Definition 1.1 Point $\boldsymbol{x}^*$ is said to be *regular* for a given set of equality constraints $\{a_i(\boldsymbol{x}) = 0, i = 1,$ $2, \ldots, p\}$ if the p vectors $\{\nabla a_i(\boldsymbol{x}^*), i = 1, 2...., p\}$ are linearly independent.

1.3.1 Optimality Conditions in Simple Cases

Definition 1.2 Let $\boldsymbol{x}$ be a feasible point for problem (1.1). Vector $\boldsymbol{d} \in R^n$ is said to be a *feasible direction* at $\boldsymbol{x}$ if there exists $\hat{\alpha} > 0$ such that $\boldsymbol{x} + \alpha \boldsymbol{d}$ remains feasible for all $\alpha \in [0, \hat{\alpha}]$.

Theorem 1.1 If $f(\boldsymbol{x}) \in C^1$ and $\boldsymbol{x}^*$ is a local minimizer of problem (1.1), then

$$\nabla f(\boldsymbol{x}^*)^T \boldsymbol{d} \geq 0 \quad \text{for every feasible direction } \boldsymbol{d} \text{ at } \boldsymbol{x}^* \tag{1.11}$$

Proof We examine Taylor's expansion of $f(\boldsymbol{x})$ at $\boldsymbol{x} = \boldsymbol{x}^* + \alpha \boldsymbol{d}$:

$$f(\boldsymbol{x}) = f(\boldsymbol{x}^*) + \alpha \nabla f(\boldsymbol{x}^*)^T \boldsymbol{d} + o(\alpha \| \boldsymbol{d} \|_2)$$

Hence

$$\alpha \nabla f(\boldsymbol{x}^*)^T \boldsymbol{d} + o(\alpha \| \boldsymbol{d} \|_2) = f(\boldsymbol{x}) - f(\boldsymbol{x}^*) \geq 0$$

which implies (1.11). ■

Example 1.2 Let $\boldsymbol{x}^*$ be a local minimizer of the equality constrained problem

$$\begin{aligned} &\text{minimize} \quad f(\boldsymbol{x}) \\ &\text{subject to: } \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \end{aligned} \tag{1.12}$$

where $\boldsymbol{A} \in R^{p \times n}$ with $p < n$. The feasible region is defined by $\mathcal{R} = \{\boldsymbol{x}: \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}\}$. Obviously, $\boldsymbol{d}$ is a feasible direction at a feasible point $\boldsymbol{x}$ if and only if $\boldsymbol{A}\boldsymbol{d} = \boldsymbol{0}$, namely $\boldsymbol{d}$ belongs to the null space of $\boldsymbol{A}$. By applying Theorem 1.1, we obtain

$$\nabla f(\boldsymbol{x}^*)^T \boldsymbol{d} \geq 0 \quad \text{for } \boldsymbol{d} \in \mathcal{N}(\boldsymbol{A}) \tag{1.13}$$

where $\mathcal{N}(\boldsymbol{A})$ denotes the null space of $\boldsymbol{A}$. Since $\boldsymbol{d} \in \mathcal{N}(\boldsymbol{A})$ implies $-\boldsymbol{d} \in \mathcal{N}(\boldsymbol{A})$, (1.13) implies that $\nabla f(\boldsymbol{x}^*)^T \boldsymbol{d} = 0$ for $\boldsymbol{d} \in \mathcal{N}(\boldsymbol{A})$. From linear algebra, we conclude that

$$\nabla f(\boldsymbol{x}^*) \in \mathcal{N}^{\perp}(\boldsymbol{A}) = \mathcal{R}(\boldsymbol{A}^T) \tag{1.14}$$

