Classical Optimization

1 var unconstrained optim.

[Necessary Condition]

when $f(x=x^*)$ $\exists$ loc. min. $\wedge \exists \frac{d}{dx} f \longrightarrow \frac{d}{dx} f(x^*) = 0$

[Sufficient Condition]

$f'(x^*) = f^{(2)}(x^*) = \cdots = f^{(n-1)}(x^*) = 0$, $f^{(n)}(x^*) \neq 0$, then

i) $f^{(m)}(x^*) > 0$, $m \in$ even $\longrightarrow f(x^*) \in$ loc. min.

ii) < 0, : $\longrightarrow \in$ loc. max.

iii) ` ` ` , $m \in$ odd $\rightarrow$ neither.

---

m vars unconstraints opt

< Def >

$$d^r f(x^*) : \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \cdots \sum_{i_r=1}^{n} h_{i_1} h_{i_2} \cdots h_{i_r} \frac{\partial^r f(x^*)}{\partial x_{i_1} \partial x_{i_2} \cdots \partial x_{i_r}}$$

H (Hessian Matrix) : $\begin{bmatrix} \frac{\partial^2}{\partial x_1^2} f & \cdots & \frac{\partial^2}{\partial x_1 \partial x_n} f \\ \vdots & \ddots & \vdots \\ \frac{\partial^2}{\partial x_n \partial x_1} f & \cdots & \frac{\partial^2}{\partial x_n^2} f \end{bmatrix}$

Pos / Neg Definite : $\forall \lambda \in$ eigen value > 0 / < 0

[Necessary Condition]

when $f(x=x^*)$ $\exists$ loc. min. $\wedge$ $\exists$ $\frac{df}{dx}$ $\longrightarrow$ $\nabla f(x^*) = 0$

[Sufficient Condition]

When $m \geq 2$, $d^r f(x^*) = 0$ $\forall\ 1 \leq r \leq m-1$, $d^m f(x^*) = 0$, then

i) $d^m f(x^*)$ $\in$ Pos Def, $m \in$ even $\longrightarrow$ $f(x^*)$ $\exists$ loc. min.

ii) 〃 $\in$ Neg Def, 〃 $\longrightarrow$ 〃 $\exists$ loc. max.

iii) $\in$ Semi-Poi Def, $\longrightarrow$ 〃 doesn't decrease around x^*

iv) ~ , $m \in$ odd $\longrightarrow$ 〃 not a extreme.

---

m Vars equality constraints opt.

<Def>

find $\arg\min(f) = x^* = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$

Subject to $h_j(x) = 0$, $j = 1, 2, \cdots m$

[Lagrange Multipliers] $L(x, \lambda) = f(x) - \lambda\, h(x)$

Assume $\nabla h_j(x^*) \in$ Linearly Independ.

$\exists!\ \lambda_j^*$ satisfying $\frac{\partial}{\partial x_i} L = \frac{\partial}{\partial x_i} f + \sum_{j=1}^{m} \lambda_j^* \frac{\partial}{\partial x_i} h_j = 0$

[Necessary Condition]

$$\frac{\partial}{\partial x_i} L = \frac{\partial}{\partial x_i} f + \sum_{j=1}^{m} \lambda_j \frac{\partial}{\partial x_i} h_j = 0, \quad \frac{\partial}{\partial \lambda_i} L = h_j(x) = 0$$

[Sufficient Condition]

1. If the objective function f(x) is convex and the constraint functions $h_j(x)$ are either convex or concave, then the necessary conditions are also sufficient conditions for local min/max points.
2. Define $Q = \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial^2 L}{\partial x_i \partial x_j} dx_i dx_j$ as the Hessian matrix of the Lagrange function. If Q is positive definite at $\underline{x}=\underline{x}^*$ and $\underline{\lambda}=\underline{\lambda}^*$, then $\underline{x}=\underline{x}^*$ is a local minimizer. If Q is negative definite at $\underline{x}=\underline{x}^*$ and $\underline{\lambda}=\underline{\lambda}^*$, then $\underline{x}=\underline{x}^*$ is a local maximizer.
3. It can be shown that, if the roots (z values) of the following determinant equation are all positive, then Q is positive definite. If the roots are all negative, then Q is negative definite.

$$\begin{vmatrix} L_{11}-z & L_{12} & L_{13} & \dots & L_{1n} & h_{11} & h_{21} & \dots & h_{m1} \\ L_{21} & L_{22}-z & L_{23} & \dots & L_{2n} & h_{12} & h_{22} & \dots & h_{m2} \\ \vdots & & & & & & & & \\ L_{n1} & L_{n2}-z & L_{n3} & \dots & L_{nn}-z & h_{1n} & h_{2n} & \dots & h_{mn} \\ h_{11} & h_{12} & \dots & & h_{1n} & 0 & 0 & \dots & 0 \\ h_{21} & h_{22} & \dots & & h_{2n} & 0 & 0 & \dots & 0 \\ & & & & & & & & \\ h_{m1} & h_{m2} & \dots & & h_{mn} & 0 & 0 & \dots & 0 \end{vmatrix} = 0$$

