Chapter 7
Multivariable Unconstrained Optimization: Applications

7.1 Competitive Firm Input Choices: Cobb-Douglas Technology

The profit function of a firm with Cobb-Douglas production function in a competitive product and inputs markets is given by

$$\boldsymbol{max}\,\pi = pL^{\alpha}K^{\beta} - wL - rK$$

First-Order Sufficient Condition:

$$\left.\begin{aligned} \frac{\partial \pi}{\partial L} &= \alpha p L^{*\alpha-1}K^{*\beta} - w = 0 \\ \frac{\partial \pi}{\partial K} &= \beta p L^{*\alpha}K^{*\beta-1} - r = 0 \end{aligned}\right\} \Rightarrow \begin{cases} VMP_L = w \\ VMP_K = r, \end{cases}$$

where VMP_L and VMP_K are the *values of marginal product of labor and capital*, respectively.

Second-Order Sufficient Condition: The Hessian at the critical point $\left(L^*, K^*\right)$ is negative definite.

$$\mathbf{H}\left(L^*, K^*\right) = \begin{bmatrix} \alpha(\alpha-1)pL^{*\alpha-2}K^{*\beta} & \alpha\beta pL^{*\alpha-1}K^{*\beta-1} \\ \alpha\beta pL^{*\alpha-1}K^{*\beta-1} & \beta(\beta-1)pL^{*\alpha}K^{*\beta-2} \end{bmatrix}$$

Test the negative definiteness:

$$\left|\mathbf{H}_1\right| = \alpha(\alpha-1)pL^{*\alpha-2}K^{*\beta} < 0, \text{ if } 0 < \alpha < 1$$

$$\begin{aligned} \left|\mathbf{H}_2\right| = \left|\mathbf{H}\right| &= \alpha(\alpha-1)pL^{*\alpha-2}K^{*\beta}\beta(\beta-1)pL^{*\alpha}K^{*\beta-2} - \left(\alpha\beta pL^{*\alpha-1}K^{*\beta-1}\right)^2 \\ &= \alpha(\alpha-1)\beta(\beta-1)p^2L^{*2\alpha-2}K^{*2\beta-2} - \alpha^2\beta^2p^2L^{*2\alpha-2}K^{*2\beta-2} > 0 \\ &\Leftrightarrow (\alpha-1)(\beta-1) - \alpha\beta > 0, \text{ if } \beta > 0 \\ &\Leftrightarrow \alpha\beta - \alpha - \beta + 1 - \alpha\beta = 1 - \alpha - \beta > 0. \end{aligned}$$

- That is, $\alpha > 0$, $\beta > 0$ and $\alpha + \beta < 1$. The production function has to be homogeneous of degree less than 1.
- That means the production function is decreasing returns to scale, and by Microeconomics, the marginal cost is positively sloped.

Comparative Static Analysis: The optimal solution $\left(L^*, K^*\right)$ can be solved as functions of parameters α, β, *w*, *r*, and *p*. That is, assuming α, β unchanged, at a given particular $\left(\overline{w}, \overline{r}, \overline{p}\right)$, the first-order sufficient conditions can be written as an implicit functions of $\left(L^*, K^*\right)$ as follows.

$$\mathbf{f}\left(L^*, K^*; \overline{w}, \overline{r}, \overline{p}\right) = \begin{bmatrix} f^1\left(L^*, K^*; \overline{w}, \overline{r}, \overline{p}\right) \\ f^2\left(L^*, K^*; \overline{w}, \overline{r}, \overline{p}\right) \end{bmatrix} = \begin{bmatrix} \alpha\overline{p}{L^*}^{\alpha-1}{K^*}^{\beta} - \overline{w} \\ \beta\overline{p}{L^*}^{\alpha}{K^*}^{\beta-1} - \overline{r} \end{bmatrix} = \mathbf{0}$$