This is to say that gradient $\nabla f(\boldsymbol{x}^*)$ is in the range of matrix $\boldsymbol{A}^T$. Therefore, there exists $\boldsymbol{\lambda} \in R^p$ (known as *Lagrange multiplier* for equality constraints) such that $\nabla f(\boldsymbol{x}^*) = -\boldsymbol{A}^T \boldsymbol{\lambda}$, namely,

$$\nabla f(\boldsymbol{x}^*) + \boldsymbol{A}^T \boldsymbol{\lambda} = \boldsymbol{0} \tag{1.15}$$

■

1.3.2 Equality Constraints

We now consider the constrained problem

$$\begin{aligned} &\text{minimize} \quad f(\boldsymbol{x}) \\ &\text{subject to: } a_i(\boldsymbol{x}) = 0 \ \text{ for } \ 1 \le i \le p \end{aligned} \tag{1.16}$$

Let $\boldsymbol{x}^*$ be a local minimizer of (1.16) and $\boldsymbol{s}$ be a feasible vector at $\boldsymbol{x}^*$. Thus we have $a_i(\boldsymbol{x}^*) = 0$ and $a_i(\boldsymbol{x}^* + \boldsymbol{s}) = 0$. The Taylor expansion of $a_i(\boldsymbol{x}^* + \boldsymbol{s})$ gives

$$0 = a_i(\boldsymbol{x}^* + \boldsymbol{s}) = a_i(\boldsymbol{x}^*) + \boldsymbol{s}^T \nabla a_i(\boldsymbol{x}^*) + o(\|\boldsymbol{s}\|) = \boldsymbol{s}^T \nabla a_i(\boldsymbol{x}^*) + o(\|\boldsymbol{s}\|)$$

which implies that

$$\boldsymbol{s}^T \nabla a_i(\boldsymbol{x}^*) = 0 \ \text{ for } \ 1 \le i \le p \tag{1.17}$$

In other words, $\boldsymbol{s}$ is feasible if and only if it is orthogonal to the gradients of the constraint functions. Now we project the gradient $\nabla f(\boldsymbol{x}^*)$ onto the space spanned by $\{\nabla a_1(\boldsymbol{x}^*), \nabla a_2(\boldsymbol{x}^*)$, $\ldots, \nabla a_p(\boldsymbol{x}^*)\}$. Let the projection be given by $-\sum_{i=1}^{p} \lambda_i^* \nabla a_i(\boldsymbol{x}^*)$, then $\nabla f(\boldsymbol{x}^*)$ can be expressed as

$$\nabla f(\boldsymbol{x}^*) = -\sum_{i=1}^{p} \lambda_i^* \nabla a_i(\boldsymbol{x}^*) + \boldsymbol{d} \tag{1.18}$$

where $\boldsymbol{d}$ is orthogonal to $\nabla a_i(\boldsymbol{x}^*)$ for i = 1, 2, …, p. Hence $\boldsymbol{s} = -\boldsymbol{d}$ is a feasible direction. From (1.17) and (1.18), it follows that $\boldsymbol{s}^T \nabla f(\boldsymbol{x}^*) = -\|\boldsymbol{d}\|^2$ which means that for $\boldsymbol{x}^*$ to be a local minimizer $\boldsymbol{d}$ must be zero (otherwise $\boldsymbol{s} = -\boldsymbol{d}$ would be a nontrivial feasible direction from $\boldsymbol{x}^*$ along which $f(\boldsymbol{x})$ would decrease, a contradiction with the assumption of $\boldsymbol{x}^*$ being a minimizer). This leads (1.18) to

$$\nabla f(\boldsymbol{x}^*) = -\sum_{i=1}^{p} \lambda_i^* \nabla a_i(\boldsymbol{x}^*) \tag{1.19}$$

