Chapter 2 Convexity and Duality

A distinguishing idea which dominates many issues in optimization theory is convexity. Before we proceed, we quote the following from R. T. Rockafellar "Lagrange multipliers and optimality", *SIAM Review*, vol. 35, no. 2, pp. 183-238, June 1993: "Convexity is a large subject which can hardly be addressed here, but much of the impetus for its growth in recent decades has come from applications in optimization. An important reason is the fact that when a convex function is minimized over a convex set every locally optimal solution is global. Also, first-order necessary conditions for optimality turn out to be sufficient. A variety of other properties conducive to computation and interpretation of solutions ride on convexity as well. In fact the great watershed in optimization isn't between linearity and nonlinearity, but convexity and nonconvexity. Even for problems that aren't themselves of convex type, convexity may enter, for instance, in setting up subproblems as part of an iterative numerical scheme."

2.0 Notation

$\mathbf{R}$: the set of all real numbers.

$\mathbf{R}_+$: the set of all nonnegative real numbers.

$\mathbf{R}_{++}$: the set of all positive real numbers.

$\mathbf{R}^p$: the set of p-dimensional real-valued (column) vectors.

$\mathbf{R}^{p\times q}$: the set of real-valued matrices of size p by q.

$\mathbf{S}^n$: the set of n by n symmetric matrices.

$\mathbf{S}_+^n$: the set of n by n symmetric positive semidefinite matrices.

$\mathbf{S}_{++}^n$: the set of n by n symmetric positive definite matrices.

dom f : domain where function $f(\boldsymbol{x})$ is defined.

2.1 Convex Sets and Convex Functions

Definition 2.1 A set $\mathcal{R}_c \subset E^n$ is said to be *convex* if for every pair of points $\boldsymbol{x}_1$, $\boldsymbol{x}_2$ in $\mathcal{R}_c$ and for every real number α in the range $0 < \alpha < 1$, the point $\boldsymbol{x} = \alpha\boldsymbol{x}_1 + (1-\alpha)\boldsymbol{x}_2$ is located in $\mathcal{R}_c$.

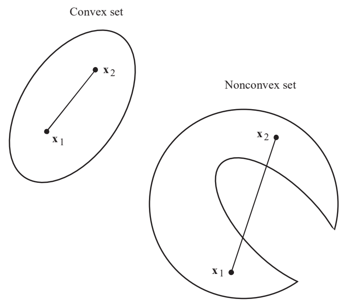

Figure 2.1 Convex and nonconvex sets.

Definition 2.2 A function $f(\boldsymbol{x})$ defined over a convex set $\mathcal{R}_c$ is said to be convex if for every pair of points in $\mathcal{R}_c$ and every real number α in the range 0 < α < 1, the following inequality holds

$$f[\alpha \boldsymbol{x}_1 + (1-\alpha)\boldsymbol{x}_2] \le \alpha f(\boldsymbol{x}_1) + (1-\alpha) f(\boldsymbol{x}_2) \tag{2.1}$$

Geometrically, (2.1) says that $f(\boldsymbol{x})$ is convex if and only if the function's graph always lies underneath (or coincides with) the corresponding line segment, see Fig. 2.2.

- $f(\boldsymbol{x})$ is said to be *strictly* convex if (2.1) holds strictly, i.e.,

$$f[\alpha \boldsymbol{x}_1 + (1-\alpha)\boldsymbol{x}_2] < \alpha f(\boldsymbol{x}_1) + (1-\alpha) f(\boldsymbol{x}_2)$$

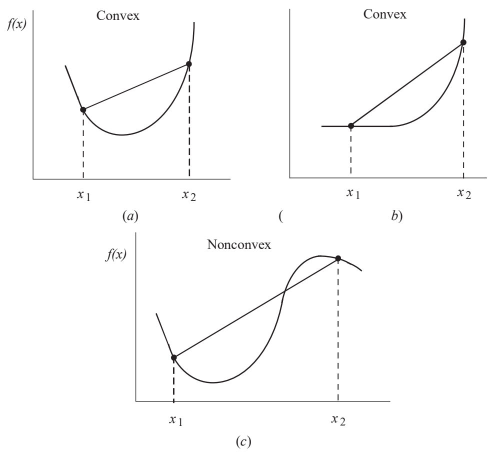

Figure 2.2 Convex and nonconvex functions

Definition 2.3 A function $f(\boldsymbol{x})$ is said to be concave if for every pair of points and every real number α in the range 0 < α < 1, the following inequality holds

$$f[\alpha \boldsymbol{x}_1 + (1-\alpha)\boldsymbol{x}_2] \ge \alpha f(\boldsymbol{x}_1) + (1-\alpha) f(\boldsymbol{x}_2) \tag{2.2}$$

- $f(\boldsymbol{x})$ is concave if and only if the function's graph always lies above (or coincides with) the corresponding line segment.
- $f(\boldsymbol{x})$ is said to be *strictly* concave if the inequality in (2.2) holds strictly.
- $f(\boldsymbol{x})$ is (strictly) convex if and only if $-f(\boldsymbol{x})$ is (strictly) concave, and vice versa.

