Optimization : Overview

March 25, 2013

1 Introduction

Normally optimization problems can be grouped into 3 type

1. Unconstrained Optimization

$$\min_{x \in \mathbb{R}^n} f(x) \qquad \text{or} \qquad \max_{x \in \mathbb{R}} f(x)$$

2. Optimization subject to equality constraints

$$\min_{x \in \mathbb{R}^n} f(x) \qquad \text{subject to } g_i(x) = 0$$

3. Optimization subject to inequality constraints

$$\min_{x \in \mathbb{R}^n} f(x) \qquad \text{subject to} \quad \begin{array}{l} g_i(x) = 0 \\ h_i(x) \geq 0 \end{array}$$

How to solve these problems ?

2 Unconstrained Optimization

2.1 Fermat's Theorem of stationary points

For function $f : (a, b) \to \mathbb{R}$, for $x_0 \in (a, b)$, it is a local extremum of f if $f'(x_0) = 0$

2.2 For Single Variable Function

This is what we have learned from high school single variable calculus :

To find the extremum,

1. First find $\dfrac{df}{dx}$
2. Then set $\dfrac{df(x)}{dx} = 0$ to solve for x_0
3. Test is x_0 the extremum of f
4. If f is convex / concave, then the x_0 is the global extremum of f

2.3 For Multivariable Function

For multiple variable function , such as$f(\mathbf{x}) = f \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$,

To find the extremum, apply the same logic

1. First find $\nabla f(\mathbf{x}) = \begin{bmatrix} \frac{\partial f(\mathbf{x})}{\partial x_1} \\ \frac{\partial f(\mathbf{x})}{\partial x_2} \\ \vdots \\ \frac{\partial f(\mathbf{x})}{\partial x_n} \end{bmatrix}$

2. Set $\nabla f(\mathbf{x}) = 0$ (With luck , this will become a system of n-linear equations, otherwise, for non-linear system, it is hard to solve it !)

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

For example 2 variable function , $\nabla^2 f(x, y) = \begin{bmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{bmatrix}$

Perform the second derivative test $|\nabla^2 f(\mathbf{x})|$, thus for $|\nabla^2 f(x, y)| = f_{xx} f_{yy} - f_{xy}^2$

If $D > 0$, $f_{xx} > 0$, then local min

If $D > 0$, $f_{xx} < 0$, then local max

If $D < 0$, saddle point

4. For local min and local max, if f is convex, then it is the global max / min.

3 Optimization subject to equality constraints

$$\min_{x \in \mathbb{R}^n} f(x) \qquad \text{subject to } g_i(x) = 0$$

3.1 Lagrangian Multiplier

For optimize the cost function $f(x)$ under one equality constraints $g(x) = 0$, let the lagrangian function be

$$L(x, \lambda) = f(x) - \lambda g(x)$$

The extremum of $L(x, \lambda)$ is the extremum of the problem.

3.2 Why Lagrangian Multiplier Works by Implicit Function

Consider 2 variable function $f(x, y)$ and constraint $g(x, y) = 0$

Since $g(x, y) = 0$, then we can treate y as a implicit function of x

$$g(x,y) = g(x,y(x)) = 0 \qquad\qquad f(x,y) = f(x,y(x))$$

Consider $\dfrac{df}{dx} = 0$

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

i.e.

$$f_x + f_y\frac{dy}{dx} = 0$$

Since $g(x,y(x)) = 0$, thus $\dfrac{dg}{dx} = 0$

$$g_x + g_y\frac{dy}{dx} = 0$$

Thus eliminate $\dfrac{dy}{dx}$

$$\begin{cases} f_x + f_y\dfrac{dy}{dx} = 0 \\ g_x + g_y\dfrac{dy}{dx} = 0 \end{cases} \quad \frac{dy}{dx} = -\frac{f_x}{f_y} \qquad \Longrightarrow \qquad f_x g_y - f_y g_x = 0 \quad \Rightarrow \quad \begin{vmatrix} f_x & f_y \\ g_x & g_y \end{vmatrix} = 0$$

This can be true if $f_x \propto g_x$ and $f_y \propto g_y$, that is

$$f_x = \lambda_1 g_x$$
$$f_y = \lambda_2 g_y$$

And $\lambda_1 = \lambda_2$ (Just plug the eqaution into the determinant)
Thus

$$\begin{bmatrix} f_x \\ f_y \end{bmatrix} = \lambda \begin{bmatrix} g_x \\ g_y \end{bmatrix}$$

i.e.

$$\nabla\Big(f(\mathbf{x}) - \lambda g(\mathbf{x})\Big) = 0 \qquad \Longrightarrow \nabla L(x,\lambda) = 0$$

3.3 Why Lagrangian Multiplier Works by Geometry

The function $g(x,y) = 0$ is the constraint, the $f(x,y) = c$ are the isoclines.

