Unconstrained Optimization

Ch.17

econMath.namun+2016su@gmail.com

July 15, 2016

1 Definitions

Definitions

Definition 1 ((strict) max/min, (strict) local max/min). *Let* $f : U \in \mathbb{R}^n \to \mathbb{R}$

1. *A point* $\mathbf{x}^*$ *is a* *(global, or absolute) max, maximizer, maximum point* *of* f *on* U *if* $f(\mathbf{x}^*) \geq f(\mathbf{x}) \quad \forall \mathbf{x} \in U$
2. $\mathbf{x}^* \in U$ *is a* *strict (global, or absolute) max* *if* $\mathbf{x}^*$ *is a max and* $f(\mathbf{x}^*) > f(\mathbf{x}) \quad \forall \mathbf{x} \in U - \{\mathbf{x}^*\}$
3. $\mathbf{x}^* \in U$ *is a* *local (relative) max* *of* f *if* $\exists \epsilon > 0 \quad s.t.\ f(\mathbf{x}^*) \geq f(\mathbf{x}) \quad \forall \mathbf{x} \in B_\epsilon(\mathbf{x}^*) \cap U$
4. $\mathbf{x}^* \in U$ *is a* *strict local (relative) max* *of* f *if* $\exists \epsilon > 0 \quad s.t.\ f(\mathbf{x}^*) > f(\mathbf{x}) \quad \forall \mathbf{x} \in B_\epsilon(\mathbf{x}^*) \cap U$

- Definition of min: $>, \geq \to <, \leq$

2 First Order Conditions

FOC

Theorem 1 (17.1). *Let* $f : U \in \mathbb{R}^n \to \mathbb{R}$ *be a* C^1 *function. If* $\mathbf{x}^*$ *is a local max or min of* f *and* $\mathbf{x}^*$ *is an interior point of* U*, then*

$$\frac{\partial f}{\partial x_i}(\mathbf{x}^*) = 0 \quad \forall i$$

In short,

$$Df_{\mathbf{x}}(\mathbf{x}^*) = \mathbf{0}$$

$\mathbf{x}^*$ *is a critical point of* f

Note: Compare with one-var version FOC (Theorem 3.3)

Theorem 2 (3.3: First Order Condition (FOC)). x_0 *is an interior max or min of* $f \Rightarrow x_0$ *is a critical point of* f*. i.e.,* $f'(x_0) = 0$ *(Inverse is not always true)*

3 Second Order Conditions

SOC (Sufficient Conditions)

Theorem 3 (17.2)**.** *Let $f : U \in \mathbb{R}^n \to \mathbb{R}$ be a C^2 function and U is open. Suppose $\mathbf{x}^*$ is a critical point of f. (i.e., $Df_{\mathbf{x}}(\mathbf{x}^*) = \mathbf{0}$) Then,*

1. *If Hessian ($D^2 f_{\mathbf{x}}(\mathbf{x}^*)$) is ND, then $\mathbf{x}^*$ is a strict local max of f*
2. *If Hessian ($D^2 f_{\mathbf{x}}(\mathbf{x}^*)$) is PD, then $\mathbf{x}^*$ is a strict local min of f*
3. *If Hessian is ID, $\mathbf{x}^*$ is neither a local max nor local min of f. (saddle point)*

Note: one-var version: (Theorem 3.4)

$$f'(x^*) = 0 \quad \wedge \quad f'' < 0 \quad \Rightarrow \quad x^* \text{ is a local max}$$

SOC (Necessary Conditions)

Theorem 4 (17.6)**.** *Let $f : U \in \mathbb{R}^n \to \mathbb{R}$ be a C^2 function and U is open. Then,*

1. *$\mathbf{x}^*$ is a local min of $f \Rightarrow Df(\mathbf{x}^*) = \mathbf{0} \quad \wedge \quad D^2 f(\mathbf{x}^*)$ is PSD*
2. *$\mathbf{x}^*$ is a local max of $f \Rightarrow Df(\mathbf{x}^*) = \mathbf{0} \quad \wedge \quad D^2 f(\mathbf{x}^*)$ is NSD*

Note: one-var version:

$$x^* \text{ is local max} \quad \Rightarrow \quad x' = 0 \quad \wedge \quad f'' \leq 0$$

4 Global Maxima and Minima

Finding Global Max/Min

Different from one-var function, condition 1 (below) is not true when f is multi-var function

Sufficient Conditions for Global Max/Min ($f : I \in \mathbb{R} \to \mathbb{R}$)

1. x^* is a local max/min and x^* is the only critical point of f in I
2. $f'' \leq 0 \quad \forall I$. *i.e.*, f is concave on I (max)
   - $f'' \leq 0 \quad \forall I$ (max)
   - $f'' \geq 0 \quad \forall I$ (min)

However, condition 2 is true even when f is multi-var function!

