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

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

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 (L^*, K^*) is negative definite.

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

Test the negative definiteness:

$$\begin{aligned} |\mathbf{H}_1| &= \alpha(\alpha-1)pL^{*\alpha-2}K^{*\beta} < 0, \\ |\mathbf{H}_2| = |\mathbf{H}| &= \alpha(\alpha-1)pL^{*\alpha-2}K^{*\beta}\beta(\beta-1)pL^{*\alpha}K^{*\beta-2} \\ &\qquad - \left(\alpha\beta pL^{*\alpha-1}K^{*\beta-1}\right)^2 \\ &= \alpha(\alpha-1)\beta(\beta-1)p^2L^{*2\alpha-2}K^{*2\beta-2} \\ &\qquad - \alpha^2\beta^2p^2L^{*2\alpha-2}K^{*2\beta-2} > 0, \\ &\Leftrightarrow (\alpha-1)(\beta-1) - \alpha\beta > 0, \\ &\Leftrightarrow \alpha\beta - \alpha - \beta + 1 - \alpha\beta = 1 - \alpha - \beta > 0. \end{aligned}$$

- 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 (L^*, K^*) can be solved as functions of parameters α, β, w, r, and p. That is, assuming α and β are unchanged, at a given particular (w_0, r_0, p_0) the first-order sufficient conditions can be written as an implicit functions of (L^*, K^*) as follows.

$$\mathbf{f}(L^*, K^*; w_0, r_0, p_0) = \begin{bmatrix} f^1(L^*, K^*; w_0, r_0, p_0) \\ f^2(L^*, K^*; w_0, r_0, p_0) \end{bmatrix} = \begin{bmatrix} \alpha p_0 L^{*\alpha_0 - 1} K^{*\beta_0} - w_0 \\ \beta p_0 L^{*\alpha_0} K^{*\beta_0 - 1} - r_0 \end{bmatrix} = \mathbf{0}.$$

The Implicit Function Theorem applies here because

$$\nabla_{\begin{bmatrix} L \\ K \end{bmatrix}} \mathbf{f}(L^*, K^*; w_0, r_0, p_0) = \mathbf{H}(L^*, K^*; w_0, r_0, p_0),$$

is nonsingular because the Hessian is negative definite. We have

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

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

for $|w - w_0| < \varepsilon$, $|r - r_0| < \varepsilon$, and $|p - p_0| < \varepsilon$ for some $\varepsilon > 0$.

b) $L(w_0, r_0, p_0) = L^*$ and $K(w_0, r_0, p_0) = K^*$

c) The gradient

$$\nabla_{\begin{bmatrix} w \\ r \\ p \end{bmatrix}} \begin{bmatrix} L(w_0, r_0, p_0) \\ K(w_0, r_0, p_0) \end{bmatrix} = -\left[\nabla_{\begin{bmatrix} L \\ K \end{bmatrix}} \mathbf{f}(L^*, K^*; w_0, r_0, p_0)\right]^{-1} \nabla_{\begin{bmatrix} w \\ r \\ p \end{bmatrix}} \mathbf{f}(L^*, K^*; w_0, r_0, p_0)$$
$$= -\mathbf{H}(L^*, K^*; w_0, r_0, p_0)^{-1} \nabla_{\begin{bmatrix} w \\ r \\ p \end{bmatrix}} \mathbf{f}(L^*, K^*; w_0, r_0, p_0)$$
$$= -\begin{bmatrix} \alpha_0(\alpha_0 - 1) p_0 L^{*\alpha_0 - 2} K^{*\beta_0} & \alpha_0 \beta_0 p_0 L^{*\alpha_0 - 1} K^{*\beta_0 - 1} \\ \alpha_0 \beta_0 p_0 L^{*\alpha_0 - 1} K^{*\beta_0 - 1} & \beta_0(\beta_0 - 1) p_0 L^{*\alpha_0} K^{*\beta_0 - 2} \end{bmatrix}^{-1} \begin{bmatrix} -1 & 0 & \alpha_0 L^{*\alpha_0 - 1} K^{*\beta_0} \\ 0 & -1 & \beta_0 L^{*\alpha_0} K^{*\beta_0 - 1} \end{bmatrix}$$

and by Cramer's Rule

$$\frac{\partial L(w_0, r_0, p_0)}{\partial w} = -\frac{\begin{vmatrix} -1 & \alpha_0 \beta_0 p_0 L^{*\alpha_0 - 1} K^{*\beta_0 - 1} \\ 0 & \beta_0(\beta_0 - 1) p_0 L^{*\alpha_0} K^{*\beta_0 - 2} \end{vmatrix}}{\begin{vmatrix} \alpha_0(\alpha_0 - 1) p_0 L^{*\alpha_0 - 2} K^{*\beta_0} & \alpha_0 \beta_0 p_0 L^{*\alpha_0 - 1} K^{*\beta_0 - 1} \\ \alpha_0 \beta_0 p_0 L^{*\alpha_0 - 1} K^{*\beta_0 - 1} & \beta_0(\beta_0 - 1) p_0 L^{*\alpha_0} K^{*\beta_0 - 2} \end{vmatrix}}$$
$$= \frac{\beta_0(\beta_0 - 1) p_0 L^{*\alpha_0} K^{*\beta_0 - 2}}{|\mathbf{H}|} < 0.$$