Definition 2.4 The extended-value extension of a convex function $f(\boldsymbol{x})$ is defined as

$$\tilde{f}(\boldsymbol{x}) = \begin{cases} f(\boldsymbol{x}) & \boldsymbol{x} \in \operatorname{dom} f \\ +\infty & \boldsymbol{x} \notin \operatorname{dom} f \end{cases} \tag{2.3}$$

Examples of convex functions of one variable:

- affine function $f(x) = ax + b$
- exponential function $f(x) = e^{\alpha x}$ for any $\alpha \in \mathrm{R}$
- power function $f(x) = x^{\alpha}$ on R_{++} for $\alpha \geq 1$ or $\alpha \leq 0$
- negative entropy $f(x) = x \log x$ on R_{++}

Examples of concave functions of one variable:

- affine function $f(x) = ax + b$
- logrithmic function $f(x) = \log x$ on R_{++}
- power function $f(x) = x^{\alpha}$ on R_{++} for $0 \leq \alpha \leq 1$

Examples of multivariable convex functions:

- affine function $f(\boldsymbol{x}) = \boldsymbol{a}^T \boldsymbol{x} + b$
- norm $f(\boldsymbol{x}) = \|\boldsymbol{x}\|_p = \left(|x_1|^p + |x_2|^p + \cdots + |x_n|^p\right)^{1/p}$ for $p \geq 1$, and $\|\boldsymbol{x}\|_\infty = \max_{1 \leq i \leq n}\{|x_i|\}$

2.2 Properties of Convex Functions

(1) ***Convexity of linear combination of convex functions*** If $f_1(\boldsymbol{x})$ and $f_2(\boldsymbol{x})$ are convex and a and b are nonnegative scalars, then $a f_1(\boldsymbol{x}) + b f_2(\boldsymbol{x})$ is convex.

(2) ***Relation between a convex function and convex sets*** If $f(\boldsymbol{x})$ is convex on a convex set $\mathcal{R}_c$, then the set $S_K = \{\boldsymbol{x} : \boldsymbol{x} \in \mathcal{R}_c, f(\boldsymbol{x}) \leq K\}$ is convex for every real number K. See Fig. 2.3 for an illustration of this property.

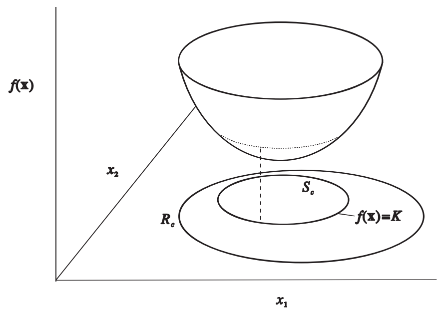

Figure 2.3

Convex functions can be characterized in different ways. The next definition of convexity turns out to be very useful.

(3) ***Property of convex function relating to gradient***

Suppose $f(\boldsymbol{x})$ is a smooth (C^1) function. Then $f(\boldsymbol{x})$ is convex over $\mathcal{R}_c$ if and only if

$$f(\boldsymbol{x}_1) \ge f(\boldsymbol{x}) + \nabla^T f(\boldsymbol{x})(\boldsymbol{x}_1 - \boldsymbol{x}) \tag{2.4}$$

for all $\boldsymbol{x}$ and $\boldsymbol{x}_1$ in $\mathcal{R}_c$. See Fig. 2.4 for an illustration of this property.

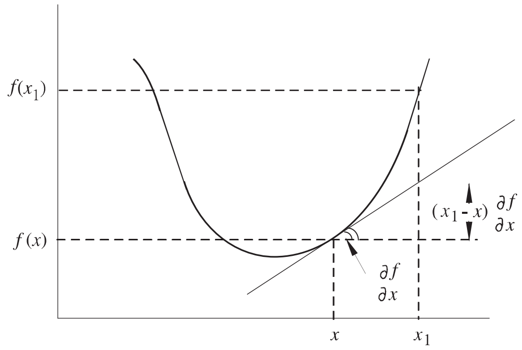

Fig. 2.4

(4) ***Property of convex functions relating to the Hessian*** A function $f(\boldsymbol{x}) \in C^2$ is convex over a convex $\mathcal{R}_c$ if and only if the Hessian $\nabla^2 f(\boldsymbol{x})$ is positive semidefinite over $\mathcal{R}_c$.

