Lecture 17.

Monday, 7 March 2022 1:47 PM

$$f(x) := \frac{1}{2} \underline{x}^T Q \underline{x} - \underline{b}^T \underline{x} + \underline{c}, \quad \boxed{\begin{array}{l} Q = Q^T \\ Q > 0 \end{array}}$$

CG Algorithm

① $\underline{d}^{(0)} = \underline{r}^{(0)} (= g^{(0)}) = Q\underline{x}^{(0)} - \underline{b}$

$\underline{k=0}$

② $\alpha_k = -\dfrac{(g^{(k)})^T \underline{r}^{(k)}}{(\underline{d}^{(k)})^T Q \underline{d}^{(k)}}$

③ $\underline{x}^{(k+1)} = \underline{x}^{(k)} + \alpha_k \underline{d}^{(k)}$

④ $\underline{r}^{(k+1)} = \underline{r}^{(k)} + \alpha_k Q \underline{d}^{(k)}$

⑤ $\beta_{k+1} = \dfrac{(\underline{r}^{(k+1)})^T \underline{r}^{(k+1)}}{(\underline{r}^{(k)})^T \underline{r}^{(k)}}$

⑥ $\underline{d}^{(k+1)} = \underline{r}^{(k+1)} + \beta_{k+1} \underline{d}^{(k)}$

→ CG for [non linear / ~~non-quadratic~~] $f(x)$

Second order Taylor Series approximation of objective function.

Quasi-Newton Methods

Newton's method →

$$\boxed{\underline{x}^{(k+1)} = \underline{x}^{(k)} - \alpha_k F(\underline{x}^{(k)})^{-1} g^{(k)}}$$

α_k is s.t. $f(\underline{x}^{(k+1)}) < f(\underline{x}^{(k)})$

$\alpha_k = \arg\min\limits_{\alpha \geq 0} f(\underline{x}^{(k)} - \alpha F(\underline{x}^{(k)})^{-1} g^{(k)})$

$\boxed{\text{Thm } 9.2}$

→ Use an approximate value of $(F(x^{(k)}))^{-1}$ in place of

↳ Use an approximate value of $\left(F(x^{(k)})\right)$ in place of the true inverse.

That is, $$\underline{x}^{(k+1)} = \underline{x}^{(k)} - \alpha_k H_k \, g^{(k)}.$$

$H_k \rightarrow n \times n$ real matrix, α_k is the search parameter

$$f(\underline{x}^{(k+1)}) = f(\underline{x}^{(k)}) + (g^{(k)})^T (\underline{x}^{(k+1)} - \underline{x}^{(k)}) + o\left(\| \underline{x}^{(k+1)} - x^{(k)} \|\right)$$

$$= f(\underline{x}^{(k)}) - \underbrace{\alpha_k \overbrace{(g^{(k)})^T H_k \, g^{(k)}}}\; + \underbrace{o\left(\alpha_k \| H_k \, g^{(k)} \|\right)}.$$

As $\alpha_k \to 0$, the second term dominates the third.

$f(\underline{x}^{(k+1)}) < f(\underline{x}^{(k)})$ is guaranteed if

$$(g^{(k)})^T H_k \, g^{(k)} > 0$$

H_k is positive-definite guarantees this.

Summarize:

$f \in \mathcal{C}^{(1)}$

$\underline{x}^{(k)} \in \mathbb{R}^{(n)}$

$g^{(k)} = \nabla f(\underline{x}^{(k)}) \neq 0$ [Why?]

H_k is $n \times n$ real-symmetric SPD.

$\underline{x}^{(k+1)} = \underline{x}^{(k)} - \alpha_k H_k \, g^{(k)}$

$$\alpha_k = \arg\min_{\alpha \geq 0} f(\underline{x}^{(k)} - \alpha H_k \, g^{(k)}) > 0$$

$f(\underline{x}^{(k+1)}) < f(\underline{x}^{(k)})$.

Hessian approximation using [objective fn

Hessian approximation using [objective fn gradient.]

$f(x) = \frac{1}{2} \underline{x}^T Q \underline{x} - \underline{x}^T \underline{b} + \underline{c}, \quad Q = Q^T > 0$

Let us understand for quadratic function.

