

Chapter 10

Inequality Constraints Optimization

The optimization problem with m inequality constraints is written as

$$\begin{aligned} \max & f(\mathbf{x}) \\ \text{st. } & \mathbf{h}(\mathbf{x}) \leq \mathbf{b}. \end{aligned}$$

where $\mathbf{h}: \mathbb{R}^n \rightarrow \mathbb{R}^m$.

Definition A constraint $h^i(\mathbf{x}) \leq b_i$ is **binding** (*active*, or *effective*) at a solution $\mathbf{x}_0$ if $h^i(\mathbf{x}_0) = b_i$, and **not binding** if $h^i(\mathbf{x}_0) < b_i$.

We will first discuss the case of one inequality constraint,

$$\begin{aligned} \max & f(\mathbf{x}) \\ \text{st. } & h(\mathbf{x}) \leq b. \end{aligned}$$

then state the results for the general case.

10. 1 First-Order Sufficient Conditions: One Inequality Constraint

Theorem If $\mathbf{x}^*$ is a local maximum point and the constraint is *binding* at $\mathbf{x}^*$, we have

$$\begin{aligned} \nabla f(\mathbf{x}^*) - \mu^* \nabla h(\mathbf{x}^*) &= \mathbf{0} \\ \mu^* &\geq 0. \end{aligned}$$

Sketch of Proof: If the constraint is binding, we have exactly the same situation as in the equality-constraint case; so the first-order condition applies. The reason that $\mu^* \geq 0$ comes from the fact that the gradient of a function points to the direction the value of the function increases at the fastest rate.

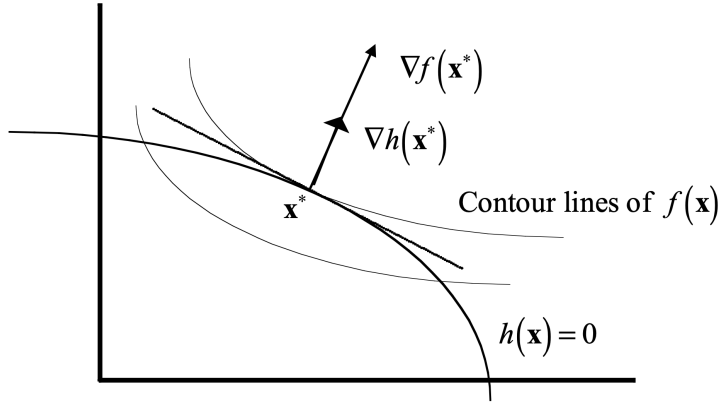


Figure 10.1 The gradients $\nabla f(\mathbf{x}^*)$ and $\nabla h(\mathbf{x}^*)$ at an maximum solution $\mathbf{x}^*$ when h is binding.

We can just write the same Lagrange function

$$\mathcal{L}(\mathbf{x}, \mu) = f(\mathbf{x}) - \mu(h(\mathbf{x}) - b),$$

as in the equality constraint case and the local maximum point $\mathbf{x}^*$ must also have the gradient $\nabla \mathcal{L}(\mathbf{x}^*, \mu^*) = \mathbf{0}$ with the value of μ^* is required to be nonnegative.

Theorem If $\mathbf{x}^*$ is a local maximum point and the constraint is *not binding* at $\mathbf{x}^*$, the gradient $\nabla f(\mathbf{x}^*) = \mathbf{0}$.

Proof When the constraint is not binding at $\mathbf{x}^*$, it means that $\mathbf{x}^*$ is in the interior of the feasible set. The point $\mathbf{x}^*$ is thus a local maximum point for the unconstrained maximization of the function f . By the first-order condition of in the unconstrained optimization, the gradient $\nabla f(\mathbf{x}^*) = \mathbf{0}$. $\square$

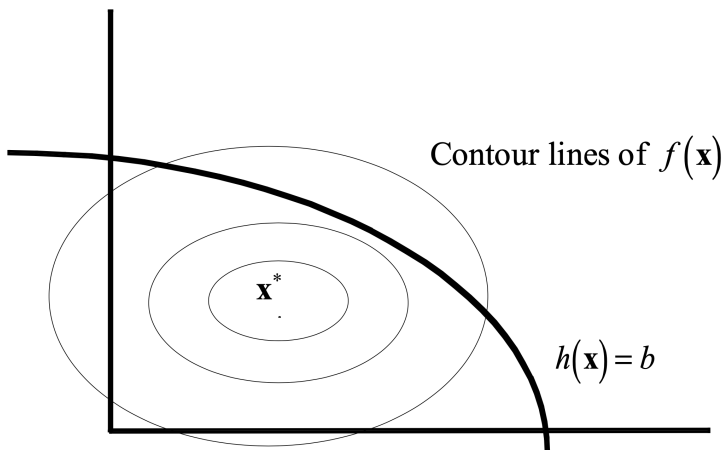


Figure 10.2 The gradient $\nabla f(\mathbf{x}^*)$ at an maximum point $\mathbf{x}^*$ when h is not binding.