To prove property 4, we use Taylor expansion to write

$$f(\boldsymbol{x}_1) = f(\boldsymbol{x}) + \nabla^T f(\boldsymbol{x})(\boldsymbol{x}_1 - \boldsymbol{x}) + \tfrac{1}{2}(\boldsymbol{x}_1 - \boldsymbol{x})^T \nabla^2 f(\boldsymbol{x})(\boldsymbol{x}_1 - \boldsymbol{x}) + o(\|\boldsymbol{x}_1 - \boldsymbol{x}\|^2)$$

for all $\boldsymbol{x}$ and $\boldsymbol{x}_1$ in $\mathcal{R}_c$. Hence

$$f(\boldsymbol{x}_1) - f(\boldsymbol{x}) - \nabla^T f(\boldsymbol{x})(\boldsymbol{x}_1 - \boldsymbol{x}) = \tfrac{1}{2}(\boldsymbol{x}_1 - \boldsymbol{x})^T \nabla^2 f(\boldsymbol{x})(\boldsymbol{x}_1 - \boldsymbol{x}) + o(\|\boldsymbol{x}_1 - \boldsymbol{x}\|^2)$$

which implies that the left-hand side of the above equation is nonnegative if and only if $\nabla^2 f(\boldsymbol{x})$ is positive semidefinite. But from property 3 the left-hand side of the above equation is nonnegative if and only if $f(\boldsymbol{x})$ is convex. ■

(5) ***From an optimization perspective, what* (2.4) *tells us?*** (2.4) is a characterization of a function being convex, but it also provides a global linear lower bound of a convex that is tightest at a given point in the function's (convex) domain. The observation made below offers yet another way to appreciate the importance of convexity.

Let $\boldsymbol{x}$ be a point in the domain (let us assume it is the entire space R^n) of a smooth convex function $f(\boldsymbol{x})$ with gradient $\nabla f(\boldsymbol{x}) \ne \boldsymbol{0}$ (this is reasonable to assume. In fact, if $\nabla f(\boldsymbol{x}) = \boldsymbol{0}$, then $\boldsymbol{x}$ would be a global minimizer of $f(\boldsymbol{x})$, hence not much left for us to do, see the next property). The hyperplane that contains point $\boldsymbol{x}$ and with gradient $\nabla f(\boldsymbol{x})$ as its normal can be described as a set of $\boldsymbol{x}_1$, each satisfies

$$\mathcal{P}_x: \qquad \nabla^T f(\boldsymbol{x})(\boldsymbol{x}_1 - \boldsymbol{x}) = 0$$

and hyperplane $\mathcal{P}_x$ cuts (divides) the entire domain into two parts, see Fig. 2.5. Let us examine the part of the domain that contains gradient $\nabla f(\boldsymbol{x})$ (point $\boldsymbol{x} + \nabla f(\boldsymbol{x})$ to be precise). From the figure it is immediate that at each point $\boldsymbol{x}_1$ in this part (excluding hyperplane $\mathcal{P}_x$ itself) satisfies

$$\nabla^T f(\boldsymbol{x})(\boldsymbol{x}_1 - \boldsymbol{x}) > 0$$

which in conjunction with (2.4) gives

$$f(\boldsymbol{x}_1) \geq f(\boldsymbol{x}) + \nabla^T f(\boldsymbol{x})(\boldsymbol{x}_1 - \boldsymbol{x}) > f(\boldsymbol{x})$$

Therefore, this part of the domain contains *no* minimizers. In other words, the minimizers are all contained in the other part of the domain. We see that (2.4) is of help in reducing a region in searching a minimizer.

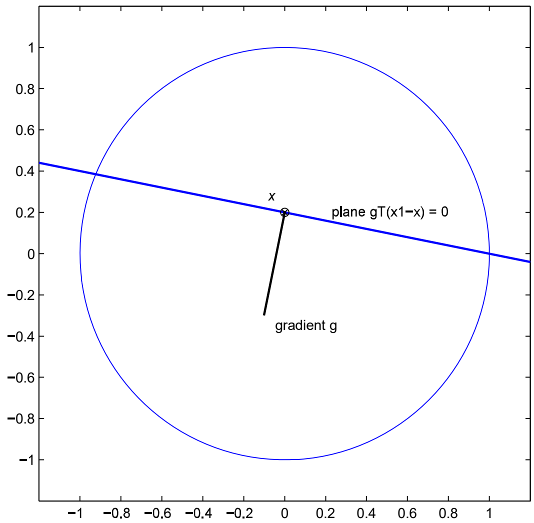

Figure 2.5

(6) ***Global solution of an unconstrained convex minimization problem***

Let $f(\boldsymbol{x})$ be a smooth convex function. If $\nabla f(\boldsymbol{x}^*) = \boldsymbol{0}$, then $\boldsymbol{x}^*$ is a global minimizer of $f(\boldsymbol{x})$. This property is an immediate consequence of (2.4):

$$f(\boldsymbol{x}) \geq f(\boldsymbol{x}^*) + \nabla^T f(\boldsymbol{x}^*)(\boldsymbol{x} - \boldsymbol{x}^*) = f(\boldsymbol{x}^*)$$

