5.CONSTRAINED OPTIMIZATION

the constrained optimization problem

$$\text{minimize } f(\mathbf{x})$$

$$\begin{aligned} \text{subject to:} \quad & a_i(\mathbf{x}) = 0 \qquad \text{for } i = 1, 2, \ldots, p \\ & c_j(\mathbf{x}) \geq 0 \qquad \text{for } j = 1, 2, \ldots, q \end{aligned}$$

$$\mathcal{R} = \{\mathbf{x} : a_i(\mathbf{x}) = 0 \text{ for } i = 1, 2, \ldots, p, \ c_j(\mathbf{x}) \geq 0 \text{ for } j = 1, 2, \ldots, q\}$$

Equality constraints

$$\mathbf{a}(\mathbf{x}) = [a_1(\mathbf{x})\ a_2(\mathbf{x})\ \cdots\ a_p(\mathbf{x})]^T$$

we have

$$\mathbf{a}(\mathbf{x}) = \mathbf{0}$$

Definition 10.1 A point $\mathbf{x}$ is called a *regular point* of the constraints in Eq. (10.2) if $\mathbf{x}$ satisfies Eq. (10.2) and column vectors $\nabla a_1(\mathbf{x}),\ \nabla a_2(\mathbf{x}),\ \ldots,\ \nabla a_p(\mathbf{x})$ are linearly independent.

$\mathbf{J}_e = [\nabla a_1(\mathbf{x})\ \ \nabla a_2(\mathbf{x})\ \ \cdots\ \ \nabla a_p(\mathbf{x})]^T$ has full row rank.

雅克比矩阵是行满秩的,通常p<n

linear equality constraints which can be expressed as

$$\mathbf{A}\mathbf{x} = \mathbf{b}$$

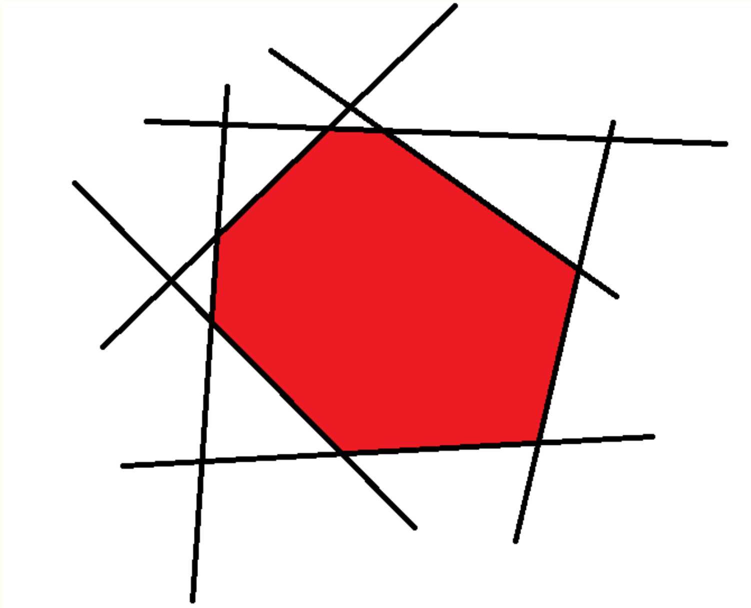
where $\mathbf{A} \in R^{p \times n}$ is numerically equal to the Jacobian, i.e., $\mathbf{A} = \mathbf{J}_e$

When $\text{rank}(\mathbf{A}) = p' < p$ we can apply singular-value decomposition (SVD) method to get

the reduced set of equality constraints (p269-270)

Inequality constraints

$$
\begin{gathered}
c_1(\mathbf{x}) \geq 0 \\
c_2(\mathbf{x}) \geq 0 \\
\vdots \\
c_q(\mathbf{x}) \geq 0
\end{gathered}
$$



不等约束q可远大于n

$c_i(\mathbf{x}) = 0$, which are called *active constraints*,
$c_i(\mathbf{x}) > 0$, which are called *inactive constraints*

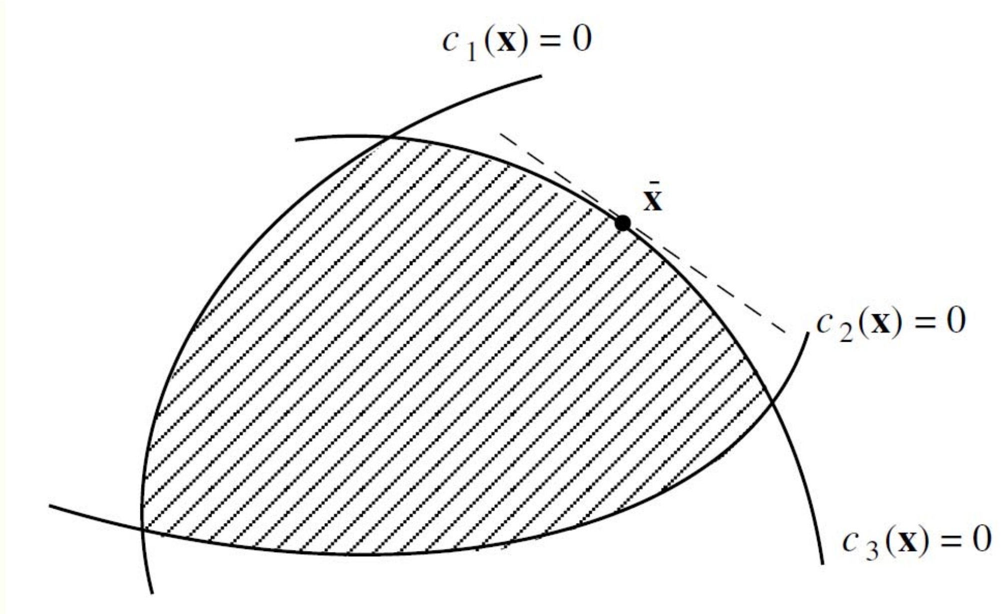


fig 5.1 c3(x) is active, c1,c2 are inactive

the local properties of x will not be affected by the inactive constraints.In other words, active constraints restrict the feasible region of the neighborhoods of x.