Let $H_0, H_1, H_2, \ldots$ be successive approximation of $F(\underline{x}^{(k)})^{-1}$.

$$\underline{g}^{(k+1)} = \left\{ \begin{array}{l} g^{(k+1)} = Q x^{(k+1)} - \underline{b} \\ g^{(k)} = Q x^{(k)} - \underline{b} \end{array} \right\}$$

$\Delta g^{(k)} = g^{(k+1)} - g^{(k)}$

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

$$\boxed{\Delta \underline{g}^{(k)} = Q \, \Delta \underline{x}^{(k)}}$$

$$Q^{-1} \Delta g^{(i)} = \Delta \underline{x}^{(i)} \qquad 0 \le i \le k$$

Impose the requirement $\rightarrow$ $\boxed{H_{k+1} \, \Delta g^{(i)} = \Delta x^{(i)}}$ $\rightarrow$ $\quad 0 \le i \le k.$

$\boxed{k=1}$ $\quad H_2 \, \Delta g^{(i)} = \Delta x^{(i)} \qquad i = 0, 1$

$[H_2]_{n \times n}$

$$\left. \begin{array}{l} H_2 \, \Delta g^{(0)} = \Delta x^{(0)} \\ H_2 \, \Delta g^{(1)} = \Delta x^{(1)}. \end{array} \right]$$

$$[H_2]_{n \times n} \left[\Delta g^{(0)} \quad \Delta g^{(1)} \right]_{n \times 2} = \left[\Delta x^{(0)} \quad \Delta x^{(1)} \right]_{n \times 2}$$

$$H_{(n)} \left[\Delta g^{0} \; \Delta g^{(1)} \; \ldots \; \Delta g^{(n-1)} \right] = \left[\Delta x^{(0)}, \; \Delta x^{(1)}, \; \ldots \; \Delta x^{(n-1)} \right].$$

Q satisfies $\quad Q^{-1} \left[\underbrace{\Delta g^{(0)} \; \ldots \; \Delta g^{(n-1)}} \right] = \left[\Delta x^{(0)}, \; \ldots \; \Delta x^{(n-1)} \right]$

If $\left[\Delta g^{(0)} \; \ldots \; \Delta g^{(n-1)} \right]$ is non-singular,

$$Q^{-1} = H_n = \left[\Delta x^{(0)} \; \ldots \; \Delta x^{(n-1)} \right] \left[\Delta g^{(0)} \ldots \Delta g^{(n-1)} \right]^{-1}. \checkmark$$

[if $[\Delta g^{0} \ldots \Delta g^{(n-1)}]$ is

$$Q^{-1} = H_n = [\Delta x^{(0)} \cdots \Delta x^{(n-1)}][\Delta g \cdots \Delta g]^{-1}$$

[if $[\Delta g^0, \ldots \Delta g^{(n-1)}]$ is invertible.]

At 'n' th iteration, H_n satisfies

$$\left[\begin{array}{l} H_n \Delta g^{(i)} = \Delta x^{(i)} \qquad 0 \le i \le n-1. \\ \underline{x}^{(k+1)} = \underline{x}^{(k)} - \alpha_k H^{(k)} g^{(k)}. \end{array}\right.$$

$$\left[\begin{array}{l} \Delta x^{(i)} = \underline{x}^{(i+1)} - x^{(i)} \\ \Delta g^{(i)} = g^{(i+1)} - g^{(i)} = Q \Delta x. \end{array}\right.$$

Result: Consider a quasi-Newton method applied to a quadratic function with Hessian $Q = Q^T$ s.t.

$0 \le k < n-1$, $\boxed{H_{k+1} \Delta g^{(i)} = \Delta x^{(i)}}$ — (1) $0 \le i \le k$,

where $H_{k+1} = H_{k+1}^T$.

If $\alpha_i \ne 0$, $0 \le i \le k$, then $\left(\underline{d}^{(k)} = -H_k g^{(k)}\right)$

$\underline{d}^{(0)}, \ldots, \underline{d}^{(k+1)}$ are Q-conjugate. (orthogonal)

[Conclude from III of Lecture 16 that Convergence happens in finite steps.]

