Lecture 19

Saturday, 12 March 2022 9:35 AM

Recall
$$\left[B_{k+1} = B_k - \underbrace{\frac{B_k (\Delta x^{(k)})(\Delta x^{(k)})^T B_k}{(\Delta x^{(k)})^T B_k (\Delta x^{(k)})} + \frac{(\Delta g^{(k)})(\Delta g^{(k)})^T}{(\Delta g^{(k)})^T \Delta x^{(k)}}} \right.$$

$$H_{k+1} = B_{k+1}^{-1}.$$

Woodbury formula

$$(A + \underbrace{UCV}_{B})^{-1} = \underbrace{A^{-1} - A^{-1} U (C^{-1} + VA^{-1}U)^{-1}}_{B^{-1}} VA^{-1}$$

[Verify $BB^{-1} = I$]

Remove 'k' for notational convenience

$$B_+ = B - \frac{B \overbrace{(\Delta x)}^{a} \overbrace{(\Delta x)^T}^{a^T} B}{\underbrace{(\Delta x)^T}_{a} B \underbrace{(\Delta x)}_{a}} + \frac{\overbrace{(\Delta g)}^{b} \overbrace{(\Delta g)^T}^{b^T}}{\underbrace{(\Delta g)^T}_{b^T} \underbrace{(\Delta x)}_{a}}$$

$$\underbrace{\begin{bmatrix} B\Delta x & \Delta g \end{bmatrix}}_{U} \underbrace{\begin{bmatrix} \frac{-1}{(\Delta x)^T B (\Delta x)} & 0 \\ 0 & \frac{1}{(\Delta g)^T (\Delta x)} \end{bmatrix}}_{C} \underbrace{\begin{bmatrix} (\Delta x)^T B \\ (\Delta g)^T \end{bmatrix}}_{V}$$

$$A = B$$
$$U = [B\Delta x \quad \Delta g]$$
$$C = \begin{bmatrix} -\frac{1}{(\Delta x)^T B (\Delta x)} & 0 \\ 0 & \frac{1}{(\Delta g)^T \Delta x} \end{bmatrix}$$
$$V = \begin{bmatrix} (\Delta x)^T B \\ (\Delta g)^T \end{bmatrix}$$

$$C = \begin{bmatrix} d & 0 \\ 0 & e \end{bmatrix} \qquad C^{-1} = \frac{1}{de}\begin{bmatrix} e & 0 \\ 0 & d \end{bmatrix} = \begin{bmatrix} d^{-1} & 0 \\ 0 & e^{-1} \end{bmatrix}$$

$$\hookrightarrow C^{-1} = \begin{bmatrix} -(\Delta x)^T B (\Delta x) & 0 \\ 0 & (\Delta g)^T (\Delta x) \end{bmatrix}$$

Plug in Woodbury formula:

$$(B_+) \; B_{k+1} = \underbrace{\left(I - \frac{(\Delta x^{(k)})(\Delta g^{(k)})^T}{(\Delta g^{(k)})^T (\Delta x^{(k)})}\right) B_k^{-1} \left(I - \frac{(\Delta g^{(k)})(\Delta x^{(k)})^T}{(\Delta g^{(k)})^T \Delta x^{(k)}}\right)}_{P} + \underbrace{\frac{\overbrace{(\Delta x^{(k)})}^{a} \overbrace{(\Delta x^{(k)})^T}^{a^T}}{(\Delta g^{(k)})^T (\Delta x^{(k)})}}_{Q}$$

$\boxed{\text{Tut 7.(1)}}$.

$\boxed{\text{Tut 7 (2)}}$ - Rank two-update preserve positive-definiteness.

$\boxed{\text{Tut 7 (2)}}$ → Rank two-update preserve positive-definiteness.