The gradient (slope) of $f(x,y)$ is $\dfrac{\partial f}{\partial x}\hat{x} + \dfrac{\partial f}{\partial y}\hat{y}$, and for $g(x,y)$, it is $\dfrac{\partial g}{\partial x}\hat{x} + \dfrac{\partial g}{\partial y}\hat{y}$

For the gradient vector, it always point in the direction of a function's steepest slope at a given point, and thus $\nabla f \perp f = c$

In the critical point $\mathbf{x}_c$, $\nabla f(\mathbf{x}_c)$ // $\nabla g(\mathbf{x}_c)$ $\rightarrow \nabla f(\mathbf{x}_c) = \lambda \nabla g(\mathbf{x}_c)$

3.4 Optimization with one equality constraints

$$\min_{x\in\mathbb{R}^n} f(\mathbf{x}) \qquad \text{subject to } g(\mathbf{x}) = 0$$

1. Let $L(\mathbf{x},\lambda) = f(\mathbf{x}) - \lambda g(\mathbf{x})$
2. Solve $\begin{cases} \nabla L(\mathbf{x},\lambda) = 0 \\ g(\mathbf{x}) = 0 \end{cases}$ to obtain $\mathbf{x}_0$

3. Test $\mathbf{x}_0$ is the extremum or not.
4. If f is convex, then $\mathbf{x}_0$ is global extremum

3.5 Optimization with multiple equality constraints

$$\min_{x\in\mathbb{R}^n} f(\mathbf{x}) \qquad \text{subject to } g_i(\mathbf{x}) = 0 \quad i = 1, 2, ..., n$$

1. Let $L(\mathbf{x}, \lambda) = f(\mathbf{x}) - \lambda_1 g_1(\mathbf{x}) - \lambda_2 g_2(\mathbf{x}) - ... - \lambda_n g_n(\mathbf{x})$
2. Solve $\begin{cases} \nabla L(\mathbf{x}, \lambda) = 0 \\ g_i(\mathbf{x}) = 0 \end{cases}$
3. Test $\mathbf{x}_0$ is the extremum or not.
4. If f is convex, then $\mathbf{x}_0$ is global extremum.

3.6 Example

$$\min_{x\in\mathbb{R}^n} f(x, y) \qquad \text{subject to } g(x, y) = c$$

1. First $g*(x, y) = g(x, y) - c = 0$
2. Then set up the Lagrangian $L(x, y, \lambda) = f(x, y) - \lambda g*(x, y)$
3. Find $L_x(x, y, \lambda)$, $L_y(x, y, \lambda)$ and $L_\lambda(x, y, \lambda)$ (i.e. find ∇L)
4. Solve for the critical point $\begin{cases} L_x(x, y, \lambda) = 0 \\ L_y(x, y, \lambda) = 0 \\ L_\lambda(x, y, \lambda) = 0 \end{cases}$, let the critical point be $\mathbf{x}_0 = (x_0, y_0)$
5. Test $\mathbf{x}_0$ is extremum of not.
6. If f is convex, then $\mathbf{x}_0$ is global extremum

$$\min_{x\in\mathbb{R}^n} f(x, y) \qquad \text{subject to } g(x, y) = c \;\; h(x, y) = d$$

1. First $g*(x, y) = g(x, y) - c = 0$, $h*(x, y) = h(x, y) - d = 0$
2. Then set up the Lagrangian $L(x, \lambda) = f(x, y) - \lambda g*(x, y) - \mu h*(x, y)$
3. Find $L_x(x, y, \lambda, \mu)$, $L_y(x, y, \lambda, \mu)$ and $L_\lambda(x, y, \lambda, \mu)$, $L_\mu(x, y, \lambda, \mu)$(i.e. find ∇L)
4. Solve for the critical point $\begin{cases} L_x(x, y, \lambda, \mu) = 0 \\ L_y(x, y, \lambda, \mu) = 0 \\ L_\lambda(x, y, \lambda, \mu) = 0 \\ L_\mu(x, y, \lambda, \mu) = 0 \end{cases}$, let the critical point be $\mathbf{x}_0$
5. Test $\mathbf{x}_0$ is extremum of not.
6. If f is convex, then $\mathbf{x}_0$ is global extremum

4 Optimization subject to inequality constraints

$$\min_{x\in\mathbb{R}^n} f(x) \qquad \text{subject to } \begin{array}{l} g_i(x) = 0 \\ h_i(x) \geq 0 \end{array}$$

4.1 Karush–Kuhn–Tucker conditions

The KKT Conditions are the generalization of Lagrangian Multiplier.

