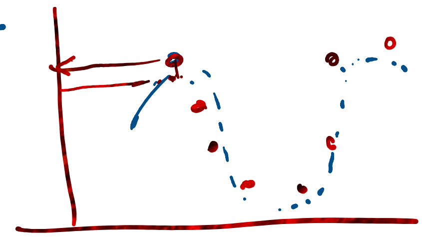
Lecture 8

Thursday, 27 January 2022 1:54 PM

Non-linear least 'square' problem

(differentiability)



$$|y_i - A\sin(\omega t_i + \phi)|$$

$$r_i(\underline{x}) = y_i - A\sin(\omega t_i + \phi)$$

$$\underline{x} = \begin{bmatrix} A \\ \omega \\ \phi \end{bmatrix}$$

Newton's method

$$f(\underline{x}) = \sum_{i=1}^{m} r_i(\underline{x})^2$$

$$= \underline{r}^T(\underline{x})\,\underline{r}(\underline{x})$$

$$f(\underline{x}) = r_1(\underline{x})^2 + r_2(\underline{x})^2 + \dots + r_m(\underline{x})^2$$

$$\boxed{\underline{x}^{(k+1)} = \underline{x}^{(k)} - \left(F(\underline{x}^{(k)})\right)^{-1} \nabla f(\underline{x}^{(k)})}$$

Applying Newton's Method: for non-linear least square

$\underline{x} \in \mathbb{R}^n$

$r_i(x)$, $1 \le i \le m$.

$$\nabla f(\underline{x}) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix} = \begin{bmatrix} 2 r_1(\underline{x}) \frac{\partial r_1}{\partial x_1} + \dots + 2 r_m(\underline{x}) \frac{\partial r_m}{\partial x_1} \\ \vdots \\ 2 r_1(\underline{x}) \frac{\partial r_1}{\partial x_n} + \dots + 2 r_m(\underline{x}) \frac{\partial r_m}{\partial x_n} \end{bmatrix}$$

$$f(\underline{x}) = r_1(x)^2 + \dots + r_m(x)^2$$

jth component of $\nabla f(x)$

$$(\nabla f(x))_j = 2\sum_{i=1}^{m} r_i(x) \frac{\partial r_i}{\partial x_j}(x)$$

Jacobian matrix

$$\underline{r}(x) = (r_1(x), \dots r_m(x))^T$$

$$J(x) = \begin{bmatrix} \frac{\partial r_1}{\partial x_1} & \frac{\partial r_1}{\partial x_2} & \dots & \frac{\partial r_1}{\partial x_n} \\ \vdots & & & \vdots \\ \frac{\partial r_m}{\partial x_1} & \frac{\partial r_m}{\partial x_2} & \dots & \frac{\partial r_m}{\partial x_n} \end{bmatrix}_{m\times n}$$

$$\boxed{\underset{n\times 1}{\nabla f(\underline{x})} = 2\, \underset{n\times m}{J(\underline{x})^T}\, \underset{m\times 1}{\underline{r}(x)}}$$

Hessian of f is computed next

... matrix of f is

Hessian of f is computed next

$(k, j)^{th}$ component of the Hessian matrix of f is

$$\frac{\partial^2 f}{\partial x_k \partial x_j}(\underline{x}) = \frac{\partial}{\partial x_k}\left(\frac{\partial f}{\partial x_j}\right)$$

$$= \frac{\partial}{\partial x_k}\left(2 \sum_{i=1}^{m} r_i(\underline{x}) \frac{\partial r_i}{\partial x_j}(\underline{x})\right)$$

$$= \underbrace{2 \sum_{i=1}^{m} \frac{\partial r_i}{\partial x_k} \frac{\partial r_i}{\partial x_j}}_{J(\underline{x})^T J(\underline{x})} + \underbrace{2 \sum_{i=1}^{m} r_i(\underline{x}) \frac{\partial^2 r_i}{\partial x_k \partial x_j}}_{}$$

$S(\underline{x})$ which has $(k,j)^{th}$ component as above

$$\boxed{F(\underline{x}) = 2\left[J(\underline{x})^T J(\underline{x}) + S(\underline{x})\right]}$$

Newton's method

$$x^{(k+1)} = x^{(k)} - \cancel{\tfrac{1}{2}}\left[J(\underline{x})^T J(\underline{x}) + S(\underline{x})\right]^{-1} \cancel{2}\, J(\underline{x})^T r(\underline{x})$$

$$\boxed{x^{(k+1)} = x^{(k)} - \left[J(\underline{x})^T J(\underline{x}) + S(\underline{x})\right]^{-1} J(\underline{x})^T r(\underline{x})}$$

Invertible → ASSUMPTION

In applications, $S(\underline{x})$ may be ignored if the components are small.

