Chapter 10
Inequality Constraints Optimization

The optimization problem with m inequality constraints is written as

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

where $\mathbf{h}: \boldsymbol{R}^n \to \boldsymbol{R}^m$.

A constraint $h^i(\mathbf{x}) \le b_i$ 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,

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

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

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

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 $\mu^* \ge 0$ comes from the fact that the gradient of a function points to the direction the value of the function increases.

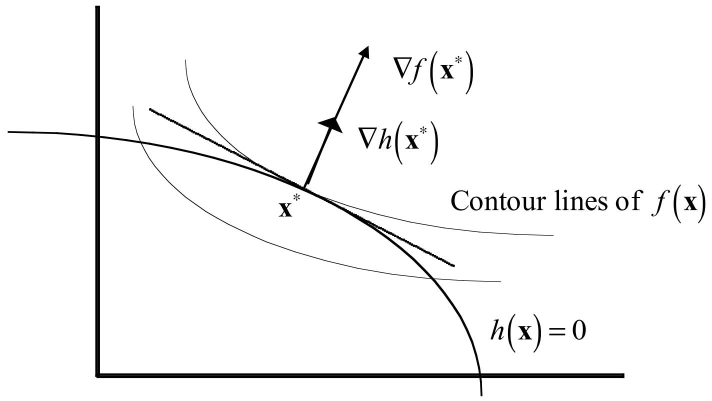

Figure 10.1 The gradients $\nabla f\left(\mathbf{x}^*\right)$ and $\nabla h\left(\mathbf{x}^*\right)$ at an maximum solution $\mathbf{x}^*$ when h is binding.

We can just write the same Lagrange function

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

as in the equality constraint case and the local maximum point $\mathbf{x}^*$ must also have the gradient $\nabla\mathcal{L}\left(\mathbf{x}^*,\mu^*\right)=\mathbf{0}$ 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 $\nabla f\left(\mathbf{x}^*\right)=\mathbf{0}$.

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 $\nabla f\left(\mathbf{x}^*\right)=\mathbf{0}$.❑

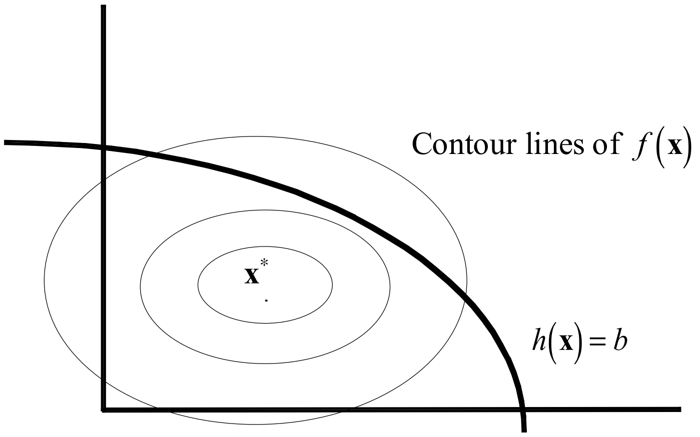

Figure 10.2 The gradient $\nabla f(\mathbf{x}^*)$ at an maximum point $\mathbf{x}^*$ when h is not binding.

Assuming that the point $\mathbf{x}^*$ is a local maximum point of the function f under the inequality constraint $h(\mathbf{x}) \le 0$, 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***

$$\mu^* \left(h(\mathbf{x}^*) - b \right) = 0 \text{ and } \mu^* \ge 0 .$$

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 $\boldsymbol{C}^1$ and $\mathbf{x}^*$ is a local maximum point under the constraint $h(\mathbf{x}) \le b$. 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) $\nabla_{\mathbf{x}} \mathcal{L}\left(\mathbf{x}^*, \mu^*\right) = \nabla f\left(\mathbf{x}^*\right) - \mu^* \nabla h\left(\mathbf{x}^*\right) = \mathbf{0}$,

b) $\mu^* \dfrac{\partial}{\partial \mu} \mathcal{L}\left(\mathbf{x}^*, \mu^*\right) = -\mu^*\left(h\left(\mathbf{x}^*\right) - b\right) = 0$

c) $\mu^* \geq 0$, and

d) $h\left(\mathbf{x}^*\right) \leq b$.

