Conjugate Gradient Method

5/26/92

§. Direction Set Method

We expand a function f around the origin $\mathbb{P} \in \mathbb{R}^N$.

$$f(\mathbb{x}) = f(\mathbb{P}) + \sum_{i=1}^{N} x_i \frac{\partial f}{\partial P_i} + \sum_{i,j=1}^{N} \frac{x_i x_j}{2} \frac{\partial^2 f}{\partial x_i \partial x_j} + \cdots \tag{1}$$

$$\approx c - \mathbb{b}\cdot\mathbb{x} + \frac{1}{2}\mathbb{x}\cdot\mathbb{A}\cdot\mathbb{x} \tag{2}$$

where

$$c = f(\mathbb{P}), \quad \mathbb{b} = -\nabla f(\mathbb{P}), \quad [\mathbb{A}]_{ij} = \frac{\partial^2}{\partial P_i \partial P_j} f(\mathbb{P}) \tag{3}$$

In the quadratic form, Eq. (2), the gradient of f at $\mathbb{x}$ is calculated as

$$\nabla f = \mathbb{A}\cdot\mathbb{x} - \mathbb{b} \tag{4}$$

Minimum point is found as follows: Suppose $\{\mathbb{e}_i \mid i = 1, \dots, N\}$ is a lineary independent set of basic vectors. Then the minimum point $\mathbb{x} = \sum_{i=1}^{N} \lambda_i \mathbb{e}_i$ satisfies

$$\mathbb{e}_j \cdot \nabla f\left(\mathbb{x} = \sum_{i=1}^{N} \lambda_i \mathbb{e}_i\right) = 0 \quad (j = 1, \dots, N) \tag{5}$$

Suppose we have found a point $\mathbb{Q}$, where
$\mathbb{u}\cdot\nabla f(\mathbb{Q}) = 0$.

We now search for the line minimum along the direction $\mathbb{Q}+\lambda\mathbb{V}$, i.e., $\mathbb{v}\cdot\nabla f(\mathbb{Q}+\lambda\mathbb{v})$. For the new point to be also the line minimum, i.e., $\mathbb{u}\cdot\nabla f(\mathbb{Q}+\lambda\mathbb{v})$, $\mathbb{u}$ & $\mathbb{v}$ must satisfy the following relation.

$$\underbrace{\mathbb{U}\cdot\nabla f(\mathbb{Q}+\lambda\mathbb{V})}_{\nabla f(\mathbb{Q})\;+\;\lambda\,\mathbb{A}\cdot\mathbb{V}\quad(\because\ \text{Eq. }(4))}$$

$$=\lambda\,\mathbb{U}\cdot\mathbb{A}\cdot\mathbb{V}$$

(Conjugate Direction)

If $\mathbb{U}$ & $\mathbb{V}$ ($\in \mathbb{R}^N$) are conjugate, i.e.,

$$\mathbb{U}\cdot\mathbb{A}\cdot\mathbb{V} = 0, \tag{6}$$

then a line minimization along $\mathbb{V}$, starting from a line minimum along $\mathbb{U}$, achieves a minimization along both $\mathbb{U}$ & $\mathbb{V}$.

§. Conjugate Gradient Method

(Th : Gram-Schmidt Bi-Orthogonalization)

Let $\mathbb{A}$ be a symmetric, positive definite, $N \times N$ matrix. Let $\forall \mathbf{g}_0 \in \mathbb{R}^N$ and $\mathbf{h}_0 = \mathbf{g}_0$. For $i = 0, 1, 2, \dots$, define the two sequences of vectors

$$\begin{cases} \mathbf{g}_{i+1} = \mathbf{g}_i - \lambda_i \mathbb{A} \cdot \mathbf{h}_i & (7) \\ \mathbf{h}_{i+1} = \mathbf{g}_{i+1} + \gamma_i \mathbf{h}_i & (8) \end{cases}$$

where

$$\begin{cases} \lambda_i = \dfrac{\mathbf{g}_i \cdot \mathbf{g}_i}{\mathbf{g}_i \cdot \mathbb{A} \cdot \mathbf{h}_i} & (9) \\ \gamma_i = - \dfrac{\mathbf{g}_{i+1} \cdot \mathbb{A} \cdot \mathbf{h}_i}{\mathbf{h}_i \cdot \mathbb{A} \cdot \mathbf{h}_i} & (10) \end{cases}$$

then for $i \neq j$,

$$\begin{cases} \mathbf{g}_i \cdot \mathbf{g}_j = 0 & (11) \\ \mathbf{h}_i \cdot \mathbb{A} \cdot \mathbf{h}_j = 0 & (12) \end{cases}$$