Assuming that the point $\mathbf{x}^*$ is a local maximum point of the function f under the inequality constraint $h(\mathbf{x}) \leq b$, the

following two exclusive and exhaustive cases according to the two theorems above,

- a) If $h(\mathbf{x}^*) = b$, then $\mu^* > 0$, and
- b) If $h(\mathbf{x}^*) < b$, then $\mu^* = 0$,

can be shown to be equivalent to the so-called **complementary slackness condition**

$$\mu^*(h(\mathbf{x}^*) - b) = 0, \text{ and } \mu^* \geq 0.$$

This condition replaces the condition that the partial derivative of $\mathcal{L}$ with respect to μ is zero at a local maximum solution $\mathbf{x}^*$ in equality constraint optimization. We state this in the following theorem.

Theorem (Simon and Blume [1994], Theorem 18.3, page 427)
Suppose f and h are functions in $\mathcal{C}^1$ and $\mathbf{x}^*$ is a local maximum point under the constraint $h(\mathbf{x}) \leq b$. Write the Lagrange function

$$\mathcal{L}(\mathbf{x}, \mu) = f(\mathbf{x}) - \mu(h(\mathbf{x}) - b).$$

Then there exists a constant μ^* such that,

- a) $\nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}^*, \mu^*) = \nabla f(\mathbf{x}^*) - \mu^* \nabla h(\mathbf{x}^*) = \mathbf{0}$,
- b) $\mu^* \frac{\partial}{\partial \mu} \mathcal{L}(\mathbf{x}^*, \mu^*) = -\mu^*(h(\mathbf{x}^*) - b) = 0$
- c) $\mu^* \geq 0$, and
- d) $h(\mathbf{x}^*) \leq b$.

HW Rewrite this theorem for the maximization of $f(\mathbf{x})$ with $h(\mathbf{x}) \geq b$; minimization of $f(\mathbf{x})$ with $h(\mathbf{x}) \geq b$; and then minimization of $f(\mathbf{x})$ with $h(\mathbf{x}) \leq b$.

Example Simon and Blume [1994], Example 18.7.
Maximizing $f(x, y) = xy$, subject to $g(x, y) = x^2 + y^2 \leq 1$,
the points $(x^*, y^*, \mu^*) =$
 $\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, \frac{1}{2}\right), \left(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, -\frac{1}{2}\right), \left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}, -\frac{1}{2}\right),$ and
 $\left(-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}, \frac{1}{2}\right)$ are found to satisfy the four conditions in the Theorem 18.1.

HW. Simon and Blume [1994], Example 18.8. If we maximize a utility function subject to the inequality budget constraint $p_x x + p_y y \leq B$, the optimal solution will be the same under equality budget constraint. Why?

10.2 First-Order Sufficient Conditions: Several Inequality Constraints

Theorem (Simon and Blume [1994], Theorem 18.4, page 430)
Suppose f and $\mathbf{h}$ are functions in $\mathcal{C}^1$, where $\mathbf{h}: \mathbb{R}^n \rightarrow \mathbb{R}^m$, $\mathbf{h}(\mathbf{x}) \leq \mathbf{b}$, and $\mathbf{x}^*$ is a local maximum point. Assume without loss of generality that the first b constraints are binding at $\mathbf{x}^*$ and the last $m - b$ are not. Write the Lagrange function,

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\mu}) = f(\mathbf{x}) - \boldsymbol{\mu}^T(\mathbf{h}(\mathbf{x}) - \mathbf{b}),$$

where $\boldsymbol{\mu} \in \mathbb{R}^m$. Then there exist a vector $\boldsymbol{\mu}^*$ such that

- a) $\nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}^*, \boldsymbol{\mu}^*) = \nabla f(\mathbf{x}^*) - \nabla \mathbf{h}(\mathbf{x}^*)^T \boldsymbol{\mu}^* = \mathbf{0}$,
- b) $\mu_i^*(h^i(\mathbf{x}^*) - b_i) = 0, i = 1, 2, \dots, m$,
- c) $\boldsymbol{\mu}^* \geq \mathbf{0}$, and
- d) $\mathbf{h}(\mathbf{x}^*) \leq \mathbf{b}$.

Example Simon and Blume [1994], Exercise 18.9.
Maximizing $f(x, y, z) = xyz$ subject to the constraints $x + y + z \leq 1$, $x \geq 0$, $y \geq 0$, and $z \geq 0$, the point $(x^*, y^*, z^*, \mu_1^*, \mu_2^*, \mu_3^*, \mu_4^*) = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, 0, 0, 0, 0)$ is found to satisfy the conditions in the theorem above.

Example Maximize $x_1^2 x_2$, subject to $x_1 + x_2 \leq 12$ and $3x_1 + x_2 \leq 18$.