(7) ***Globalness and convexity of minimizers in convex programming (CP) problems.*** Recall a constrained problem is a CP problem if it assumes the form

$$\begin{aligned} &\text{minimize} \quad f(\boldsymbol{x}) \\ &\text{subject to:} \quad \boldsymbol{a}_i^T \boldsymbol{x} = b_i \quad \text{for } 1 \leq i \leq p \\ &\qquad\qquad\quad c_j(\boldsymbol{x}) \leq 0 \quad \text{for } 1 \leq j \leq q \end{aligned}$$

where $f(\boldsymbol{x})$ and $c_j(\boldsymbol{x})$ are convex. That is, a CP problem minimizes a *convex objective* *function* over *a convex feasible region*. We can state that

(a) $\boldsymbol{x}^*$ is a minimizer if and only if

$$\nabla f(\boldsymbol{x}^*)^T(\boldsymbol{x}-\boldsymbol{x}^*) \geq 0 \quad \text{for any feasible } \boldsymbol{x} \tag{2.5}$$

Proof The necessity of (2.5) immediately follows from Theorem 1.1, and the sufficiency of (2.5) follows from (2.4), namely,

$$f(\boldsymbol{x}) \geq f(\boldsymbol{x}^*) + \nabla^T f(\boldsymbol{x}^*)(\boldsymbol{x}-\boldsymbol{x}^*)$$

which in conjuction with (2.5) gives

$$f(\boldsymbol{x}) - f(\boldsymbol{x}^*) \geq \nabla^T f(\boldsymbol{x}^*)(\boldsymbol{x}-\boldsymbol{x}^*) \geq 0$$

hence $f(\boldsymbol{x}) \geq f(\boldsymbol{x}^*)$. ■

(b) If $\boldsymbol{x}^*$ is a local minimizer, then $\boldsymbol{x}^*$ is also a global minimizer.

(c) The set of minimizers of a CP problem is a convex set.

(d) If the objective function is strictly convex, then the minimizer is unique.

(8) ***Sufficiency of KKT Conditions for Convex Problems*** If $\boldsymbol{x}^*$ is a regular point of the constraints in (1.44) and satisfies the KKT conditions, then it is a global minimizer.

Proof For a feasible point $\bar{\boldsymbol{x}}$ with $\bar{\boldsymbol{x}} \neq \boldsymbol{x}^*$, we have $a_i(\bar{\boldsymbol{x}}) = 0$ for $1 \leq i \leq p$ and $c_j(\bar{\boldsymbol{x}}) \leq 0$ for $1 \leq j \leq q$. Thus we can write

$$f(\bar{\boldsymbol{x}}) \geq f(\bar{\boldsymbol{x}}) + \sum_{j=1}^{q} \mu_j^* c_j(\bar{\boldsymbol{x}}).$$

Since $f(\boldsymbol{x})$ and $c_j(\boldsymbol{x})$ are convex, we have

$$f(\bar{\boldsymbol{x}}) \geq f(\boldsymbol{x}^*) + \nabla^T f(\boldsymbol{x}^*)(\bar{\boldsymbol{x}}-\boldsymbol{x}^*) \quad \text{and} \quad c_j(\bar{\boldsymbol{x}}) \geq c_j(\boldsymbol{x}^*) + \nabla^T c_j(\boldsymbol{x}^*)(\bar{\boldsymbol{x}}-\boldsymbol{x}^*)$$

Hence

$$f(\bar{\boldsymbol{x}}) \geq f(\boldsymbol{x}^*) + \nabla^T f(\boldsymbol{x}^*)(\bar{\boldsymbol{x}}-\boldsymbol{x}^*) + \sum_{j=1}^{q} \mu_j^* \nabla^T c_j(\boldsymbol{x}^*)(\bar{\boldsymbol{x}}-\boldsymbol{x}^*) + \sum_{j=1}^{q} \mu_j^* c_j(\boldsymbol{x}^*)$$

In the light of the complementarity conditions, the last term in the above inequality vanishes, hence we have

$$f(\bar{\boldsymbol{x}}) \geq f(\boldsymbol{x}^*) + \left[\nabla f(\boldsymbol{x}^*) + \sum_{j=1}^{q} \mu_j^* \nabla c_j(\boldsymbol{x}^*)\right]^T (\bar{\boldsymbol{x}}-\boldsymbol{x}^*) \tag{2.6}$$

Since $a_i(\bar{\boldsymbol{x}}) = a_i(\boldsymbol{x}^*) = 0$, we can write

$$0 = a_i(\bar{\boldsymbol{x}}) - a_i(\boldsymbol{x}^*) = \boldsymbol{a}_i^T(\bar{\boldsymbol{x}}-\boldsymbol{x}^*) = \nabla^T a_i(\boldsymbol{x}^*)(\bar{\boldsymbol{x}}-\boldsymbol{x}^*)$$

Multiplying the above by λ_i^* and adding it to (2.6) for $1 \le i \le p$ gives

$$f(\bar{\boldsymbol{x}}) \ge f(\boldsymbol{x}^*) + \left[\nabla f(\boldsymbol{x}^*) + \sum_{i=1}^{p} \lambda_i^* \nabla a_i(\boldsymbol{x}^*) + \sum_{j=1}^{q} \mu_j^* \nabla c_j(\boldsymbol{x}^*) \right]^T (\bar{\boldsymbol{x}} - \boldsymbol{x}^*)$$

By the KKT conditions, the last term in the above inequality is zero, which leads to $f(\bar{\boldsymbol{x}}) \ge f(\boldsymbol{x}^*)$. This shows that $\boldsymbol{x}^*$ is a global minimizer. ■

2.3 How to Verify Convexity of a Function?

(1) By definitions. See Eqs. (2.1) and (2.4).