Another approach to deal with inequality constraints is to convert them into equality constraints.

$$\text{minimize } f(\mathbf{x}) \qquad \mathbf{x} \in R^n$$

$$\text{subject to:} \quad c_i(\mathbf{x}) \geq 0 \qquad \text{for } i = 1, 2, \ldots, q$$

$$\begin{aligned}
\hat{c}_1 &= c_1(\mathbf{x}) - y_1 = 0 \\
\hat{c}_2 &= c_2(\mathbf{x}) - y_2 = 0 \\
&\vdots \\
\hat{c}_q &= c_q(\mathbf{x}) - y_q = 0 \\
y_i &\geq 0 \qquad \text{for } 1 \leq i \leq q
\end{aligned}$$

where $y_1, y_2, \ldots, y_q$ are called *slack variables*

Another approach to deal with inequality constraints is to convert them into equality constraints.

$$y_i = \hat{y}_i^2 \qquad \text{for } 1 \le i \le q$$

If we let

$$\hat{\mathbf{x}} = [x_1 \ \cdots \ x_n \ \hat{y}_1 \ \cdots \ \hat{y}_q]^T$$

then the problem in Eq. (10.14) can be formulated as

$$\text{minimize } f(\hat{\mathbf{x}}) \qquad \hat{\mathbf{x}} \in E^{n+q}$$

$$\text{subject to: } \hat{c}_i(\hat{\mathbf{x}}) = 0 \qquad \text{for } i = 1, 2, \ldots, q$$

The idea of introducing slack variables to reformulate an optimization problem has been used successfully in the past.

Classification of Constrained Optimization Problems

Constrained optimization problems can be classified according to the nature of the objective function and the constraints.

1. LP
2. QP
3. QCQP
4. Convex Programming
5. Semidefinite Programming
6. SOCP
7. Non-Linear Programming

Linear programming

The standard form of a linear programming (LP) problem can be stated as

$$\begin{aligned} \text{minimize } & f(\mathbf{x}) = \mathbf{c}^T\mathbf{x} \\ \text{subject to: } & \mathbf{A}\mathbf{x} = \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0} \end{aligned}$$

LP problems may also be encountered in the nonstandard form

$$\begin{aligned} \text{minimize } & \mathbf{c}^T\mathbf{x} \\ \text{subject to: } & \mathbf{A}\mathbf{x} \geq \mathbf{b} \end{aligned}$$

Quadratic programming

the objective function is quadratic and the constraints are linear

$$\text{minimize } f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{H}\mathbf{x} + \mathbf{x}^T\mathbf{p} + c$$
$$\text{subject to: } \quad \mathbf{A}\mathbf{x} = \mathbf{b}$$
$$\mathbf{C}\mathbf{x} \geq \mathbf{d}$$

Lagrange Multipliers

Lagrange multipliers play a crucial role in the study of constrained optimization，they provide a natural connection between constrained and corresponding unconstrained optimization problems;each individual Lagrange multiplier can be interpreted as the rate of change in the objective function with respect to changes in the associated constraint function。

if x□ is a local minimizer of a constrained minimization problem, then in addition to x□ being a feasible point, the gradient of the objective function at x□ has to be a linear combination of the gradients of the constraint functions, and the Lagrange multipliers are the coefficients in that linear combination。

Moreover, the Lagrange multipliers associated with inequality constraints have to be **nonneg-ative** and the multipliers associated with **inactive inequality constraints have to be zero**. Collectively, these conditions are known as the **Karush-Kuhn-Tucker conditions** (KKT)

Lagrange Multipliers

consider the minimization of the objective function $f(x_1,\ x_2,\ x_3,\ x_4)$

subject to :

$$a_1(x_1,\ x_2,\ x_3,\ x_4) = 0$$
$$a_2(x_1,\ x_2,\ x_3,\ x_4) = 0$$

these constraints can be expressed as

$$x_3 = h_1(x_1,\ x_2)$$
$$x_4 = h_2(x_1,\ x_2)$$

the original constrained optimization problem changes to unconstrained optimization problem

$$\text{minimize } f[x_1,\ x_2,\ h_1(x_1,\ x_2),\ h_2(x_1,\ x_2)]$$

therefore, follows that at x□ we have

$$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \\ \frac{\partial f}{\partial x_2} \end{bmatrix} = \mathbf{0}$$

$$\frac{\partial f}{\partial x_1} + \frac{\partial f}{\partial x_3}\frac{\partial h_1}{\partial x_1} + \frac{\partial f}{\partial x_4}\frac{\partial h_2}{\partial x_1} = 0$$
$$\frac{\partial f}{\partial x_2} + \frac{\partial f}{\partial x_3}\frac{\partial h_1}{\partial x_2} + \frac{\partial f}{\partial x_4}\frac{\partial h_2}{\partial x_2} = 0$$