The concept of Lagrange multiplier may also be understood from a different perspective. To see this we introduce the *Lagrangian* of problem (1.16) as

$$L(\boldsymbol{x}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) + \sum_{i=1}^{p} \lambda_i a_i(\boldsymbol{x}) \tag{1.20}$$

where $\boldsymbol{\lambda} = \begin{bmatrix} \lambda_1 & \lambda_2 & \cdots & \lambda_p \end{bmatrix}^T$. In terms of the Lagrangian, (1.19) becomes

$$\nabla_x L(\boldsymbol{x},\boldsymbol{\lambda}) = \mathbf{0} \;\text{ for }\; \{\boldsymbol{x},\boldsymbol{\lambda}\} = \{\boldsymbol{x}^*,\boldsymbol{\lambda}^*\} \tag{1.21}$$

Eq. (1.21) offers n equations while the equality constraints in (1.16) contain p equations. Note that the number of equations available, $n + p$, matches the number of unknowns in $\boldsymbol{x}$ and $\boldsymbol{\lambda}$. Moreover, if we introduce the gradient operator ∇ as

$$\nabla = \begin{bmatrix} \nabla_x \\ \nabla_\lambda \end{bmatrix}$$

then these equations can be expressed as

$$\nabla L(\boldsymbol{x},\boldsymbol{\lambda}) = \mathbf{0} \;\text{ for }\; \{\boldsymbol{x},\boldsymbol{\lambda}\} = \{\boldsymbol{x}^*,\boldsymbol{\lambda}^*\} \tag{1.22}$$

In summary, the Lagrangian incorporates the constraints into a modified objective function in a way such that constrained minimizer $\boldsymbol{x}^*$ is connected to an unconstrained minimizer $\{\boldsymbol{x}^*,\boldsymbol{\lambda}^*\}$ for the augmented objective function $L(\boldsymbol{x},\boldsymbol{\lambda})$ where the augmentation is achieved with the p Lagrange multipliers.

Example 1.3 Solve the convex quadratic problem

$$\begin{aligned} &\text{minimize} \quad f(\boldsymbol{x}) = \tfrac{1}{2}\boldsymbol{x}^T\boldsymbol{H}\boldsymbol{x} + \boldsymbol{x}^T\boldsymbol{p} \\ &\text{subject to:} \;\; \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \end{aligned}$$

where $\boldsymbol{H}$ is positive definite and $\boldsymbol{A} \in R^{p\times n}$ with $p \le n$ has full row rank.

Solution The Lagrangian of the problem is given by $L(\boldsymbol{x},\boldsymbol{\lambda}) = \tfrac{1}{2}\boldsymbol{x}^T\boldsymbol{H}\boldsymbol{x} + \boldsymbol{x}^T\boldsymbol{p} + \boldsymbol{\lambda}^T(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b})$. Applying (1.22), we obtain

$$\nabla L(\boldsymbol{x},\boldsymbol{\lambda}) = \begin{bmatrix} \boldsymbol{H}\boldsymbol{x} + \boldsymbol{p} + \boldsymbol{A}^T\boldsymbol{\lambda} \\ \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b} \end{bmatrix} = \begin{bmatrix} \boldsymbol{H} & \boldsymbol{A}^T \\ \boldsymbol{A} & \mathbf{0} \end{bmatrix}\begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{\lambda} \end{bmatrix} + \begin{bmatrix} \boldsymbol{p} \\ -\boldsymbol{b} \end{bmatrix} = \mathbf{0}$$

Because $\boldsymbol{H} \succ \mathbf{0}$ and $\boldsymbol{A}$ has full row rank, one can show that matrix

$$\begin{bmatrix} \boldsymbol{H} & \boldsymbol{A}^T \\ \boldsymbol{A} & \mathbf{0} \end{bmatrix}$$

is nonsingular (try to prove this fact as an exercise!). Therefore,

$$\begin{bmatrix} \boldsymbol{x}^* \\ \boldsymbol{\lambda}^* \end{bmatrix} = \begin{bmatrix} \boldsymbol{H} & \boldsymbol{A}^T \\ \boldsymbol{A} & \mathbf{0} \end{bmatrix}^{-1}\begin{bmatrix} -\boldsymbol{p} \\ \boldsymbol{b} \end{bmatrix}$$

