Lecture 22:

Observation: Let A = symmetric and eigenvalues:

$$|\lambda_1| > |\lambda_2| > \cdots > |\lambda_n| > 0$$

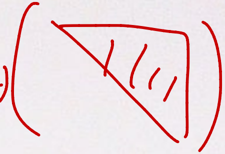
- $\bar{Q}^{(0)} = I$
- $A^k = \bar{Q}^{(k)} \bar{R}^{(k)} \quad \left(\Leftrightarrow A^{-k} = \left(\bar{R}^{(k)}\right)^{-1} \bar{Q}^{(k)\,T} \right)$
- $A_{QR}^{(k)} = A^{(k)} = \left(\bar{Q}^{(k)}\right)^T A \, \bar{Q}^{(k)}$

$$\bar{Q}^{(k)} \downarrow \begin{pmatrix} | & | & & | \\ \vec{q}_1 & \vec{q}_2 & \cdots & \vec{q}_n \\ | & | & & | \end{pmatrix}; \quad \vec{q}_j = \text{eigenvector of } A$$

w/ eigenvalue λ_j.

Note: $$A^{-k} = (A^{k})^{-1} = \underbrace{\left(\bar{R}^{(k)}\right)^{-1}}_{\text{upper triangular}} \left(\bar{Q}^{(k)}\right)^{T}$$

Then: $$\underset{(0,0,\dots,0,1)}{\underset{\shortparallel}{\vec{e}_n^{\,T}}} A^{-k} = \vec{e}_n^{\,T} \left(\bar{R}^{(k)}\right)^{-1} \left(\bar{Q}^{(k)}\right)^{T}$$

$(0,0,\dots,0,1)$ 

$$= \Big(\underbrace{\tilde{r}_{nn}^{(k)}}_{\text{(scalar)}} \vec{e}_n^{\,T}\Big) \left(\bar{Q}^{(k)}\right)^{T} = \tilde{r}_{nn}^{(k)} \left(\bar{Q}^{(k)} \vec{e}_n\right)^{T}$$

$\tilde{r}_{nn}^{k}$ = (n,n) - entry of $\left(\bar{R}^{(k)}\right)^{-1}$.

$$\therefore \quad \left(A^{-k}\right)^{T} \vec{e}_n = \tilde{r}_{nn}^{(k)} \left(\bar{Q}^{(k)} \vec{e}_n\right)$$

$$\Rightarrow \quad \boxed{\underbrace{A^{-k} \vec{e}_n}_{\text{Inverse Power method for last col}} = \tilde{r}_{nn}^{(k)} \left(\bar{Q}^{(k)} \vec{e}_n\right)}$$

converges to // eigenvector q_n

Initial matrix for simultaneous iteration:

$$X^{(0)} = \begin{pmatrix} | & | & & | \\ \vec{e}_1 & \vec{e}_2 & \dots & \vec{e}_n \\ | & | & & | \end{pmatrix}$$

QR method with shift

(By applying QR method on $A - \mu^{(k)} I$)

At k^{th} iteration (given a sequence of real numbers $\{\mu^{(k)}\}_{k=1}^{\infty}$)

① $A^{(k-1)} - \mu^{(k)} I = Q^{(k)} R^{(k)}$

② Let $A^{(k)} - \mu^{(k)} I = R^{(k)} Q^{(k)}$

$$\therefore \quad A^{(k)} = R^{(k)} Q^{(k)} + \mu^{(k)} I.$$

Choice of $\mu^{(k)}$: Rayleigh quotient:

$$\mu^{(k)} = \frac{(\vec{q}_n^{(k)})^T A \, \vec{q}_n^{(k)}}{(\vec{q}_n^{(k)})^T \cdot (\vec{q}_n^{(k)})} \quad ; \quad \vec{q}_n^{(k)} = n^{th} \text{ col of } Q^{(k)}.$$

Energy minimization method

Consider the system $A\vec{x} = \vec{b}$ where A = symmetric and positive-definite

(Let $A = B^T B$ and $\vec{b} = B^T \vec{c}$.)