from the equality constraints

$$\frac{\partial a_1}{\partial x_1} + \frac{\partial a_1}{\partial x_3}\frac{\partial h_1}{\partial x_1} + \frac{\partial a_1}{\partial x_4}\frac{\partial h_2}{\partial x_1} = 0$$
$$\frac{\partial a_1}{\partial x_2} + \frac{\partial a_1}{\partial x_3}\frac{\partial h_1}{\partial x_2} + \frac{\partial a_1}{\partial x_4}\frac{\partial h_2}{\partial x_2} = 0$$
$$\frac{\partial a_2}{\partial x_1} + \frac{\partial a_2}{\partial x_3}\frac{\partial h_1}{\partial x_1} + \frac{\partial a_2}{\partial x_4}\frac{\partial h_2}{\partial x_1} = 0$$
$$\frac{\partial a_2}{\partial x_2} + \frac{\partial a_2}{\partial x_3}\frac{\partial h_1}{\partial x_2} + \frac{\partial a_2}{\partial x_4}\frac{\partial h_2}{\partial x_2} = 0$$

The above six equations can now be expressed as

$$\begin{bmatrix} \nabla^T f(\mathbf{x}) \\ \nabla^T a_1(\mathbf{x}) \\ \nabla^T a_2(\mathbf{x}) \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ \frac{\partial h_1}{\partial x_1} & \frac{\partial h_1}{\partial x_2} \\ \frac{\partial h_2}{\partial x_1} & \frac{\partial h_2}{\partial x_2} \end{bmatrix} = \mathbf{0}$$

$$\alpha \nabla f(\mathbf{x}^*) + \beta \nabla a_1(\mathbf{x}^*) + \gamma \nabla a_2(\mathbf{x}^*) = \mathbf{0}$$

$$\nabla f(\mathbf{x}^*) - \lambda_1^* \nabla a_1(\mathbf{x}^*) - \lambda_2^* \nabla a_2(\mathbf{x}^*) = \mathbf{0}$$

$$\nabla f(\mathbf{x}^*) = \lambda_1^* \nabla a_1(\mathbf{x}^*) + \lambda_2^* \nabla a_2(\mathbf{x}^*)$$

at a local minimizer of the constrained optimization problem, the gradient of the objective function is a linear combination of the gradients of the constraints,Constants λ_1^* and λ_2^* are called the Lagrange multipliers of the the equality constraints,

Equality constraints

We now consider the constrained optimization problem

$$\text{minimize } f(\mathbf{x})$$

$$\text{subject to:} \quad a_i(\mathbf{x}) = 0 \qquad \text{for } i = 1, 2, \ldots, p$$

Let x□ be a local minimizer of the problem, By using the Taylor series of constraint function ai(x) at x□ , we can write

$$\begin{aligned} a_i(\mathbf{x}^* + \mathbf{s}) &= a_i(\mathbf{x}^*) + \mathbf{s}^T \nabla a_i(\mathbf{x}^*) + o(||\mathbf{s}||) \\ &= \mathbf{s}^T \nabla a_i(\mathbf{x}^*) + o(||\mathbf{s}||) \end{aligned}$$

If s is a feasible vector at $\mathbf{x}^*$, then $a_i(\mathbf{x}^* + \mathbf{s}) = 0$

$$\mathbf{s}^T \nabla a_i(\mathbf{x}^*) = 0 \qquad \text{for } i = 1, 2, \ldots, p$$

In other words, $\mathbf{s}$ is feasible if it is orthogonal to the gradients of the constraint functions. Now we project the gradient $\nabla f(\mathbf{x}^*)$ orthogonally onto the space spanned by $\{\nabla a_1(\mathbf{x}^*),\ \nabla a_2(\mathbf{x}^*),\ \ldots,\ \nabla a_p(\mathbf{x}^*)\}$. If we denote the projection as

$$\sum_{i=1}^{p} \lambda_i^* \nabla a_i(\mathbf{x}^*)$$

then $\nabla f(\mathbf{x}^*)$ can be expressed as

$$\nabla f(\mathbf{x}^*) = \sum_{i=1}^{p} \lambda_i^* \nabla a_i(\mathbf{x}^*) + \mathbf{d} \tag{10.60}$$

where $\mathbf{d}$ is orthogonal to $\nabla a_i(\mathbf{x}^*)$ for $i = 1,\ 2,\ \ldots,\ p$.

In what follows, we show that if $\mathbf{x}^*$ is a local minimizer then $\mathbf{d}$ must be zero. The proof is accomplished by contradiction. Assume that $\mathbf{d} \neq \mathbf{0}$ and let $\mathbf{s} = -\mathbf{d}$. Since $\mathbf{s}$ is orthogonal to $\nabla a_i(\mathbf{x}^*)$ by virtue of Eq. (10.59), $\mathbf{s}$ is feasible at $\mathbf{x}^*$. Now we use Eq. (10.60) to obtain

$$\mathbf{s}^T \nabla f(\mathbf{x}^*) = \mathbf{s}^T \left(\sum_{i=1}^{p} \lambda_i^* \nabla a_i(\mathbf{x}^*) + \mathbf{d} \right) = -||\mathbf{d}||^2 < 0$$

This means that $\mathbf{s}$ is a descent direction at $\mathbf{x}^*$ which contradicts the fact that $\mathbf{x}^*$ is a minimizer. Therefore, $\mathbf{d} = \mathbf{0}$ and Eq. (10.60) becomes

$$\nabla f(\mathbf{x}^*) = \sum_{i=1}^{p} \lambda_i^* \nabla a_i(\mathbf{x}^*) \tag{10.61}$$

In effect, for an arbitrary constrained problem with equality constraints, the gradient of the objective function at a local minimizer *is equal to the linear combination of the gradients of the equality constraint functions with the Lagrange multipliers as the coefficients*.