Solution:

Write the Lagrange function and the first-order sufficient conditions as

$$\begin{aligned} \mathcal{L}(x_1, x_2, \mu_1, \mu_2) &= x_1^2 x_2 - \mu_1(x_1 + x_2 - 12) \\ &\quad - \mu_2(3x_1 + x_2 - 18) \end{aligned}$$

10.3 First-Order Sufficient Condition: Mixed Constraints

The first-order conditions for the inequality and equality constraints are then combined to be the conditions for the mixed constraints case. Then we will state the second-order sufficient conditions for the inequality and mixed constraints later in one theorem.

Theorem (Simon and Blume [1994], Theorem 18.5, page 435)
Suppose $f: \mathbb{R}^n \rightarrow \mathbb{R}$, $\mathbf{g}: \mathbb{R}^n \rightarrow \mathbb{R}^k$ and $\mathbf{h}: \mathbb{R}^n \rightarrow \mathbb{R}^m$ are functions in $\mathcal{C}^1$, and $\mathbf{x}^*$ is a local maximum point of f on the feasible set $\{\mathbf{x} | \mathbf{x} \in \mathbb{R}^n, \mathbf{g}(\mathbf{x}) = \mathbf{c}, \mathbf{h}(\mathbf{x}) \leq \mathbf{b}\}$. Write the Lagrange function,

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\mathbf{x}) - \boldsymbol{\lambda}^T(\mathbf{g}(\mathbf{x}) - \mathbf{c}) - \boldsymbol{\mu}^T(\mathbf{h}(\mathbf{x}) - \mathbf{b}),$$

where $\boldsymbol{\lambda} \in \mathbb{R}^k$ and $\boldsymbol{\mu} \in \mathbb{R}^m$. Then there exist vectors $\boldsymbol{\lambda}^*$ and $\boldsymbol{\mu}^*$ such that

- a) $\nabla_{\mathbf{x}}\mathcal{L}(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = \nabla f(\mathbf{x}^*) - \nabla \mathbf{g}(\mathbf{x}^*)^T \boldsymbol{\lambda}^* - \nabla \mathbf{h}(\mathbf{x}^*)^T \boldsymbol{\mu}^* = \mathbf{0}$,
- b) $\nabla_{\boldsymbol{\lambda}}\mathcal{L}(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = -(\mathbf{g}(\mathbf{x}^*) - \mathbf{c}) = \mathbf{0}$,
- c) $\mu_i^*(h^i(\mathbf{x}^*) - b_i) = 0, i = 1, 2, \dots, m$,
- d) $\boldsymbol{\mu}^* \geq \mathbf{0}$, and
- e) $\mathbf{h}(\mathbf{x}^*) \leq \mathbf{b}$.

Example Simon and Blume [1994], Exercise 18.10. Maximizing $f(x, y) = x - y^2$, subject to $x^2 + y^2 = 4, x \geq 0$ and $y \geq 0$, the point $(x^*, y^*, \mu_1^*, \mu_2^*, \mu_3^*) = (2, 0, \frac{1}{4}, 0, 0)$ is found to satisfy the conditions in the theorem above.

10.4 First-Order Sufficient Conditions: Minimization under Mixed Constraints

If we maximize $f(\mathbf{x})$ subject to $\mathbf{h}(\mathbf{x}) \leq \mathbf{b}$, we require that the multipliers $\boldsymbol{\mu}^* \geq \mathbf{0}$. The following theorem states that when we minimize $f(\mathbf{x})$ and want $\boldsymbol{\mu}^* \geq \mathbf{0}$ we have to specify that the inequality constraints are in the form $\mathbf{h}(\mathbf{x}) \geq \mathbf{b}$.

Theorem (Simon and Blume [1994], Theorem 18.6, page 437) Suppose $f: \mathbb{R}^n \rightarrow \mathbb{R}$, $\mathbf{g}: \mathbb{R}^n \rightarrow \mathbb{R}^k$ and $\mathbf{h}: \mathbb{R}^n \rightarrow \mathbb{R}^m$ are functions in $\mathcal{C}^1$ and $\mathbf{x}^*$ is a local minimum point of f on the feasible set $\{\mathbf{x} | \mathbf{x} \in \mathbb{R}^n, \mathbf{g}(\mathbf{x}) = \mathbf{c}, \mathbf{h}(\mathbf{x}) \geq \mathbf{b}\}$. Write the Lagrange function,

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\mathbf{x}) - \boldsymbol{\lambda}^T(\mathbf{g}(\mathbf{x}) - \mathbf{c}) - \boldsymbol{\mu}^T(\mathbf{h}(\mathbf{x}) - \mathbf{b}),$$

where $\boldsymbol{\lambda} \in \mathbb{R}^k$ and $\boldsymbol{\mu} \in \mathbb{R}^m$. Then there exist vectors $\boldsymbol{\lambda}^*$ and $\boldsymbol{\mu}^*$ such that

- a) $\nabla_{\mathbf{x}}\mathcal{L}(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = \nabla f(\mathbf{x}^*) - \nabla \mathbf{g}(\mathbf{x}^*)^T \boldsymbol{\lambda}^* - \nabla \mathbf{h}(\mathbf{x}^*)^T \boldsymbol{\mu}^* = \mathbf{0}$,
- b) $\nabla_{\boldsymbol{\lambda}}\mathcal{L}(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = -(\mathbf{g}(\mathbf{x}^*) - \mathbf{c}) = \mathbf{0}$,
- c) $\mu_i^*(h^i(\mathbf{x}^*) - b_i) = 0, i = 1, 2, \dots, m$,
- d) $\boldsymbol{\mu}^* \geq \mathbf{0}$, and
- e) $\mathbf{h}(\mathbf{x}^*) \geq \mathbf{b}$.

