Lecture 23:

Recall:

Conjugate gradient method

Goal: Minimize a quadratic functional.

$$\vec{x}_* = \underset{\vec{x} \in \mathbb{R}^n}{\operatorname{argmin}} \varphi(\vec{x}) \quad ; \quad \varphi(\vec{x}) = \frac{1}{2}\vec{x}^T A \vec{x} - \vec{b}^T \vec{x}$$

where A = symmetric positive definite matrix in $M_{n \times n}(\mathbb{R})$ and $\vec{b} \in \mathbb{R}^n$.

Recall: $\nabla \varphi(\vec{x}) = A\vec{x} - \vec{b}$ and $\underbrace{\varphi''(\vec{x})}_{\text{Hessian}} = A$

Minimizer $\vec{x}^*$ of $\varphi(\vec{x})$ satisfies $A\vec{x}^* = \vec{b}$

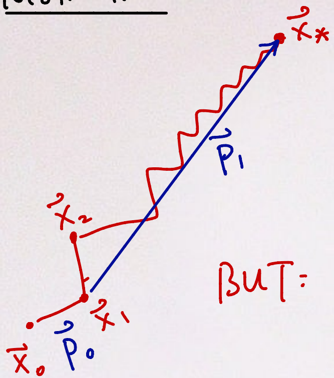
$\left(\frac{\partial^2 \varphi}{\partial x_i \partial x_j}\right)$

Strategy: Given a current approximation $\vec{x}_k$, find a new approximation

by: $\vec{x}_{k+1} = \vec{x}_k + \alpha_k \vec{p}_k$ $\left(\begin{array}{l} \vec{p}_k = \text{search direction} \\ \alpha_k = \text{time step} \end{array} \right)$

But we want to choose: time step α_k to be the optimal and search direction such that $\vec{p}_i \cdot A \vec{p}_j = 0$ for $i \neq j$.

Motivation:



For $A \in M_{2\times 2}(\mathbb{R})$, if $\vec{x}^*$ = sol of $A\vec{x} = \vec{b}$.

Ideally, we want to find $\vec{p}_1$ such that the direction allows us to move directly to $\vec{x}_*$.

$$\therefore \ \vec{p}_1 \parallel \vec{x}^* - \vec{x}_1 \Rightarrow \vec{p}_1 = c(\vec{x}^* - \vec{x}_1)$$

BUT: $$A\vec{p}_1 = \underset{\substack{\uparrow \\ \mathbb{R}}}{c}\,(\underbrace{A\vec{x}_*}_{\vec{b}} - A\vec{x}_1) = \underset{\substack{\uparrow \\ \mathbb{R}}}{c}\,(\vec{b} - A\vec{x}_1)$$

$$A\vec{p}_1 \cdot \vec{p}_0 = c\,(\underbrace{\vec{b} - A\vec{x}_1}_{-\nabla\varphi(\vec{x}_1)}) \cdot \vec{p}_0 = -c\,\nabla\varphi(\vec{x}_0 + \alpha_0\vec{p}_0)\cdot\vec{p}_0$$

$$= -c\,\frac{d}{d\alpha}\Big|_{\alpha=\alpha_0}\varphi(\vec{x}_0 + \alpha\vec{p}_0) = 0$$

$$\therefore \ A\vec{p}_1 \cdot \vec{p}_0 = 0$$

Get convergence in JUST 2 steps!!

Summary: (Goal:) Find search directions $\vec{p}_j$ $(j=0,1,2,\ldots)$ such that $\vec{p}_j^{\,T} A \vec{p}_k = 0$ for $j \neq k$ and find optimal time step α_j.

Definition: We say that the set of directions $\{\vec{p}_j\}_{j=0}^{k-1}$ is a conjugate set of directions (with respect to A) if:

$$\vec{p}_j^{\,T} A \vec{p}_i = 0 \quad \text{for all } 1 \le i, j \le k-1 \text{ and } i \neq j.$$

Remark: $\because$ A is symmetric

$$\therefore \quad \vec{p}_j^{\,T} A \vec{p}_i = \vec{p}_i^{\,T} A^T \vec{p}_j = \vec{p}_i^{\,T} A \vec{p}_j$$

Notation: $W_k \overset{\text{def}}{:=} \operatorname{span}\{\vec{p}_0, \vec{p}_1, \ldots, \vec{p}_{k-1}\}$