(2) Check if its Hessian is positive semidefinite, see Perperty (4) in Sec. 2.2.

(3) Check the function's convexity along lines: $f(\boldsymbol{x})$ is convex if and only if $g(t) = f(\boldsymbol{x} + t\boldsymbol{v})$ is convex in t in $\text{dom } g = \{t : \boldsymbol{x} + t\boldsymbol{v} \in \text{dom } f)$ for any $\boldsymbol{x} \in \text{dom } f$.

- Example: $f(\boldsymbol{X}) = -\log\det(\boldsymbol{X})$ on $\text{dom} f = \mathrm{S}_{++}^n$ is convex.

(4) Composition with affine function: if $f(\boldsymbol{x})$ is convex, then so is $f(\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b})$.

- Examples:
  (i) log barrier for linear inequalities, i.e.,

$f(\boldsymbol{x}) = -\sum_{i=1}^{m} \log(b_i - \boldsymbol{a}_i^T \boldsymbol{x})$, $\text{dom} f = \{\boldsymbol{x} : \boldsymbol{a}_i^T \boldsymbol{x} < b_i, i = 1, 2, ..., m\}$ is convex;

  (ii) any norm of affine function $f(\boldsymbol{x}) = \|\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}\|$ is convex.

(5) If all $f_i(\boldsymbol{x})$ are convex, then $f(\boldsymbol{x}) = \max\{f_1(\boldsymbol{x}), f_2(\boldsymbol{x}), \ldots, f_m(\boldsymbol{x})\}$ is convex.

- Example: piecewise-linear function $f(\boldsymbol{x}) = \max_{1 \le i \le m}\{\boldsymbol{a}_i^T \boldsymbol{x} + b_i\}$ is convex.

(6) If $g(\boldsymbol{x})$ is convex, $h(\boldsymbol{x})$ is convex and $\tilde{h}(\boldsymbol{x})$ is nondecreasing, then $f(\boldsymbol{x}) = h(g(\boldsymbol{x}))$ is convex. (here $\tilde{h}(\boldsymbol{x})$ denotes the extended-value extension of $h(\boldsymbol{x})$, see Definition 2.4)

- Example: if $f(\boldsymbol{x})$ is convex, then so is $e^{f(x)}$.

(7) If $g(\boldsymbol{x})$ is concave, $h(\boldsymbol{x})$ is convex and $\tilde{h}(\boldsymbol{x})$ is nonincreasing, then $f(\boldsymbol{x}) = h(g(\boldsymbol{x}))$ is convex.

- Examples:
  (i) if $f(\boldsymbol{x})$ is concave and positive, then $1/f(\boldsymbol{x})$ is convex.

(ii) if $f(\boldsymbol{x})$ is concave and positive, then $-\log f(\boldsymbol{x})$ is convex.

(8) A vector version of property (6): If all $g_i(\boldsymbol{x})$ are convex and $h(\boldsymbol{x}) = h(x_1, x_2, ..., x_m)$ is convex and $\tilde{h}(\boldsymbol{x})$ is nondecreasing in each variable, then $f(\boldsymbol{x}) = h(g_1(\boldsymbol{x}), g_2(\boldsymbol{x}), ..., g_m(\boldsymbol{x}))$ is convex.

- Example: if all $f_i(\boldsymbol{x})$ are convex, then $\log \sum_{i=1}^{m} e^{f_i(\boldsymbol{x})}$ is convex (Prob. 2.5).

(9) A vector version of property (7): If all $g_i(\boldsymbol{x})$ are concave and $h(\boldsymbol{x}) = h(x_1, x_2, ..., x_m)$ is convex and $\tilde{h}(\boldsymbol{x})$ is nonincreasing in each variable, then $f(\boldsymbol{x}) = h(g_1(\boldsymbol{x}), g_2(\boldsymbol{x}), ..., g_m(\boldsymbol{x}))$ is convex.

- Example: if all $f_i(\boldsymbol{x})$ are concave and positive, then $-\sum_{i=1}^{m} \log f_i(\boldsymbol{x})$ is convex.

2.4 Duality

The concept of duality as applied to optimization is essentially a problem transformation that leads to an indirect but sometimes more efficient solution method. In a duality-based method, the original problem, which is referred to as the *primal* problem, is transformed into a problem in which the parameters are the Lagrange multipliers of the primal. The transformed problem is called the *dual* problem. In the case where the number of inequality constraints is much greater than the dimension of $\boldsymbol{x}$, solving the dual problem to find the Lagrange multipliers and then finding $\boldsymbol{x}^*$ for the primal problem becomes an attractive alternative.