等式约束的拉格朗日函数：

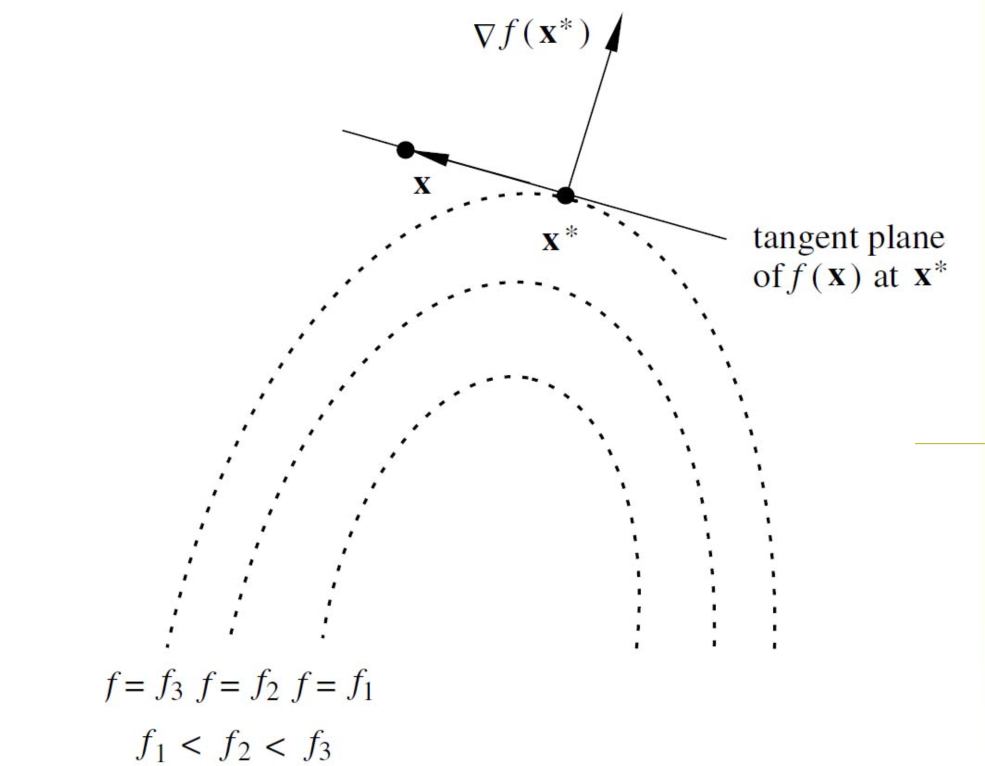
$$L(\mathbf{x},\ \boldsymbol{\lambda}) = f(\mathbf{x}) - \sum_{i=1}^{p} \lambda_i a_i(\mathbf{x})$$

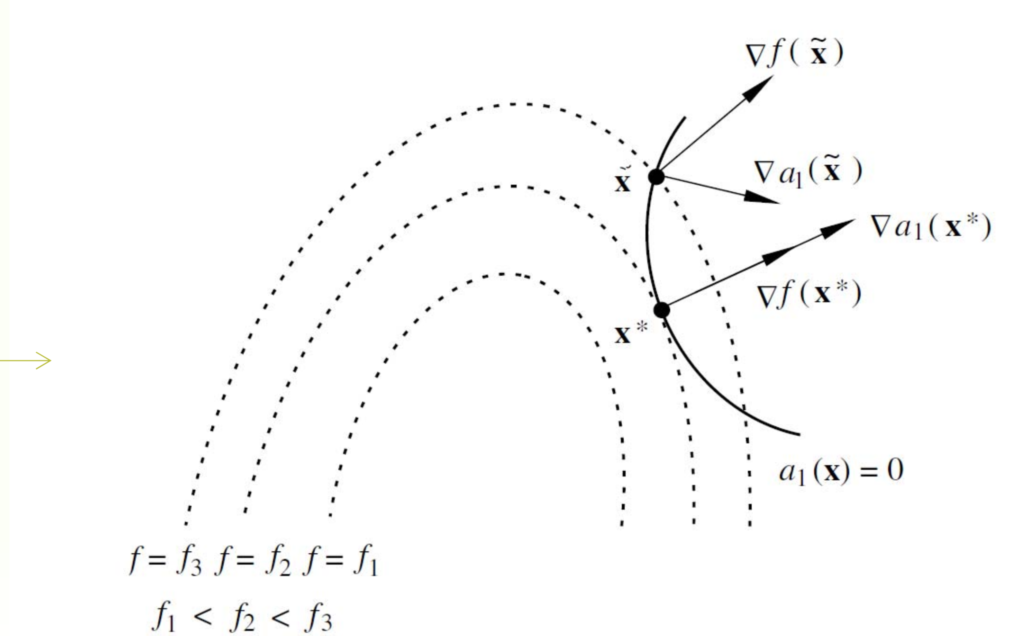
$$\nabla_x L(\mathbf{x},\ \boldsymbol{\lambda}) = \mathbf{0} \qquad \text{for } \{\mathbf{x}, \boldsymbol{\lambda}\} = \{\mathbf{x}^*,\ \boldsymbol{\lambda}^*\}$$