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

Example **Simon and Blume** [1994], Example 18.7. Maximizing $f\left(x, y\right) = xy$, subject to , the points $\left(x^*, y^*, \mu^*\right) = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, \frac{1}{2}\right)$, $\left(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, -\frac{1}{2}\right)$, $\left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}, -\frac{1}{2}\right)$, and $\left(-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}, +\frac{1}{2}\right)$ are found to satisify 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 $\boldsymbol{C}^1$, where $\mathbf{h}: \boldsymbol{R}^n \to \boldsymbol{R}^m$, $\mathbf{h}\left(\mathbf{x}\right) \leq \mathbf{b}$, 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,

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

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

a) $\nabla_{\mathbf{x}}\mathcal{L}\left(\mathbf{x}^*,\boldsymbol{\mu}^*\right)=\nabla f\left(\mathbf{x}^*\right)-\ \nabla\mathbf{h}\left(\mathbf{x}^*\right)^{\mathrm{T}}\boldsymbol{\mu}^*=\mathbf{0}$,

b) $\mu_i^*\left(h^i\left(\mathbf{x}^*\right)-b_i\right)=0$, $i=1,2,\ldots,m$,

c) $\boldsymbol{\mu}^*\geq\mathbf{0}$, and

d) $\mathbf{h}\left(\mathbf{x}^*\right)\leq\mathbf{b}$.

Example Simon and Blume [1994], Exercise 18.9. Maximizing $f\left(x,y,z\right)=xyz$ subject to the constraints $x+y+z\leq 1$, $x\geq 0$, $y\geq 0$, and $z\geq 0$, the point $\left(x^*,y^*,z^*,\mu_1^*,\mu_2^*,\mu_3^*,\mu_4^*\right)=\left(\frac{1}{3},\frac{1}{3},\frac{1}{3},\frac{1}{9},0,0,0\right)$ is found to satisfy the conditions in the theorem above.

Example Maximize $x_1^2x_2$

$$\text{st. } x_1+x_2\leq 12$$
$$3x_1+x_2\leq 18$$

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

$$\mathcal{L}\left(x_1,x_2,\mu_1,\mu_2\right)=x_1^2x_2-\mu_1\left(x_1+x_2-12\right)-\mu_2\left(3x_1+x_2-18\right)$$

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 , and $\mathbf{h}:\boldsymbol{R}^n\to\boldsymbol{R}^m$ are functions in $\boldsymbol{C}^1$, and $\mathbf{x}^*$ is a local maximum point of *f* on the feasible set $\left\{\mathbf{x}\middle|\mathbf{x}\in\boldsymbol{R}^n,\ \mathbf{g}\left(\mathbf{x}\right)=\mathbf{c},\ \mathbf{h}\left(\mathbf{x}\right)\leq\mathbf{b}\right\}$. Write the Lagrange function,

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

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

a) $\nabla_{\mathbf{x}} \mathcal{L}\left(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*\right) = \nabla f\left(\mathbf{x}^*\right) - \nabla \mathbf{g}\left(\mathbf{x}^*\right)^{\mathrm{T}} \boldsymbol{\lambda}^* - \nabla \mathbf{h}\left(\mathbf{x}^*\right)^{\mathrm{T}} \boldsymbol{\mu}^* = \mathbf{0}$,

b) $\nabla_{\boldsymbol{\lambda}} \mathcal{L}\left(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*\right) = -\left(\mathbf{g}\left(\mathbf{x}^*\right) - \mathbf{c}\right) = \mathbf{0}$,

c) $\mu_i^*\left(h^i\left(\mathbf{x}^*\right) - b_i\right) = 0$, $i = 1,2,\ldots,m$,

d) $\boldsymbol{\mu}^* \geq \mathbf{0}$, and

e) $\mathbf{h}\left(\mathbf{x}^*\right) \leq \mathbf{b}$.

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 is $\left(x^*, y^*, \mu_1^*, \mu_2^*, \mu_3^*\right) = \left(2, 0, \frac{1}{4}, 0, 0\right)$ is found to satisfy the conditions in the theorem above.