2.4.1 The Lagrange Dual of a Convex Programming Problem

In this section we introduce another concept known as the Lagrange dual which turns out to be more useful in the study of convex programming. To this end we need a related concept called Lagrange dual function. Consider the general convex programming (CP) problem

$$\begin{aligned} &\text{minimize} \quad f(\boldsymbol{x}) \\ &\text{subject to:} \ \boldsymbol{a}_i^T \boldsymbol{x} = b_i \quad \text{for} \ 1 \le i \le p \\ &\qquad\qquad\quad c_j(\boldsymbol{x}) \le 0 \quad \text{for} \ 1 \le j \le q \end{aligned} \tag{2.7}$$

and recall its Lagrangian

$$L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\boldsymbol{x}) + \sum_{i=1}^{p} \lambda_i \left(\boldsymbol{a}_i^T \boldsymbol{x} - b_i \right) + \sum_{j=1}^{q} \mu_i \, c_j(\boldsymbol{x})$$

Definition 2.5 The *Lagrange dual function* of problem is defined as

$$q(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \inf_{\boldsymbol{x}} L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) \tag{2.8}$$

for $\boldsymbol{\lambda} \in R^p$ and $\boldsymbol{\mu} \in R^q$ with $\boldsymbol{\mu} \ge \boldsymbol{0}$. Note that the Lagrangian is *convex* with respect to $\boldsymbol{x}$. If $L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu})$ is unbounded below for some $\{\boldsymbol{\lambda}, \ \boldsymbol{\mu}\}$, then the value of $q(\boldsymbol{\lambda}, \boldsymbol{\mu})$ is assigned to $-\infty$.

Property 1 $q(\boldsymbol{\lambda},\boldsymbol{\mu})$ is a concave function with respect to $\{\boldsymbol{\lambda},\ \boldsymbol{\mu}\}$. The property follows from that fact that for $\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2 \in R^p$ and $\boldsymbol{\mu}_1,\boldsymbol{\mu}_2 \in R^q$ with $\boldsymbol{\mu}_1,\boldsymbol{\mu}_2 \ge \mathbf{0}$ and for $t \in (0,1)$, we have

$$q(t\boldsymbol{\lambda}_1+(1-t)\boldsymbol{\lambda}_2, t\boldsymbol{\mu}_1+(1-t)\boldsymbol{\mu}_2) = \inf_{\boldsymbol{x}} L(\boldsymbol{x}, t\boldsymbol{\lambda}_1+(1-t)\boldsymbol{\lambda}_1, t\boldsymbol{\mu}_1+(1-t)\boldsymbol{\mu}_2)$$
$$= \inf_{\boldsymbol{x}}\left[(t+1-t)f(\boldsymbol{x})+\sum_{i=1}^{p}(t\lambda_{1,i}+(1-t)\lambda_{2,i})(a_i^T\boldsymbol{x}-b_i)+\sum_{j=1}^{q}(t\mu_{1,j}+(1-t)\mu_{2,j})c_j(\boldsymbol{x})\right]$$
$$\ge t\cdot\inf_{\boldsymbol{x}}\left[f(\boldsymbol{x})+\sum_{i=1}^{p}\lambda_{1,i}(a_i^T\boldsymbol{x}-b_i)+\sum_{j=1}^{q}\mu_{1,j}c_j(\boldsymbol{x})\right]+(1-t)\cdot\inf_{\boldsymbol{x}}\left[f(\boldsymbol{x})+\sum_{i=1}^{p}\lambda_{2,i}(a_i^T\boldsymbol{x}-b_i)+\sum_{j=1}^{q}\mu_{2,j}c_j(\boldsymbol{x})\right]$$
$$= t\cdot q(\boldsymbol{\lambda}_1,\boldsymbol{\mu}_1)+(1-t)\cdot q(\boldsymbol{\lambda}_2,\boldsymbol{\mu}_2)$$

Definition 2.6 The *Lagrange dual problem* with respect to problem (2.7) is defined as

$$\begin{aligned}&\underset{\boldsymbol{\lambda},\boldsymbol{\mu}}{\text{maximize}} \quad q(\boldsymbol{\lambda},\boldsymbol{\mu})\\ &\text{subject to:}\ \ \boldsymbol{\mu}\ge\mathbf{0}\end{aligned} \tag{2.9}$$

Property 2 For any $\boldsymbol{x}$ feasible for problem (2.7) and $\{\boldsymbol{\lambda},\ \boldsymbol{\mu}\}$ feasible for problem (2.9), we have

$$f(\boldsymbol{x}) \ge q(\boldsymbol{\lambda},\boldsymbol{\mu}) \tag{2.10}$$

This is because

$$L(\boldsymbol{x},\boldsymbol{\lambda},\boldsymbol{\mu}) = f(\boldsymbol{x})+\sum_{i=1}^{p}\lambda_i\left(\boldsymbol{a}_i^T\boldsymbol{x}-b_i\right)+\sum_{j=1}^{q}\mu_i c_j(\boldsymbol{x}) = f(\boldsymbol{x})+\sum_{j=1}^{q}\mu_i c_j(\boldsymbol{x}) \le f(\boldsymbol{x})$$

thus

$$q(\boldsymbol{\lambda},\boldsymbol{\mu}) = \inf_{\boldsymbol{x}} L(\boldsymbol{x},\boldsymbol{\lambda},\boldsymbol{\mu}) \le L(\boldsymbol{x},\boldsymbol{\lambda},\boldsymbol{\mu}) \le f(\boldsymbol{x})$$