Here L_{ij} and g_{ij} are defined as the following and evaluated at $\underline{x}=\underline{x}^*$ and $\underline{\lambda}=\underline{\lambda}^*$ (see Rao's book for more detail):

$$L_{ij} = \frac{\partial^2 L}{\partial x_i \partial x_j} dx_i dx_j ; \quad h_{ij} = \frac{\partial h_i}{\partial x_j}$$

} Example }

$\text{argmin}(f)$, $f(x_1, x_2) = x_1^2 + x_1 x_2$, subject to $x_1 - 2x_2 = 3$

$$L(x, \lambda) = x_1^2 + x_1 x_2 - \lambda(x_1 - 2x_2 - 3)$$

$$\begin{cases} \frac{\partial}{\partial x_1} L = 2x_1 + x_2 + \lambda = 0 \\ \frac{\partial}{\partial x_2} L = x_1 - 2\lambda = 0 \\ \frac{\partial}{\partial \lambda} L = x_1 - 2x_2 - 3 = 0 \end{cases}$$

m Vars with Inequality constraints opt.

《Method》

$\text{let } G_j(x) = g_j(x) + \underbrace{"y_j^2"}_{\text{always} \geq 0} = 0 \Rightarrow g_j(x) = -y_j^2 \Rightarrow g_j(x) \leq 0$

[KKT method]

$$\begin{cases} \dfrac{\partial}{\partial x_i} f + \sum_{j=1}^{m} \lambda_j \dfrac{\partial}{\partial x_i} g_j + \sum_{k=1}^{p} \beta_k \dfrac{\partial}{\partial x_i} h_k = 0 \\ \lambda_j g_j = 0 \\ g_j \leq 0 \\ \lambda_j \geq 0 \\ h_k = 0 \end{cases}$$

Unconstrained n-d

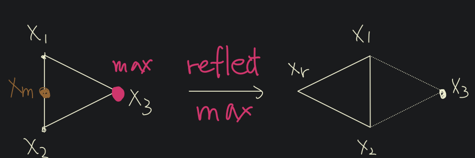
Simplex method

[Simplex]

The geometric figure formed by a set of $n+1$ points, e.g. 2D → triangle, 3D − tetrahedron

《Method》

step 1) reflect = $X_r = X_m + R(X_m - X_{n+1})$



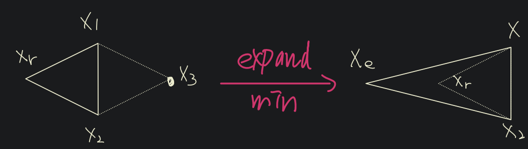
if $f(x_1) < f(x_r) < f(x_2)$ → next iter step 2) Contraction

if $f(x_r)$ is max

next step

if $f(x_r)$ is min

Expand $= X_e = x_m + E(x_r - x_m)$



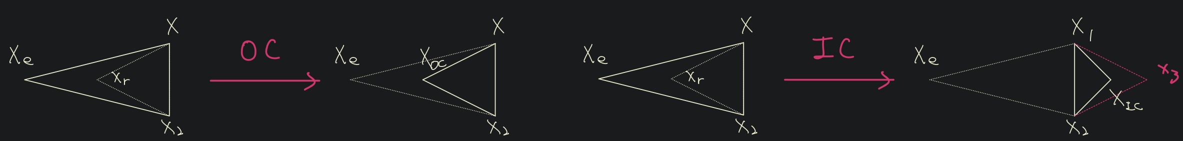
step 3)

if $f(x_r) < f(x_{n+1})$: $f(x_r) <$ origin

Outside $= X_{oc} = X_m + OC(X_r - X_m)$

if $f(x_r) > f(x_{n+1})$: $f(x_r) >$ origin

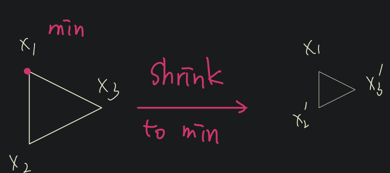
Inside $= X_{ic} = X_m - IC(X_m - X_{n+1})$



if $f(X_{oc}) < f_r$, then do next iter

if $f(x_{ic}) < f(x_{n+1})$, then do next iter
o.w. next step

step 4) Shrink



$\Rightarrow$ do next iter.

Powell's Conjugate Direction Method

[Thm 1]

let $A \in$ Sym. Matrix . if directions $S = \{s_1, \cdots s_n\}$ are nonzero and A-conjugate,

then S are linearly Indep.

[Thm 2]

$\forall x, \quad x = \sum_{i=1}^{n} \lambda_i S_i \;, \quad \lambda_i = \dfrac{S_i^T A x}{S_i^T A S_i}$

[Thm 3]

$Q(x) = \frac{1}{2} x^T A x + B^T x + C$ of n vars, and 2 start points X_a, X_b are given

From X_a, X_b, min is searched along the same S, reaching the min X_1, X_2, respectively.

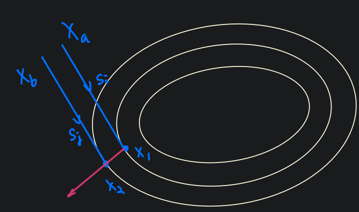
Then the line joining X_1, X_2 is A-conjugate to S.