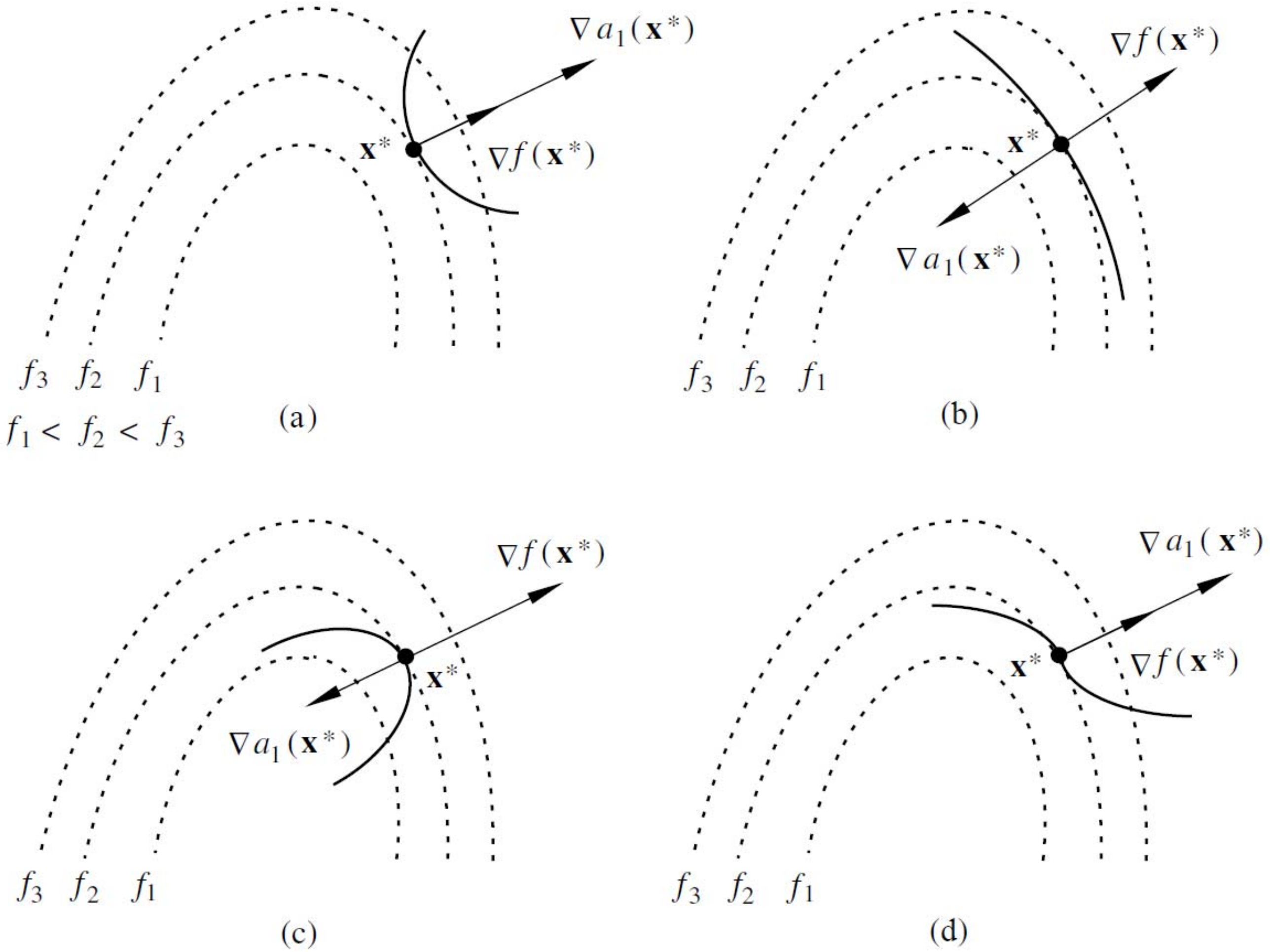
and

$$\nabla_\lambda L(\mathbf{x},\ \boldsymbol{\lambda}) = \mathbf{0} \qquad \text{for } \{\mathbf{x}, \boldsymbol{\lambda}\} = \{\mathbf{x}^*,\ \boldsymbol{\lambda}^*\}$$

几何解释


$\nabla f(\mathbf{x}^*)$
$\mathbf{x}$
$\mathbf{x}^*$
tangent plane
of $f(\mathbf{x})$ at $\mathbf{x}^*$
$f = f_3$ $f = f_2$ $f = f_1$
$f_1 < f_2 < f_3$


$\nabla f(\tilde{\mathbf{x}})$
$\nabla a_1(\tilde{\mathbf{x}})$
$\nabla a_1(\mathbf{x}^*)$
$\nabla f(\mathbf{x}^*)$
$\tilde{\mathbf{x}}$
$\mathbf{x}^*$
$a_1(\mathbf{x}) = 0$
$f = f_3$ $f = f_2$ $f = f_1$
$f_1 < f_2 < f_3$


$\nabla a_1(\mathbf{x}^*)$
$\mathbf{x}^*$
$\nabla f(\mathbf{x}^*)$
f_3 f_2 f_1
$f_1 < f_2 < f_3$
(a)
$\nabla f(\mathbf{x}^*)$
$\mathbf{x}^*$
$\nabla a_1(\mathbf{x}^*)$
f_3 f_2 f_1
(b)
$\nabla f(\mathbf{x}^*)$
$\mathbf{x}^*$
$\nabla a_1(\mathbf{x}^*)$
f_3 f_2 f_1
(c)
$\nabla a_1(\mathbf{x}^*)$
$\mathbf{x}^*$
$\nabla f(\mathbf{x}^*)$
f_3 f_2 f_1
(d)