Hint: $\underline{z}^T B_{k+1} \underline{z} = \underline{z}^T P \underline{z} + \underline{z}^T Q \underline{z}$

$$\left(\underline{z}^T - \frac{z^T (\Delta x^{(k)})(\Delta g^{(k)})^T}{(\Delta g^{(k)})^T (\Delta x^{(k)})} \right) B_k^{-1} \left(\underline{z} - \frac{z \Delta g^{(k)} (\Delta x^{(k)})^T}{(\Delta g^{(k)})^T (\Delta x^{(k)})} \right)$$

$$= \underbrace{\left(\underline{z} - \frac{\Delta g^{(k)} \Delta x^{(k)} z}{(\Delta g^{(k)})^T (\Delta x^{(k)})} \right)^T}_{\underline{r}} B_k^{-1} \underbrace{\left(\quad \right)}_{\underline{r}}$$

$\underline{r}^T B_k^{-1} \underline{r} > 0$ [Induction]

Use B_k is positive-definite.

$$z^T \frac{\underline{a}\, \underline{a}^T}{\underline{b}\, \underline{a}^T} z = \frac{(\underline{a}^T \underline{z})^2}{\underline{b}\, \underline{a}} > 0$$

→ $(\Delta g^{(k)})^T (\Delta x^{(k)}) \geq 0$?

Hint $\Delta x^{(k)} = x^{(k+1)} - x^{(k)}$
$= \alpha \underline{d}^{(k)}$

$$\boxed{\min_{B_{k+1}} \| B_{k+1}^{-1} - B_k^{-1} \| \quad \text{s.t. symmetry},\ B_{k+1}^{-1} \Delta g^{(i)} = \Delta x^{(i)}}$$ ✓

Tut. Sheet 7 (3) $\left[H_{k+1} \Delta \underline{g}^{(i)} = \Delta \underline{x}^{(i)} \quad 0 \leq i \leq k. \right.$

→ $\boxed{\text{Induction.}}$

$\boxed{\text{BFGS}}$ → Quasi-Newton method

→ positive-definiteness
→ $H_{k+1} \Delta g^{(i)} = \Delta x^{(i)}$
$0 \leq i \leq k.$
→ [Conjugate direction property.

Lagrange / KKT Conditions

Optimization problems with constraints ✓

Equality constraints (Lagrange)

(Karush-Kuhn-Tucker) (KKT) Inequality constraints ✓

↓

$$\begin{aligned} \min \quad & f(\underline{x}) \\ \text{s.t.} \quad & \underline{h}(\underline{x}) = 0 \end{aligned}$$

$\underline{x} \in \mathbb{R}^n$, $\quad f : \mathbb{R}^n \to \mathbb{R}$

$\underline{h} : \mathbb{R}^n \longrightarrow \mathbb{R}^m \qquad (m \le n)$

$$\underline{h} = \begin{bmatrix} h_1(\underline{x}) \\ h_2(\underline{x}) \\ \vdots \\ h_m(\underline{x}) \end{bmatrix} \qquad \boxed{h \in C^{(1)}}$$

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & \underline{h}(x) = 0 \\ & \underline{g}(x) \le 0 \end{aligned}$$

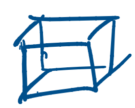
$f : \mathbb{R}^n \to \mathbb{R}$

$\underline{h} : \mathbb{R}^n \to \mathbb{R}^m \qquad (m \le n)$

$\underline{g} : \mathbb{R}^n \to \mathbb{R}^p$, $\quad \boxed{h, g \in C^{(1)}}$

Revise: Surface, Curve on S, Tangent space, Normal space. [Section 20.3].

Examples (1) Given a fixed area of a cardboard, construct a closed cardboard box with max. volume.



$\begin{bmatrix} m = 1 \\ n = 3 \end{bmatrix}$

$$\begin{aligned} \max \quad & x_1 x_2 x_3 \qquad \boxed{f(x)} = -x_1 x_2 x_3 \\ \text{s.t.} \quad & x_1 x_2 + x_2 x_3 + x_1 x_3 = A/2 \end{aligned} \qquad \boxed{x_1, x_2, x_3 > 0} \qquad \boxed{\begin{aligned} f &: \mathbb{R}^3 \to \mathbb{R}. \\ h &: \mathbb{R}^3 \to \mathbb{R}. \end{aligned}}$$

$\min \ f(\underline{x}) \quad \text{s.t.} \quad h(\underline{x}) = (x_1 x_2 + x_2 x_3 + x_1 x_3 - A/2) = 0$

$\boxed{\nabla f(\underline{x}) + \lambda \, \nabla h(\underline{x}) = 0} \longrightarrow$ Lagrange condition.