[Thm 4]

If $Q(x)$ is minized sequentially, once along a dir $\in$ A-conjugate dir,

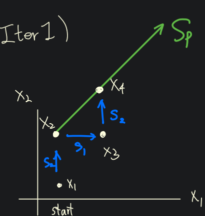
then the min will be found before the n^{th} steps irrespective of the start point.

let $S_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, $S_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$



$S_i^T A (X_1 - X_2) = 0$ $\quad \therefore$ S_i and $(X_1 - X_2)$ are A-conjugate

Iter 1)

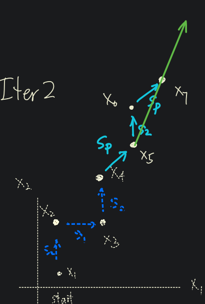


S_p is A-Conjugare to S_1

So now the new dirs are S_2 and S_p

S_p' is A-Conjugate to S_p

Iter 2



So now the new dirs are S_p and S_p

Unconstrained Nonlinear Programming : n-d min. using function 1^{st} order derivative

Gradient descent with constant step length (learning rate)

{Example}

$$f\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = 2x_1^2 + 2x_1x_2 + x_2^2 + x_1 - x_2$$

1. ∇f

$$\nabla f = \begin{bmatrix} 4x_1 + 2x_2 + 1 \\ 2x_1 + 2x_2 - 1 \end{bmatrix}$$

2. $S_1 = -\nabla f(X_1)$

$$S_1 = -\nabla f(X_1) = -\begin{bmatrix} 3 \\ 1 \end{bmatrix}$$

3. Find λ^* to min $f(X_1 + \lambda S_1) = f(X_2)$

$$f\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix} - \begin{bmatrix} 3\lambda \\ \lambda \end{bmatrix}\right) = f\left(\begin{bmatrix} -3\lambda \\ 1-\lambda \end{bmatrix}\right) = 25\lambda^2 - 10\lambda$$

$$\frac{d}{d\lambda} f = 50\lambda - 10 = 0 \quad \therefore \lambda^* = \frac{1}{5} \quad \Rightarrow \quad X_2 = \begin{bmatrix} -\frac{3}{5} \\ \frac{4}{5} \end{bmatrix}$$

4. Check $|\nabla f(X_2)|$ or $|X_2 - X_1|$ close to 0, $\rightarrow$ to 2. $S_2 = -\nabla f(X_2)$; $\rightarrow$ stop

Steepest Descent Method (GD with optimal step length obtained through line search)

<Idea>

Search along the dir of $-\nabla f$,

and find opt. step len. along this dir.

{Example}

$$\min \ f\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = 2x_1^2 + 2x_1x_2 + x_2^2 + x_1 - x_2$$

1. To $Q(x) = \frac{1}{2} x^T A x + Bx + C$

$$\frac{1}{2}\begin{bmatrix} x_1 & x_2 \end{bmatrix}\begin{bmatrix} 4 & 2 \\ 2 & 2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 1 & -1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + 0$$

2. $\nabla f = \begin{bmatrix} \frac{\partial}{\partial x_1} f \\ \frac{\partial}{\partial x_2} f \end{bmatrix} = Ax + b$

$$\nabla f = \begin{bmatrix} 4 & 2 \\ 2 & 2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 4x_1 + 2x_2 + 1 \\ 2x_1 + 2x_2 - 1 \end{bmatrix}$$

3. $S_1 = -\nabla f(X_1)$, X_1 is init

$$-\nabla f\left(X_1 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) = -\begin{bmatrix} 3 \\ 1 \end{bmatrix}$$

4. Find λ^* to min $f(X_1 + \lambda S_1) = f(X_2)$

$$f\left(\begin{bmatrix} 0 - 3\lambda \\ 1 - \lambda \end{bmatrix}\right) = 25\lambda^2 - 10\lambda \quad \therefore \lambda^* = \frac{1}{5}$$

Back to 3.

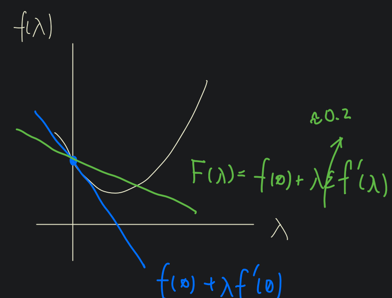
$$X_2 = X_1 + \lambda^* S_1 = \begin{bmatrix} 0 + \frac{1}{5}(-3) \\ 1 + \frac{1}{5}(-1) \end{bmatrix} = \begin{bmatrix} -\frac{3}{5} \\ \frac{4}{5} \end{bmatrix} \longrightarrow \text{stop if } |\nabla f(X_2)| \le \varepsilon$$

GD with Armijo's inexact line search

<Idea 1>

n-D opt can be transform to 1-D search

$f(x) \Rightarrow f(\lambda)$



<Idea 2>

$f(\lambda)$ can be approximated near to $\lambda = 0$

as $f(\lambda) \approx f(0) + f'(0)(\lambda - 0)$

$= f(0) + \lambda f'(0)$

{Example}

$f\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = x_1^2 + x_2^2 + 1$ start from $\underbrace{\begin{bmatrix} 1 \\ 1 \end{bmatrix}}_{x_1}$