Proof By direct application of previous theorem and see the next problem. $\square$

HW Show that we can obtain the above theorem by applying the previous theorem to the problem of maximizing $-f$ subject to $-\mathbf{h}(\mathbf{x}) \leq -\mathbf{b}$ and $\mathbf{g}(\mathbf{x}) = \mathbf{c}$.

Example Simon and Blume [1994], Exercise 18.10. Minimize $f(x, y) = 2y - x^2$ subject to $x^2 + y^2 \leq 1, x \geq 0$ and $y \geq 0$.

10.5 Second-Order Sufficient Conditions: Mixed Constraints

This result obviously is applicable also for the optimization problems with only equality or inequality constraints.

Theorem (Simon and Blume [1994], Theorem 19.8, page 466)
Consider the maximization of f in the feasible set $\{\mathbf{x} | \mathbf{x} \in \mathbb{R}^n, \mathbf{g}(\mathbf{x}) = \mathbf{c}, \mathbf{h}(\mathbf{x}) \leq \mathbf{b}\}$, where $f: \mathbb{R}^n \rightarrow \mathbb{R}$, $\mathbf{g}: \mathbb{R}^n \rightarrow \mathbb{R}^k$ and $\mathbf{h}: \mathbb{R}^n \rightarrow \mathbb{R}^m$ are twice-differentiable functions. Write the Lagrange function

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\mathbf{x}) - \boldsymbol{\lambda}^T(\mathbf{g}(\mathbf{x}) - \mathbf{c}) - \boldsymbol{\mu}^T(\mathbf{h}(\mathbf{x}) - \mathbf{b}).$$

Suppose at a given point $\mathbf{x}^*$, the following hold.

- There exist vectors $\boldsymbol{\lambda}^*$ and $\boldsymbol{\mu}^*$ such that all the first-order conditions are satisfied.
- Without loss of generality, assume that the first r inequality constraints are binding at $\mathbf{x}^*$ and that the last $m - r$ are not. The Hessian $\nabla_{\mathbf{x}}^2 \mathcal{L}(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}_r^*)$ is negative (positive) definite under the restriction of the linear constraints

$$\begin{bmatrix} \nabla \mathbf{g}(\mathbf{x}^*) \\ \nabla \mathbf{h}^r(\mathbf{x}^*) \end{bmatrix} \mathbf{d} = \begin{bmatrix} \nabla g^1(\mathbf{x}^*)^T \\ \vdots \\ \nabla g^k(\mathbf{x}^*)^T \\ \nabla h^1(\mathbf{x}^*)^T \\ \vdots \\ \nabla h^r(\mathbf{x}^*)^T \end{bmatrix} \mathbf{d} = \mathbf{0} \in \mathbb{R}^{k+r}$$

That is, $\mathbf{d}^T \nabla_{\mathbf{x}}^2 \mathcal{L}(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}_r^*) \mathbf{d} < (>) 0$, for all $\mathbf{d} \in \mathbb{R}^n$, $\mathbf{d} \neq \mathbf{0}$, and $\begin{bmatrix} \nabla \mathbf{g}(\mathbf{x}^*) \\ \nabla \mathbf{h}^r(\mathbf{x}^*) \end{bmatrix} \mathbf{d} = \mathbf{0}$, where $\boldsymbol{\mu}_r^*$ are the Lagrange Multipliers associated with the r binding constraints, i.e., $\boldsymbol{\mu}^* = \begin{bmatrix} \boldsymbol{\mu}_r^* \\ \mathbf{0} \end{bmatrix}$.

Then the solution $\mathbf{x}^*$ is a strict local constrained maximum (minimum) of f .

Definition The point $\mathbf{x}^*$ that satisfies the condition (a) of the theorem above is called a **critical point**.

To check the condition (b), construct the Bordered Hessian $\bar{\mathbf{H}}$

$$\bar{\mathbf{H}} = \nabla_{\begin{bmatrix} \boldsymbol{\lambda} \\ \boldsymbol{\mu}_r \\ \mathbf{x} \end{bmatrix}}^2 \mathcal{L}(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}_r^*) = \begin{bmatrix} \mathbf{0} & \mathbf{0} & -\nabla \mathbf{g}(\mathbf{x}^*) \\ \mathbf{0} & \mathbf{0} & -\nabla \mathbf{h}^r(\mathbf{x}^*) \\ -\nabla \mathbf{g}(\mathbf{x}^*)^T & -\nabla \mathbf{h}^r(\mathbf{x}^*)^T & \nabla_{\mathbf{x}}^2 \mathcal{L}(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}_r^*) \end{bmatrix} \in \mathbb{R}^{(k+r+n) \times (k+r+n)}$$

where $\boldsymbol{\mu}_r^*$ is the vector of the Lagrange multipliers of first r binding inequality constraints.