Example
Consider only one constraint, an inequality constraint

$$\min_{x\in\mathbb{R}^n} f(x) \qquad \text{subject to } g(x) \geq 0$$

The Lagrangian

$$L(x,\lambda) = f(x) - \lambda g(x)$$

The KKT conditions for the optimal solution are

$$\begin{cases} \nabla L(x,\lambda) = 0 \\ \lambda \geq 0 \\ g(x) \geq 0 \\ \lambda g(x) = 0 \end{cases} \quad \text{expand} \begin{cases} L_x(x,\lambda) = 0 \\ L_\lambda(x,\lambda) = 0 \\ \lambda \geq 0 \\ g(x) \geq 0 \\ \lambda g(x) = 0 \end{cases}$$

Example
Consider only a two variable function with 3 constraints, all inequality constraint

$$\min_{x\in\mathbb{R}^n} f(x,y) \qquad \text{subject to } g_i(x,y) \geq 0 \qquad i = 1,2,3$$

The Lagrangian , in vector form

$$L(\mathbf{x},\lambda) = f(\mathbf{x}) - \sum_{i=1}^{3} \lambda_i g_i(\mathbf{x})$$

Expand,

$$L(x,y,\lambda) = f(x,y) - \sum_{i=1}^{3} \lambda_i g_i(x,y)$$

$$L(x,y,\lambda) = f(x,y) - \lambda_1 g_1(x,y) - \lambda_2 g_2(x,y) - \lambda_3 g_3(x,y)$$

The KKT conditions for the optimal solution are

$$\begin{cases} \nabla L(\mathbf{x},\lambda) = 0 \\ \lambda_i \geq 0 \\ g_i(x) \geq 0 \\ \lambda_i g_i(x) = 0 \end{cases} \quad \text{expand} \begin{cases} L_x(x,y,\lambda) = 0 \\ L_y(x,y,\lambda) = 0 \\ L_\lambda(x,y,\lambda) = 0 \\ \lambda_i \geq 0 \\ g_i(x) \geq 0 \\ \lambda_i g_i(x) = 0 \end{cases}$$

The most general form , in both vector form and expanded form

$$\min_{x\in\mathbb{R}^n} f(\mathbf{x}) \quad \text{subject to } \begin{matrix} g_i(\mathbf{x}) = 0 \\ h_i(\mathbf{x}) \geq 0 \end{matrix} \qquad \text{or} \qquad \min_{x\in\mathbb{R}^n} f(x_1,x_2,...,x_m) \quad \text{subject to } \begin{matrix} g_i(x_1,x_2,...,x_m) = 0 \\ h_i(x_1,x_2,...,x_m) \geq 0 \end{matrix}$$

The Lagrangian

$$L(\mathbf{x}, \lambda) = f(\mathbf{x}) - \sum_{i=1}^{l} \lambda_i g_i(\mathbf{x}) - \sum_{i=1}^{k} \mu_i h_i(\mathbf{x})$$

The KKT conditions for the optimal solution are

$$\left\{ \begin{array}{c} \nabla L(\mathbf{x}, \lambda) = \left\{ \begin{array}{l} \nabla_{\mathbf{x}} L(\mathbf{x}, \lambda) = 0 \\ \nabla_{\lambda} L(\mathbf{x}, \lambda) = 0 \end{array} \right. \\ g_i(\mathbf{x}) = 0 \\ h_i(\mathbf{x}) \geq 0 \\ \mu_i \geq 0 \\ \mu_i h_i(\mathbf{x}) = 0 \end{array} \right. \quad \text{expand} \quad \left\{ \begin{array}{c} L_{x_1}(x_1, x_2, ..., x_m, \lambda) = 0 \\ \vdots \\ L_{x_m}(x_1, x_2, ..., x_m, \lambda) = 0 \\ L_{\lambda}(x_1, x_2, ..., x_m, \lambda) = 0 \\ g_i(x_1, x_2, ..., x_m) = 0 \\ h_i(x_1, x_2, ..., x_m) \geq 0 \\ \mu_i \geq 0 \\ \mu_i h_i(x_1, x_2, ..., x_m) = 0 \end{array} \right.$$

$-END-$