1. ∇f

$$\nabla f = \begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix}$$

2. $S_1 = -\nabla f(x_1)$

$$S_1 = \begin{bmatrix} -2 \\ -2 \end{bmatrix}$$

3. $f(\lambda) \approx f(0) + \lambda f'(\lambda)$

$$f(x_1 + \lambda S_1) = \begin{bmatrix} 1 - 2\lambda \\ 1 - 2\lambda \end{bmatrix} = (1-2\lambda)^2 + (1-2\lambda)^2 + 1$$

$$f'(\lambda) = 16\lambda - 8$$

$$f(\lambda) = f(0) + \lambda f'(0)$$

$$= 3 + \lambda(-8) = 3 - 8\lambda$$

4. Define $F(\lambda) = f(0) + \lambda \varepsilon f'(\lambda)$

$$F(\lambda) = 3 - 8\varepsilon\lambda$$

5. Get approximated opt. λ by comparing $f(\lambda)$ and $F(\lambda)$

Assume $\lambda = 1$

shrink

$$\begin{cases} f(\lambda=1) = 3 \\ F(\lambda=1) = 1.4 \end{cases}, \quad f(\lambda) > F(\lambda) \quad \therefore \lambda' = \frac{\lambda}{2} = \frac{1}{2}$$

until $f(\lambda) < F(\lambda)$ is acceptable

Conjugate Gradient Method

《Method》

1. start from X_1

2. search along $S_1 = -\nabla f(X_1)$, $X_2 = X_1 + \lambda^* S_1$

3. $S_{j+1} = -\nabla f(X_{j+1}) + \beta_j S_j$, where $\beta_j = \dfrac{\nabla f_{j+1}^T \nabla f_{j+1}}{\nabla f_j^T \nabla f_j}$

{Example}

$$\min f\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = 2x_1^2 + 2x_1x_2 + x_2^2 + x_1 - x_2$$

1. ∇f

$$\nabla f = \begin{bmatrix} 4x_1 + 2x_2 + 1 \\ 2x_1 + 4x_2 - 1 \end{bmatrix}$$

2. $S_1 = -\nabla f(x_1)$, $x_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

$$S_1 = -\begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

3. Find λ^* s.t. $f(x_1 + \lambda^* S_1) = f(x_2)$

$$f\left(\begin{bmatrix} -\lambda \\ \lambda \end{bmatrix}\right) = \cancel{2\lambda^2} - \cancel{2\lambda^2} + \lambda^2 - \lambda - \lambda = \lambda^2 - 2\lambda$$

$$\$ f'\left(\begin{bmatrix} \lambda \\ \lambda \end{bmatrix}\right) = 0 \Rightarrow 2\lambda - 2 = 0 \quad \therefore \lambda^* = 1 \quad X_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

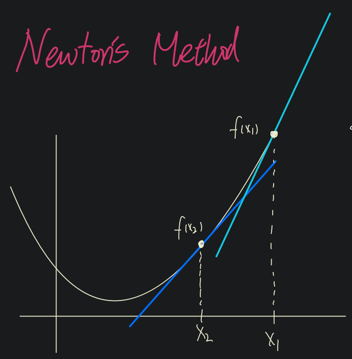
4. $S_2 = -\nabla f_2 + \beta_1 S_1$, where $\beta_1 = \dfrac{\nabla f_2^T \nabla f_2}{\nabla f_1^T \nabla f_1}$

$$\begin{cases} \nabla f_1 = \nabla f(x_1) = -S_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \\ \nabla f_2 = \nabla f(x_2) = \begin{bmatrix} -4+2+1 \\ -2+4-1 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \end{bmatrix} \end{cases} , \quad \beta_1 = \frac{2}{2}$$

$$S_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix} + \begin{bmatrix} -1 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ -2 \end{bmatrix}$$

Unconstrained Nonlinear Programming : N-d min using 2^{nd} order derivative

Newton's Method



$$\text{slope} = \frac{f(x_1) - 0}{x_1 - x_2} = f'(x_1)$$

$$\therefore\ x_1 - x_2 = \frac{f(x_1)}{f'(x_1)} \Rightarrow x_2 = x_1 - \frac{f(x_1)}{f'(x_1)}$$

$$\therefore\ x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)}$$

For opt. $x_{i+1} = x_i - \dfrac{f'(x_i)}{f''(x_i)}$ (Find $f'(x) = 0$)

$$x_{i+1} = x_i - H^{-1}(x_i)\,\nabla f(x_i)$$

{Example}

$f\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = 2x_1^2 + 2x_1x_2 + x_2^2 + x_1 - x_2$, start from $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$