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:

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

Second-Order Sufficient Condition:

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

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

By Implicit Function Theorem

$$\nabla_{\begin{bmatrix} w \\ r \\ p \end{bmatrix}} \begin{bmatrix} L(w_0,r_0,p_0) \\ K(w_0,r_0,p_0) \end{bmatrix}$$
$$= -\left[\nabla_{\begin{bmatrix} L \\ K \end{bmatrix}} \mathbf{f}(L^*,K^*;w_0,r_0,p_0)\right]^{-1} \nabla_{\begin{bmatrix} w \\ r \\ p \end{bmatrix}} \mathbf{f}(L^*,K^*;w_0,r_0,p_0)$$
$$= -\mathbf{H}(L^*,K^*;w_0,r_0,p_0)^{-1} \nabla_{\begin{bmatrix} w \\ r \\ p \end{bmatrix}} \mathbf{f}(L^*,K^*;w_0,r_0,p_0)$$
$$= -\begin{bmatrix} p_0 f_{LL} & p_0 f_{LK} \\ p_0 f_{KL} & p_0 f_{KK} \end{bmatrix}^{-1} \begin{bmatrix} -1 & 0 & f_L \\ 0 & -1 & f_K \end{bmatrix}$$

and by Cramer's Rule

$$\frac{\partial L(w_0,r_0,p_0)}{\partial w} = -\frac{\begin{vmatrix} -1 & p_0 f_{LK} \\ 0 & p_0 f_{KK} \end{vmatrix}}{\begin{vmatrix} p_0 f_{LL} & p_0 f_{LK} \\ p_0 f_{KL} & p_0 f_{KK} \end{vmatrix}}$$
$$= \frac{p_0 f_{KK}}{|\mathbf{H}|} < 0.$$

- 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 $\frac{\partial L(w_0,r_0,p_0)}{\partial r}$ and $\frac{\partial L(w_0,r_0,p_0)}{\partial p}$.

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.

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

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.,

$$\begin{aligned} C_1'(q_1^*) = C_2'(q_2^*) = \cdots = C_n'(q_n^*) \\ = P'(Q^*)Q^* + P(Q^*) \\ = R'(Q^*) = MR(Q^*) \end{aligned}$$

The Hessian is given by

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

where $R'' = P''(Q^*)Q^* + 2P'(Q^*)$. The Hessian is negative definite if,

$$\begin{aligned} |\mathbf{H}_1| &= R'' - C_1'' < 0 \\ |\mathbf{H}_2| &= (R'' - C_1'')(R'' - C_2'') - R''^2 \\ &= C_1''C_2'' - R''(C_1'' + C_2'') > 0 \\ (-1)^i|\mathbf{H}_i| &> 0, i = 3,4, \dots, n. \end{aligned}$$

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

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

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

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

HW Show that the Hessian \

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

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

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

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

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

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

$$\begin{aligned} |\mathbf{H}_2| &= (R'' + C_1'')(R'' + C_2'') - R''^2 \\ &= C_1''C_2'' + R''(C_1'' + C_2'') > 0. \end{aligned}$$

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

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

By a property of determinant,

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

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

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

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

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

where

$$R_1'' = \alpha P''(q_1^*)q_1^* + 2\alpha P'(q_1^*)$$
$$R_2'' = P''(q_2^*)q_2^* + 2P'(q_2^*).$$

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

$$|\mathbf{H}_1(q_1^*, q_2^*)| = R_1'' - C'' < 0$$
$$|\mathbf{H}(q_1^*, q_2^*)| = (R_1'' - C'')(R_2'' - C'') - C''^2$$
$$= R_1''R_2'' - R_1''C'' - R_2''C'' > 0.$$

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

$$\nabla_{\mathbf{q}}\pi(q_1^*, q_2^*; \alpha_0, t_0) = \begin{bmatrix} \pi_1(q_1^*, q_2^*; \alpha_0, t_0) \\ \pi_2(q_1^*, q_2^*; \alpha_0, t_0) \end{bmatrix}$$
$$= \begin{bmatrix} \alpha_0 P'(q_1^*)q_1^* + \alpha_0 P(q_1^*) - C'(q_1^* + q_2^*) \\ P'(q_2^*)q_2^* + \alpha_0 P(q_2^*) - C'(q_1^* + q_2^*) - t_0 \end{bmatrix} = \mathbf{0}$$