First-Order Necessary Conditions

- Equality constraints
- Inequality constraints

Equality constraints

Theorem 10.1 ***First-order necessary conditions for a minimum, equality constraints*** *If* $\mathbf{x}^*$ *is a constrained local minimizer of the problem in Eq. (10.57) and is a regular point of the constraints in Eq. (10.57b), then*

(a) $a_i(\mathbf{x}^*) = 0 \ \ for \ \ i = 1, 2, \ldots, p,$ *and* (10.73)

(b) *there exist Lagrange multipliers* $\lambda_i^* \ \ for \ \ i = 1, 2, \ldots, p$ *such that*

$$\nabla f(\mathbf{x}^*) = \sum_{i=1}^{p} \lambda_i^* \nabla a_i(\mathbf{x}^*) \tag{10.74}$$

也可以写矩阵形式

$$\mathbf{g}(\mathbf{x}^*) - \mathbf{J}_e^T(\mathbf{x}^*)\boldsymbol{\lambda}^* = \mathbf{0} \longrightarrow \boldsymbol{\lambda}^* = [\mathbf{J}_e^T(\mathbf{x}^*)]^+\mathbf{g}^*$$

结合等式约束

$$\mathbf{J}_e(\mathbf{x}) = [\nabla a_1(\mathbf{x})\ \nabla a_2(\mathbf{x})\ \cdots\ \nabla a_p(\mathbf{x})]^T$$

$$\begin{bmatrix} \mathbf{g}(\mathbf{x}^*) - \mathbf{J}_e^T(\mathbf{x}^*)\boldsymbol{\lambda}^* \\ \mathbf{a}(\mathbf{x}^*) \end{bmatrix} = \mathbf{0}$$

n+p个方程，求解n+p个变量

Example 10.9 Solve the problem

$$\text{minimize } f(\mathbf{x}) = \tfrac{1}{2}\mathbf{x}^T\mathbf{H}\mathbf{x} + \mathbf{x}^T\mathbf{p}$$

$$\text{subject to:} \quad \mathbf{A}\mathbf{x} = \mathbf{b}$$

where $\mathbf{H} \succ \mathbf{0}$ and $\mathbf{A} \in R^{p\times n}$ has full row rank.

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

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

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

例如：

$$\text{minimize } f(\mathbf{x}) = x_1^2 + x_2^2 + \tfrac{1}{4}x_3^2$$

$$\begin{aligned} \text{subject to:} \quad a_1(\mathbf{x}) &= -x_1 + x_3 - 1 = 0 \\ a_2(\mathbf{x}) &= x_1^2 + x_2^2 - 2x_1 = 0 \end{aligned}$$

Solution We have

$$\mathbf{g}(\mathbf{x}) = \begin{bmatrix} 2x_1 \\ 2x_2 \\ \frac{1}{2}x_3 \end{bmatrix}, \quad \mathbf{J}_e^T(\mathbf{x}) = \begin{bmatrix} -1 & 2x_1 - 2 \\ 0 & 2x_2 \\ 1 & 0 \end{bmatrix}$$

Hence Eq. (10.75) becomes

$$\begin{aligned} 2x_1 + \lambda_1 - \lambda_2(2x_1 - 2) &= 0 \\ 2x_2 - 2\lambda_2 x_2 &= 0 \\ x_3 - 2\lambda_1 &= 0 \\ -x_1 + x_3 - 1 &= 0 \\ x_1^2 + x_2^2 - 2x_1 &= 0 \end{aligned}$$

Solving the above system of equations, we obtain two solutions, i.e.,

$$\mathbf{x}_1^* = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \quad \text{and} \quad \boldsymbol{\lambda}_1^* = \begin{bmatrix} \frac{1}{2} \\ \\ -\frac{1}{4} \end{bmatrix}$$

and

$$\mathbf{x}_2^* = \begin{bmatrix} 2 \\ 0 \\ 3 \end{bmatrix} \quad \text{and} \quad \boldsymbol{\lambda}_2^* = \begin{bmatrix} \frac{3}{2} \\ \\ \frac{11}{4} \end{bmatrix}$$

Inequality constraints

If there are K active inequality constraints at $\mathbf{x}^*$ and

$$\mathcal{J}(\mathbf{x}^*) = \{j_1, j_2, \ldots, j_K\}$$

$$\nabla f(\mathbf{x}^*) = \sum_{i=1}^{p} \lambda_i^* \nabla a_i(\mathbf{x}^*) + \sum_{k=1}^{K} \mu_{j_k}^* \nabla c_{j_k}(\mathbf{x}^*)$$

In words, Eq. (10.77) states that *the gradient* at $\mathbf{x}^*$, $\nabla f(\mathbf{x}^*)$, *is a linear combination of the gradients of all the constraint functions that are active at* $\mathbf{x}^*$.