$\because$ $\mathbf{g}_{i+1} \cdot \mathbf{g}_i = \mathbf{h}_{i+1} \cdot \mathbb{A} \cdot \mathbf{h}_i = 0$ holds by construction:

$$\mathbf{g}_{i+1} \cdot \mathbf{g}_i = \mathbf{g}_i \cdot \mathbf{g}_i - \frac{\mathbf{g}_i \cdot \mathbf{g}_i}{\cancel{\mathbf{g}_i \cdot \mathbb{A} \cdot \mathbf{h}_i}} \cancel{\mathbf{h}_i \cdot \mathbb{A} \cdot \mathbf{g}_i} = 0$$

$$\mathbf{h}_{i+1} \cdot \mathbb{A} \cdot \mathbf{h}_i = \mathbf{g}_{i+1} \cdot \mathbb{A} \cdot \mathbf{h}_i - \frac{\mathbf{h}_i \cdot \mathbb{A} \cdot \mathbf{g}_{i+1}}{\cancel{\mathbf{h}_i \cdot \mathbb{A} \cdot \mathbf{h}_i}} \cancel{\mathbf{h}_i \cdot \mathbb{A} \cdot \mathbf{h}_i} = 0$$

Suppose Eqs. (11) & (12) hold for $i, j \leq n$, and we construct $\mathbb{g}_{n+1}$ & $\mathbb{h}_{n+1}$ as Eqs. (7) & (8). Then, for $i < n$,

① $$\mathbb{g}_{n+1} \cdot \mathbb{g}_i = \cancel{\mathbb{g}_n \cdot \mathbb{g}_i} - \lambda_n \mathbb{h}_n \cdot \mathbb{A} \cdot \mathbb{g}_i \quad (\because \text{Eq. } (7))$$

$$= \begin{cases} -\lambda_n \mathbb{h}_n \cdot \mathbb{A} \cdot \left(\cancel{\mathbb{h}_i} - \gamma_{i-1} \cancel{\mathbb{h}_{i-1}} \right) = 0 & (i \neq 0, \because \text{Eq. } (8)) \\ -\lambda_n \mathbb{h}_n \cdot \mathbb{A} \cdot \underbrace{\mathbb{g}_0}_{\mathbb{h}_0} = 0 & (i = 0) \end{cases}$$

② $$\mathbb{h}_{n+1} \cdot \mathbb{A} \cdot \mathbb{h}_i = \mathbb{g}_{n+1} \cdot \mathbb{A} \cdot \mathbb{h}_i + \cancel{\gamma_n \mathbb{h}_n \cdot \mathbb{A} \cdot \mathbb{h}_i} \quad (\because \text{Eq. } (8))$$

$$= \mathbb{g}_{n+1} \cdot \frac{\mathbb{g}_i - \mathbb{g}_{i+1}}{\lambda_i} \quad (\because \text{Eq. } (7))$$

$$= 0 \quad (\because n+1 > n > i \ \& \ n+1 > i+1 \text{ from assumption})$$

//

(Lemma)

$$\gamma_i = \frac{\mathbf{g}_{i+1}\cdot\mathbf{g}_{i+1}}{\mathbf{g}_i\cdot\mathbf{g}_i} = \frac{(\mathbf{g}_{i+1}-\mathbf{g}_i)\cdot\mathbf{g}_{i+1}}{\mathbf{g}_i\cdot\mathbf{g}_i} \qquad (13)$$

$$\lambda_i = \frac{\mathbf{g}_i\cdot\mathbf{h}_i}{\mathbf{h}_i\cdot\mathbf{A}\cdot\mathbf{h}_i} \qquad (14)$$

$\because$ ①

$$\gamma_i = -\frac{\mathbf{g}_{i+1}}{\mathbf{h}_i\cdot\mathbf{A}\cdot\mathbf{h}_i}\cdot\frac{\mathbf{g}_i-\mathbf{g}_{i+1}}{\lambda_i} \qquad (\because \text{Eqs. (10) \& (7)})$$

$$= \frac{(\mathbf{g}_{i+1}-\mathbf{g}_i)\cdot\mathbf{g}_{i+1}}{\mathbf{h}_i\cdot\mathbf{A}\cdot\mathbf{h}_i}\,\frac{\mathbf{g}_i\cdot\mathbf{A}\cdot\mathbf{h}_i}{\mathbf{g}_i\cdot\mathbf{g}_i} \qquad (\because \text{Eq. (9)})$$

Here,

$$\mathbf{h}_i\cdot\mathbf{A}\cdot\mathbf{h}_i = (\mathbf{g}_i - \gamma_{i-1}\cancel{\mathbf{h}_{i-1}})\cdot\mathbf{A}\cdot\mathbf{h}_i \qquad (\because \text{Eq. (8)})$$

$$= \mathbf{g}_i\cdot\mathbf{A}\cdot\mathbf{h}_i$$

$$\therefore\ \gamma_i = \frac{(\mathbf{g}_{i+1}-\mathbf{g}_i)\cdot\mathbf{g}_{i+1}}{\mathbf{g}_i\cdot\mathbf{g}_i}$$