1. ∇f

$$\nabla f = \begin{bmatrix} 4x_1 + 2x_2 + 1 \\ 2x_1 + 2x_2 - 1 \end{bmatrix}$$

2. $H = \begin{bmatrix} \frac{\partial^2}{\partial x^2} f & \frac{\partial^2}{\partial x \partial y} f \\ \frac{\partial^2}{\partial y \partial x} f & \frac{\partial^2}{\partial y^2} f \end{bmatrix}$

$$H = \begin{bmatrix} 4 & 2 \\ 2 & 2 \end{bmatrix}$$

3. $x_2 = x_1 - H^{-1}(x_1)\,\nabla f(x_1)$

$$x_2 = \begin{bmatrix} 0 \\ 0 \end{bmatrix} - \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

$$= \begin{bmatrix} -1 \\ \frac{3}{2} \end{bmatrix}$$

Marquardt Method (Levenberg - Marquardt Method)

Marquardt Method / Levenberg - Marquardt Method

⟨Idea⟩

let $J_i = H_i + \alpha_i I$

$X_{i+1} = X_i - J_i^{-1} \nabla f(X_i)$

Quasi - Newton Method (approximated Hessian Matrix)

⟨Idea⟩

Using approximate H or H^{-1}

By updating a Sym. Matrix B_i (usually $B_i = I_i$)

$$B_{i+1} = B_i + \Delta B_i$$

⟨Method⟩

1. start from 2 diff. points (X_i, X_{i+1}) approximate ∇f using the same ref point X_0

$$\begin{cases} \nabla f_i = \nabla f_0 + H(X_0)(X_i - X_0) & - ① \\ \nabla f_{i+1} \approx \nabla f_0 + H(X_0)(X_{i+1} - X_0) & - ② \end{cases}$$

② - ① ⇒ $\nabla f_{i+1} - \nabla f_i = H(X_0)(X_{i+1} - X_i)$

let $d_i \triangleq X_{i+1} - X_i$, $g_i \triangleq \nabla f_{i+1} - \nabla f_i$

⇒ $g_i = H(X_0)\, d_i$ ⇒ Quasi - Newton Condition

DFP Method

<Idea>

Make B_i to approximat H_i^{-1}

$$\Rightarrow \begin{cases} B_i g_i = d_i \\ B_{i+1}\, g_i = d_i \end{cases}$$

To find ΔB_i, first try Rank 1 update

Assume $\Delta B_i = c\, z z^T$, $c \in$ Const., $z \in M_{n\times 1}$

$$\Rightarrow d_i = (B_i + \Delta B_i)\, g_i$$

$$= B_i g_i + c\, z z^T g_i \qquad \therefore\ z = d_i - B_i g_i\ ,\quad c = \frac{1}{z^T g_i}$$

$$\Delta B_i = \frac{1}{z^T g_i}(d_i - B_i g_i)(d_i - B_i g_i)^T$$ $\Rightarrow$ rank 1 update isn't enough.

Assume $\Delta B_i = c_1 z_1 z_1^T + c_2 z_2 z_2^T$ (In this way, B_i become Sym & Pos-Definite)

$$d_i = (B_i + \Delta B_i)\, g_i$$

$$= B_i g_i + \overbrace{c_1 z_1 z_1^T}^{d_i} + \overbrace{c_2 z_2 z_2^T}^{-B_i g_i}$$

By choose $z_1 = d_i$, $z_2 = B_i g_i$, $c_1 = \frac{1}{z_1^T g_1}$, $c_2 = \frac{-1}{z_2^T g_2}$

$$\Rightarrow \Delta B_i = \frac{d_i d_i^T}{z_1^T g_i} - \frac{B_i g_i}{z_2^T g_i}(B_i g_i)^T$$

⟨⟨ Method ⟩⟩

$$B_{i+1} = B_i + M_i + N_i$$

$$M_i = \lambda_i^* \frac{S_i S_i^T}{S_i^T g_i}\ ,\quad N_i = -\frac{(B_i g_i)(B_i g_i)^T}{g_i^T B_i g_i}$$

{Example}

$$\min f\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = 2x_1^2 + 2x_1x_2 + x_2^2 + x_1 - x_2, \quad \text{start point } \overset{\overset{x_1 =}{\vee}}{\begin{bmatrix} 0 \\ 0 \end{bmatrix}}, \text{ with } B_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

1. To $Q(x)$

$$Q(x) = \frac{1}{2} x^T A x + Bx + C$$

$$A = \begin{bmatrix} 4 & 2 \\ 2 & 2 \end{bmatrix}, \quad B = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad C = 0$$

2. $S_1 = -B_1 \nabla f(x_1)$, $\nabla f = Ax + b$

$$S_1 = -\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\left(\begin{bmatrix} 4 & 2 \\ 2 & 2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 1 \\ -1 \end{bmatrix}\right) = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

3. Find λ_1^* to min. $f(x_1 + \lambda_1 S_1)$

$$f(x_1 + \lambda_1 S_1) = f\left(\begin{bmatrix} 0 - \lambda_1 \\ 0 + \lambda_1 \end{bmatrix}\right) = 2\cancel{\lambda_1^2} - \cancel{2\lambda_1^2} + \lambda_1^2 - \lambda_1 - \lambda_1 = \lambda_1^2 - 2\lambda_1$$

let $f'(x_1 + \lambda_1 S_1) = 0 \Rightarrow 2\lambda_1 - 2 = 0 \quad \therefore \lambda_1^* = 1$

$$\therefore x_2 = \begin{bmatrix} 0 - 1 \\ 0 + 1 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