We call the convex minimization problem in (2.7) the *primal problem* and the concave maximization problem in (2.9) the *dual problem*. From (2.10), it is natural to introduce a *duality gap* between the primal and dual objectives as

$$\delta(\boldsymbol{x},\boldsymbol{\lambda},\boldsymbol{\mu}) = f(\boldsymbol{x}) - q(\boldsymbol{\lambda},\boldsymbol{\mu}) \tag{2.11}$$

It follows that for feasible $\{\boldsymbol{x},\boldsymbol{\lambda},\ \boldsymbol{\mu}\}$ the duality gap is always nonnegative.

Property 3 Let $\boldsymbol{x}^*$ be a solution of the primal problem in (2.7). Then the dual function at any feasible $\{\boldsymbol{\lambda},\ \boldsymbol{\mu}\}$ serves as a lower bound of the optimal value of the primal objective, $f(\boldsymbol{x}^*)$, namely,

$$f(\boldsymbol{x}^*) \ge q(\boldsymbol{\lambda},\boldsymbol{\mu}) \tag{2.12}$$

This property follows immediately from (2.10) by taking the minimum of $f(\boldsymbol{x})$ on its left-hand side.

- Now a question naturally arises: what will be the tightest lower bound the dual function $q(\boldsymbol{\lambda}, \boldsymbol{\mu})$ can offer? From (2.12), obviously the answer is found by maximizing the dual function $q(\boldsymbol{\lambda}, \boldsymbol{\mu})$ on the right-hand side of (2.12) subject to $\boldsymbol{\mu} \geq \mathbf{0}$. This is exactly the Lagrange dual problem as formulated in (2.9). Therefore, if we denote the solution of (2.9) by $(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$, then we have

$$f(\boldsymbol{x}^*) \geq q(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) \tag{2.13}$$

Based on (2.13), we now introduce the concept of *strong* and *weak* duality as follows.

Definition 2.7 Let $\boldsymbol{x}^*$ and $(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ be solutions of primal problem (2.7) and dual problem (2.9), respectively. We say strong duality holds if $f(\boldsymbol{x}^*) = q(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$, i.e., the optimal duality gap is zero; and a weak duality holds if $f(\boldsymbol{x}^*) > q(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$.

- It can be shown that if the primal problem is strictly feasible, i.e., there exists $\boldsymbol{x}$ satisfying

$$\begin{aligned} &\boldsymbol{a}_i^T \boldsymbol{x} = b_i \quad \text{for } 1 \leq i \leq p \\ &c_j(\boldsymbol{x}) < 0 \quad \text{for } 1 \leq j \leq q \end{aligned}$$

(this is to say that the interior of the feasible region of problem (2.7) is nonempty), then strong duality holds, i.e., the optimal duality gap is zero.

Example 2.1 Find the Lagrange dual of the LP problem

$$\begin{aligned} &\text{minimize} \quad \boldsymbol{c}^T \boldsymbol{x} \\ &\text{subject to:} \;\; \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \;\; \boldsymbol{x} \geq \mathbf{0} \end{aligned} \tag{1.3}$$

Solution We write $\boldsymbol{x} \geq \mathbf{0}$ as $-\boldsymbol{x} \leq \mathbf{0}$ hence the Lagrangian of the LP problem is given by

$$L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \boldsymbol{c}^T \boldsymbol{x} + (\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b})^T \boldsymbol{\lambda} - \boldsymbol{x}^T \boldsymbol{\mu}$$

thus

$$q(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \inf_{\boldsymbol{x}} \left\{ \boldsymbol{c}^T \boldsymbol{x} + (\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b})^T \boldsymbol{\lambda} - \boldsymbol{x}^T \boldsymbol{\mu} \right\} = \inf_{\boldsymbol{x}} \left\{ (\boldsymbol{c} + \boldsymbol{A}^T \boldsymbol{\lambda} - \boldsymbol{\mu})^T \boldsymbol{x} - \boldsymbol{b}^T \boldsymbol{\lambda} \right\} \tag{2.14}$$