The Implicit Function Theorem applies here because

$$\nabla_{\left[\begin{smallmatrix} L \\ K \end{smallmatrix}\right]}\mathbf{f}\left(L^*, K^*; \overline{w}, \overline{r}, \overline{p}\right) = \mathbf{H}\left(L^*, K^*; \overline{w}, \overline{r}, \overline{p}\right),$$

is nonsingular because the Hessian is negative definite. We have

a) there are functions $L\left(w, r, p\right)$ and $K\left(w, r, p\right)$ such that

$$\mathbf{f}\left(L\left(w, r, p\right), K\left(w, r, p\right); w, r, p\right) = \mathbf{0},$$

for $\left|w - \overline{w}\right| < \varepsilon$, $\left|r - \overline{r}\right| < \varepsilon$, and $\left|p - \overline{p}\right| < \varepsilon$ for some $\varepsilon > 0$.

b) $L\left(\overline{w}, \overline{r}, \overline{p}\right) = L^*$ and $K\left(\overline{w}, \overline{r}, \overline{p}\right) = K^*$

c) The gradient

$$\nabla_{\left[\begin{smallmatrix} w \\ r \\ p \end{smallmatrix}\right]}\begin{bmatrix} L\left(\overline{w}, \overline{r}, \overline{p}\right) \\ K\left(\overline{w}, \overline{r}, \overline{p}\right) \end{bmatrix} = -\left[\nabla_{\left[\begin{smallmatrix} L \\ K \end{smallmatrix}\right]}\mathbf{f}\left(L^*, K^*; \overline{w}, \overline{r}, \overline{p}\right)\right]^{-1}\nabla_{\left[\begin{smallmatrix} w \\ r \\ p \end{smallmatrix}\right]}\mathbf{f}\left(L^*, K^*; \overline{w}, \overline{r}, \overline{p}\right)$$

$$= -\mathbf{H}\left(L^*, K^*; \overline{w}, \overline{r}, \overline{p}\right)^{-1}\nabla_{\left[\begin{smallmatrix} w \\ r \\ p \end{smallmatrix}\right]}\mathbf{f}\left(L^*, K^*; \overline{w}, \overline{r}, \overline{p}\right)$$

$$= -\begin{bmatrix} \alpha\left(\alpha-1\right)\overline{p}{L^*}^{\alpha-2}{K^*}^{\beta} & \alpha\beta\overline{p}{L^*}^{\alpha-1}{K^*}^{\beta-1} \\ \alpha\beta\overline{p}{L^*}^{\alpha-1}{K^*}^{\beta-1} & \beta\left(\beta-1\right)\overline{p}{L^*}^{\alpha}{K^*}^{\beta-2} \end{bmatrix}^{-1}\begin{bmatrix} -1 & 0 & \alpha{L^*}^{\alpha-1}{K^*}^{\beta} \\ 0 & -1 & \beta{L^*}^{\alpha}{K^*}^{\beta-1} \end{bmatrix}$$

and by Cramer's Rule

$$\frac{\partial L(\bar{w},\bar{r},\bar{p})}{\partial w} = -\frac{\begin{vmatrix} -1 & \alpha\beta\bar{p}{L^*}^{\alpha-1}{K^*}^{\beta-1} \\ 0 & \beta(\beta-1)\bar{p}{L^*}^{\alpha}{K^*}^{\beta-2} \end{vmatrix}}{\begin{vmatrix} \alpha(\alpha-1)\bar{p}{L^*}^{\alpha-2}{K^*}^{\beta} & \alpha\beta\bar{p}{L^*}^{\alpha-1}{K^*}^{\beta-1} \\ \alpha\beta\bar{p}{L^*}^{\alpha-1}{K^*}^{\beta-1} & \beta(\beta-1)\bar{p}{L^*}^{\alpha}{K^*}^{\beta-2} \end{vmatrix}}$$

$$= \frac{\beta(\beta-1)\bar{p}{L^*}^{\alpha}{K^*}^{\beta-2}}{|\mathbf{H}|} < 0.$$

We can similarly find all other partial derivatives.