10.4 First-Order Sufficient Conditions: Minimization under Mixed Constraints

If we maximize *f* 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* 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 $f : \boldsymbol{R}^n \to \boldsymbol{R}$, $\mathbf{g} : \boldsymbol{R}^n \to \boldsymbol{R}^k$ and $\mathbf{h} : \boldsymbol{R}^n \to \boldsymbol{R}^m$ are functions in $\boldsymbol{C}^1$ and $\mathbf{x}^*$ is a local minimum point of f on the feasible set $\left\{\mathbf{x} \middle| \mathbf{x} \in \boldsymbol{R}^n,\ \mathbf{g}(\mathbf{x}) = \mathbf{c},\ \mathbf{h}(\mathbf{x}) \geq \mathbf{b}\right\}$. Without loss of generality, assume that the first b inequality constraints are binding at $\mathbf{x}^*$ and that the last $m - b$ are not. Write the Lagrange function,

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

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

a) $\nabla_{\mathbf{x}} \mathcal{L}\left(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*\right) = \nabla f\left(\mathbf{x}^*\right) - \nabla \mathbf{g}\left(\mathbf{x}^*\right)^{\mathrm{T}} \boldsymbol{\lambda}^* - \nabla \mathbf{h}\left(\mathbf{x}^*\right)^{\mathrm{T}} \boldsymbol{\mu}^* = \mathbf{0}$,

b) $\nabla_{\boldsymbol{\lambda}} \mathcal{L}\left(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*\right) = -\left(\mathbf{g}\left(\mathbf{x}^*\right) - \mathbf{c}\right) = \mathbf{0}$,

c) $\mu_i^*\left(h^i\left(\mathbf{x}^*\right) - b_i\right) = 0$, $i = 1,2,\ldots,m$,

d) $\boldsymbol{\mu}^* \geq \mathbf{0}$, and

e) $\mathbf{h}\left(\mathbf{x}^*\right) \geq \mathbf{b}$.

Proof By direct application of previous theorem and see the next problem.❑

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

Example Simon and Blume [1994], Exercise 18.10. Minimize $f\left(x, y\right) = 2y - x^2$ 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 $\left\{\mathbf{x} \middle| \mathbf{x} \in \boldsymbol{R}^n,\ \mathbf{g}\left(\mathbf{x}\right) = \mathbf{c},\ \mathbf{h}\left(\mathbf{x}\right) \leq \mathbf{b}\right\}$, where $f\colon \boldsymbol{R}^n \to \boldsymbol{R}$, $\mathbf{g}\colon \boldsymbol{R}^n \to \boldsymbol{R}^k$ and $\mathbf{h}\colon \boldsymbol{R}^n \to \boldsymbol{R}^m$ are twice-differentiable functions. Write the Lagrange function

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

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

a) There exist vectors $\boldsymbol{\lambda}^*$ and $\boldsymbol{\mu}^*$ such that all the first-order conditions are satisfied.

b) Without loss of generality, assume that the first b inequality constraints are binding at $\mathbf{x}^*$ and that the last $m-b$ are not. The Hessian $\nabla^2_{\mathbf{x}}\mathcal{L}\left(\mathbf{x}^*,\boldsymbol{\lambda}^*,\boldsymbol{\mu}^*\right)$ is negative (positive) definite under the restriction of the linear constraints

$$\begin{bmatrix} \nabla\mathbf{g}\left(\mathbf{x}^*\right) \\ \nabla\mathbf{h}^b\left(\mathbf{x}^*\right) \end{bmatrix}\mathbf{d} = \begin{bmatrix} \nabla g^1\left(\mathbf{x}^*\right)^{\mathrm{T}} \\ \vdots \\ \nabla g^k\left(\mathbf{x}^*\right)^{\mathrm{T}} \\ \nabla h^1\left(\mathbf{x}^*\right)^{\mathrm{T}} \\ \vdots \\ \nabla h^b\left(\mathbf{x}^*\right)^{\mathrm{T}} \end{bmatrix}\mathbf{d} = \mathbf{0} \in \boldsymbol{R}^{k+b} .$$

That is, $\mathbf{d}^{\mathrm{T}}\nabla^2_{\mathbf{x}}\mathcal{L}\left(\mathbf{x}^*,\boldsymbol{\lambda}^*,\boldsymbol{\mu}^*\right)\mathbf{d} < (>) 0$, for all $\mathbf{d} \in \boldsymbol{R}^n$, $\mathbf{d} \neq \mathbf{0}$, and $\begin{bmatrix} \nabla\mathbf{g}\left(\mathbf{x}^*\right) \\ \nabla\mathbf{h}^b\left(\mathbf{x}^*\right) \end{bmatrix}\mathbf{d} = \mathbf{0}$.