Condition **(b)** holds if and only if the Bordered Hessian $\bar{\mathbf{H}}$ is negative definite. The Bordered Hessian $\bar{\mathbf{H}}$ is negative definite if the determinant of $\bar{\mathbf{H}}$ has the same sign of $(-1)^n$ and the last $n - (k + r)$ leading principal minors alternate in sign, and positive definite if the determinant of $\bar{\mathbf{H}}$ and all the last $n - (k + r)$ leading principal minors have the same sign as $(-1)^{k+r}$.

If the Bordered Hessian $\bar{\mathbf{H}}$ at the critical point $(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ is neither negative nor positive definite, then the $\mathbf{x}^*$ is not guaranteed to be a local constrained maximum nor minimum point.

10.6 Second-Order Necessary Conditions

The corresponding necessary condition that a constrained maximum solution must satisfy is, of course, that the Hessian $\nabla_{\mathbf{x}}^2 \mathcal{L}(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ is negative semidefinite for all direction $\mathbf{d}$ under the restriction $\begin{bmatrix} \nabla \mathbf{g}(\mathbf{x}^*) \\ \nabla \mathbf{h}^r(\mathbf{x}^*) \end{bmatrix} \mathbf{d} = \mathbf{0}$. The characterization of constrained positive (negative) semidefiniteness in terms of principal minors of the Bordered Hessian $\bar{\mathbf{H}}$ can be found in Theorem 5.5 of **Sundaram** [1996], page 120.

10.7 Comparative Static Analysis: Sensitivity Analysis

Suppose that the constrained optimization contains a vector of parameters $\mathbf{d}_0 \in \mathbb{R}^p$. We write

$$\begin{aligned} & \max f(\mathbf{x}; \mathbf{d}_0) \\ & \text{st. } \mathbf{g}(\mathbf{x}; \mathbf{d}_0) = \mathbf{0} \\ & \quad \mathbf{h}(\mathbf{x}; \mathbf{d}_0) \leq \mathbf{0}, \end{aligned}$$

where the parameters $\mathbf{d}_0$ can be both in the objective function and the constraints. The right-hand-side of the constraints are now set to zero so that the parameters $\mathbf{d}_0$ here could also be the right-hand-side itself.

The Lagrange function is, for a given vector $\mathbf{d}$,

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}; \mathbf{d}) = f(\mathbf{x}; \mathbf{d}) - \boldsymbol{\lambda}^T (\mathbf{g}(\mathbf{x}; \mathbf{d}) - \mathbf{c}) - \boldsymbol{\mu}^T (\mathbf{h}(\mathbf{x}; \mathbf{d}) - \mathbf{b}).$$

Assuming the first r inequality constraints are binding, the first-order condition is thus a set of implicit functions:

$$\begin{aligned} \nabla_{\begin{bmatrix} \lambda \\ \mu_r \\ \mathbf{x} \end{bmatrix}} \mathcal{L}(\mathbf{x}(\mathbf{d}), \lambda(\mathbf{d}), \mu_r(\mathbf{d}); \mathbf{d}) &= \begin{bmatrix} \nabla_{\lambda} \mathcal{L}(\mathbf{x}(\mathbf{d}), \lambda(\mathbf{d}), \mu_r(\mathbf{d}); \mathbf{d}) \\ \nabla_{\mu_r} \mathcal{L}(\mathbf{x}(\mathbf{d}), \lambda(\mathbf{d}), \mu_r(\mathbf{d}); \mathbf{d}) \\ \nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}(\mathbf{d}), \lambda(\mathbf{d}), \mu_r(\mathbf{d}); \mathbf{d}) \end{bmatrix} \\ &= \begin{bmatrix} -\mathbf{g}(\mathbf{x}(\mathbf{d}); \mathbf{d}) \\ -\mathbf{h}^r(\mathbf{x}(\mathbf{d}); \mathbf{d}) \\ \nabla_{\mathbf{x}} f(\mathbf{x}(\mathbf{d}); \mathbf{d}) - \nabla_{\mathbf{x}} \mathbf{g}(\mathbf{x}(\mathbf{d}); \mathbf{d})^T \lambda(\mathbf{d}) - \nabla_{\mathbf{x}} \mathbf{h}^r(\mathbf{x}(\mathbf{d}); \mathbf{d})^T \mu_r(\mathbf{d}) \end{bmatrix} = \mathbf{0}. \end{aligned}$$