4. $S_2 = -B_2 \nabla f(x_2)$, $B_2 = B_1 + M_1 + N_1$

$$B_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \lambda_1^* \frac{S_1 S_1^T}{S_1^T g_1} - \frac{(B_1 g_1)(B_1 g_1)^T}{g_1^T B_1 g_1} \qquad \therefore B_2 = \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{3}{2} \end{bmatrix}$$

$$g_1 = \nabla f(x_2) - \nabla f(x_1)$$

$$= \begin{bmatrix} -1 \\ -1 \end{bmatrix} - \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} -2 \\ 0 \end{bmatrix}$$

$$M_1 = \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{bmatrix}, \quad N_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

$$S_2 = -\begin{bmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{3}{2} \end{bmatrix} \begin{bmatrix} -1 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Regression

Known : $X = [X_1, \cdots X_n]$
$y = [y_1, \cdots y_n]$

linear regression: $y_i = mX_i + b$

By $\min f(m,b) = \sum_{i=1}^{n} \left(y_i - (mX_i + b)\right)^2$

<Method>

$$\begin{cases} \frac{\partial}{\partial m} f = 2\sum (y_i - (mX_i + b))(-X_i) = 0 & — ① \\ \frac{\partial}{\partial b} f = 2\sum (y_i - (mX_i + b))(-1) = 0 & — ② \end{cases}$$

def $\sigma_x = \sum X_i$, $\sigma_y = \sum y_i$, $\sigma_{xy} = \sum X_i y_i$, $\sigma_{xx} = \sum X_i^2$

$$\Rightarrow \begin{cases} 2(m\sigma_{xx} + b\sigma_x + \sigma_{xy}) = 0 \\ 2(m\sigma_x + bn - \sigma_y) = 0 \end{cases}, \quad m = \frac{\sigma_x\sigma_y - n\sigma_{xy}}{\sigma_x^2 - n\sigma_{xx}}, \quad b = \frac{\sigma_y - m\sigma_x}{n}$$

Polynomial Regression : $y_i = a_0 x^0 + a_1 x^1 + a_2 x^2 + \cdots + a_n x^n$

By $\min f(a) = \sum_{i=1}^{n} \left(y_i - (a_0 + a_1 x + \cdots + a_n x^n)\right)^2$

<Method>

$$
\begin{cases}
\frac{\partial}{\partial a_0} f = 2\sum \left(y_1 - (a_0 + \cdots + a_n x^n)\right)(-1) \\
\frac{\partial}{\partial a_1} f = 2\sum \left(\cdots \right)(-x_i) = 0 \\
\vdots \\
\frac{\partial}{\partial a_n} f = 2\sum \left(\cdots \right)(-x_i^n) = 0
\end{cases}
$$

$$
\Rightarrow
\begin{cases}
a_0 n + a_1 \sum x_1 + \cdots + a_n \sum x_1^n = \sum y_i \\
a_0 \sum x_1 + a_1 \sum_1^2 + \cdots + a_n \sum x_1^{n+1} = \sum x_i y_i \\
a_n \sum x_i^n + a_1 \sum x_i^{n+1} + \cdots + a_n \sum x^{2n} = \sum x_i^n y_j
\end{cases}
$$

$$
\Rightarrow
\begin{bmatrix}
n & \sum x_i & \sum x_i^2 & \cdots & \sum x_i^n \\
\sum x_i & & \ddots & & \vdots \\
\vdots & & & & \\
\sum x_i^n & \cdots & & & \sum x_i^{2n}
\end{bmatrix}
\begin{bmatrix}
a_0 \\ \vdots \\ a_n
\end{bmatrix}
=
\begin{bmatrix}
\sum y_i \\ \sum x_i y_i \\ \vdots \\ \sum x_i^n y_i
\end{bmatrix}
$$

let
$$
X = \begin{bmatrix}
1 & x_1 & x_1^2 & \cdots & x_1^k \\
1 & & \ddots & & \\
1 & & & \ddots & \\
1 & x_n & x_n^2 & \cdots & x_n^k
\end{bmatrix}, \quad
a = \begin{bmatrix}
a_0 \\ \vdots \\ a_n
\end{bmatrix}
$$

$$X^T X a = X y, \quad a = (X^T X)^{-1} X^T$$

Indirect methods for constrained opt.

Transformation Method

1. $\min f(x)$ s.t. $x \geq 0$

   def $x = y^2$ $\Rightarrow \min f(y)$ without constraint

2. $\min f(x)$ s.t. $-1 \leq x \leq 1$

   def $x = \sin(y)$ $\Rightarrow$ . . .

Penalty func. method

<Idea>

In min., constraint violation cause the obj func to incr.