Gauss- Newton's method

$$\underline{x}^{(k+1)} = \underline{x}^{(k)} - \underbrace{\left(J(\underline{x})^T J(\underline{x})\right)}_{}^{-1} J(\underline{x})^T r(\underline{x})$$

don't know if this is positive-definite.

Levenberg- Marquardt Algorithm

Levenberg-Marquardt Algorithm

$$x^{(k+1)} = x^{(k)} - \left(\underbrace{J(x)^T J(x) + \mu_k I}_{\text{(apprx to } S(x)\text{?)}}\right)^{-1} J(x)^T r(x)$$

$(\mu_k \geq 0)$

(Why will this work?
Why is this matrix positive-definite?)

Can be used also if $F(x)$ is not positive-definite

(Hessian)

Let $F = F^T$, $F \not> 0$ (not positive-definite). real but need not be positive.

Let $\boxed{\lambda_1, \ldots, \lambda_n}$ be the eigenvalues of F and $v_1, \ldots, v_n$ be the corresponding eigenvectors.

Let $G = F + \mu I$, $\quad \mu \geq 0$.

$$G v_i = (F + \mu I) v_i$$

$$= F v_i + \mu v_i$$

$$= \lambda_i v_i + \mu v_i$$

$$G v_i = (\lambda_i + \mu) v_i$$

G has eigenvalues $(\lambda_i + \mu)$ and eigenvectors v_i.

Hence if μ is chosen sufficiently large, all eigenvalues of G are positive and $G = F + \mu I$ is positive definite.

$(G = F + \mu_k I)$

definite.

$(G = \nabla^2 f + \mu_k I)$

$$x^{(k+1)} = \underline{x}^{(k)} - G(x^{(k)})^{-1} \nabla f(x^{(k)}).$$

$\boxed{\text{Convergence of Newton's Method}}$

Pre-requisites $\boxed{\text{Chapter 5}}$

① Let $A \in \mathbb{R}^{n\times n}$. Then

$$\lim_{k\to\infty} A^k = 0 \iff \text{eigenvalues } \lambda_i \text{ of } A \text{ satisfy } |\lambda_i(A)| < 1 \quad i=1,\dots,n.$$

(Exercise)

$\Updownarrow$

② The series of $n\times n$ matrices $\underbrace{I_n + A + A^2 + \dots + A^k + \dots}$ converges

$$\iff \lim_{k\to\infty} A^k = 0.$$

In this case, the sum is equal to $\underbrace{(I_n - A)^{-1}}$.

Pf. $\Rightarrow$ is obvious.

$\Leftarrow$ let $\underline{\lim_{k\to\infty} A^k = 0} \Rightarrow |\lambda_i(A)| < 1 \quad i=1,\dots,n$

$\Rightarrow |A - I_n| \neq 0 \Rightarrow (I_n - A)^{-1}$ exists.

Consider

$$(I_n + A + A^2 + \dots + A^k)(I_n - A)$$

$$= I_n + A + A^2 + \dots + A^k - A - A^2 - \dots - A^{k+1} = (I_n - A^{k+1})$$

$$(I_n + A + A^2 + \dots + A^k) = (I_n - A^{k+1})(I_n - A)^{-1}$$

$$(I_n + A + A^2 + \cdots + A^k) = (I_n - A^{k+1})(I_n - A)$$

$$= (I_n - A)^{-1} - \underbrace{A^{k+1}(I_n - A)^{-1}}$$

$$\sum_{j=0}^{k} A^j = (I_n - A)^{-1} - \underbrace{A^{k+1}}(I_n - A)^{-1}$$

$$\boxed{\lim_{k\to\infty} \sum_{j=0}^{\infty} A^j = (I_n - A)^{-1}}$$

Defn: A matrix-valued function $A : \mathbb{R}^n \to \mathbb{R}^{n\times n}$ is **continuous** at $\xi_0 \in \mathbb{R}^n$ if

$$\lim_{\|\xi - \xi_0\| \to 0} \| A(\xi) - A(\xi_0) \| = 0.$$

$\boxed{3}$ Let $A : \mathbb{R}^n \to \mathbb{R}^{n\times n}$ be a $n\times n$ matrix-valued function that is continuous at ξ_0. If $(A(\xi_0))^{-1}$ exists, then $(A(\xi))^{-1}$ exists for ξ sufficiently close to ξ_0 and $(A(\xi_0))^{-1}$ is continuous at ξ_0.

Newton's Convergence → 3 → 2 → 1

Prerequisite

Prerequisite

$f \in C^{(3)}$, $\boxed{x^{(0)} \leftrightarrow x^{(*)}}$

Proof:

Reference → Newton Chapter 9.