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 $\overline{\mathbf{H}}$,

$$\overline{\mathbf{H}} = \nabla^2_{\begin{bmatrix}\boldsymbol{\lambda}\\ \boldsymbol{\mu}_b\\ \mathbf{x}\end{bmatrix}}\mathcal{L}\left(\mathbf{x}^*,\boldsymbol{\lambda}^*,\boldsymbol{\mu}^*\right) = \begin{bmatrix} \mathbf{0} & \mathbf{0} & -\nabla\mathbf{g}\left(\mathbf{x}^*\right) \\ \mathbf{0} & \mathbf{0} & -\nabla\mathbf{h}^b\left(\mathbf{x}^*\right) \\ -\nabla\mathbf{g}\left(\mathbf{x}^*\right)^{\mathrm{T}} & -\nabla\mathbf{h}^b\left(\mathbf{x}^*\right)^{\mathrm{T}} & \nabla^2_{\mathbf{x}}\mathcal{L}\left(\mathbf{x}^*,\boldsymbol{\lambda}^*,\boldsymbol{\mu}^*\right) \end{bmatrix} \in \boldsymbol{R}^{(k+b+n)\times(k+b+n)}$$

where $\boldsymbol{\mu}_b$ is the vector of the Lagrange multipliers of first b binding inequality constraints.

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

If the gradient $\nabla_{\mathbf{x}}\mathcal{L}(\mathbf{x}^*,\boldsymbol{\lambda}^*,\boldsymbol{\mu}^*)=\mathbf{0}$ but the Bordered Hessian $\overline{\mathbf{H}}$ is neither negative nor positive definite, then the critical point $\mathbf{x}^*$ 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 $\nabla^2_{\mathbf{x}}\mathcal{L}(\mathbf{x}^*,\boldsymbol{\lambda}^*,\boldsymbol{\mu}^*)$ is negative semidefinite for all direction $\mathbf{d}$ under the restriction $\begin{bmatrix}\nabla\mathbf{g}(\mathbf{x}^*)\\ \nabla\mathbf{h}^b(\mathbf{x}^*)\end{bmatrix}\mathbf{d}=\mathbf{0}$. The characterization of constrained positive (negative) semidefiniteness in terms of principal minors of the Bordered Hessian $\overline{\mathbf{H}}$ 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 $\mathbf{d}_0 \in \boldsymbol{R}^p$. We write

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

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 $\mathbf{d}$,

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

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

$$\nabla_{\begin{bmatrix}\boldsymbol{\lambda}\\ \boldsymbol{\mu}_b\\ \mathbf{x}\end{bmatrix}}\mathcal{L}\left(\mathbf{x}(\mathbf{d}),\boldsymbol{\lambda}(\mathbf{d}),\boldsymbol{\mu}_b(\mathbf{d});\mathbf{d}\right) = \begin{bmatrix}\nabla_{\boldsymbol{\lambda}}\mathcal{L}\left(\mathbf{x}(\mathbf{d}),\boldsymbol{\lambda}(\mathbf{d}),\boldsymbol{\mu}_b(\mathbf{d});\mathbf{d}\right)\\ \nabla_{\boldsymbol{\mu}_b}\mathcal{L}\left(\mathbf{x}(\mathbf{d}),\boldsymbol{\lambda}(\mathbf{d}),\boldsymbol{\mu}_b(\mathbf{d});\mathbf{d}\right)\\ \nabla_{\mathbf{x}}\mathcal{L}\left(\mathbf{x}(\mathbf{d}),\boldsymbol{\lambda}(\mathbf{d}),\boldsymbol{\mu}_b(\mathbf{d});\mathbf{d}\right)\end{bmatrix}$$

$$= \begin{bmatrix}-\mathbf{g}\left(\mathbf{x}(\mathbf{d});\mathbf{d}\right)\\ -\mathbf{h}^b\left(\mathbf{x}(\mathbf{d});\mathbf{d}\right)\\ \nabla_{\mathbf{x}}f\left(\mathbf{x}(\mathbf{d});\mathbf{d}\right) - \nabla_{\mathbf{x}}\mathbf{g}\left(\mathbf{x}(\mathbf{d});\mathbf{d}\right)^{\mathrm{T}}\boldsymbol{\lambda}(\mathbf{d}) - \nabla_{\mathbf{x}}\mathbf{h}^b\left(\mathbf{x}(\mathbf{d});\mathbf{d}\right)^{\mathrm{T}}\boldsymbol{\mu}_b(\mathbf{d})\end{bmatrix} = \mathbf{0}.$$

