Equality Constraint Optimization
Lagrangian Multiplier

April 26, 2013

1 Review on single variable calculus

The problem

$$\min_{x \in \mathbb{R}^1} f(x)$$

can be solved by finding the minimizer $x*$ that minimize $f(x)$, that is, $f(x*) < f(x)\ \forall x \in \mathbb{R}^1$

Example

$$\min_{x \in \mathbb{R}^1}\ f(x) = e^{x^2}$$

To find the minimizer, apply the first order condition (first derivative test) to find the critical point

$$\frac{dg(x)}{dx} = 2xe^{x^2} = 0 \iff x = 0$$

Second order condition (second derivative test) can be used to test the critical value is optimizer

$$\left.\frac{d^2 g(x)}{dx^2}\right|_{x=0} = \left.\left(2e^{x^2} + 4x^2 e^{x^2}\right)\right|_{x=0} = 2 > 0$$

Thus $x = 0$ is a local (global) minimizer of $f(x)$

2 Optimization on multi-variable function

Then what about

$$\min_{(x,y) \in \mathbb{R}^2} f(x,y)\ ,\ \min_{(x,y,z) \in \mathbb{R}^3} f(x,y,z)\ ,\ \min_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x})$$

The idea is the same, first find out those critical points using first order condition : $\nabla f(\mathbf{x}) = 0$

Then test these points by using second order condition $\nabla^2 f(\mathbf{x})$

Review on ∇

For $f(x,y)$

$$\nabla f(x,y) = \begin{bmatrix} \dfrac{\partial f(x,y)}{\partial x} \\ \dfrac{\partial f(x,y)}{\partial y} \end{bmatrix}$$

$$\nabla^2 f(x,y) = \begin{bmatrix} \begin{bmatrix} \dfrac{\partial}{\partial x}\dfrac{\partial f(x,y)}{\partial x}, & \dfrac{\partial}{\partial y}\dfrac{\partial f(x,y)}{\partial x} \\ \dfrac{\partial}{\partial x}\dfrac{\partial f(x,y)}{\partial y}, & \dfrac{\partial}{\partial y}\dfrac{\partial f(x,y)}{\partial y} \end{bmatrix} \end{bmatrix} = \begin{bmatrix} \dfrac{\partial^2 f(x,y)}{\partial x^2} & \dfrac{\partial^2 f(x,y)}{\partial y \partial x} \\ \dfrac{\partial^2 f(x,y)}{\partial x \partial y} & \dfrac{\partial^2 f(x,y)}{\partial y^2} \end{bmatrix}$$

For $f(x,y,z)$

$$\nabla f(x,y,z) = \begin{bmatrix} \dfrac{\partial f(x,y,z)}{\partial x} \\ \dfrac{\partial f(x,y,z)}{\partial y} \\ \dfrac{\partial f(x,y,z)}{\partial z} \end{bmatrix}$$

$$\nabla^2 f(x,y,z) = \begin{bmatrix} \dfrac{\partial^2 f(x,y,z)}{\partial x^2} & \dfrac{\partial^2 f(x,y,z)}{\partial y \partial x} & \dfrac{\partial^2 f(x,y,z)}{\partial z \partial x} \\ \dfrac{\partial^2 f(x,y,z)}{\partial x \partial y} & \dfrac{\partial^2 f(x,y,z)}{\partial y^2} & \dfrac{\partial^2 f(x,y,z)}{\partial z \partial y} \\ \dfrac{\partial^2 f(x,y,z)}{\partial x \partial z} & \dfrac{\partial^2 f(x,y,z)}{\partial y \partial z} & \dfrac{\partial^2 f(x,y,z)}{\partial z^2} \end{bmatrix}$$

For $f(\mathbf{x}) = f(x_1, x_2, ..., x_n)$

$$\nabla f(\mathbf{x}) = \begin{bmatrix} \dfrac{\partial f(x_1, x_2, ..., x_n)}{\partial x_1} \\ \dfrac{\partial f(x_1, x_2, ..., x_n)}{\partial x_2} \\ \vdots \\ \dfrac{\partial f(x_1, x_2, ..., x_n)}{\partial x_n} \end{bmatrix}$$