It follows that

$$\begin{aligned} \boldsymbol{x}^* &= -\boldsymbol{H}^{-1}(\boldsymbol{A}^T\boldsymbol{\lambda}^* + \boldsymbol{p}) \\ \boldsymbol{\lambda}^* &= -(\boldsymbol{A}\boldsymbol{H}^{-1}\boldsymbol{A}^T)^{-1}(\boldsymbol{A}\boldsymbol{H}^{-1}\boldsymbol{p} + \boldsymbol{b}) \end{aligned} \tag{1.23}$$

1.3.3 Inequality Constraints

First, a concept concerning inequality constraints. Suppose $\boldsymbol{x}^*$ is a minimizer of problem (1.1) and let $I(\boldsymbol{x}^*)$ be an index set $I(\boldsymbol{x}^*) = \{j_1, j_2, \ldots, j_K\} \subseteq \{1, 2, \ldots, q\}$ for *active constraints* at $\boldsymbol{x}^*$, i.e., $c_j(\boldsymbol{x}^*) = 0$ for $j = j_1, j_2, \ldots, j_K$. As an example, Figure 1.2 shows a feasible region defined by R = $\{\boldsymbol{x} : c_j(\boldsymbol{x}) \le 0, \text{for } j = 1, 2, 3\}$ and a point $\bar{\boldsymbol{x}}$ at which constraint $c_3(\boldsymbol{x}) \le 0$ is active because $c_3(\bar{\boldsymbol{x}}) = 0$. We see that a point lies on the boundary of the feasible region if at the point some constraints become active.

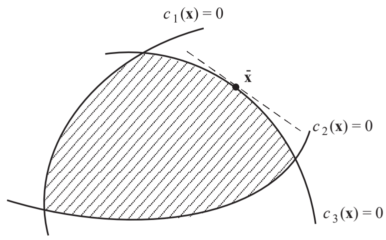

Figure 1.2

Now let us consider the general constrained problem (1.1):

$$\text{minimize} \quad f(\boldsymbol{x}) \tag{1.1a}$$

$$\text{subject to:} \quad a_i(\boldsymbol{x}) = 0 \quad \text{for } i = 1, 2, \ldots, p \tag{1.1b}$$

$$c_j(\boldsymbol{x}) \le 0 \quad \text{for } j = 1, 2, \ldots, q \tag{1.1c}$$

and let $\boldsymbol{x}^*$ be a local minimizer of (1.1) at which there are K active inequality constraints, namely,

$$c_j(\boldsymbol{x}^*) = 0 \quad \text{for} \quad j \in I(\boldsymbol{x}^*) = \{j_1, j_2, \ldots, j_K\} \tag{1.24}$$

The K active constraints at $\boldsymbol{x}^*$ act like K additional equality constraints. Consequently, the equation in (1.19) need to be modified to

$$\nabla f(\boldsymbol{x}^*) = -\sum_{i=1}^{p} \lambda_i^* \nabla a_i(\boldsymbol{x}^*) - \sum_{k=1}^{K} \mu_{j_k}^* \nabla c_{j_k}(\boldsymbol{x}^*) \tag{1.25}$$

where $\mu_{j_k}^*$ are called Lagrange multipliers of the active inequality constraints.

An important property of the Lagrange multipliers associated with inequality constraints is that they are nonnegative, i.e.

$$\mu_{j_k}^* \ge 0 \quad \text{for} \quad 1 \le k \le K \tag{1.26}$$

This property is illustrated in Fig. 1.3 for the case where the minimizer is in the interior of the feasible region (case (a)) and the case where the minimizer is on the boundary of the feasible region (case (b)).