Choice of time step α_k (Given $\vec{p}_0, \vec{p}_1, \dots, \vec{p}_k$)

Optimal α_k:

$$\varphi(\vec{x}_k + \alpha \vec{p}_k) = \varphi(\vec{x}_k) + \alpha \nabla\varphi(\vec{x}_k) \cdot \vec{p}_k + \frac{\alpha^2}{2} \vec{p}_k^{\,T} A \vec{p}_k$$

(Taylor expansion)

Minimum attained at $\alpha = \dfrac{-\nabla\varphi(\vec{x}_k) \cdot \vec{p}_k}{\vec{p}_k^{\,T} A \vec{p}_k}$

$\therefore$ we choose $\boxed{\alpha_k = \dfrac{-\vec{r}_k \cdot \vec{p}_k}{\langle \vec{p}_k, \vec{p}_k \rangle_A}}$

where $\vec{r}_k = A\vec{x}_k - \vec{b}$ and $\langle \vec{u}, \vec{v} \rangle_A \overset{\text{def}}{=} \vec{u} \cdot A\vec{v}$.

Choice of the search directions $\vec{p}_j$ $(j = 0, 1, 2, \ldots)$

Suppose $\vec{p}_0, \vec{p}_1, \ldots, \vec{p}_{k-1}$ are known. Then, we proceed to find the search direction $\vec{p}_k$ in the form:

$$\vec{p}_k = -\vec{r}_k - \beta_{k-1}\vec{p}_{k-1} \quad \text{where} \quad \beta_{k-1} = -\frac{\vec{p}_{k-1}^{\,T} A \vec{r}_k}{\vec{p}_{k-1}^{\,T} A \vec{p}_{k-1}}$$

$$\Rightarrow A\vec{p}_k = -A\vec{r}_k - \beta_{k-1} A\vec{p}_{k-1}$$

$$\Rightarrow \underbrace{\vec{p}_{k-1}^{\,T} A \vec{p}_k}_{=\,0} = -\vec{p}_{k-1}^{\,T} A \vec{r}_k - \beta_{k-1}\vec{p}_{k-1}^{\,T} A \vec{p}_{k-1}$$

$$\Rightarrow \quad \beta_{k-1} = -\frac{\vec{p}_{k-1}^{\,T} A \vec{r}_k}{\vec{p}_{k-1}^{\,T} A \vec{p}_{k-1}}$$

Next: We need to show: with the above recursive scheme to obtain $\{\vec{p}_j\}_{j=0}^{\infty}$;

(i) $\{\vec{p}_j\}$ are conjugate to each others.

(ii) $\vec{r}_i \cdot \vec{r}_j = 0$ for $i \neq j$.

Next time.

$(\vec{r}_i = A\vec{x}_i - \vec{b})$

If the above are true, then we have:

Theorem: Consider the iterative scheme:

$$\vec{x}_{k+1} = \vec{x}_k + \alpha_k \vec{p}_k \quad (k = 0, 1, 2, \ldots)$$

Suppose $\{\vec{r}_k\}_{k=0}^{\infty}$ are orthogonal to each others ($\vec{r}_i \cdot \vec{r}_j = 0$ for $i \neq j$)

Then, the iterative scheme converges to the sol $A\vec{x} = \vec{b}$ in less than or equal to n iterations.

Proof: Suppose $\vec{r}_0, \vec{r}_1, \ldots, \vec{r}_n$ are all non-zero.

Then: $\{\vec{r}_0, \ldots, \vec{r}_n\}$ form an orthogonal $\quad (\vec{r}_i \overset{\text{def}}{=} A\vec{x}_i - \vec{b})$

and lin. independent set in $\mathbb{R}^n$. Contradiction.
with $n+1$ elements

$\therefore \quad \vec{r}_i = 0$ for some $i \leq n$.

$\therefore \quad A\vec{x}_i - \vec{b} = \vec{0}$ for some $i \leq n$.

$\therefore \quad \vec{x}_i =$ sol of $A\vec{x} = \vec{b}$ for some $i \leq n$.