If $(\mathbf{x}^*, \lambda^*, \mu^*)$ is the optimal solution for the parameter vector $\mathbf{d}_0$, under the condition that

$$\nabla_{\begin{bmatrix} \lambda \\ \mu_r \\ \mathbf{x} \end{bmatrix}} \left[\nabla_{\begin{bmatrix} \lambda \\ \mu_r \\ \mathbf{x} \end{bmatrix}} \mathcal{L}(\mathbf{x}^*, \lambda^*, \mu^*; \mathbf{d}_0) \right] = \nabla_{\begin{bmatrix} \lambda \\ \mu_r \\ \mathbf{x} \end{bmatrix}}^2 \mathcal{L}(\mathbf{x}^*, \lambda^*, \mu^*; \mathbf{d}_0) = \bar{\mathbf{H}}(\mathbf{x}^*, \lambda^*, \mu^*; \mathbf{d}_0)$$

is nonsingular, which is true if the second-order sufficient condition holds, then there is a differentiable

function $\begin{bmatrix} \lambda(\mathbf{d}) \\ \mu_r(\mathbf{d}) \\ \mathbf{x}(\mathbf{d}) \end{bmatrix}$ such that

- a) $\nabla_{\begin{bmatrix} \lambda \\ \mu_r \\ \mathbf{x} \end{bmatrix}} \mathcal{L}(\mathbf{x}(\mathbf{d}), \lambda(\mathbf{d}), \mu_r(\mathbf{d}); \mathbf{d}) = \mathbf{0}$, for $\|\mathbf{d} - \mathbf{d}_0\| < \varepsilon$
- b) $\begin{bmatrix} \lambda(\mathbf{d}_0) \\ \mu_r(\mathbf{d}_0) \\ \mathbf{x}(\mathbf{d}_0) \end{bmatrix} = \begin{bmatrix} \lambda^* \\ \mu_r^* \\ \mathbf{x}^* \end{bmatrix}$, and
- c) the gradient

$$\begin{aligned} \nabla_{\mathbf{d}} \begin{bmatrix} \lambda(\mathbf{d}_0) \\ \mu_r(\mathbf{d}_0) \\ \mathbf{x}(\mathbf{d}_0) \end{bmatrix} &= -[\bar{\mathbf{H}}(\mathbf{x}^*, \lambda^*, \mu^*; \mathbf{d}_0)]^{-1} \nabla_{\mathbf{d}} \left[\nabla_{\begin{bmatrix} \lambda \\ \mu_r \\ \mathbf{x} \end{bmatrix}} \mathcal{L}(\mathbf{x}^*, \lambda^*, \mu^*; \mathbf{d}_0) \right] \\ &= \begin{bmatrix} \mathbf{0} & \mathbf{0} & \nabla_{\mathbf{d}} \mathbf{g}(\mathbf{x}^*; \mathbf{d}_0) \\ \mathbf{0} & \mathbf{0} & \nabla_{\mathbf{d}} \mathbf{h}^r(\mathbf{x}^*; \mathbf{d}_0) \\ \nabla_{\mathbf{d}} [\nabla_{\mathbf{x}} f(\mathbf{x}(\mathbf{d}); \mathbf{d}) - \nabla_{\mathbf{x}} \mathbf{g}(\mathbf{x}(\mathbf{d}); \mathbf{d})^T \lambda(\mathbf{d}) - \nabla_{\mathbf{x}} \mathbf{h}^r(\mathbf{x}(\mathbf{d}); \mathbf{d})^T \mu_r(\mathbf{d})] \end{bmatrix}^{-1} \begin{bmatrix} -\nabla_{\mathbf{d}} \mathbf{g}(\mathbf{x}(\mathbf{d}); \mathbf{d}) \\ -\nabla_{\mathbf{d}} \mathbf{h}^r(\mathbf{x}(\mathbf{d}); \mathbf{d}) \\ \nabla_{\mathbf{d}} [\nabla_{\mathbf{x}} f(\mathbf{x}(\mathbf{d}); \mathbf{d}) - \nabla_{\mathbf{x}} \mathbf{g}(\mathbf{x}(\mathbf{d}); \mathbf{d})^T \lambda(\mathbf{d}) - \nabla_{\mathbf{x}} \mathbf{h}^r(\mathbf{x}(\mathbf{d}); \mathbf{d})^T \mu_r(\mathbf{d})] \end{bmatrix}. \end{aligned}$$

We may write interchangeably $\nabla_{\mathbf{d}} \begin{bmatrix} \lambda(\mathbf{d}_0) \\ \mu_r(\mathbf{d}_0) \\ \mathbf{x}(\mathbf{d}_0) \end{bmatrix} = \nabla_{\mathbf{d}} \begin{bmatrix} \lambda^* \\ \mu_r^* \\ \mathbf{x}^* \end{bmatrix}$.