HW Baldani, p. 216, #8.2, 8.3, 8.4

7.2 Competitive Firm Input Choices: General Production Technology

The same firm as in 7.1 but now with a generic production function $f(L,K)$ will maximize the profit function

$$\max \pi = pf(L,K) - wL - rK.$$

First-Order Sufficient Condition:

$$\left.\begin{aligned} \frac{\partial \pi}{\partial L} &= pf_L - w = 0 \\ \frac{\partial \pi}{\partial K} &= pf_K - r = 0 \end{aligned}\right\} \Rightarrow \begin{cases} VMP_L = w \\ VMP_K = r \end{cases}$$

Second-Order Sufficient Condition:

$$\mathbf{H}(L^*,K^*) = \begin{bmatrix} pf_{LL} & pf_{LK} \\ pf_{KL} & pf_{KK} \end{bmatrix}$$

$$|\mathbf{H}_1| = pf_{LL} < 0$$
$$|\mathbf{H}_2| = |\mathbf{H}| = p^2\left(f_{LL}f_{KK} - f_{KL}^2\right) > 0$$

By Implicit Function Theorem:

$$\nabla_{\begin{bmatrix} w \\ r \\ p \end{bmatrix}} \begin{bmatrix} L(\overline{w},\overline{r},\overline{p}) \\ K(\overline{w},\overline{r},\overline{p}) \end{bmatrix} = -\left[\nabla_{\begin{bmatrix} L \\ K \end{bmatrix}} \mathbf{f}\left(L^*, K^*; \overline{w},\overline{r},\overline{p}\right)\right]^{-1} \nabla_{\begin{bmatrix} w \\ r \\ p \end{bmatrix}} \mathbf{f}\left(L^*, K^*; \overline{w},\overline{r},\overline{p}\right)$$

$$= -\mathbf{H}\left(L^*, K^*; \overline{w},\overline{r},\overline{p}\right)^{-1} \nabla_{\begin{bmatrix} w \\ r \\ p \end{bmatrix}} \mathbf{f}\left(L^*, K^*; \overline{w},\overline{r},\overline{p}\right)$$

$$= -\begin{bmatrix} pf_{LL} & pf_{LK} \\ pf_{KL} & pf_{KK} \end{bmatrix}^{-1} \begin{bmatrix} -1 & 0 & f_L \\ 0 & -1 & f_K \end{bmatrix}$$

$$\frac{\partial L(\overline{w},\overline{r},\overline{p})}{\partial w} = -\frac{\begin{vmatrix} -1 & pf_{LK} \\ 0 & pf_{KK} \end{vmatrix}}{\begin{vmatrix} pf_{LL} & pf_{LK} \\ pf_{KL} & pf_{KK} \end{vmatrix}}$$

$$= \frac{pf_{KK}}{|\mathbf{H}|} < 0.$$

- The demand for labor has a negative slope with respect to wage rate if the capital exhibits diminishing returns-- $f_{KK} < 0$ means the slope of the marginal product is negative.

HW Determine the signs of $\dfrac{\partial L(\overline{w},\overline{r},\overline{p})}{\partial r}$ and $\dfrac{\partial L(\overline{w},\overline{r},\overline{p})}{\partial p}$.

HW Baldani, p. 216, #8.5.

7.3 Multi-plant Firm A firm with n plants.

$TC_i(q_i) = C_i(q_i)$, q_i = quantity produced at plant i.

$$TR = R(Q) = P(Q)Q$$

$$Q = \sum_{i=1}^{n} q_i$$

$$\pi(\mathbf{q}) = R(Q) - \sum_{i=1}^{n} C_i(q_i) = P(Q)Q - \sum_{i=1}^{n} C_i(q_i)$$

$$\pi_i(\mathbf{q}^*) = P'(Q^*)Q^* + P(Q^*) - C_i'(q_i^*) = 0,\ i = 1,2,\ldots,n.$$