Conjugate gradient method (Solve: $A\vec{x} = \vec{b}$)

Given $\vec{x}_0 \in \mathbb{R}^n$, $\vec{p}_0 = -\vec{r}_0 \overset{\text{def}}{=} -(A\vec{x}_0 - \vec{b})$, find $\vec{x}_k$ and $\vec{p}_k$ $(k = 1, 2, \dots)$ such that:

(a) $\vec{x}_{k+1} = \vec{x}_k + \alpha_k \vec{p}_k$

$\vec{r}_k = A\vec{x}_k - \vec{b}$

(b) $\alpha_k = -\dfrac{\vec{r}_k \cdot \vec{p}_k}{\langle \vec{p}_k, \vec{p}_k \rangle_A}$

$\langle \vec{u}, \vec{v} \rangle_A = \vec{u} \cdot A\vec{v}$.

(c) $\vec{p}_{k+1} = -\vec{r}_{k+1} - \beta_k \vec{p}_k$

(d) $\beta_k = -\dfrac{\langle \vec{r}_{k+1}, \vec{p}_k \rangle_A}{\langle \vec{p}_k, \vec{p}_k \rangle_A}$

Iteration converges in less than n iterations.
$(\vec{r}_i \cdot \vec{r}_j = 0 \ \forall i \neq j)$

Theorem: In the conjugate gradient method,

(i) $\vec{r}_i \cdot \vec{r}_j = 0$ for $i \neq j$

(ii) $\langle \vec{p}_i, \vec{p}_j \rangle_A \stackrel{def}{=} \vec{p}_i \cdot A\vec{p}_j = 0$ for $i \neq j$.

Lemma: $\mathrm{Span}\{\vec{p}_0, \dots, \vec{p}_{k-1}\} = \mathrm{span}\{\vec{r}_0, \dots, \vec{r}_{k-1}\}$

$= \mathrm{span}\{\vec{r}_0, A\vec{r}_0, \dots, A^{k-1}\vec{r}_0\}$

Proof: (i) and (ii) are true for $i, j \leq 1$

($\vec{r}_0 \cdot \vec{r}_1 = 0$ because $0 = \left.\frac{d}{d\alpha}\right|_{\alpha=\alpha_0} \varphi(\vec{x}_0 + \alpha \underbrace{\vec{p}_0}_{=-\vec{r}_0}) = \underbrace{\nabla\varphi(\overbrace{\vec{x}_0 + \alpha_0\vec{p}_0}^{\vec{x}_1})}_{(A\vec{x}_1 - \vec{b}) = \vec{r}_1} \cdot \underbrace{\vec{p}_0}_{=-\vec{r}_0}$

$\langle \vec{p}_1, \vec{p}_0 \rangle_A = 0$ follows from the definition)

Suppose the statement is true for $i, j \le k$. For $k+1$,

$\because \; \mathrm{Span}\{\vec{p}_0, \dots, \vec{p}_j\} = \mathrm{Span}\{\vec{r}_0, \dots, \vec{r}_j\}$

$\therefore$ we get $\vec{r}_k \cdot \underbrace{\vec{p}_j}_{\in\, \mathrm{Span}\{\vec{r}_0, \dots, \vec{r}_j\}} = 0$ for $j = 0, 1, 2, \dots, k-1$ (By induction hypothesis)

Now, $\vec{r}_{k+1} = \vec{r}_k + \alpha_k A \vec{p}_k$

$\left(\because \; A\vec{x}_{k+1} - \vec{b} = A\vec{x}_k - \vec{b} + \alpha_k A \vec{p}_k \right)$

$$\Rightarrow \vec{r}_{k+1} \cdot \vec{p}_j = \cancelto{0}{\vec{r}_k \cdot \vec{p}_j} + \alpha_k \cancelto{0 \text{ (induction hypothesis)}}{\overbrace{A\vec{p}_k \cdot \vec{p}_j}^{\langle \vec{p}_k, \vec{p}_j \rangle_A}} = 0 \quad \text{for } j = 0, 1, 2, \dots, k-1$$

Also,

$$0 = \frac{d}{d\alpha}\Big|_{\alpha = \alpha_0} \varphi(\vec{x}_k + \alpha \vec{p}_k) = \underbrace{\nabla \varphi(\overbrace{\vec{x}_k + \alpha \vec{p}_k}^{\vec{x}_{k+1}})}_{A\vec{x}_{k+1} - \vec{b} \,=\, \vec{r}_{k+1}} \cdot \vec{p}_k$$

All together, we get $\vec{r}_{k+1} \cdot \vec{p}_j = 0$ for $j = 0, 1, 2, \ldots, k$

$\because$ $\text{Span}\{\vec{p}_0, \ldots, \vec{p}_k\} = \text{Span}\{\vec{r}_0, \ldots, \vec{r}_k\}$

$\therefore$ $\vec{r}_{k+1} \cdot \vec{r}_j = 0$ for $j = 0, 1, 2, \ldots, k$

$\therefore$ (i) is true for the case $k+1$.