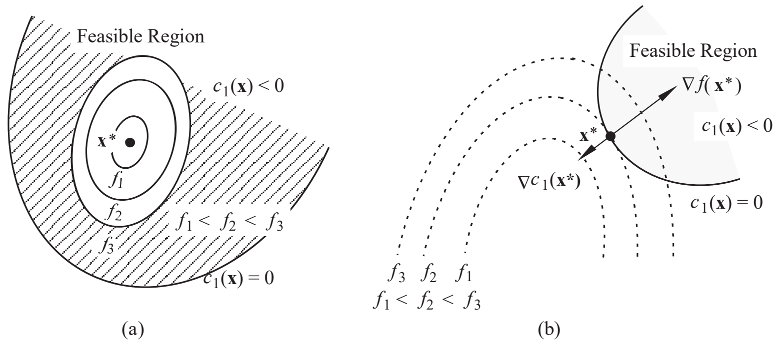

Figure 1.3

1.3.4 Karush-Kuhn-Tucker (KKT) Conditions

The KKT conditions are first-order necessary conditions that a local minimizer of problem (1.1) must satisfy.

Theorem 1.2 ***Karush-Kuhn-Tucker conditions*** *If* $\boldsymbol{x}^*$ *is a local minimizer of problem* (1.1) *and is regular for the constraints that are active at* $\boldsymbol{x}^*$*, then*

(*a*) $$a_i(\boldsymbol{x}^*) = 0 \text{ for } i = 1, 2, \ldots, p$$

(*b*) $$c_j(\boldsymbol{x}^*) \le 0 \text{ for } i = 1, 2, \ldots, q$$

(*c*) *There exist Lagrange multipliers* λ_i^* for $1 \le i \le p$ and μ_i^* for $1 \le i \le q$ *such that*

$$\nabla f(\boldsymbol{x}^*) + \sum_{i=1}^{p} \lambda_i^* \nabla a_i(\boldsymbol{x}^*) + \sum_{j=1}^{q} \mu_j^* \nabla c_j(\boldsymbol{x}^*) = \mathbf{0} \tag{1.27}$$

(*d*) *Complementarity conditions*

$$\lambda_i^* a_i(\boldsymbol{x}^*) = 0 \text{ for } 1 \le i \le p \tag{1.28a}$$

$$\mu_j^* c_j(\boldsymbol{x}^*) = 0 \text{ for } 1 \le j \le q \tag{1.28b}$$

(*e*) $$\mu_j^* \ge 0 \text{ for } 1 \le j \le q \tag{1.29}$$

Definition 1.3 The Lagrangian for the general constrained problem (1.1) is defined by

$$L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\boldsymbol{x}) + \sum_{i=1}^{p} \lambda_i a_i(\boldsymbol{x}) + \sum_{j=1}^{q} \mu_j c_j(\boldsymbol{x}) \tag{1.30}$$

Note that the condition in (1.27) can be expressed in terms of Lagrangian as

$$\nabla_x L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \mathbf{0}$$

which is a set of n equations. These in combination with the p equations in condition (*a*) and q

equations in (1.28b) form a system of ($n + p + q$) equations for the same number of "unknowns" in $\boldsymbol{x}, \boldsymbol{\lambda}$ and $\boldsymbol{\mu}$. However, these equations in many cases are nonlinear and solving nonlinear system of equations is often no easier than solving the optimization problem itself. As a matter of fact, nonlinear equations are often solved by optimization techniques such as nonlinear least squares these days.

Example 1.4 Solve the constrained minimization problem

$$\begin{aligned}&\text{minimize} && f(\boldsymbol{x}) = x_1^2 + x_2^2 - 14x_1 - 6x_2 \\ &\text{subject to:} && c_1(\boldsymbol{x}) = -2 + x_1 + x_2 \le 0 \\ & && c_2(\boldsymbol{x}) = -3 + x_1 + 2x_2 \le 0\end{aligned}$$

by applying the KKT conditions.