$$\nabla^2 f(\mathbf{x}) = \begin{bmatrix} \dfrac{\partial^2 f(x_1, x_2, ..., x_n)}{\partial x_1^2} & \dfrac{\partial^2 f(x_1, x_2, ..., x_n)}{\partial x_2 \partial x_1} & ... & \dfrac{\partial^2 f(x_1, x_2, ..., x_n)}{\partial x_n \partial x_1} \\ \dfrac{\partial^2 f(x_1, x_2, ..., x_n)}{\partial x_1 \partial x_2} & \dfrac{\partial^2 f(x_1, x_2, ..., x_n)}{\partial x_2^2} & ... & \dfrac{\partial^2 f(x_1, x_2, ..., x_n)}{\partial x_n \partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial^2 f(x_1, x_2, ..., x_n)}{\partial x_1 \partial x_n} & \dfrac{\partial^2 f(x_1, x_2, ..., x_n)}{\partial x_2 \partial x_n} & ... & \dfrac{\partial^2 f(x_1, x_2, ..., x_n)}{\partial x_n^2} \end{bmatrix}$$

Where $\nabla^2 f = \left[\dfrac{\partial^2 f}{\partial x_i \partial x_j}\right]_{ij}$ is called the $Hessian\ Matrix\ H$

Same as single variable calculus, if $H(\mathbf{x}*) \begin{smallmatrix} > \\ < \end{smallmatrix} 0$ then $\mathbf{x}*$ is a local $\begin{matrix}\text{minimizer} \\ \text{maximizer}\end{matrix}$ of f

3 Derivation of Lagrangian Multiplier

$$\min_{(x,y)\in\mathbb{R}^2} f(x,y) \text{ s.t. } c(x,y)=0$$

For the constraint $c(x,y)=0$, it can be transformed into a implicit function of $y=g(x)$

Thus $f(x,y)$ becomes $f(x,g(x))$, a function only depend on x , let this function $f(x,g(x))=h(x)$

Then to find the extreme value of $h(x)$, it is same as before, apply the first order condition to find the critical point

$$\frac{dh(x)}{dx}=0$$

$$\iff \frac{df(x,g(x))}{dx}=\frac{\partial f(x,g(x))}{\partial x}\frac{dx}{dx}+\frac{\partial f(x,g(x))}{\partial y}\frac{dy}{dx} \qquad (Chain\ Rule)$$

$$\iff \frac{\partial f(x,g(x))}{\partial x}+\frac{\partial f(x,g(x))}{\partial y}\frac{dg(x)}{dx}=0$$

Consider $g'(x)$ and $c(x,y)=0$

$$c(x,y)=0$$

by chain rule on ∂x

$$\frac{\partial c(x,y)}{\partial x}+\frac{\partial c(x,y)}{\partial y}\frac{dy}{dx}=\frac{\partial 0}{\partial x}=0$$

Thus

$$g'(x)=\frac{dy}{dx}=-\frac{\frac{\partial c(x,y)}{\partial x}}{\frac{\partial c(x,y)}{\partial y}}$$

Therefore $\frac{\partial f(x,g(x))}{\partial x}+\frac{\partial f(x,g(x))}{\partial y}\frac{dg(x)}{dx}=0$ becomes

$$\frac{\partial f(x,y)}{\partial x}+\frac{\partial f(x,y)}{\partial y}\left[-\frac{\frac{\partial c(x,y)}{\partial x}}{\frac{\partial c(x,y)}{\partial y}}\right]=0$$

i.e.

$$\frac{\frac{\partial f(x,y)}{\partial x}}{\frac{\partial c(x,y)}{\partial x}}=\frac{\frac{\partial f(x,y)}{\partial y}}{\frac{\partial c(x,y)}{\partial y}}$$

Since $c(x,y)=0$ can be transformed into $\begin{cases} y=g(x) & \text{function of } x \text{ only} \\ x=s(y) & \text{function of } y \text{ only}\end{cases}$

$$\frac{\frac{\partial f(x,y)}{\partial x}}{\frac{\partial c(x,y)}{\partial x}}=\frac{\frac{\partial f(x,y)}{\partial y}}{\frac{\partial c(x,y)}{\partial y}} \qquad \iff \qquad \underbrace{\frac{\frac{\partial f(x,g(x))}{\partial x}}{\frac{\partial c(x,g(x))}{\partial x}}}_{\text{x only}}=\underbrace{\frac{\frac{\partial f(s(y),y)}{\partial y}}{\frac{\partial c(s(y),y)}{\partial y}}}_{\text{y only}}$$

Since moving the variable in one side does not change the value of the other side, therefore the other side can be treated as a constant λ

$$\frac{\frac{\partial f(x,y)}{\partial x}}{\frac{\partial c(x,y)}{\partial x}} = \frac{\frac{\partial f(x,y)}{\partial y}}{\frac{\partial c(x,y)}{\partial y}} = \lambda$$

i.e.