$\boxed{G = f + \lambda h} = 0$

$$\left\{ \begin{aligned} & -\begin{bmatrix} x_2 x_3 \\ x_1 x_3 \\ x_1 x_2 \end{bmatrix} + \lambda \begin{bmatrix} x_2 + x_3 \\ x_1 + x_3 \\ x_1 + x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \\ & x_1 x_2 + x_2 x_3 + x_1 x_3 = A/2. \\ & \lambda \in \mathbb{R}. \end{aligned} \right.$$

$x_1 = 0$

$\lambda \ne 0 \qquad \lambda = 0$

Contradicts Constraint

$\ldots(+\lambda) \qquad x_2 x_3 = A/2.$

$x_1, x_2, x_3, \lambda \neq 0$

$\underline{x_1 = 0}$ | $(\lambda \neq 0)$ $\underline{x_2 x_3 = A/2}$.

$\lambda \neq 0$ $\lambda = 0$ Contradicts Constraints

$\lambda x_3 = 0$
$\lambda x_2 = 0$ || $\underline{x_2 x_3 = 0}$ Contradicts Constraint $x_2 x_3 = A/2$.

$x_1 \neq 0$.

$\lambda = 0$ $x_2 x_3 = 0$ Contradicts the Constraint $x_2 x_3 = A/2$.

$\lambda = 0$ $\Rightarrow \underline{x_1 x_2 + x_2 x_3 + x_1 x_3 = 0}$

Contradicts the Constraint.

$\downarrow$

$\left.\begin{array}{l} x_2 x_3 = 0 \\ x_1 x_3 = 0 \\ x_1 x_2 = 0 \end{array}\right]$

$[\, x_2 x_3 - \lambda (x_2 + x_3) = 0 \,]\, x_1$

$[\, x_1 x_3 - \lambda (x_1 + x_3) = 0 \,]\, x_2$

$[\, x_1 x_2 - \lambda (x_1 + x_2) = 0 \,]\, x_3$

$x_1 x_2 + x_2 x_3 + x_3 x_1 = A/2$

$x_1 x_2 x_3 - \lambda (x_1 x_2 + x_1 x_3) = 0$
$x_1 x_2 x_3 - \lambda (x_1 x_2 + x_2 x_3) = 0$

$\lambda (x_1 - x_2)\, x_3 = 0$

$\Downarrow$

$\boxed{x_1 = x_2}$

III^{ly} $x_2 = x_3$

$3x_1^2 = A/2 \Rightarrow x_1^2 = A/6 \Rightarrow \boxed{x_1 = \sqrt{A/6} = x_2 = x_3}$

[$\underline{x_1, x_2, x_3 > 0 \text{ is ignored}}$; however, this is taken taken care of !]

Recall:

Second order "necessary" vs Second order "Sufficient."

$\boxed{1 \text{ variable}}$

$f'(x^*) = 0$ [first order necessary] || [x^* can be a minimum, maximum, } "critical point"

x^* is minimum

$f''(x^*) \geq 0$ | $f''(x^*) > 0$.

second-order necessary || $\downarrow$

guarantee's x^* is a

$\downarrow$

... doesn't guarantee

Counter-example

$f(x) = x^3$

$f''(0) = 0$

second-order necessary doesn't guarantee that x^* is a minimum

guarantee's x^* is a minimum.

We want to do this in the context of "Lagrangian".

$\boxed{\text{Ex. 2}}$ $\min f(x) := x_1^2 + x_2^2$

s.t. $\left\{ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} : \underbrace{x_1^2 + 2x_2^2 - 1}_{h(x)} = 0 \right\} \to$ (ellipse).

$$\left[\begin{array}{l} \nabla f(x) + \lambda h(x) = 0 \\ x_1^2 + 2x_2^2 - 1 = 0 \end{array} \right]$$

$$\left[x_1 = \begin{pmatrix} 0 \\ 1/\sqrt{2} \end{pmatrix} \quad x_2 = \begin{pmatrix} 0 \\ -1/\sqrt{2} \end{pmatrix} \quad x_3 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad x_4 = \begin{pmatrix} -1 \\ 0 \end{pmatrix} \right\}$$

Justification of $\to \left[\begin{array}{l} \nabla f(x) + \lambda \nabla h(x) = 0 \\ h(x) = 0 \end{array} \right]$

$\to$ Do we have second-order conditions?

[Example where $f \notin C^1$, $h \notin C^1$ but still we can consider min/max.]