<u>Lecture 21:</u> Recall:

Gradient descent method

<u>Goal</u>: Look for an iterative scheme:

$$\underbrace{\vec{\xi}^{\,k+1}}_{\in \mathbb{R}^n} = \underbrace{\vec{\xi}^{\,k}}_{\in \mathbb{R}^n} + \underbrace{\overset{\text{time step}}{\alpha_k}}_{\in \mathbb{R}} \underbrace{\overset{\text{search direction}}{\vec{d}^{\,k}}}_{\in \mathbb{R}^n} \quad , \quad k = 0, 1, 2, \ldots$$

such that:

$$f(\vec{\xi}^{\,1}) > f(\vec{\xi}^{\,2}) > \cdots > f(\vec{\xi}^{\,k}) > f(\vec{\xi}^{\,k+1})$$

where $f(\vec{x}) = \vec{x}^{\,T} A \vec{x} - \vec{b}^{\,T} \vec{x}$.

<u>Remark</u>:

- $\vec{d}^{\,k}$ is chosen as $-\nabla f(\vec{\xi}^{\,k}) = -(A\vec{\xi}^{\,k} - \vec{b})$
- Optimal α_k is: $\alpha_k = \dfrac{-(A\vec{\xi}^{\,k} - \vec{b}) \cdot \vec{d}^{\,k}}{\vec{d}^{\,k} \cdot A\vec{d}^{\,k}} = \dfrac{-\vec{r}^{\,k} \cdot \vec{d}^{\,k}}{\langle \vec{d}^{\,k}, \vec{d}^{\,k} \rangle_A} = \dfrac{\vec{d}^{\,k} \cdot \vec{d}^{\,k}}{\langle \vec{d}^{\,k}, \vec{d}^{\,k} \rangle_A}$

  where $\vec{r}^{\,k} = A\vec{\xi}^{\,k} - \vec{b}$ and $\langle \vec{u}, \vec{v} \rangle_A = \vec{u} \cdot A\vec{v}$.

Convergence analysis

Assume A is symmetric positive definite.

We consider the gradient descent method with constant α (small enough α)

Then:

- $\vec{e}^{\,k} = (\underline{I} - \alpha A)^k \vec{e}_0$, where $\vec{e}^{\,n} = \vec{x}^{\,n} - \vec{x}^{\,*}$ and $\vec{x}^{\,*}$ is the solution of $A\vec{x} = \vec{b}$.
- Converges if $\alpha < \dfrac{2}{\lambda_{max}}$ $\lambda_{max} = \max\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$

In practice, we choose $\alpha = \frac{1}{\lambda_{max}}$

Then: $\rho(I - \alpha A) = 1 - \frac{\lambda_{min}}{\lambda_{max}}$ $\quad (\lambda_{min} = \min\{\lambda_1, \dots, \lambda_n\})$

Define: $\frac{\lambda_{max}}{\lambda_{min}} = K(A) =$ <u>condition number</u> of A

$\therefore \quad \rho(I - \alpha A) = 1 - \frac{1}{K(A)} < 1.$

$\therefore$ Gradient descent method converges

<u>Remark</u>: Convergence depends on the condition number.

If condition number is BIG, the convergence is slow!!

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 $\underbrace{\varphi''(\vec{x})}_{\text{Hessian}} = A$

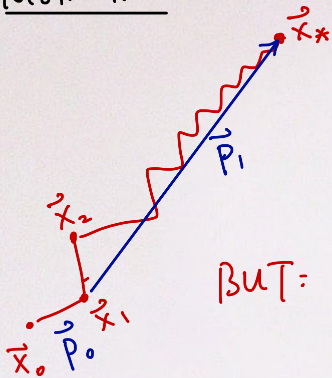
Minimizer $\vec{x}^*$ of $\varphi(\vec{x})$ satisfies $A\vec{x}^* = \vec{b}$ "$\left(\frac{\partial^2\varphi}{\partial x_i\partial x_j}\right)$"

<u>Strategy:</u> 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: $$A\vec{p}_1 = \underset{\substack{\uparrow \\ \mathbb{R}}}{c}\,(\underbrace{A\vec{x}_*}_{\vec{b}} - A\vec{x}_1) = \underset{\substack{\Uparrow \\ \mathbb{R}}}{c}\,(\vec{b} - A\vec{x}_1)$$

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

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

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

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.

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 $\vec{p}_j$ $(j = 0, 1, 2, \dots)$

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

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

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

$$\Rightarrow \quad \underset{\overset{\shortparallel}{0}}{\vec{p}_{k-1}^T A \vec{p}_k} = -\vec{p}_{k-1}^T A \vec{r}_k - \beta_{k-1} \vec{p}_{k-1}^T A \vec{p}_{k-1}$$

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

Next time: we can prove:

$$\vec{r}_i \cdot \vec{r}_j = 0 \quad \text{for } i \neq j.$$

$\Downarrow$

Conjugate gradient method converges in n iterations

because:

$$\{ \vec{r}_0, \vec{r}_1, \vec{r}_2, \ldots, \vec{r}_{n-1}, \vec{r}_n, \vec{r}_{n+1}, \ldots \}$$