②

$$\lambda_i = \frac{\mathbf{g}_i\cdot\mathbf{g}_i}{\mathbf{g}_i\cdot\mathbf{A}\cdot\mathbf{h}_i}$$

$\hookrightarrow \mathbf{h}_i\cdot\mathbf{A}\cdot\mathbf{h}_i$ (see above)

Here,

$$\mathbf{g}_i\cdot\mathbf{g}_i = \mathbf{g}_i\cdot(\mathbf{h}_i - \gamma_{i-1}\mathbf{h}_{i-1}) \qquad (\because \text{Eq. (8)})$$

$$= \mathbf{g}_i\cdot\mathbf{h}_i - \gamma_{i-1}\,\mathbf{g}_i\cdot\underbrace{\mathbf{h}_{i-1}}_{\cancel{\mathbf{g}_{i-1}} + \gamma_{i-2}\mathbf{h}_{i-2}}$$

$$= \mathbf{g}_i\cdot\mathbf{h}_i - \gamma_{i-1}\gamma_{i-2}\cdots\gamma_0\,\underbrace{\mathbf{g}_i\cdot\underbrace{\mathbf{h}_0}_{\mathbf{g}_0}}_{=0}$$

$$= \mathbf{g}_i\cdot\mathbf{h}_i$$

$$\therefore\ \lambda_i = \frac{\mathbf{g}_i\cdot\mathbf{h}_i}{\mathbf{h}_i\cdot\mathbf{A}\cdot\mathbf{h}_i} \qquad //$$

(Th: Conjugate Gradient Method)

Suppose we have sequences $\mathbb{g}_i$ & $\mathbb{h}_i$ for $i \leq n$ which are constructed as in Eqs. (7) & (8). Suppose that $\mathbb{g}_n = -\nabla f(\mathbb{P}_n) = -\mathbb{A}\cdot\mathbb{P}_n + \mathbb{b}$. We determine $\mathbb{P}_{n+1}$ as a line minimum along the direction $\mathbb{P}_{n+1} = \mathbb{P}_n + \lambda\, \mathbb{h}_n$, i.e., $\mathbb{h}_n \cdot \nabla f(\mathbb{P}_{n+1}) = 0$, then we calculate $\mathbb{g}_{n+1} = -\nabla f(\mathbb{P}_{n+1})$. Then, this $\mathbb{g}_{n+1}$ is equivalent to one calculated from Eq. (7).

$$\because \quad \mathbb{h}_n \cdot \underbrace{\nabla f(\mathbb{P}_n + \lambda\, \mathbb{h}_n)}_{\underbrace{\nabla f(\mathbb{P}_n)}_{-\mathbb{g}_n} + \lambda\, \mathbb{A}\cdot\mathbb{h}_n \quad (\because \text{Eq. (4)})} = 0$$

$$-\mathbb{g}_n \cdot \mathbb{h}_n + \lambda\, \mathbb{h}_n \cdot \mathbb{A} \cdot \mathbb{h}_n = 0$$

$$\therefore \lambda = \frac{\mathbb{g}_n \cdot \mathbb{h}_n}{\mathbb{h}_n \cdot \mathbb{A} \cdot \mathbb{h}_n}$$

$$\therefore \mathbb{g}_{n+1} = -\nabla f\left(\mathbb{P}_n + \frac{\mathbb{g}_n \cdot \mathbb{h}_n}{\mathbb{h}_n \cdot \mathbb{A} \cdot \mathbb{h}_n}\, \mathbb{h}_n\right)$$

$$= \mathbb{g}_n - \frac{\mathbb{g}_n \cdot \mathbb{h}_n}{\mathbb{h}_n \cdot \mathbb{A} \cdot \mathbb{h}_n}\, \mathbb{A} \cdot \mathbb{h}_n$$

This is equivalent to Eqs. (7) & (9). //

8. Algorithm

① Start from $\mathbb{P}_0 \in \mathbb{R}^N$

② $\mathbb{g}_0 = -\nabla f(\mathbb{P}_0)$, $\mathbb{h}_0 = \mathbb{g}_0$

③ do $i = 0, n_{cgmax}$

Line minimize $f(\mathbb{P}_{i+1} \leftarrow \mathbb{P}_i + \lambda \mathbb{h}_i)$

if $(|\mathbb{P}_{i+1} - \mathbb{P}_i| < \varepsilon)$ exit

$$\begin{cases} \mathbb{g}_{i+1} \leftarrow -\nabla f(\mathbb{P}_{i+1}) \\ \mathbb{h}_{i+1} \leftarrow \mathbb{g}_{i+1} + \underbrace{\dfrac{(\mathbb{g}_{i+1} - \mathbb{g}_i)\cdot \mathbb{g}_{i+1}}{\mathbb{g}_i \cdot \mathbb{g}_i}}_{\gamma_i} \mathbb{h}_i \end{cases}$$

enddo

write 'CG iteration exceeds n_{cgmax}'