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

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

is nonsingular, which is true if the second-order sufficient condition holds, then there is a differentiable function $\begin{bmatrix}\boldsymbol{\lambda}(\mathbf{d})\\ \boldsymbol{\mu}_b(\mathbf{d})\\ \mathbf{x}(\mathbf{d})\end{bmatrix}$ such that

a) $\nabla_{\begin{bmatrix}\boldsymbol{\lambda}\\ \boldsymbol{\mu}_b\\ \mathbf{x}\end{bmatrix}}\mathcal{L}\left(\mathbf{x}(\mathbf{d}),\boldsymbol{\lambda}(\mathbf{d}),\boldsymbol{\mu}_b(\mathbf{d});\mathbf{d}\right) = \mathbf{0}$, for $\left\|\mathbf{d}-\mathbf{d}_0\right\| < \boldsymbol{\varepsilon}$,

b) $\begin{bmatrix}\boldsymbol{\lambda}(\mathbf{d}_0)\\ \boldsymbol{\mu}_b(\mathbf{d}_0)\\ \mathbf{x}(\mathbf{d}_0)\end{bmatrix} = \begin{bmatrix}\boldsymbol{\lambda}^*\\ \boldsymbol{\mu}_b^*\\ \mathbf{x}^*\end{bmatrix}$, and

c) the gradient

$$\nabla_{\mathbf{d}}\begin{bmatrix}\boldsymbol{\lambda}(\mathbf{d}_0)\\ \boldsymbol{\mu}_b(\mathbf{d}_0)\\ \mathbf{x}(\mathbf{d}_0)\end{bmatrix} = -\left[\bar{\mathbf{H}}\left(\mathbf{x}^*,\boldsymbol{\lambda}^*,\boldsymbol{\mu}_b^*;\mathbf{d}_0\right)\right]^{-1}\nabla_{\mathbf{d}}\left[\nabla_{\begin{bmatrix}\boldsymbol{\lambda}\\ \boldsymbol{\mu}_b\\ \mathbf{x}\end{bmatrix}}\mathcal{L}\left(\mathbf{x}^*,\boldsymbol{\lambda}^*,\boldsymbol{\mu}_b^*;\mathbf{d}_0\right)\right]$$

$$= -\begin{bmatrix}\mathbf{0} & \mathbf{0} & -\nabla\mathbf{g}\left(\mathbf{x}^*;\mathbf{d}_0\right)\\ \mathbf{0} & \mathbf{0} & -\nabla\mathbf{h}^b\left(\mathbf{x}^*;\mathbf{d}_0\right)\\ -\nabla\mathbf{g}\left(\mathbf{x}^*;\mathbf{d}_0\right)^{\mathrm{T}} & -\nabla\mathbf{h}^b\left(\mathbf{x}^*;\mathbf{d}_0\right)^{\mathrm{T}} & \nabla_{\mathbf{x}}^2\mathcal{L}\left(\mathbf{x}^*,\boldsymbol{\lambda}^*,\boldsymbol{\mu}_b^*;\mathbf{d}_0\right)\end{bmatrix}^{-1}\begin{bmatrix}-\nabla_{\mathbf{d}}\mathbf{g}\left(\mathbf{x}^*;\mathbf{d}_0\right)\\ -\nabla_{\mathbf{d}}\mathbf{h}^b\left(\mathbf{x}^*;\mathbf{d}_0\right)\\ \nabla_{\mathbf{d}}\left[\nabla_{\mathbf{x}}f\left(\mathbf{x}^*;\mathbf{d}_0\right)-\nabla_{\mathbf{x}}\mathbf{g}\left(\mathbf{x}^*;\mathbf{d}_0\right)^{\mathrm{T}}\boldsymbol{\lambda}^*-\nabla_{\mathbf{x}}\mathbf{h}^b\left(\mathbf{x}^*;\mathbf{d}_0\right)^{\mathrm{T}}\boldsymbol{\mu}_b\left(\mathbf{d}_0\right)\right]\end{bmatrix}$$

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

10.8 Kuhn-Tucker Formulation

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

$$\begin{aligned}&\max f(\mathbf{x})\\ &\text{st}\quad \mathbf{h}(\mathbf{x}) \le \mathbf{b}\\ &\qquad\quad \mathbf{x} \ge \mathbf{0}.\end{aligned}$$