To show (ii) for the case $k+1$ (given the induction hypothesis),

note that: $\vec{r}_{j+1} = \vec{r}_j + \alpha_j A \vec{p}_j \Rightarrow A\vec{p}_j \in \text{Span}\{\vec{r}_j, \vec{r}_{j+1}\}$

$$\therefore \vec{r}_{k+1} \cdot \underbrace{A\vec{p}_j}_{\in\, \text{Span}\{\vec{r}_j,\, \vec{r}_{j+1}\}} = \langle \vec{r}_{k+1}, \vec{p}_j \rangle_A = 0 \quad \text{for } j = 0, 1, 2, \ldots, k-1$$

Now, $\quad \vec{p}_{k+1} \cdot A\vec{p}_j = -\cancelto{0}{\vec{r}_{k+1} \cdot A\vec{p}_j} - \beta_k \cancelto{0}{\vec{p}_k \cdot A\vec{p}_j}$

$\quad\quad \parallel$

$\langle \vec{p}_{k+1}, \vec{p}_j \rangle_A = 0 \quad$ for $\quad j = 0, 1, 2, \ldots, k-1$

Also, $\quad \langle \vec{p}_{k+1}, \vec{p}_k \rangle_A = 0 \quad$ by definition.

$\therefore \quad \langle \vec{p}_{k+1}, \vec{p}_j \rangle_A = 0 \quad$ for $\quad j = 0, 1, 2, \ldots, k$

$\therefore \quad \langle \vec{p}_i, \vec{p}_j \rangle_A = 0 \quad$ for $\quad i, j \leq k+1$.

By M.I., the theorem is generally true!!

$$\therefore \quad \vec{r}_{k+1} \cdot \vec{p}_j = 0 \quad \text{for} \quad j = 0, 1, 2, \ldots, k-1, k$$

$$\because \quad \operatorname{Span}\{\vec{p}_0, \ldots, \vec{p}_k\} = \operatorname{Span}\{\vec{r}_0, \ldots, \vec{r}_k\}$$

$$\therefore \quad \vec{r}_{k+1} \cdot \vec{r}_j = 0 \quad \text{for} \quad j = 0, 1, 2, \ldots, k$$

$$\operatorname{Span}\{\vec{p}_0, \ldots, \vec{p}_k\}$$

$\therefore$ (i) is true for the case $k+1$.

To show (ii) for the case $k+1$ (given the induction hypothesis)

Note that: $\vec{r}_{j+1} = \vec{r}_j + \alpha_j A\vec{p}_j \Rightarrow A\vec{p}_j \in \text{Span}\{\vec{r}_j, \vec{r}_{j+1}\}$

$\therefore \vec{r}_{k+1} \cdot \underbrace{A\vec{p}_j}_{\subset\, \text{Span}\{\vec{r}_j, \vec{r}_{j+1}\}} = \langle \vec{r}_{k+1}, \vec{p}_j \rangle_A = 0$ for $j = 0, 1, 2, \ldots, k-1$

$\therefore \langle \vec{r}_{k+1}, \vec{p}_j \rangle_A = 0$ for $j = 0, 1, 2, \ldots, k-1$ induction hypothesis

Now, $\vec{p}_{k+1} \cdot A\vec{p}_j = -\cancelto{0}{\vec{r}_{k+1} \cdot A\vec{p}_j} - \beta_k \cancelto{0}{\vec{p}_k \cdot A\vec{p}_j} = 0$

$\Big(\vec{p}_{k+1} = -\vec{r}_{k+1} - \beta_k \vec{p}_k \Big)$

$\parallel$ for $j = 0, 1, 2, \ldots, k-1$

$\therefore \langle \vec{p}_{k+1}, \vec{p}_j \rangle_A = 0$ for $j = 0, 1, 2, \ldots, k-1$

Also, $\langle \vec{p}_{k+1}, \vec{p}_k \rangle_A = 0$ (by definition)

All together,

$$\langle \vec{p}_{k+1}, \vec{p}_j \rangle_A = 0 \quad \text{for } j = 0, 1, 2, \ldots, k.$$

By M. I., the thm is true!!