For given $\{\boldsymbol{\lambda}, \boldsymbol{\mu}\}$ such that $\boldsymbol{c} + \boldsymbol{A}^T \boldsymbol{\lambda} - \boldsymbol{\mu} \neq \mathbf{0}$, from (2.14) we have $q(\boldsymbol{\lambda}, \boldsymbol{\mu}) = -\infty$. Therefore, to deal with a well-defined dual function $q(\boldsymbol{\lambda}, \boldsymbol{\mu})$ we assume $\boldsymbol{c} + \boldsymbol{A}^T \boldsymbol{\lambda} - \boldsymbol{\mu} = \mathbf{0}$ which leads to

$$q(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \inf_{\boldsymbol{x}} (-\boldsymbol{b}^T \boldsymbol{\lambda}) = -\boldsymbol{b}^T \boldsymbol{\lambda}$$

and the Lagrange dual of (1.3) is given by

$$\begin{array}{ll} \underset{\boldsymbol{\lambda},\boldsymbol{\mu}}{\text{maximize}} & -\boldsymbol{b}^T\boldsymbol{\lambda} \\ \text{subject to:} & \boldsymbol{\mu} \geq \boldsymbol{0} \end{array}$$

Since $\boldsymbol{c} + \boldsymbol{A}^T\boldsymbol{\lambda} - \boldsymbol{\mu} = \boldsymbol{0}$, $\boldsymbol{\mu} = \boldsymbol{c} + \boldsymbol{A}^T\boldsymbol{\lambda}$ so the above problem becomes

$$\begin{array}{ll} \underset{\boldsymbol{\lambda}}{\text{maximize}} & -\boldsymbol{b}^T\boldsymbol{\lambda} \\ \text{subject to:} & -\boldsymbol{c} - \boldsymbol{A}^T\boldsymbol{\lambda} \leq \boldsymbol{0} \end{array}$$

which is equivalent to

$$\begin{array}{ll} \underset{\boldsymbol{\lambda}}{\text{minimize}} & \boldsymbol{b}^T\boldsymbol{\lambda} \\ \text{subject to:} & (-\boldsymbol{A}^T)\boldsymbol{\lambda} \leq \boldsymbol{c} \end{array}$$

■

Example 2.2 Find the Lagrange dual of the QP problem

$$\begin{array}{ll} \text{minimize} & \tfrac{1}{2}\boldsymbol{x}^T\boldsymbol{H}\boldsymbol{x} + \boldsymbol{p}^T\boldsymbol{x} \\ \text{subject to:} & \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b} \end{array} \tag{1.5}$$

where $\boldsymbol{H}$ is positive definite.

Solution The Lagrangian of the QP problem is given by

$$L(\boldsymbol{x},\boldsymbol{\mu}) = \tfrac{1}{2}\boldsymbol{x}^T\boldsymbol{H}\boldsymbol{x} + \boldsymbol{p}^T\boldsymbol{x} + \boldsymbol{\mu}^T(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b})$$

Hence

$$q(\boldsymbol{\mu}) = \inf_{\boldsymbol{x}}\left\{\tfrac{1}{2}\boldsymbol{x}^T\boldsymbol{H}\boldsymbol{x} + \boldsymbol{p}^T\boldsymbol{x} + \boldsymbol{\mu}^T(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b})\right\} \tag{2.15}$$

where the infimum is attained at $\boldsymbol{x} = -\boldsymbol{H}^{-1}(\boldsymbol{p} + \boldsymbol{A}^T\boldsymbol{\mu})$. By substituting this solution into (2.15), we obtain

$$q(\boldsymbol{\mu}) = -\tfrac{1}{2}\boldsymbol{\mu}^T\boldsymbol{A}\boldsymbol{H}^{-1}\boldsymbol{A}^T\boldsymbol{\mu} - \boldsymbol{\mu}^T(\boldsymbol{A}\boldsymbol{H}^{-1}\boldsymbol{p} + \boldsymbol{b}) - \tfrac{1}{2}\boldsymbol{p}^T\boldsymbol{H}^{-1}\boldsymbol{p} \tag{2.16}$$

If we let $\boldsymbol{P} = \boldsymbol{A}\boldsymbol{H}^{-1}\boldsymbol{A}^T$, $\boldsymbol{t} = \boldsymbol{A}\boldsymbol{H}^{-1}\boldsymbol{p} + \boldsymbol{b}$ and neglect the constant term in (2.16), the dual function becomes $q(\boldsymbol{\mu}) = -\tfrac{1}{2}\boldsymbol{\mu}^T\boldsymbol{P}\boldsymbol{\mu} - \boldsymbol{\mu}^T\boldsymbol{t}$, hence the Lagrange dual of (1.5) is given by

$$\begin{array}{ll} \text{minimize} & \tfrac{1}{2}\boldsymbol{\mu}^T\boldsymbol{P}\boldsymbol{\mu} + \boldsymbol{\mu}^T\boldsymbol{t} \\ \text{subject to:} & \boldsymbol{\mu} \geq \boldsymbol{0} \end{array} \tag{2.17}$$

Note that by definition matrix $\boldsymbol{P}$ in (2.17) is positive definite, therefore the dual problem is also a convex QP problem, but with simpler constraints in comparison with the primal problem in (1.5). In addition, if the number of constraints involved in the primal problem is smaller than n, so is the size of the dual problem. ■