<Method>

Transform the $f(x)$ into an unconstrained one

by adding the penalty func

$$\varphi(x, r_k) = \overset{\text{original}}{\underline{f(x)}} + r_k \cdot G(\overset{\text{constraint}}{\underline{g(x)}})$$

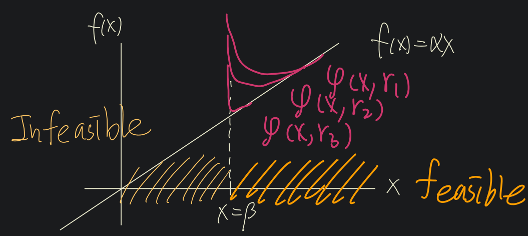
Interior method : $\varphi(x, r_k) = f(x) - \dfrac{r_k}{g(x)}$, $r \to \infty$

Exterior method : $\varphi(x, r_k) = f(x) + r_k [\max(g(x), 0)]^2$, $r \to \infty$

Example

min $\boxed{f(x) = \alpha x}$ s.t. $x \geq \beta$

def $\boxed{g(x) = \beta - x} \leq 0$



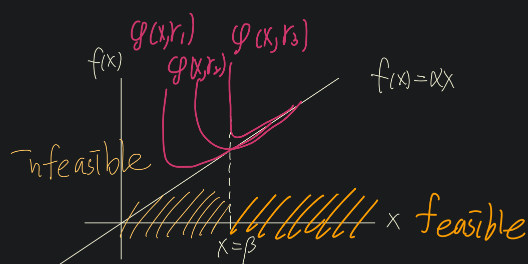
Interior method

$$\varphi(x, r_k) = \underbrace{\alpha x}_{f(x)} - \frac{r_k}{\underbrace{\boxed{\beta - x}}_{g(x)}}$$

To find the min let $\frac{\partial}{\partial x}\varphi = 0$

$$\frac{\partial}{\partial x}\varphi = \alpha - \frac{r_1}{(\beta - x)^2} = 0 \quad \therefore x = \beta \pm \sqrt{\frac{r_1}{\alpha}} \quad (\text{neg. is not feasible})$$

$$= \beta + \sqrt{\frac{r_1}{\alpha}}$$



Exterior method

$$\varphi(x, r_k) = \underbrace{\alpha x}_{f(x)} + r_k\left[\max(\underbrace{\boxed{\beta - x}}_{g(x)}, 0)\right]^2$$

$$\varphi = \alpha x + r_k(\beta - x)^2$$

$$\frac{\partial}{\partial x}\varphi = \alpha + 2 r_k(\beta - x)(-1) = 0$$

$$x = \frac{2r_k\beta - \alpha}{2r_k}, \quad x \to \beta \text{ for } r_k \to \infty$$

< Method >

start from $k=1$, $r_1=1$ choose x_i

$$\min \varphi(x, r_1) = f(x) + r_1 \Sigma (\max(g_j(x), 0))^2$$

find $x_{min}^{(k=1)}$

Next, $k=2$, $r_2 = 10 r_1$ choose $x_1 = x_{min}^{(k=1)}$

$$\min \varphi(x, r_2) = f(x) + r_2 \Sigma (\max(g_j(x), 0))^2$$

find $x_{min}^{(k=2)}$

$\vdots$

stop when stop criterion is meet

ALM (Augmented Lagrange Multiplier)

find $x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$, to min $f(x)$ s.t. $h_j(x) = 0$, $j = 1, \cdots p$

$$L(x, \lambda) = f(x) + \sum_{j=1}^{p} \lambda_j h_j(x) = f(x) + \lambda^T h$$

$$\begin{cases} \frac{\partial}{\partial x_i} L = \frac{\partial}{\partial x_i} f + \sum_{j=1}^{p} \lambda_j \frac{\partial}{\partial x_i} h_j = 0, \quad i = 1, \cdots n & - 1 \\ \frac{\partial}{\partial \lambda_j} L = h_j(x) = 0, \quad j = 1, \cdots p & - 2 \end{cases}$$

$$A(x, \lambda, r_k) = f(x) + \sum_{j=1}^{p} \lambda_j h_j(x) + r_k \sum_{j=1}^{p} (h_j(x))^2$$

$$\frac{\partial}{\partial x_i} A = \frac{\partial}{\partial x_i} f + \sum_{j=1}^{p} (\lambda_j + 2 r_k h_j) \frac{\partial}{\partial x} h_j = 0, \quad i = 1, \cdots n \quad - 3$$

①&③ ⇒ $\underline{\lambda_j \leftarrow \lambda_j + 2r_k h_j}$

In each iter, r_k doesn't need to big very large

{Example}

$f(x) = 6x_1^2 + 4x_1x_2 + 3x_2^2$ s.t. $h(x) = x_1 + x_2 - 5 = 0$

$$A(x, \lambda, r_k) = f(x) + \lambda h(x) + r_k (h(x))^2$$

$$= 6x_1^2 + 4x_1x_2 + 3x_2^2 + \lambda(x_1 + x_2 - 5) + r_k(x_1 + x_2 - 5)^2$$

$$\begin{cases} \frac{\partial}{\partial x_1} A = 12x_1 + 4x_2 + \lambda + 2r_k(x_1 + x_2 - 5) = 0 \\ \quad = (12 + 2r_k)x_1 + (4 + 2r_k)x_2 = 10r_k - \lambda \\ \frac{\partial}{\partial x_2} A = (4 + 2r_k)x_1 + (6 + 2r_k)x_2 = 10r_k - \lambda \end{cases}$$

