eta}) : \boldsymbol{\eta} \in Y_0\}$
- $\nu = \inf\{f(\mathbf{x}) : \boldsymbol{g}(\boldsymbol{x}) \leq \mathbf{0}, \boldsymbol{h}(\boldsymbol{x}) = \mathbf{0}\}$
- By Theorem 3.6, $\delta = \nu$
- A possible procedure for finding $\mathbf{x}^*$, a point at which the minimum is achieved, is to first find all solutions to the equation $f(\mathbf{x}^*) = \nu$ and then retain those solutions that satisfy the constraints.

Let $Y = \{\boldsymbol{\eta} : \boldsymbol{\eta} \in Y_0, \theta(\boldsymbol{\eta}) > -\infty\}$. Define the dual problem of **CP** (**DCP**) as follows:

$$\text{maximize } \theta(\boldsymbol{\eta})$$
$$\text{subject to } \boldsymbol{\eta} \in Y$$

Example for Lagrange Dual

Consider the following primal problem

$$\begin{aligned} \text{minimize } & x_1^2 + x_2^2 \\ \text{subject to } & x_1 + x_2 - 4 \geq 0, \\ & x_1, x_2 \geq 0. \end{aligned}$$

Note that the optimal solution occurs at $(x_1, x_2) = (2, 2)$, whose objective value is equal to 8.

Example for Lagrange Dual continued

Let $X_0 = \{\mathbf{x} \in \mathbb{R}^2 | x_1, x_2 \geq 0\}$. The dual objective function is given by

$$\begin{aligned}\theta(u) &= \inf\{x_1^2 + x_2^2 - u(x_1 + x_2 - 4) : (x_1, x_2) \in X_0\} \\ &= \inf\{x_1^2 - ux_1 : x_1 \geq 0\} + \inf\{x_2^2 - ux_2 : x_2 \geq 0\} + 4u \\ &= \begin{cases} -\frac{1}{2}u^2 + 4u & \text{for } u \geq 0 \\ 4u & \text{for } u < 0. \end{cases}\end{aligned}$$

Then its Lagrangian dual problem is

$$\begin{aligned} &\text{maximize } -\frac{1}{2}u^2 + 4u \\ &\text{subject to } u \geq 0, \end{aligned}$$

whose optimal solution is $u = 4$ and the objective value is also 8.

Example for Lagrange Dual continued

Let $X_0 = \mathbb{R}^2$. The dual objective function is given by

$$\begin{aligned} \theta(u,v,w) &= \inf\{x_1^2 + x_2^2 - u(x_1 + x_2 - 4) - vx_1 - wx_2 : (x_1, x_2) \in X_0\} \\ &= \inf\{x_1^2 - (u+v)x_1\} + \inf\{x_2^2 - (u+w)x_2\} + 4u \end{aligned}$$

For $u, v, w \geq 0$, we have $\theta(u,v,w) = -\frac{1}{4}(u+v)^2 - \frac{1}{4}(u+w)^2 + 4u$. Then its Lagrangian dual problem is

$$\begin{aligned} &\text{maximize } -\frac{1}{4}(u+v)^2 - \frac{1}{4}(u+w)^2 + 4u \\ &\text{subject to } u, v, w \geq 0, \end{aligned}$$

whose optimal solution is $u = 4, v = 0, w = 0$ and the objective value is also 8.

Weak Lagrangian Duality Theorem

Theorem 4.1 (Lagrangian Duality Theorem)

(i) *If $X \neq \phi$ and $Y \neq \phi$, then for each $\mathbf{x} \in X$ and $\boldsymbol{\eta} \in Y$*

$$\theta(\boldsymbol{\eta}) \leq f(\mathbf{x})$$

Moreover, if δ and ν are defined as before, then both are finite and $\delta \leq \nu$.

(ii) *Let $Y \neq \phi$. If $\theta(\boldsymbol{\eta})$ is unbounded above on Y, then $X = \phi$.*

(iii) *Let $X \neq \phi$. If $f(\boldsymbol{x})$ is unbounded below on X, then $Y = \phi$.*

(iv) *If there exists an $\mathbf{x}^* \in X$ and $\boldsymbol{\eta}^* \in Y$ such that $f(\mathbf{x}^*) = \theta(\boldsymbol{\eta}^*)$, then $\delta = f(\mathbf{x}^*) = \theta(\boldsymbol{\eta}^*) = \nu$. Thus $\mathbf{x}^*$ is a solution of **CP** and $\boldsymbol{\eta}^*$ is a solution of **DCP**.*

Proof.

Since $\boldsymbol{\lambda} \geq \mathbf{0}$ for $\boldsymbol{\eta} \in Y$ and $\mathbf{g}(\mathbf{x}) \leq \mathbf{0}$ and $\mathbf{h}(\mathbf{x}) = \mathbf{0}$ for $\mathbf{x} \in X$, it follows that for $\mathbf{x} \in X$ and $\boldsymbol{\eta} \in Y$, $f(\mathbf{x}) + \langle \boldsymbol{\lambda}, \mathbf{g}(\mathbf{x}) \rangle + \langle \boldsymbol{\mu}, \mathbf{h}(\mathbf{x}) \rangle \leq f(\mathbf{x})$. □

Duality Gap

Definition 4.2

Duality Gap: Problems **CP** and **DCP** are said to exhibit a duality gap if $\delta < \nu$.