Assuming $\nabla_{\mathbf{q}}^2\pi(q_1^*, q_2^*; \alpha_0, t_0) = \mathbf{H}(q_1^*, q_2^*; \alpha_0, t_0)$ being a nonsingular matrix, the Implicit Function Theorem yields

$$\nabla_{\begin{bmatrix}\alpha \\ t\end{bmatrix}}\mathbf{q}(\alpha_0, t_0)$$
$$= -\left[\nabla_{\mathbf{q}}^2\pi(q_1^*, q_2^*; \alpha_0, t_0)\right]^{-1}\nabla_{\begin{bmatrix}\alpha \\ t\end{bmatrix}}\left(\nabla_{\mathbf{q}}\pi(q_1^*, q_2^*; \alpha_0, t_0)\right)$$
$$= -\mathbf{H}(q_1^*, q_2^*; \alpha_0, t_0)^{-1}\begin{bmatrix} P'(q_1^*)q_1^* + P(q_1^*) & 0 \\ 0 & -1 \end{bmatrix}$$
$$= -\begin{bmatrix} R_1'' - C'' & -C'' \\ -C'' & R_2'' - C'' \end{bmatrix}^{-1}\begin{bmatrix} R_1'/\alpha & 0 \\ 0 & -1 \end{bmatrix}$$

By Cramer's Rule, if $P'(q_1^*)q_1^* + P(q_1^*) > 0$ and $C'' > 0$

$$\frac{\partial q_1^*}{\partial \alpha} = -\frac{\begin{vmatrix} \frac{R_1'}{\alpha} & -C'' \\ 0 & R_2'' - C'' \end{vmatrix}}{|\mathbf{H}(q_1^*, q_2^*; \alpha_0, t_0)|} = -\frac{\frac{R_1'}{\alpha}(R_2'' - C'')}{|\mathbf{H}(q_1^*, q_2^*; \alpha_0, t_0)|} > 0,$$
$$\frac{\partial q_1^*}{\partial t} = -\frac{\begin{vmatrix} 0 & -C'' \\ -1 & R_2'' - C'' \end{vmatrix}}{|\mathbf{H}(q_1^*, q_2^*; \alpha_0, t_0)|} = \frac{C''}{|\mathbf{H}(q_1^*, q_2^*; \alpha_0, t_0)|} > 0$$

Thus,

$$dq_1^* = \frac{\partial q_1^*}{\partial \alpha}d\alpha + \frac{\partial q_1^*}{\partial t}dt$$

$$= \frac{-\frac{R_1'}{\alpha}(R_2'' - C'')d\alpha + C''dt}{|\mathbf{H}(q_1^*, q_2^*; \alpha_0, t_0)|}.$$

and similarly,

$$dq_2^* = \frac{\partial q_2^*}{\partial \alpha} d\alpha + \frac{\partial q_2^*}{\partial t} dt$$
$$= \frac{\frac{R_1'}{\alpha} C'' d\alpha + (R_1'' - C'')dt}{|\mathbf{H}(q_1^*, q_2^*; \alpha_0, t_0)|}.$$

We have

$$dQ^* = dq_1^* + dq_2^* = \frac{-\frac{R_1'}{\alpha} R_2'' + R_1'' dt}{|\mathbf{H}(q_1^*, q_2^*; \alpha_0, t_0)|}.$$

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 $x_j, j = 1,2,\dots,k,$ with some random error ε. Suppose there are n observations, we have for $i = 1,2,\dots,n,$

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

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

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

where $\mathbf{y}, \boldsymbol{\varepsilon} \in \mathbb{R}^n, \boldsymbol{\beta} \in \mathbb{R}^{k+1}$, and $\mathbf{X} \in \mathbb{R}^{n\times(k+1)}$. If we estimate $\boldsymbol{\beta}$ by some $\mathbf{b} \in \mathbb{R}^{k+1}$, the estimate of $\mathbf{y}$ is thus $\hat{\mathbf{y}} = \mathbf{X}\mathbf{b}$, and the error of estimation of $\mathbf{y}$ is $\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}}$. The determination of $\mathbf{b}$ by the Least Squares Method is the choice of $\mathbf{b}$ 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 problem is thus given by

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

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

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

and the second-order sufficient condition requires that

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

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 $\hat{\mathbf{b}} = (\mathbf{X}^{\mathrm{T}}\,\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y}$, as given by the first-order sufficient condition is uniquely defined because $(\mathbf{X}^{\mathrm{T}}\,\mathbf{X})^{-1}$ 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 $x^1 = 1.5x^2$?