Unlike the Lagrange multipliers associated with equality constraints, which can be either positive or negative, those associated with *active* inequality constraints must be nonnegative, i.e.,

$$\mu_{j_k}^* \geq 0 \qquad \text{for } 1 \leq k \leq K \tag{10.80}$$

$$\mathbf{g}(\mathbf{x}^*) - \mathbf{J}_e^T(\mathbf{x}^*)\boldsymbol{\lambda}^* - \hat{\mathbf{J}}_{ie}^T(\mathbf{x}^*)\hat{\boldsymbol{\mu}}^* = \mathbf{0} \tag{10.86a}$$

$$\mathbf{a}(\mathbf{x}^*) = \mathbf{0} \tag{10.86b}$$

$$\hat{\mathbf{c}}(\mathbf{x}^*) = \mathbf{0} \tag{10.86c}$$

where

$$\hat{\boldsymbol{\mu}}^* = [\mu_{j1}^* \ \mu_{j2}^* \ \cdots \ \mu_{jK}^*]^T \tag{10.87a}$$

$$\hat{\mathbf{J}}_{ie}(\mathbf{x}) = [\nabla c_{j1}(\mathbf{x}) \ \nabla c_{j2}(\mathbf{x}) \ \cdots \ \nabla c_{jK}(\mathbf{x})]^T \tag{10.87b}$$

$$\hat{\mathbf{c}}(\mathbf{x}) = [c_{j1}(\mathbf{x}) \ c_{j2}(\mathbf{x}) \ \cdots \ c_{jK}(\mathbf{x})]^T \tag{10.87c}$$

Theorem 10.2 ***Karush-Kuhn-Tucker conditions*** *If $\mathbf{x}^*$ is a local minimizer of the problem in Eq. (10.1) and is regular for the constraints that are active at $\mathbf{x}^*$, then*

(a) $a_i(\mathbf{x}^*) = 0 \;\; for \;\; 1 \leq i \leq p,$

(b) $c_j(\mathbf{x}^*) \geq 0 \;\; for \;\; 1 \leq j \leq q,$

(c) *there exist Lagrange multipliers* λ_i^* for $1 \leq i \leq p$ *and* μ_j^* for $1 \leq j \leq q$ *such that*

$$\nabla f(\mathbf{x}^*) = \sum_{i=1}^{p} \lambda_i^* \nabla a_i(\mathbf{x}^*) + \sum_{j=1}^{q} \mu_j^* \nabla c_j(\mathbf{x}^*) \tag{10.81}$$

(d) $\lambda_i^* a_i(\mathbf{x}^*) = 0 \;\; for \;\; 1 \leq i \leq p,$ (10.82a)

$\mu_j^* c_j(\mathbf{x}^*) = 0 \;\; for \;\; 1 \leq j \leq q$, and (10.82b)

(e) $\mu_j^* \geq 0 \;\; for \;\; 1 \leq j \leq q.$ (10.83)

Example 10.12 Solve the constrained minimization problem

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

by applying the KKT conditions.

Solution The KKT conditions imply that

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

Case 1 No active constraints

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

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

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

Case 2 One constraint 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

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

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

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

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

The solution of this system of equations is given by

$$\mathbf{x}^* = \begin{bmatrix} \frac{14}{3} \\ \\ -\frac{5}{3} \end{bmatrix} \quad \text{and} \quad \mu_2^* = \frac{14}{3}$$

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

Case 3 Both constraints 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

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

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

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

$$\mathbf{x}^* = \begin{bmatrix} 3 \\ -1 \end{bmatrix}, \quad \boldsymbol{\mu}^* = \begin{bmatrix} 8 \\ 0 \end{bmatrix}$$

满足KKT条件不一定就是最优解，但最优解一定满足KKT条件。

Second-order necessary conditions

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

$$\downarrow$$

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

$$\downarrow$$

$$\begin{aligned} f(\mathbf{x}^* + \mathbf{s}) &= L(\mathbf{x}^* + \mathbf{s},\ \boldsymbol{\lambda}^*,\ \boldsymbol{\mu}^*) \\ &= L(\mathbf{x}^*,\ \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) + \mathbf{s}^T \nabla_x L(\mathbf{x}^*,\ \boldsymbol{\lambda}^*,\ \boldsymbol{\mu}^*) \\ &\quad + \tfrac{1}{2}\mathbf{s}^T \nabla_x^2 L(\mathbf{x}^*,\ \boldsymbol{\lambda}^*,\ \boldsymbol{\mu}^*)\mathbf{s} + o(||\mathbf{s}||^2) \\ &= f(\mathbf{x}^*) + \tfrac{1}{2}\mathbf{s}^T \nabla_x^2 L(\mathbf{x}^*,\ \boldsymbol{\lambda}^*,\ \boldsymbol{\mu}^*)\mathbf{s} + o(||\mathbf{s}||^2) \end{aligned}$$

$$\downarrow$$

$$\mathbf{s}^T \nabla_x^2 L(\mathbf{x}^*,\ \boldsymbol{\lambda}^*,\ \boldsymbol{\mu}^*)\mathbf{s} \geq 0$$

Second-order necessary conditions

Theorem 10.3 ***Second-order necessary conditions for a minimum, equality constraints*** *If* $\mathbf{x}^*$ *is a constrained local minimizer of the problem in Eq. (10.57) and is a regular point of the constraints in Eq. (10.57b), then*

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

(b) there exist λ_i^* *for* $i = 1, 2, \ldots, p$ *such that*

$$\nabla f(\mathbf{x}^*) = \sum_{i=1}^{p} \lambda_i^* \nabla a_i(\mathbf{x}^*)$$

(c) $\mathbf{N}^T(\mathbf{x}^*)\nabla_x^2 L(\mathbf{x}^*, \boldsymbol{\lambda}^*)\mathbf{N}(\mathbf{x}^*) \succeq \mathbf{0}$. (10.91)

