Lecture 14

Thursday, 17 February 2022 1:51 PM

Problem (Tut. Sheet 5) x^* is a minimizer of $f(\underline{x}) = \frac{1}{2} x^T Q \underline{x} - b^T \underline{x} + c$, $Q > 0$ (SPD). If $(\underline{x}^{(0)} - x^*)$ is an eigenvector of Q, then the steepest descent converges in one step.

Pf. $\nabla f(\underline{x}) = Q\underline{x} - \underline{b}$

$\nabla f(\underline{x}^*) = Q\underline{x}^* - b = 0$

$x^{(0)} - x^* = e^{(0)}$

$$\boxed{g^{(0)} = Q e^{(0)} = \lambda e^{(0)}}$$

$\nabla f(x^{(0)}) = Q x^{(0)} - \underline{b}$

$= Q(x^{(0)} - x^*)$

$\nabla f(x^0) = Q e^{(0)} = g^{(0)}$

$$\alpha_k = \frac{(g^{(k)})^T g^{(k)}}{(g^{(k)})^T Q g^{(k)}}$$

$$x^{(1)} = x^{(0)} - \alpha_0 g^{(0)}$$

$$= x^{(0)} - \frac{(g^{(0)})^T g^{(0)}}{(g^{(0)})^T Q \underbrace{g^{(0)}}_{\lambda e^{(0)}}} \overbrace{g^{(0)}}^{\lambda e^{(0)}}$$

$$= x^{(0)} - \frac{(g^{(0)})^T g^{(0)}}{(g^{(0)})^T \underbrace{Q e^{(0)}}_{(g^{(0)})}} e^{(0)} = x^{(0)} - (x^{(0)} - x^{(*)}) = \underline{\underline{x^*}}.$$

Steepest descent

[Conjugate directions Newton's

Conjugated gradient.

Drawback of Steepest descent

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

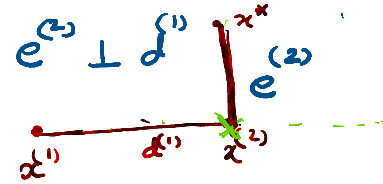
Takes same steps as earlier steps.

I. Idea: " Every time you take a step, take it right."

Pick up a set of mutually orthogonal search directions $\{d^{(1)}, d^{(2)}, \dots, d^{(n)}\}$. In each direction, take exactly ... After 'n' steps, we

Pick up ... $\{d^{(1)}, d^{(2)}, \dots, d^{(n)}\}$. In each direction, take exactly one step with the right length. After 'n' steps, we will be done.

$e^{(2)} \perp d^{(1)}$



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

$$\underline{e^{(k+1)} = e^{(k)} + \alpha_k d^{(k)}.}$$

$$e^{(k)} = x^{(k)} - x^{(*)}$$

$$Q e^{(k)} = Q x^{(k)} - \underline{b}$$
$$= \nabla f(x^{(k)}) = g^{(k)}$$

Wish: $e^{(k+1)} \perp d^{(k)}$

$$0 = (d^{(k)})^T e^{(k+1)} = (d^{(k)})^T e^{(k)} + \alpha_k (d^{(k)})^T \underline{d}^{(k)}$$

Residual equation $\leftarrow$ $\boxed{Q e^{(k)} = r^{(k)}}$

$$Q \underline{x}^* = \underline{b}$$

$$\Rightarrow \boxed{\alpha_k = - \frac{(d^{(k)})^T e^{(k)}}{(d^{(k)})^T d^{(k)}}}$$

We don't know $e^{(k)}$!

II. Rather than choosing orthogonal search directions $\{\underline{d}^{(1)}, \dots, \underline{d}^{(n)}\}$, can we choose 'Q' orthogonal search directions?

(i) What does This mean? (Q orthogonality)

(ii) Can we compute α_k easily?

(i) $$(d^{(i)})^T Q d^{(j)} = 0 \qquad i \neq j. \quad [\text{Q orthogonality}]$$

$$(d^{(k)})^T Q \underline{e}^{(k+1)} = 0 \qquad \underline{e}^{k+1} \overset{Q}{\perp} \underline{d}^{(k)}$$

$$\Rightarrow (d^{(k)})^T r^{(k+1)} = 0 \qquad \underline{x}^{k+1} = x^{(k)} + \alpha \underline{\underline{d}}^{(k)}$$

$$\Rightarrow (\nabla f(x^{k+1}))^T d^{(k)} = 0$$

$$\Rightarrow \nabla f(x^{(k+1)})^T \frac{d}{d\alpha}\left(x^{(k+1)}\right) = 0$$

$$\Rightarrow \quad \nabla f(x^{\cdots}) \quad \frac{d}{d\alpha}(\cdots)$$

$$\Rightarrow \quad \frac{d}{d\alpha}\left(\underline{f(x^{(k+1)})} \right) = 0$$

Q-orthogonality is equivalent to finding a minimum in the direction $d^{(k)}$. [as in steepest descent, instead of finding min. $g^{(k)}$, we find in the direction of $d^{(k)}$]

(iii) <u>How to compute α_k ?</u>

$$(d^{(k)})^T Q e^{(k+1)} = 0$$

$$(d^{(k)})^T Q \left(e^{(k)} + \alpha_k d^{(k)} \right) = 0$$

$$\Rightarrow \alpha_k = \frac{-(d^{(k)})^T \overbrace{Q e^{(k)}}^{g^{(k)}}}{(d^{(k)})^T Q d^{(k)}} = -\frac{(d^{(k)})^T g^{(k)}}{(d^{(k)})^T Q d^{(k)}}.$$

$$\boxed{\begin{aligned} x^{(k+1)} &= x^{(k)} + \alpha_k d^{(k)} \\ \alpha_k &= -\frac{(d^{(k)})^T g^{(k)}}{(d^{(k)})^T Q d^{(k)}} \end{aligned}}$$

$$g^{(k)} = \nabla f(x^{(k)})$$

III. If we choose 'Q' orthogonal directions can we have convergence after $(n+1)$ steps? $\boxed{\text{YES}}$

IV. How to construct 'Q' orthogonal search directions?

✓ Gram-Schmidt $\boxed{\text{Drawback}}$ Storage & computation.

I. Choose ... → <u>Residuals</u> $\{g^{(k)}\}$. ... such a way that most of

V. Choose Smart directions $\xrightarrow{}$ Residuals $\{r^{(k)}\}$. in such a way that most of the terms in Gram-Schmidt process vanish.

↳ Conjugate'd' Gradient method.

CG method:

Sketch.

- $\underline{d}^{(1)} = \underline{r}^{(1)} = Q x^{(1)} - \underline{b}$
- $\alpha_k = \cdots\cdots$
- $x^{(k+1)} = x^{(k)} + \alpha_k d^{(k)}$
- $r^{(k+1)} = r^{(k)} + \alpha_k A d^{(k)}$.
- $\beta_{k+1} = \cdots\cdots$ [Gram-Schmidt]
- $\underline{d}^{(k+1)} = \underline{r}^{(k+1)} + \underline{\beta_{k+1}} d^{(k)}$