The last equality is the first-order sufficient condition and it implies that the marginal cost of each plant is equal to the marginal revenue, i.e.,

$$C_1'\left(q_1^*\right) = C_2'\left(q_2^*\right) = \cdots = C_n'\left(q_n^*\right) = P'\left(Q^*\right)Q^* + P\left(Q^*\right) = R'\left(Q\right) = MR\left(Q\right)$$

The Hessian is given by

$$\mathbf{H}\left(\mathbf{q}^*\right) = \begin{bmatrix} R'' - C_1'' & R'' & \cdots & R'' \\ R'' & R'' - C_2'' & \cdots & \vdots \\ \vdots & \cdots & \ddots & R'' \\ R'' & \cdots & R'' & R'' - C_n'' \end{bmatrix},$$

where $R'' = P''\left(Q^*\right)Q^* + 2P'\left(Q^*\right)$. Test of negative definiteness:

$$\left|\mathbf{H}_1\right| = R'' - C_1'' < 0$$

$$\left|\mathbf{H}_2\right| = \left(R'' - C_1''\right)\left(R'' - C_2''\right) - R''^2 = C_1''C_2'' - R''\left(C_1'' + C_2''\right) > 0, \text{ and}$$

$$\left(-1\right)^i\left|\mathbf{H}_i\right| > 0,\ i = 3,4,\ldots,n.$$

For perfect competition, $R'' = 0$ and

$$\mathbf{H}\left(\mathbf{q}^*\right) = \begin{bmatrix} -C_1'' & 0 & \cdots & 0 \\ 0 & -C_2'' & \cdots & \vdots \\ \vdots & \cdots & \ddots & 0 \\ 0 & \cdots & 0 & -C_n'' \end{bmatrix},$$

which is negative definite if $C_i'' > 0$, $i = 1,2,\ldots,n$.

For monopoly, this is not necessarily true if the plants have increasing return to scale $\left(C_i'' < 0\right)$.

HW Show that the Hessian

$$\mathbf{H}\left(\mathbf{q}^*\right) = \begin{bmatrix} R'' - C_1'' & R'' & \cdots & R'' \\ R'' & R'' - C_2'' & \cdots & \vdots \\ \vdots & \cdots & \ddots & R'' \\ R'' & \cdots & R'' & R'' - C_n'' \end{bmatrix}$$

is not necessarily negative definite if $C_i'' < 0$, $i = 1,2,\ldots,n$. (Hint: Show for $n = 2$)

HW Show that the Hessian $\mathbf{H}\left(\mathbf{q}^*\right)$ above is negative definite if $R'' < 0$ and $C_i'' > 0$, $i = 1,2,\ldots,n$.

Solution: If we assume instead that $R'' > 0$ and $C_i'' > 0$ then
we can equivalently show that

$$\mathbf{H}\left(\mathbf{q}^*\right) = \begin{bmatrix} R'' + C_1'' & R'' & \cdots & R'' \\ R'' & R'' + C_2'' & \cdots & \vdots \\ \vdots & \cdots & \ddots & R'' \\ R'' & \cdots & R'' & R'' + C_n'' \end{bmatrix}$$

is positive definite. We will prove by induction. If $n = 2$, then $\left|\mathbf{H}_1\right| = R'' + C_1'' > 0$ and

$$\left|\mathbf{H}_2\right| = \left(R'' + C_1''\right)\left(R'' + C_2''\right) - R''^2 = C_1''C_2'' + R''\left(C_1'' + C_2''\right) > 0 .$$

Note that $\left|\mathbf{H}_2\right| > 0$ even when $C_1'' = 0$. We can then state the induction hypothesis that $\left|\mathbf{H}\left(\mathbf{q}^*\right)_{n\times n}\right| > 0$ when $R'' > 0$, $C_1'' \geq 0$ and $C_i'' > 0$, $i = 2,3,\ldots,n$. We need to prove

$$\left|\mathbf{H}\left(\mathbf{q}^*\right)_{(n+1)\times(n+1)}\right| = \begin{vmatrix} R'' + C_1'' & R'' & \cdots & R'' \\ R'' & R'' + C_2'' & \cdots & \vdots \\ \vdots & \cdots & \ddots & R'' \\ R'' & \cdots & R'' & R'' + C_{n+1}'' \end{vmatrix} > 0 .$$