Theorem 5 (17.8)**.** *Let $f : U \in \mathbb{R}^n \to \mathbb{R}$ be a C^2 function with convex open domain U.*

1. *$DF(\mathbf{x}^*) = \mathbf{0}$ and $D^2 f_{\mathbf{x}}$ is PSD on $U \Rightarrow \mathbf{x}^*$ is a global max of f on U*
2. *$DF(\mathbf{x}^*) = \mathbf{0}$ and $D^2 f_{\mathbf{x}}$ is NSD on $U \Rightarrow \mathbf{x}^*$ is a global min of f on U*

5 Economic Applications

Ordinary Least Squares (OLS)

OLS

Find $y = \mathbf{x}\beta + c$ for given N data $X = \begin{pmatrix} \mathbf{x_1} \\ \vdots \\ \mathbf{x_N} \end{pmatrix}$, $Y = \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix}$ satisfying:

$$\arg\min_{\beta,c} \sum_i^N ((\mathbf{x_i}\beta + c) - y_i)^2 \tag{Least Square}$$

$$y = \mathbf{x}\beta + c = x_1\beta_1 + \cdots + x_m\beta_m + c$$

Note: given points – $(y_1, \mathbf{x_1}) = (y_1, x_{11}, x_{12}, \cdots, x_{1m})$, $(y_2, \mathbf{x_2}), \cdots, (y_N, \mathbf{x_N})$ – are not variables. Our object is to find β^*, c^* (linear equation) from given data X, Y.

OLS: Solution

Let our object function $f(\beta_1, \cdots, \beta_n, c) := \sum_i^N ((\mathbf{x_i}\beta + c) - y_i)^2$. Then FOC is:

$$Df_{\beta,c}(\beta^*, c^*) = \mathbf{0} \tag{FOC}$$

This leads to $m+1$ equations:

$$\frac{\partial f}{\partial \beta_1}(\beta^*, c^*) = 2(\mathbf{x_1}\beta^* + c^* - y_1)x_{11} + 2(\mathbf{x_2}\beta^* + c^* - y_2)x_{21} + \cdots + 2(\mathbf{x_N}\beta^* + c^* - y_N)x_{N1} = 0$$

$$\cdots$$

$$\frac{\partial f}{\partial \beta_m}(\beta^*, c^*) = 2(\mathbf{x_1}\beta^* + c^* - y_1)x_{1m} + 2(\mathbf{x_2}\beta^* + c^* - y_2)x_{2m} + \cdots + 2(\mathbf{x_N}\beta^* + c^* - y_N)x_{Nm} = 0$$

$$\Rightarrow 2(\mathbf{x_1}\beta^* + c^* - y_1)\mathbf{x_1}^T + 2(\mathbf{x_2}\beta^* + c^* - y_2)\mathbf{x_2}^T + \cdots + 2(\mathbf{x_N}\beta^* + c^* - y_N)\mathbf{x}_N^T = \mathbf{0}_{m\times 1} \tag{B}$$

$$\frac{\partial f}{\partial c}(\beta^*, c^*) = 2(\mathbf{x_1}\beta^* + c^* - y_1)1 + 2(\mathbf{x_2}\beta^* + c^* - y_2)1 + \cdots + 2(\mathbf{x_N}\beta^* + c^* - y_N)1 = 0 \tag{C}$$

OLS (2)

$$\textit{Remember} \quad X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{Nm} \end{pmatrix} = \begin{pmatrix} C_1 & \cdots & C_m \end{pmatrix} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_N \end{pmatrix}$$