$$\frac{\partial f(x,y)}{\partial x} - \lambda\frac{\partial c(x,y)}{\partial x} = 0 \quad \text{and} \quad \frac{\partial f(x,y)}{\partial y} - \lambda\frac{\partial c(x,y)}{\partial y} = 0$$

Further re-arrange

$$\frac{\partial}{\partial x}\left[f(x,y) - \lambda c(x,y)\right] = 0 \quad \text{and} \quad \frac{\partial}{\partial y}\left[f(x,y) - \lambda c(x,y)\right] = 0$$

Together with the constraint

$$\begin{aligned} & c(x,y) = 0 \\ \iff & 0 - c(x,y) = 0 \\ \iff & \frac{\partial}{\partial \lambda}\left[f(x,y) - \lambda c(x,y)\right] = 0 \end{aligned}$$

Thus, introduce the *Lagrangian*

$$L(x,y,\lambda) = f(x,y) - \lambda c(x,y)$$

The λ is the *Lagrangian Multiplier*

The system of equations

$$\begin{cases} \frac{\partial}{\partial x}\left[f(x,y) - \lambda c(x,y)\right] = 0 \\ \frac{\partial}{\partial y}\left[f(x,y) - \lambda c(x,y)\right] = 0 \\ c(x,y) = 0 \end{cases}$$

Is equivalent to

$$\begin{cases} \frac{\partial L}{\partial x} = 0 \\ \frac{\partial L}{\partial y} = 0 \\ \frac{\partial L}{\partial \lambda} = 0 \end{cases} \qquad \begin{cases} L_x = 0 \\ L_y = 0 \\ L_\lambda = 0 \end{cases} \qquad \nabla L = 0$$

Using short hand notation — Using super short hand notation

Therefore, to solve

$$\min_{(x,y)\in\mathbb{R}^2} f(x,y) \text{ s.t. } c(x,y) = 0$$

Step 1. Form Lagrangian $L = f(x,y) - \lambda c(x,y)$

Step 2. Find the critical point $(x*, y*)$ by using first order condition

$$\nabla L = 0$$

Step 3. Test the critical point by using second order condition

$$\text{If } \nabla^2 L\,(x*,y*) \begin{matrix} > \\ = \\ < \end{matrix} 0\text{, then point } (x*,y*) \text{ is } \begin{matrix} \text{minimizer} \\ \text{saddle point} \\ \text{maximizer} \end{matrix}$$

Multiple Constraints

If there is only 1 constraint g , then there is one multipler λ, and

$$L = f - \lambda g$$

If there are 2 constraints g_1 g_2 , then there are 2 multipler λ_1 , λ_2 , and

$$L = f - \lambda_1 g_1 - \lambda_2 g_2$$

If there are m constraints $\mathbf{g} = [g_1, g_2, ..., g_m]^T$, then there are m multipler $\bar{\lambda} = [\lambda_1, \lambda_2, ..., \lambda_m]^T$, and

$$L = f - \bar{\lambda} \cdot \mathbf{g} = f - \sum \lambda_i g_i$$

4 Examples

1-variable function $f(x)$ with no constraint

$$L = f(x) \qquad \nabla L = \frac{df(x)}{dx} = 0 \qquad \text{(same as before)}$$

2-variable function $f(x, y)$ with no constraint

$$L = f(x, y) \qquad \nabla L = \begin{bmatrix} \frac{\partial f(x,y)}{\partial x} \\ \frac{\partial f(x,y)}{\partial y} \end{bmatrix} = 0 \qquad \text{(same as before)}$$

1-variable function $f(x)$ with 1-constraint $g(x) = 0$

$$L = f(x) - \lambda g(x) \qquad \nabla L = \begin{bmatrix} \frac{\partial L(x,\lambda)}{\partial x} \\ \frac{\partial L(x,\lambda)}{\partial \lambda} \end{bmatrix} = \begin{bmatrix} f'(x) - \lambda g'(x) \\ g(x) \end{bmatrix} = 0$$

* - sign in $g(x)$ is omitted because$-g(x) = 0$ is same as $g(x) = 0$

2-variable function $f(x, y)$ with 1-constraint $g(x, y) = 0$

$$L = f(x,y) - \lambda g(x,y) \qquad \nabla L = \begin{bmatrix} \frac{\partial f(x,y,\lambda)}{\partial x} \\ \frac{\partial f(x,y,\lambda)}{\partial y} \\ \frac{\partial f(x,y,\lambda)}{\partial \lambda} \end{bmatrix} = \begin{bmatrix} f_x(x,y) - \lambda g_x(x,y) \\ f_y(x,y) - \lambda g_y(x,y) \\ g(x,y) \end{bmatrix} = 0$$