10.8 Kuhn-Tucker Formulation

The most common constrained optimization problems in economics are of the form,

$$\begin{aligned} \max & f(\mathbf{x}) \\ \text{st. } & \mathbf{h}(\mathbf{x}) \leq \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0}. \end{aligned}$$

Kuhn-Tucker formulation is a special Lagrange method for this form optimization. Assume the rows of the Jacobian $\nabla \mathbf{h}^r(\mathbf{x}^*)$ of binding constraints are linearly independent. Write the usual Lagrange Function:

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\mu}, \mathbf{v}) = f(\mathbf{x}) - \boldsymbol{\mu}^T(\mathbf{h}(\mathbf{x}) - \mathbf{b}) + \mathbf{v}^T \mathbf{x}.$$

The first-order conditions according to the theorem in Section 10.2 are

- a) $\nabla \mathcal{L}(\mathbf{x}^*, \boldsymbol{\mu}^*, \mathbf{v}^*) = \nabla f(\mathbf{x}^*) - \nabla \mathbf{h}(\mathbf{x}^*)^T \boldsymbol{\mu}^* + \mathbf{v}^* = \mathbf{0}$,
- b) $\mu_i^*(h^i(\mathbf{x}^*) - b_i) = 0, i = 1, 2, \dots, m$,
- c) $v_j^* x_j^* = 0, j = 1, 2, \dots, n$,
- d) $\mathbf{h}(\mathbf{x}^*) \leq \mathbf{b}$,
- e) $\mathbf{x}^* \geq \mathbf{0}$,
- f) $\boldsymbol{\mu}^* \geq \mathbf{0}$, and
- g) $\mathbf{v}^* \geq \mathbf{0}$.

Definition The *Kuhn-Tucker Lagrange function* $\hat{\mathcal{L}}$ is a modified Lagrange function that does not include the nonnegativity constraint portion. That is,

$$\hat{\mathcal{L}}(\mathbf{x}, \boldsymbol{\mu}) = f(\mathbf{x}) - \boldsymbol{\mu}^T(\mathbf{h}(\mathbf{x}) - \mathbf{b}).$$

Since $\mathcal{L}(\mathbf{x}, \boldsymbol{\mu}, \mathbf{v}) = \hat{\mathcal{L}}(\mathbf{x}, \boldsymbol{\mu}) + \mathbf{v}^T \mathbf{x}$, the condition (a) at $(\mathbf{x}^*, \boldsymbol{\mu}^*)$ becomes

$$\text{A) } \nabla_{\mathbf{x}} \hat{\mathcal{L}}(\mathbf{x}^*, \boldsymbol{\mu}^*) = -\mathbf{v}^*,$$

this, together with (c) and (g), we have

$$\text{C) } x_j^* \frac{\partial \hat{\mathcal{L}}}{\partial x_j} = 0, j = 1, 2, \dots, n$$

$$\text{G) } \nabla_{\mathbf{x}} \hat{\mathcal{L}}(\mathbf{x}, \boldsymbol{\mu}) \leq \mathbf{0},$$

For any solution $\mathbf{x}$,

$$\nabla_{\boldsymbol{\mu}} \mathcal{L}(\mathbf{x}, \boldsymbol{\mu}, \mathbf{v}) = \nabla_{\boldsymbol{\mu}} \hat{\mathcal{L}}(\mathbf{x}, \boldsymbol{\mu}) = -\mathbf{h}(\mathbf{x}) + \mathbf{b},$$

therefore, at $(\mathbf{x}^*, \boldsymbol{\mu}^*)$, the conditions (b) and (d) become

$$\begin{aligned} \text{B)} \quad & \mu_i^* (h^i(\mathbf{x}^*) - b_i) = \mu_i^* \frac{\partial \hat{\mathcal{L}}}{\partial \mu_i} = 0, i = \\ & 1, 2, \dots, m, \\ \text{D)} \quad & \nabla_{\mu} \hat{\mathcal{L}}(\mathbf{x}^*, \mu^*) \geq \mathbf{0}, \end{aligned}$$

Thus, we have proved the first-order condition of the Kuhn-Tucker Lagrange function as follows.

Theorem (Kuhn-Tucker First-Order Conditions)
Consider the maximization of f in the feasible set $\{\mathbf{x} | \mathbf{x} \in \mathbb{R}^n, \mathbf{h}(\mathbf{x}) \leq \mathbf{b}, \mathbf{x} \geq \mathbf{0}\}$, where $f: \mathbb{R}^n \rightarrow \mathbb{R}$ and $\mathbf{h}: \mathbb{R}^n \rightarrow \mathbb{R}^m$ are differentiable functions. Write the Kuhn-Tucker Lagrange function

$$\hat{\mathcal{L}}(\mathbf{x}, \mu) = f(\mathbf{x}) - \mu^T (\mathbf{h}(\mathbf{x}) - \mathbf{b})$$

Suppose $\mathbf{x}^*$ is a local maximum solution to the maximization problem. Assume that the rows of the Jacobian matrix of these binding constraints are linearly independent. Then, there exists a nonnegative vector such that

$$\begin{aligned} \text{a)} \quad & \nabla_{\mathbf{x}} \hat{\mathcal{L}}(\mathbf{x}^*, \mu^*) \leq \mathbf{0}, \\ \text{b)} \quad & x_j^* \frac{\partial \hat{\mathcal{L}}}{\partial x_j} = 0, j = 1, 2, \dots, n, \\ \text{c)} \quad & \nabla_{\mu} \hat{\mathcal{L}}(\mathbf{x}^*, \mu^*) \geq \mathbf{0}, \text{ and} \\ \text{d)} \quad & \mu_i^* \frac{\partial \hat{\mathcal{L}}}{\partial \mu_i} = 0, i = 1, 2, \dots, m, \\ \text{e)} \quad & \mathbf{x}^* \geq \mathbf{0}, \\ \text{f)} \quad & \mu^* \geq \mathbf{0}, \end{aligned}$$

HW Write the Kuhn-Tucker First-Order Conditions for the maximization problem under the mixed constraints.