Pf. Induction $k = 0$.

$\underline{d}^{(0)}$ and $\underline{d}^{(1)}$ are Q orthogonal. $\left[\underline{d}^{(k)} - H_k g^{(k)}\right]$

$\underline{d}^{(0)}$ and $\underline{d}^{(1)}$ are Q orthogonal.

$$\boxed{\underline{d}^{0} = \frac{\Delta x^{(0)}}{\alpha_0}}$$

$$\begin{aligned}
\underline{d}^{(k)} &= -H_k \underline{g}^{(k)} \\
\Delta x^{(k)} &= \underline{x}^{(k+1)} - \underline{x}^{(k)} \\
\Delta \underline{g}^{(k)} &= \underline{g}^{(k+1)} - \underline{g}^{(k)} \\
\Delta \underline{g}^{(k)} &= Q \Delta x^{(k)}
\end{aligned}$$

$$\begin{aligned}
\underline{x}^{(1)} &= x^{(0)} - \alpha_0 H_0 \underline{g}^{(0)} \\
\underline{x}^{(1)} &= \underline{x}^{(0)} + \alpha_0 \underline{d}^{(0)} \\
\Rightarrow \Delta \underline{x}^{(0)} &= \alpha_0 \underline{d}^{(0)}
\end{aligned}$$

$$\begin{aligned}
\underline{x}^{(k+1)} &= \underline{x}^{(k)} - \alpha_k H_k \underline{g}^{(k)} \\
\underline{x}^{(k+1)} &= \underline{x}^{(k)} + \alpha_k \underline{d}^{(k)} \\
\Delta \underline{x}^{(k)} &= \alpha_k d^{(k)}
\end{aligned}$$

$$\begin{aligned}
(\underline{d}^{(1)})^T Q \underline{d}^{(0)} &= -(H_1 \underline{g}^{(1)})^T Q \underline{d}^{(0)} \\
&= -(\underline{g}^{(1)})^T H_1 Q \frac{\Delta x^{(0)}}{\alpha_0} \\
&= -(\underline{g}^{(1)})^T H_1 \frac{\Delta \underline{g}^{(0)}}{\alpha_0} \\
&= -(\underline{g}^{(1)})^T \frac{\Delta \underline{x}^{(0)}}{\alpha_0} \\
&= -(\underline{g}^{(1)})^T \underline{d}^{(0)} \qquad (\underline{g}^{(1)} d^{(0)} \\
&= 0
\end{aligned}$$

$[H_1 \Delta \underline{g}^{(0)} = \Delta \underline{x}^{(0)}$ from ①$]$

[from CG method proof.]

Ex. 11.1 Tut Sheet 6

Assume for $\underline{k-1}$ $(k < n-1)$.

We prove that $d^{(0)}, d^{(1)}, \ldots, d^{(k+1)}$ are Q-conjugate.

$$(\underline{d}^{(k+1)})^T Q \underline{d}^{(i)} = 0 \qquad i = 0, \ldots, k.$$

$$\begin{aligned}
(\underline{d}^{(k+1)})^T Q d^{(i)} &= -(H_{k+1} \underline{g}^{(k+1)})^T Q \underline{d}^{(i)} \\
&= -(\underline{g}^{(k+1)})^T H_{k+1} Q \frac{\Delta x^{(i)}}{\alpha_i} \\
&= -(\underline{g}^{(k+1)})^T H_{k+1} \frac{\Delta \underline{g}^{(i)}}{\alpha_i} \qquad \text{(use ①)}
\end{aligned}$$

$$= -(g \ldots \frac{\ldots}{\alpha_i}$$

$$= -\left(g^{(k+1)}\right)^T \frac{\Delta \underline{x}^{(i)}}{\alpha_i}$$

$f(x^{(k+1)}) < f(x^{(k)})$

$$= -\left(g^{(k+1)}\right)^T \underline{d}^{(i)} = 0 \qquad i = 0, \ldots, k$$

[residuals are orthogonal to $d^{(i)}$].

$$= 0$$

Qn: How to update H_k's ?

$$\boxed{H_{k+1} \Delta g^{(i)} = \Delta \underline{x}^{(i)}}$$