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

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

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

a) $\nabla_{\mathbf{x}}\mathcal{L}\left(\mathbf{x}^*,\boldsymbol{\mu}^*,\boldsymbol{\nu}^*\right) = \nabla f\left(\mathbf{x}^*\right) - \nabla\mathbf{h}\left(\mathbf{x}^*\right)^{\mathrm{T}}\boldsymbol{\mu}^* + \boldsymbol{\nu}^* = \mathbf{0}$,

b) $\mu_i^*\left(h^i\left(\mathbf{x}^*\right) - b_i\right) = 0$, $i = 1,2,\ldots,m$,

c) $\nu_j^* x_j^* = 0$, $j = 1,2,\ldots,n$,

d) $\mathbf{h}\left(\mathbf{x}^*\right) \le \mathbf{b}$,

e) $\mathbf{x}^* \ge \mathbf{0}$,

f) $\boldsymbol{\mu}^* \ge \mathbf{0}$, and

g) $\boldsymbol{\nu}^* \ge \mathbf{0}$.

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

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

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

A) $\nabla_{\mathbf{x}}\bar{\mathcal{L}}(\mathbf{x}^*,\boldsymbol{\mu}^*) = -\boldsymbol{\nu}^*$,

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

C) $x_j^* \dfrac{\partial \bar{\mathcal{L}}}{\partial x_j} = 0,\ j = 1,2,\ldots,n$.

G) $\nabla_{\mathbf{x}}\bar{\mathcal{L}}(\mathbf{x}^*,\boldsymbol{\mu}^*) \le \mathbf{0}$,

For any solution **x**,

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

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

B) $\mu_i^* \dfrac{\partial \bar{\mathcal{L}}}{\partial \mu_i} = 0$,

D) $\nabla_{\boldsymbol{\mu}}\bar{\mathcal{L}}(\mathbf{x}^*,\boldsymbol{\mu}^*) \ge \mathbf{0}$,

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 $\{\mathbf{x} \mid \mathbf{x} \in \boldsymbol{R}^n,\ \mathbf{h}(\mathbf{x}) \le \mathbf{b},\ \mathbf{x} \ge \mathbf{0}\}$, where $f: \boldsymbol{R}^n \to \boldsymbol{R}$ and $\mathbf{h}: \boldsymbol{R}^n \to \boldsymbol{R}^m$ are differentiable functions. Write the Kuhn-Tucker Lagrange function

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

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

a) $\nabla_{\mathbf{x}} \overline{\mathcal{L}}\left(\mathbf{x}^*, \boldsymbol{\mu}^*\right) \leq \mathbf{0}$,

b) $x_j^* \dfrac{\partial \overline{\mathcal{L}}}{\partial x_j} = 0$, $j = 1, 2, \ldots, n$,

c) $\nabla_{\boldsymbol{\mu}} \overline{\mathcal{L}}\left(\mathbf{x}^*, \boldsymbol{\mu}^*\right) \geq \mathbf{0}$, and

d) $\mu_i^* \dfrac{\partial \overline{\mathcal{L}}}{\partial \mu_i} = 0$, $i = 1, 2, \ldots, m$.

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

$$\begin{aligned} \max\ & f(\mathbf{x}) = \mathbf{c}^{\mathrm{T}}\mathbf{x} \\ \text{st}\quad & \mathbf{h}(\mathbf{x}) = \mathbf{A}\mathbf{x} - \mathbf{b} \le \mathbf{0} \\ & \mathbf{x} \ge \mathbf{0}. \end{aligned}$$

The Kuhn-Tucker Lagrange function is

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

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

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

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

the Kuhn-Tucker conditions are

a) $\mathbf{c} - \mathbf{A}^{\mathrm{T}}\boldsymbol{\mu}^* \le \mathbf{0}$,

b) $x_j^*\left(c_j - \sum_{i=1}^{m} a_{ij}\mu_i^*\right) = x_j^*\left(c_j - \mathbf{a}_{\cdot j}^{\mathrm{T}}\boldsymbol{\mu}^*\right) = 0$,
$j = 1,2,\ldots,n$,

c) $-\mathbf{A}\mathbf{x}^* + \mathbf{b} \ge 0$, which is just $\mathbf{A}\mathbf{x}^* \le \mathbf{b}$, and

d) $\mu_i^*\left(\sum_{j=1}^{n} a_{ij}x_j^* - b_i\right) = \mu_i^*\left(\mathbf{a}_{i\cdot}\mathbf{x}^* - b_i\right) = 0$,
$i = 1,2,\ldots,m$.

The variables 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

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

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

a) one greater than or equal constraint,
b) one equality constraint,
c) one variable is restricted to be nonpositive
d) one variable is unrestricted and it can take positive or negative value.