By a property of determinant,

$$\left|\mathbf{H}\left(\mathbf{q}^*\right)_{(n+1)\times(n+1)}\right| = \begin{vmatrix} C_1'' & R'' & \cdots & \cdots & R'' \\ -C_2'' & R''+C_2'' & \cdots & \cdots & \vdots \\ 0 & R'' & \ddots & \cdots & \vdots \\ \vdots & \cdots & \cdots & R''+C_{n+1}'' & R'' \\ 0 & R'' & \cdots & R'' & R''+C_{n+1}'' \end{vmatrix}$$

$$= C_1'' \begin{vmatrix} R''+C_2'' & R'' & \cdots & R'' \\ R'' & R''+C_3'' & \cdots & \vdots \\ \vdots & \cdots & \ddots & R'' \\ R'' & \cdots & R'' & R''+C_{n+1}'' \end{vmatrix}$$

$$+(-1)^{2+1}\left(-C_2''\right) \begin{vmatrix} R'' & R'' & \cdots & R'' \\ R'' & R''+C_3'' & \cdots & \vdots \\ \vdots & \cdots & \ddots & R'' \\ R'' & \cdots & R'' & R''+C_{n+1}'' \end{vmatrix},$$

which is positive by the induction hypothesis.

- If we assume the functional forms of TR and TC explicitly with parameters, we can perform the sensitivity analysis using the Implicit Function Theorem.

HW Baldani, p. 217, #8.9

7.4 Multi-Market Monopoly

A monopoly has two separate markets with similar demands $D_1(q_1) = \alpha P(q_1)$ and $D_2(q_2) = P(q_2)$, where q_1 and q_2 are quantities sold in the two markets. Assume that $\alpha > 1$ so that market 1 is more important. The total revenues earned are:

$$R_1(q_1) = \alpha P(q_1) q_1$$
$$R_2(q_2) = P(q_2) q_2.$$

The total cost of output $Q = q_1 + q_2$ is

$$TC(Q) = C(Q) + tq_2,$$

where \$$t$ is the extra cost per unit to sell in market 2. The profit function is thus

$$\begin{aligned}\pi(q_1,q_2) &= R_1(q_1)+R_2(q_2)-C(Q)-tq_2\\ &= \alpha P(q_1)q_1+P(q_2)q_2-C(Q)-tq_2.\end{aligned}$$

First-order Sufficient Condition:

$$\begin{aligned}\pi_1(q_1^*,q_2^*) &= R_1'(q_1^*)-C'(q_1^*+q_2^*)\\ &= \alpha P'(q_1^*)q_1^*+\alpha P(q_1^*)-C'(q_1^*+q_2^*)=0\\ \pi_2(q_1^*,q_2^*) &= R_2'(q_2^*)-C'(q_1^*+q_2^*)-t\\ &= P'(q_2^*)q_2^*+P(q_2^*)-C'(q_1^*+q_2^*)-t=0.\end{aligned}$$

We have $q_1^*>q_2^*$. (Why?)

Second-order Sufficient Condition: The Hessian is given by

$$\mathbf{H}(q_1^*,q_2^*)=\begin{bmatrix}\alpha R_1''-C'' & -C''\\ -C'' & R_2''-C''\end{bmatrix},$$

where

$$\begin{aligned}\alpha R_1'' &= \alpha P''(q_1^*)q_1^*+2\alpha P'(q_1^*)\\ R_2'' &= P''(q_2^*)q_2^*+2P'(q_2^*).\end{aligned}$$