■

Second-order necessary conditions

Theorem 10.4 ***Second-order necessary conditions for a minimum, general constrained problem*** *If* $\mathbf{x}^*$ *is a constrained local minimizer of the problem in Eq. (10.1) and is a regular point of the constraints in Eqs. (10.1b) and (10.1c), then*

(a) $a_i(\mathbf{x}^*) = 0$ *for* $1 \le i \le p$,
(b) $c_j(\mathbf{x}^*) \ge 0$ *for* $1 \le j \le q$,
(c) there exist Lagrange multipliers λ_i^**'s and* μ_j^**'s such that*

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

(d) $\lambda_i^* a_i(\mathbf{x}^*) = 0$ *for* $1 \le i \le p$ *and* $\mu_j^* c_j(\mathbf{x}^*) = 0$ *for* $1 \le j \le q$,
(e) $\mu_j^* \ge 0$ *for* $1 \le j \le q$, *and*
(f) $\mathbf{N}^T(\mathbf{x}^*)\nabla_x^2 L(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)\mathbf{N}(\mathbf{x}^*) \succeq \mathbf{0}$. (10.98)

wolf -Duality

Theorem 10.9 ***Duality in convex programming*** *Let* $\mathbf{x}^*$ *be a minimizer, and* $\boldsymbol{\lambda}^*$, $\boldsymbol{\mu}^*$ *be the associated Lagrange multipliers of the problem in Eq. (10.107). If* $\mathbf{x}^*$ *is a regular point of the constraints, then* $\mathbf{x}^*$, $\boldsymbol{\lambda}^*$, *and* $\boldsymbol{\mu}^*$ *solve the dual problem*

$$\underset{\mathbf{x},\,\boldsymbol{\lambda},\,\boldsymbol{\mu}}{\text{maximize}} \;\; L(\mathbf{x},\, \boldsymbol{\lambda},\, \boldsymbol{\mu}) \tag{10.109a}$$

$$\text{subject to:} \quad \nabla_x L(\mathbf{x},\, \boldsymbol{\lambda},\, \boldsymbol{\mu}) = \mathbf{0} \tag{10.109b}$$

$$\boldsymbol{\mu} \geq \mathbf{0} \tag{10.109c}$$

In addition, $f(\mathbf{x}^*) = L(\mathbf{x}^*,\, \boldsymbol{\lambda}^*,\, \boldsymbol{\mu}^*)$.

proof:

$$
\begin{aligned}
L(\mathbf{x}^*,\ \boldsymbol{\lambda}^*,\ \boldsymbol{\mu}^*) &= f(\mathbf{x}^*) \\
&\geq f(\mathbf{x}^*) - \sum_{i=1}^{p} \lambda_i a_i(\mathbf{x}^*) - \sum_{j=1}^{q} \mu_j c_j(\mathbf{x}^*) = L(\mathbf{x}^*,\ \boldsymbol{\lambda},\ \boldsymbol{\mu})
\end{aligned}
$$

With $\boldsymbol{\mu} \geq \mathbf{0}$, the Lagrangian $L(\mathbf{x},\ \boldsymbol{\lambda},\ \boldsymbol{\mu})$ is convex and, therefore,

$$
L(\mathbf{x}^*,\ \boldsymbol{\lambda},\ \boldsymbol{\mu}) \geq L(\mathbf{x},\ \boldsymbol{\lambda},\ \boldsymbol{\mu}) + (\mathbf{x}^* - \mathbf{x})^T \nabla_x L(\mathbf{x},\ \boldsymbol{\lambda},\ \boldsymbol{\mu}) = L(\mathbf{x},\ \boldsymbol{\lambda},\ \boldsymbol{\mu})
$$

Hence $L(\mathbf{x}^*,\ \boldsymbol{\lambda}^*,\ \boldsymbol{\mu}^*) \geq L(\mathbf{x},\ \boldsymbol{\lambda},\ \boldsymbol{\mu})$, i.e., set $\{\mathbf{x}^*,\ \boldsymbol{\lambda}^*,\ \boldsymbol{\mu}^*\}$ solves the problem in Eq. (10.109).

Example

standard-form LP problem

$$\text{minimize } \mathbf{c}^T\mathbf{x}$$

$$\text{subject to:} \quad \mathbf{A}\mathbf{x} = \mathbf{b} \qquad \mathbf{A} \in R^{p\times n}$$
$$\mathbf{x} \geq \mathbf{0}$$

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

the Wolfe dual

$$\underset{\cancel{\mathbf{x}},\ \boldsymbol{\lambda},\ \boldsymbol{\mu}}{\text{maximize}}\ \mathbf{x}^T(\mathbf{c} - \mathbf{A}^T\boldsymbol{\lambda} - \boldsymbol{\mu}) + \mathbf{b}^T\boldsymbol{\lambda}$$

$$\text{subject to:} \quad \mathbf{c} - \mathbf{A}^T\boldsymbol{\lambda} - \boldsymbol{\mu} = \mathbf{0}$$
$$\boldsymbol{\mu} \geq \mathbf{0}$$

$$\underset{\boldsymbol{\lambda},\ \boldsymbol{\mu}}{\text{maximize}}\ \mathbf{b}^T\boldsymbol{\lambda}$$

$$\text{subject to:} \quad \mathbf{c} - \mathbf{A}^T\boldsymbol{\lambda} - \boldsymbol{\mu} = \mathbf{0}$$
$$\boldsymbol{\mu} \geq \mathbf{0}$$