$$\therefore x_1 = \frac{10r_k - \lambda}{28 + 10r_1}, \quad x_2 = 4x_1$$

Iter 1 $(k=1)$, $r_1 = 1$

let $\lambda^{k=1} = 0$ $\quad \therefore x_1 = \frac{10r_k}{28+10} = \frac{10}{38}$

$$\therefore x_2 = \frac{40}{38}$$

$$h^{(k=1)} = x_1 + x_2 - 5 = -3.684$$

$$\lambda^{(k=2)} = \lambda^{(k=1)} + 2r_1 h_1^{(k=1)} = 0 + 2 \cdot 1 \cdot (-3.684) = -7.368$$

Iter 2 $(k=2)$, $r_2 = 10 \cdot r_1 = 10$

$$x_1 = \frac{10r_2 - \lambda^{k=2}}{28 + 10r_2} = 0.839 \quad , \quad x_2 = 3.355$$

$$h_1^{(k=2)} = x_1 + x_2 - 5 = -0.806$$

$$\lambda^{(k=3)} = \lambda^{(k=2)} + 2r_2 h_1^{(k=2)}$$

$$= -23.488$$

Iter 3 $(k=3)$

$$x_1 = 0.998, \quad x_2 = 4x_1$$

$$h_1^{(k=3)} = 0 \qquad \Rightarrow \quad (x_1, x_2) = (1, 4) \quad \#$$

< Inequality constraints >

$$\min f(x) \quad \text{s.t.} \quad g_j(x) \le 0 \quad , \quad j = 1, \cdots m$$

$$A(x, \lambda, r_k) = f(x) + \sum_{j=1}^{m} \lambda_j \alpha_j + r_k \sum_{j=1}^{m} \alpha_j^2$$

$$\alpha_j = \max\left(g_j(x), \frac{-\lambda_j}{2r_k}\right)$$

$$\lambda_j^{(k+1)} = \lambda_j^{(k)} + 2r_k \alpha_j^{(k)}$$

For $g_j(x) \le 0 \Rightarrow g_j(x) + y_j^2 = 0$

$$A(x, \lambda, r_k) = f(x) + \Sigma \lambda_j (g_j(x) + y_j^2) + r_k \Sigma (g_j(x) + y_j)^2$$

{Example}

$\min f(x) = (x_1-1)^2 + (x_2-2)^2$ s.t. $x_1 + x_2 \le 2$

$$A(x, \lambda, r_k) = (x_1-1)^2 + (x_2-2)^2 + \left\langle \frac{\lambda}{2r_k} + x_1 + x_2 - 2 \right\rangle^2$$

$$\begin{cases} \frac{\partial}{\partial x_1} A = 2(x_1-1) + 2r_k \left\langle \frac{\lambda}{2r_k} + x_1 + x_2 - 2 \right\rangle = 0 & \text{— ①} \\ \frac{\partial}{\partial x_2} A = 2(x_2-2) + 2r_k \left\langle \frac{\lambda}{2r_k} + x_1 + x_2 - 2 \right\rangle = 0 & \text{— ②} \end{cases}$$

② − ① $\Rightarrow x_2 - x_1 - 1 = 0$ — ③

Iter 1 ($k=1$)

let $r_1 = 1$, $\lambda^{(k=1)} = 0$

Assume $\frac{\lambda}{2r_k} + x_1 + x_2 - 2 \ge 0$

① $\Rightarrow 2(x_1 - 1) + 2(x_1 + x_2 - 2) = 0$

$\Rightarrow 2x_1 - 2 + 2x_1 + 2x_2 - 4 = 0$

$\Rightarrow x_1 = \frac{2}{3}, \; x_2 = \frac{5}{3}$

Contradiction!

if $\frac{\lambda}{2r_k} + x_1 + x_2 - 2 < 0 \Rightarrow \begin{cases} 2(x_1-1) = 0 \\ 2(x_2-2) = 0 \end{cases} \Rightarrow \cancel{\begin{cases} x_1 = 1 \\ x_2 = 2 \end{cases}}$

$$\lambda^{(k=2)} = \langle \lambda^{k=1} + 2r_1 g(x) \rangle$$

$$= \langle 0 + 2 \cdot 1 \cdot \tfrac{1}{3} \rangle = \frac{2}{3}$$

Iter 2 ($k=2$)

let $r_2 = 10r_1 = 10$

$$\therefore \text{①} \Rightarrow 2(x_1-1) + 2 \cdot 10 \cdot \langle \frac{\frac{2}{3}}{20} + x_1 + x_2 - 2 \rangle = 0$$

$$\Rightarrow 2x_1 - 2 + \frac{2}{3} + 20x_1 + 20x_2 - 40 = 0$$

$$\text{①,③} \Rightarrow x_1 = \frac{22 - \frac{2}{3}}{42}, \quad x_2 = x_1 + 1$$