HW Write the Kuhn-Tucker First-Order Conditions for the minimization problem under the inequality ($\geq$) constraints, as derived from the first-order conditions of the ordinary Lagrange function.

HW Write the second-order sufficient conditions, given a point satisfies the Kuhn-Tucker First-Order Conditions for the maximization problem under the inequality constraints.

There are two advantages over the usual Lagrange formulation. First it involves only $n + m$ equations and $n + m$ unknowns, compared with $2n + m$ equations and $2n + m$ unknowns in the usual formulation. Second it shows a symmetry between the

primal variables $\mathbf{x}$ and the **dual** variables $\boldsymbol{\mu}$. This leads naturally to the dual analysis in optimization. See Chapter 22 of **Simon and Blume** [1994] and **Cornes** [1992] for more duality discussion.

Example Duality in Linear Programming. Consider the maximization problem

$$\begin{aligned} \max f(\mathbf{x}) &= \mathbf{c}^T \mathbf{x} \\ \text{st. } \mathbf{h}(\mathbf{x}) &= \mathbf{Ax} - \mathbf{b} \leq \mathbf{0} \\ \mathbf{x} &\geq \mathbf{0}. \end{aligned}$$

The Kuhn-Tucker Lagrange function is

$$\hat{\mathcal{L}}(\mathbf{x}, \boldsymbol{\mu}) = \mathbf{c}^T \mathbf{x} - \boldsymbol{\mu}^T (\mathbf{Ax} - \mathbf{b})$$

If $\mathbf{x}^*$ is a local maximum solution, it is a global maximum solution (see the problem below), and noting that

$$\mathbf{A} = [\mathbf{a}_{\cdot 1} \quad \cdots \quad \mathbf{a}_{\cdot j} \quad \cdots \quad \mathbf{a}_{\cdot n}] = \begin{bmatrix} \mathbf{a}_{1\cdot} \\ \vdots \\ \mathbf{a}_{i\cdot} \\ \vdots \\ \mathbf{a}_{m\cdot} \end{bmatrix}$$

$$\mathbf{A}^T = \begin{bmatrix} \mathbf{a}_{\cdot 1}^T \\ \vdots \\ \mathbf{a}_{\cdot j}^T \\ \vdots \\ \mathbf{a}_{\cdot n}^T \end{bmatrix},$$

the Kuhn-Tucker conditions are

- a) $\nabla_{\mathbf{x}} \hat{\mathcal{L}}(\mathbf{x}^*, \boldsymbol{\mu}^*) = \mathbf{c} - \mathbf{A}^T \boldsymbol{\mu}^* \leq \mathbf{0}$,
- b) $x_j^* (c_j - \mathbf{a}_{\cdot j}^T \boldsymbol{\mu}^*) = 0, j = 1, 2, \dots, n$,
- c) $\nabla_{\boldsymbol{\mu}} \hat{\mathcal{L}}(\mathbf{x}^*, \boldsymbol{\mu}^*) = -\mathbf{Ax}^* + \mathbf{b} \geq \mathbf{0}$, which is just $\mathbf{Ax}^* \leq \mathbf{b}$, and
- d) $\mu_i^* \frac{\partial \hat{\mathcal{L}}}{\partial \mu_i} = \mu_i^* (\mathbf{a}_{i\cdot} \mathbf{x}^* - b_i) = 0, i = 1, 2, \dots, m$,
- e) $\mathbf{x}^* \geq \mathbf{0}$
- f) $\boldsymbol{\mu}^* \geq \mathbf{0}$.

The variables $\boldsymbol{\mu}$ are called the dual variables. The conditions (b) and (d) are the **complementary slackness conditions**.

HW Show that if $\mathbf{x}^*$ is a local maximum solution of the above Linear Programming problem, it is a global maximum solution.

HW Show that the same four conditions in the example of Duality in Linear Programming above can be obtained from the minimization problem

$$\begin{aligned} \min \quad & \mathbf{b}^T \boldsymbol{\mu} \\ \text{st.} \quad & \mathbf{A}^T \boldsymbol{\mu} \geq \mathbf{c} \\ & \boldsymbol{\mu} \geq \mathbf{0}, \end{aligned}$$

using $\mathbf{x}$ as the dual variables. We can conclude therefore that if we can solve one problem, we solve the other.

HW Write the dual LP of a maximizing primal LP having

- one greater than or equal constraint,
- one equality constraint,
- one variable is restricted to be nonpositive
- one variable is unrestricted and it can take positive or negative value.