Rearrange FOCs:

$$\begin{pmatrix} \mathbf{x_1}^T & \mathbf{x_2}^T & \cdots & \mathbf{x_N}^T \end{pmatrix} \begin{pmatrix} \mathbf{x_1}\beta^* + c^* - y_1 \\ \vdots \\ \mathbf{x_N}\beta^* + c^* - y_N \end{pmatrix} = X^T(X\beta^* + \mathbf{1}_{N\times 1}c^* - Y) = \mathbf{0}_{m\times 1} \tag{B2}$$

$$c^* = \frac{1}{N}\begin{pmatrix} 1 & 1 & \cdots & 1 \end{pmatrix} \begin{pmatrix} y_1 - \mathbf{x_1}\beta^* \\ \vdots \\ y_N - \mathbf{x_N}\beta^* \end{pmatrix} = \frac{1}{N}\mathbf{1}_{1\times N}(Y - X\beta^*) \tag{C2}$$

OLS (3)

From (C2) and (B2),

$$X^T\left(X\beta^* + \mathbf{1}_{N\times 1}\frac{1}{N}\mathbf{1}_{1\times N}(Y - X\beta^*) - Y\right) = \mathbf{0}_{m\times 1} \tag{D}$$

Rearrange (D) with regard to β^* yields:

$$X^T(X - \frac{1}{N}\mathbf{1}_{N\times 1}\mathbf{1}_{1\times N}X)\beta^* = X^T(I_N - \frac{1}{N}\mathbf{1}_{N\times 1}\mathbf{1}_{1\times N})Y$$

Let $\mathbf{1}_{N\times 1}\mathbf{1}_{1\times N} = \mathbf{1}_N$. ($N \times N$ matrix with all elements are 1)

$$\begin{aligned}\beta^* &= \left(X^T\left(X - \frac{1}{N}\mathbf{1}_N X\right)\right)^{-1}\left(X^T\left(Y - \frac{1}{N}\mathbf{1}_N Y\right)\right)\\ &= \left(X^T\left(I_N - \frac{1}{N}\mathbf{1}_N\right)X\right)^{-1}\left(X^T\left(I_N - \frac{1}{N}\mathbf{1}_N\right)Y\right)\end{aligned}$$

OLS (4)

Sample Mean $\bar{X}, \bar{Y}$

$$\frac{1}{N}\mathbf{1}_N X = \frac{1}{N}\begin{pmatrix}1 & \cdots & 1\\ 1 & \cdots & 1\\ \vdots & \vdots & \vdots\\ 1 & \cdots & 1\end{pmatrix}\begin{pmatrix}x_{11} & \cdots & x_{1m}\\ x_{21} & \cdots & x_{2m}\\ \vdots & \vdots & \vdots\\ x_{N1} & \cdots & x_{Nm}\end{pmatrix} = \begin{pmatrix}\bar{\mathbf{x}}_1 & \cdots & \bar{\mathbf{x}}_m\\ \bar{\mathbf{x}}_1 & \cdots & \bar{\mathbf{x}}_m\\ \cdots & & \\ \bar{\mathbf{x}}_1 & \cdots & \bar{\mathbf{x}}_m\end{pmatrix} = \bar{X}$$

$$\frac{1}{N}\mathbf{1}_N Y = \frac{1}{N}\begin{pmatrix}1 & 1 & \cdots & 1\\ \vdots & \vdots & \vdots & \vdots\\ 1 & 1 & \cdots & 1\end{pmatrix}\begin{pmatrix}y_1\\ \vdots\\ y_N\end{pmatrix} = \begin{pmatrix}\bar{y}\\ \vdots\\ \bar{y}\end{pmatrix} = \bar{Y}$$

Here $\bar{\mathbf{x}}_j$, $\bar{y}$ means sample mean of x_{ij}, y_i

$$\bar{\mathbf{x}}_j := \frac{1}{N}\sum_i^N x_{ij}, \quad \bar{y} := \frac{1}{N}\sum_i^N y_i$$

OLS (5)

Therefore, β^* is:

$$\beta^* = (X^T(X - \bar{X}))^{-1}X^T(Y - \bar{Y})$$

Note1: If $N \to \infty$, then $I_N - \frac{1}{N}\mathbf{1}_N \to I_N$ and

$$\beta^* \to (X^TX)^{-1}X^TY$$

Note2: We should check SOC: whether $H = D^2 f_{\beta,c}(\beta^*, c^*)$ is PD or not. Our object function has quadratic form with positive sign with regard to β, c when $\mathbf{x_j}$ is independent with each other and this means f is PD (when $\mathbf{x_j}$ is independent with each other: covariance with other variables are 0).

Note3: Some researchers denote X^T by X'