Test of negative definiteness of the Hessian: If $R_1''<0$, $R_2''<0$ and $C''>0$

$$\begin{aligned}\left|\mathbf{H}_1(q_1^*,q_2^*)\right| &= \alpha R_1''-C''<0\\ \left|\mathbf{H}(q_1^*,q_2^*)\right| &= (\alpha R_1''-C'')(R_2''-C'')-C''^2=\alpha R_1''R_2''-\alpha R_1''C''-R_2''C''>0.\end{aligned}$$

Comparative Static Analysis: Write the first-order condition as implicit functions.

$$\nabla_{\mathbf{q}}\pi(q_1^*,q_2^*;\alpha,t)=\begin{bmatrix}\pi_1(q_1^*,q_2^*;\alpha,t)\\ \pi_2(q_1^*,q_2^*;\alpha,t)\end{bmatrix}=\begin{bmatrix}\alpha R_1'(q_1^*)-C'(q_1^*+q_2^*)\\ R_2'(q_2^*)-C'(q_1^*+q_2^*)-t\end{bmatrix}=\mathbf{0}$$

Assuming $\nabla^2_{\mathbf{q}}\pi\left(q_1^*,q_2^*;\alpha,t\right)=\mathbf{H}\left(q_1^*,q_2^*;\alpha,t\right)$ being a nonsingular matrix, the Implicit Function Theorem yields

$$\begin{aligned}\nabla_{\left[\begin{smallmatrix}\alpha\\ t\end{smallmatrix}\right]}\mathbf{q}^* &= -\nabla^2_{\mathbf{q}}\pi\left(q_1^*,q_2^*;\alpha,t\right)^{-1}\nabla_{\left[\begin{smallmatrix}\alpha\\ t\end{smallmatrix}\right]}\left(\nabla_{\mathbf{q}}\pi\left(q_1^*,q_2^*;\alpha,t\right)\right)\\ &= -\mathbf{H}\left(q_1^*,q_2^*\right)^{-1}\begin{bmatrix}R_1'\left(q_1^*\right) & 0\\ 0 & -1\end{bmatrix}\\ &= -\begin{bmatrix}\alpha R_1''-C'' & -C''\\ -C'' & R_2''-C''\end{bmatrix}^{-1}\begin{bmatrix}R_1'\left(q_1^*\right) & 0\\ 0 & -1\end{bmatrix}.\end{aligned}$$

By Cramer's Rule, if $R_1'\left(q_1^*\right)>0$, $C''>0$

$$\frac{\partial q_1^*}{\partial\alpha}=-\frac{\begin{vmatrix}R_1'\left(q_1^*\right) & -C''\\ 0 & R_2''-C''\end{vmatrix}}{\left|\mathbf{H}\left(q_1^*,q_2^*\right)\right|}=-\frac{R_1'\left(q_1^*\right)\left(R_2''-C''\right)}{\left|\mathbf{H}\left(q_1^*,q_2^*\right)\right|}>0,$$

$$\frac{\partial q_1^*}{\partial t}=-\frac{\begin{vmatrix}0 & -C''\\ -1 & R_2''-C''\end{vmatrix}}{\left|\mathbf{H}\left(q_1^*,q_2^*\right)\right|}=\frac{C''}{\left|\mathbf{H}\left(q_1^*,q_2^*\right)\right|}>0.$$

Thus,

$$\begin{aligned}dq_1^* &= \frac{\partial q_1^*}{\partial\alpha}d\alpha+\frac{\partial q_1^*}{\partial t}dt\\ &= \frac{-R_1'\left(q_1^*\right)\left(R_2''-C''\right)d\alpha+C''dt}{\left|\mathbf{H}\left(q_1^*,q_2^*\right)\right|},\end{aligned}$$

and similarly,

$$\begin{aligned}dq_2^* &= \frac{\partial q_2^*}{\partial\alpha}d\alpha+\frac{\partial q_2^*}{\partial t}dt\\ &= \frac{-R_1'\left(q_1^*\right)C''d\alpha+\left(\alpha R_1''-C''\right)dt}{\left|\mathbf{H}\left(q_1^*,q_2^*\right)\right|}.\end{aligned}$$