Consider the following primal problem

$$\begin{aligned} \text{minimize } & -2x_1 + x_2 \\ \text{subject to } & x_1 + x_2 - 3 = 0, \\ & (x_1, x_2) \in X_0, \end{aligned}$$

where $X_0 = \{(0,0), (0,4), (4,4), (4,0), (1,2), (2,1)\}$. It is easy to verify that $(2,1)$ is the optimal solution with objective value equal to -3.
The dual objective function is given by

$$\begin{aligned} \theta(v) &= \inf\{-2x_1 + x_2 - v(x_1 + x_2 - 3) : (x_1, x_2) \in X_0\} \\ &= \begin{cases} -4 - 5v & \text{for } v \geq 1 \\ -8 - v & \text{for } -2 \leq v \leq 1 \\ 3v & \text{for } v \leq -2 \end{cases} \end{aligned}$$

Then the dual optimal solution is $v^* = -2$ with objective value -6. Note that there exists a duality gap in this example.

Strong Lagrangian Duality Theorem

Theorem 4.3

*Let $X_0 = \mathbb{R}^n$, A have full rank, the primal problem **CP** be strongly consistent and have a solution $\mathbf{x}^*$. Then:*

(i) *There is no duality gap.*

(ii) *If $\boldsymbol{\eta}^*$ is a multiplier for **CP**, then $\boldsymbol{\eta}^*$ is a solution of the dual problem.*

(iii) *If $\boldsymbol{\eta}_0 = (\boldsymbol{\lambda}_0, \boldsymbol{\mu}_0)$ is a solution of the dual problem, then $(\mathbf{x}^*, \boldsymbol{\eta}_0)$ is a KKT pair for the primal problem.*

Proof.

Now let $\boldsymbol{\eta}_0 = (\boldsymbol{\lambda}_0, \boldsymbol{\mu}_0)$ be a solution of the dual problem. From the absence of a duality gap, we get that $\inf\{f(\mathbf{x}) + \langle \boldsymbol{\lambda}_0, \mathbf{g}(\mathbf{x})\rangle + \langle \boldsymbol{\mu}_0, \mathbf{h}(\mathbf{x})\rangle : \mathbf{x} \in \mathbb{R}^n\} = \theta(\boldsymbol{\eta}_0) = f(\mathbf{x}^*)$. Specially, $f(\mathbf{x}^*) + \langle \boldsymbol{\lambda}_0, \mathbf{g}(\mathbf{x}^*)\rangle + \langle \boldsymbol{\mu}_0, \mathbf{h}(\mathbf{x}^*)\rangle \geq f(\mathbf{x}^*)$. Since $\boldsymbol{\lambda}_0 \geq \mathbf{0}, \mathbf{g}(\mathbf{x}^*) \leq \mathbf{0}$ and $\mathbf{h}(\mathbf{x}^*) = \mathbf{0}$, then $\langle \boldsymbol{\lambda}_0, \mathbf{g}(\mathbf{x}^*)\rangle = 0$. □

Example

Example 4.4

Let $X_0 = \mathbb{R}^2$, minimize

$$f(\mathbf{x}) = x_1,$$

subject to

$$g_1(\mathbf{x}) = (x_1 + 1)^2 + (x_2)^2 - 1 \le 0, \quad g_2(\mathbf{x}) = -x_1 \le 0.$$

Solution: The problem is clearly a convex programming problem. The feasible set consists of the origin $\mathbf{0}$ in $\mathbb{R}^2$, and the problem is not strongly consistent. The solution to the primal problem is $\mathbf{x}^* = \mathbf{0}$ and $\nu = 0$.

The dual objective function is given by $\theta(\boldsymbol{\lambda}) = \inf\{x_1 + \lambda_1[(x_1 + 1)^2 + (x_2)^2 - 1] - \lambda_2 x_1 : \mathbf{x} \in \mathbb{R}^2\}$, where $\boldsymbol{\lambda} \ge \mathbf{0}$. If we take $\boldsymbol{\lambda} = \boldsymbol{\lambda}^* = (0, 1)$ we get that $\theta(\boldsymbol{\lambda}^*) = 0$. It then follows from (iv) of Theorem 4.1 that there is no duality gap and that $\boldsymbol{\lambda}^*$ is a solution of the dual problem.

Geometric Interpretation

- $I = \{(\mathbf{z}, \xi), \xi = f(\mathbf{x}), \mathbf{z} = \mathbf{g}(\mathbf{x}) : \mathbf{x} \in \mathbb{R}^n\}$
- $\theta(\boldsymbol{\lambda}) = \inf\{\xi + \langle \boldsymbol{\lambda}, \mathbf{z} \rangle : \xi = f(\mathbf{x}), \mathbf{z} = \mathbf{g}(\mathbf{x})\} = \inf\{\alpha : \alpha = \xi + \langle \boldsymbol{\lambda}, \mathbf{z} \rangle : (\mathbf{z}, \xi) \in I\}$
- $\langle (-\boldsymbol{\lambda}, -1), (\mathbf{z}, \xi) \rangle \leq -\theta(\boldsymbol{\lambda})$
- $\langle (-\boldsymbol{\lambda}, -1), (\mathbf{z}, \xi) \rangle = -\theta(\boldsymbol{\lambda})$ is a supporting hyperplane for the set I
- $\xi = \langle -\boldsymbol{\lambda}, \mathbf{z} \rangle + \theta(\boldsymbol{\lambda})$