2-variable function $f(x, y)$ with 1-constraint $g(x) = 0$

$$L = f(x,y) - \lambda g(x) \qquad \nabla L = \begin{bmatrix} \frac{\partial f(x,y,\lambda)}{\partial x} \\ \frac{\partial f(x,y,\lambda)}{\partial y} \\ \frac{\partial f(x,y,\lambda)}{\partial \lambda} \end{bmatrix} = \begin{bmatrix} f_x(x,y) - \lambda g'(x) \\ f_y(x,y) \\ g(x) \end{bmatrix} = 0$$

Short hand notation will be used from now on

3-variable function $f(x,y,z)$ with 1-constraint $g(x,y,z)=0$

$$L = f - \lambda g \qquad \nabla L = \begin{bmatrix} L_x \\ L_y \\ L_z \\ L_\lambda \end{bmatrix} = \begin{bmatrix} f_x - \lambda g_x \\ f_y - \lambda g_y \\ f_z - \lambda g_z \\ g \end{bmatrix} = 0$$

3-variable function $f(x,y,z)$ with 2-constraint $g_1(x,y,z)=0$ and $g_2(x,y,z)=0$

$$L = f - \lambda_1 g_1 - \lambda g_2 \qquad \nabla L = \begin{bmatrix} L_x \\ L_y \\ L_z \\ L_{\lambda_1} \\ L_{\lambda_2} \end{bmatrix} = \begin{bmatrix} f_x - \lambda_1 g_{1x} - \lambda_2 g_{2x} \\ f_y - \lambda_1 g_{1y} - \lambda_2 g_{2y} \\ f_z - \lambda_1 g_{1z} - \lambda_2 g_{2z} \\ g_1 \\ g_2 \end{bmatrix} = 0$$

7-variable function $f(\mathbf{x}) = f(x_1, x_2, ..., x_7)$ with 4 constraints $\begin{cases} a(x_1,x_2,x_3,x_4,x_5,x_6,x_7)=0 \\ b(x_1,x_3,x_5,x_7)=0 \\ c(x_2,x_4,x_6)=0 \\ d(x_7)=0 \end{cases}$

$$L = f - \sum_{i=1}^{4} \lambda_i g_i \qquad \nabla L = \begin{bmatrix} L_{x_1} \\ L_{x_2} \\ L_{x_3} \\ L_{x_4} \\ L_{x_5} \\ L_{x_6} \\ L_{x_7} \\ L_{\lambda_1} \\ L_{\lambda_2} \\ L_{\lambda_3} \\ L_{\lambda_4} \end{bmatrix} = \begin{bmatrix} f_{x_1} - \lambda_1 a_{x_1} - \lambda_2 b_{x_1} \\ f_{x_2} - \lambda_1 a_{x_2} - \lambda_3 c_{x_2} \\ f_{x_3} - \lambda_1 a_{x_3} - \lambda_2 b_{x_3} \\ f_{x_4} - \lambda_1 a_{x_4} - \lambda_3 c_{x_4} \\ f_{x_5} - \lambda_1 a_{x_5} - \lambda_2 b_{x_5} \\ f_{x_6} - \lambda_1 a_{x_6} - \lambda_3 c_{x_6} \\ f_{x_7} - \lambda_1 a_{x_7} - \lambda_4 d' \\ a \\ b \\ c \\ d \end{bmatrix} = 0$$

General case : n variable function $f(\mathbf{x}) = f(x_1, x_2, ..., x_n)$ with m constraint $\mathbf{g} = [g_1, g_2, ..., g_m]$

$$L = f(\mathbf{x}) - \bar{\lambda} \cdot \mathbf{g} = f(\mathbf{x}) - \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x})$$

$$\nabla L = \begin{bmatrix} \nabla_{\mathbf{x}} L \\ \nabla_{\bar{\lambda}} L \end{bmatrix} = 0$$

Note : the $\mathbf{x}$ inside g_i need not to be same as the $\mathbf{x}$ inside f .

Therefore, the Lagrangian Multiplier method can be used to tackle the problems of optimization with equality constraints.

And the major task to do is to solve the huge system of equations of $\nabla L = 0$ that involve a lots of algebra.

–END–