We have

$$dQ^* = dq_1^* + dq_2^* = \frac{-R_1'\left(q_1^*\right)R_2''d\alpha + \alpha R_1''dt}{\left|\mathbf{H}\left(q_1^*, q_2^*\right)\right|}.$$

HW Baldani, p. 217, #8.10, 8.11

7.5 Statistical Estimation: Linear Regression

Recall matrix differentiation,

$$\nabla \mathbf{c}^{\mathrm{T}}\mathbf{x} = \mathbf{c}$$
$$\nabla \alpha\mathbf{x}^{\mathrm{T}}\mathbf{x} = 2\alpha\mathbf{x}$$
$$\nabla \mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x} = 2\mathbf{A}\mathbf{x},$$

where **A** is a symmetric matrix.

Linear regression model: *Least Squares* The dependent variable y is determined linearly by the independent variables x_j, $j = 1,2,\ldots,k$, with some random error ε. Suppose there are n observation, we have for $i = 1,2,\ldots,n$,

$$y_i = \beta_0 + \beta_1 x_i^1 + \beta_2 x_i^2 + \cdots + \beta_k x_i^k + \varepsilon_i,$$

where β_j, $j = 0,1,2,\ldots,k$, are unknown parameters. In matrix form,

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon,$$

where $\mathbf{y}, \varepsilon \in \boldsymbol{R}^n$, $\boldsymbol{\beta} \in \boldsymbol{R}^{k+1}$, and $\mathbf{X} \in \boldsymbol{R}^{n\times(k+1)}$. If we estimate $\boldsymbol{\beta}$ by some $\mathbf{b} \in \boldsymbol{R}^{k+1}$, the estimate of **y** is thus , and the error of estimation of **y** is $\hat{\mathbf{y}} = \mathbf{X}\mathbf{b}$. The determination of **b** by the Least Squares Method is the choice of **b** such that the sum of squares of the error of estimation $SSR = \mathbf{e}^{\mathrm{T}}\mathbf{e} = \sum_{i=1}^{n} e_i^2$ is minimized. The minimization is thus given by

$$\min_{\mathbf{b}\in \mathrm{R}^{k+1}} f(\mathbf{b}) = \mathbf{e}^{\mathrm{T}}\mathbf{e} = (\mathbf{y}-\hat{\mathbf{y}})^{\mathrm{T}}(\mathbf{y}-\hat{\mathbf{y}})$$
$$= (\mathbf{y}-\mathbf{X}\mathbf{b})^{\mathrm{T}}(\mathbf{y}-\mathbf{X}\mathbf{b})$$
$$= \mathbf{y}^{\mathrm{T}}\mathbf{y} - 2\mathbf{y}^{\mathrm{T}}\mathbf{X}\mathbf{b} + \mathbf{b}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{X}\mathbf{b}.$$

By the first-order sufficient condition, the critical solution is the solution that

$$\nabla f(\hat{\mathbf{b}}) = -2\mathbf{X}^{\mathrm{T}}\mathbf{y} + 2\mathbf{X}^{\mathrm{T}}\mathbf{X}\hat{\mathbf{b}} = \mathbf{0} \ ,$$

and thus the critical solution is given by

$$\hat{\mathbf{b}} = \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y} \ ,$$

and the second-order sufficient condition requires that

$$\nabla^2 f(\mathbf{b}) = 2\mathbf{X}^{\mathrm{T}}\mathbf{X} \, ,$$

be positive definite. The square symmetric matrix $\mathbf{X}^{\mathrm{T}}\mathbf{X}$ is always positive semidefinite. Why? Now, if it is also positive definite, the solution , as given by the first-order sufficient condition is uniquely defined because $\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}$ exists (Why?), is a strict local minimum point. Can we say that is a strict global minimum?

HW What will happen if one of the independent variable is just a linear combination of some of the other independent variable? For example, what if $x^1 = 1.5x^2$?