Solution The KKT conditions imply that

$$\begin{aligned}&2x_1 - 14 + \mu_1 + \mu_2 = 0 \\ &2x_2 - 6 + \mu_1 + 2\mu_2 = 0 \\ &\mu_1(-2 + x_1 + x_2) = 0 \\ &\mu_2(-3 + x_1 + 2x_2) = 0 \\ &\mu_1 \ge 0 \\ &\mu_2 \ge 0\end{aligned}$$

One way to find the solution in this simple case is to consider all possible cases with regard to active constraints and verify the nonnegativity of the μ_j's obtained.

Case 1: *No active constraints*. If there are no active constraints, we have $\mu_1^* = \mu_2^* = 0$, which leads to

$$\boldsymbol{x}^* = \begin{bmatrix} 7 \\ 3 \end{bmatrix}$$

Obviously, this $\boldsymbol{x}^*$ violates both constraints and it is not a solution.

Case 2: *One constraint is active.* If only the first constraint is active, then we have $\mu_2^* = 0$, and

$$\begin{aligned}&2x_1 - 14 + \mu_1 = 0 \\ &2x_2 - 6 + \mu_1 = 0 \\ &-2 + x_1 + x_2 = 0\end{aligned}$$

Solving this system of equations, we obtain

$$\boldsymbol{x}^* = \begin{bmatrix} 3 \\ -1 \end{bmatrix} \quad \text{and} \quad \mu_1^* = 8$$

Since $\boldsymbol{x}^*$ also satisfies the second constraint, $\boldsymbol{x}^* = [3 \quad -1]^T$ and $\boldsymbol{\mu}^* = [8 \quad 0]^T$ satisfy the KKT conditions.

If only the second constraint is active, then $\mu_1^* = 0$ and the KKT conditions become

$$\begin{aligned} 2x_1 - 14 + \mu_2 &= 0 \\ 2x_2 - 6 + 2\mu_2 &= 0 \\ -3 + x_1 + 2x_2 &= 0 \end{aligned}$$

The solution of this system of equations is given by

$$\boldsymbol{x}^* = \begin{bmatrix} 5 \\ -1 \end{bmatrix} \text{ and } \mu_2^* = 4$$

As $\boldsymbol{x}^*$ violates the first constraint, the above $\boldsymbol{x}^*$ and $\boldsymbol{\mu}^*$ do not satisfy the KKT conditions.

Case 3: *Both constraints are active*. If both constraints are active, we have

$$\begin{aligned} 2x_1 - 14 + \mu_1 + \mu_2 &= 0 \\ 2x_2 - 6 + \mu_1 + 2\mu_2 &= 0 \\ -2 + x_1 + x_2 &= 0 \\ -3 + x_1 + 2x_2 &= 0 \end{aligned}$$

The solution to this system of equations is given by

$$\boldsymbol{x}^* = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \text{ and } \boldsymbol{\mu}^* = \begin{bmatrix} 20 \\ -8 \end{bmatrix}$$

Since $\mu_2^* < 0$, this is not a solution of the optimization problem.

Therefore, the only candidate for a minimizer of the problem is given by

$$\boldsymbol{x}^* = \begin{bmatrix} 3 \\ -1 \end{bmatrix} \text{ and } \boldsymbol{\mu}^* = \begin{bmatrix} 8 \\ 0 \end{bmatrix}$$

As can be observed in Fig. 1.4, the above point is actually the global minimizer.

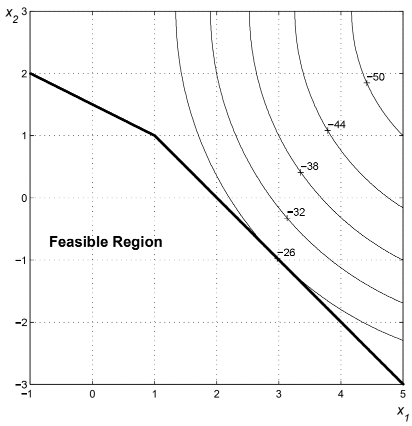

Figure 1.4