We'll show: solving $A\vec{x} = \vec{b}$ is equivalent to minimizing:

$$f(\vec{\eta}) = \frac{1}{2}\vec{\eta}\cdot A\vec{\eta} - \vec{b}\cdot\vec{\eta} = \frac{1}{2}\vec{\eta}^{\,T} A\vec{\eta} - \vec{b}^{\,T}\vec{\eta}.$$

$\vec{\eta} = (\eta_1, \eta_2, \ldots, \eta_n)$

We can easily show:

$$f(\vec{\eta}) = \frac{1}{2}\|B\vec{\eta} - \vec{c}\|^2 - \vec{c}\cdot\vec{c}$$

Gradient descent method

Goal: Look for an iterative scheme:

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

such that:

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

From calculus, we can check:

$$\nabla f \overset{\text{def}}{=} \left(\frac{\partial f}{\partial \eta_1}, \frac{\partial f}{\partial \eta_2}, \dots, \frac{\partial f}{\partial \eta_n} \right)^T = A\vec{\eta} - \vec{b}$$

Hessian of f $\overset{\text{def}}{=}$ $\left(\dfrac{\partial^2 f}{\partial \eta_i \partial \eta_j} \right) = A.$ $\nwarrow$ $n \times n$ matrix

〃 $f''(\vec{\eta})$

Taylor's expansion:

$$f(\vec{\zeta}^{k+1}) = f(\vec{\zeta}^k) + \alpha_k \nabla f(\vec{\zeta}^k) \cdot \vec{d}_k + \frac{\alpha_k^2}{2} \vec{d}_k \cdot f''(\vec{\zeta}^k) \vec{d}_k$$

$\parallel$

$\vec{\zeta}^k + \alpha_k \vec{d}_k$

If α_k is small enough, we choose

$$\vec{d}_k = -\nabla f(\vec{\zeta}^k) \qquad \text{(Steepest descent direction)}$$

How about α_k?

Goal: Choose α_k such that $f(\vec{\xi}^k + \alpha_k \vec{d}_k) = \min_{\alpha > 0} f(\vec{\xi}^k + \alpha \vec{d}_k)$

If α_k is optimal, $\frac{d}{d\alpha} f(\vec{\xi}^k + \alpha \vec{d}_k) = 0$ at $\alpha = \alpha_k$

$$\Leftrightarrow \nabla f(\underbrace{\vec{\xi}^k + \alpha_k \vec{d}_k}_{\vec{\xi}^{k+1}}) \cdot \vec{d}_k = 0$$

$$\Leftrightarrow (A \vec{\xi}^{k+1} - \vec{b}) \cdot \vec{d}_k = 0$$

$$\Leftrightarrow (A(\vec{\xi}^k + \alpha_k \vec{d}_k) - \vec{b}) \cdot \vec{d}_k = 0$$

$$\Leftrightarrow (A\vec{\xi}^k - \vec{b}) \cdot \vec{d}_k + \alpha_k \vec{d}_k \cdot A\vec{d}_k = 0$$

$$\therefore \text{Optimal } \alpha_k = -\frac{(A\vec{\xi}^k - \vec{b}) \cdot \vec{d}_k}{\vec{d}_k \cdot A\vec{d}_k}$$

Convergence analysis

We consider the gradient descent method with constant α

(small enough α)

$$\begin{cases} (*) & \vec{z}^{k+1} = \vec{z}^{k} + \alpha\, \vec{d}^{k} = -(A\vec{z}^{k} - \vec{b}) \\ (**) & \vec{d}^{k} = -(A\vec{z}^{k} - \vec{b}) \end{cases}$$

Let $\vec{z}$ be the sol of $A\vec{x} = \vec{b}$. $\therefore$ $\vec{z} = \vec{z} - \alpha(A\vec{z} - \vec{b})$ $(***)$

$(*) - (***)$: $\vec{e}^{k+1} = (I - \alpha A)\vec{e}^{k}$ $\quad$ ($\vec{e}^{k} = \vec{z}^{k} - \vec{z}$ = error vector)

In order that the method converges, we need $\rho(I - \alpha A) < 1$.

Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n > 0$ be the eigenvalues of A (SPD)

Then: $1 - \alpha\lambda_1, 1 - \alpha\lambda_2, \ldots, 1 - \alpha\lambda_n$ are the eigenvalues of $I - \alpha A$.

$\therefore$ $\rho(I - \alpha A) < 1$ iff $|1 - \alpha\lambda_j| < 1$ for $\forall j$

iff $1 - \alpha\lambda_j < 1$ and $1 - \alpha\lambda_j > -1$ for all j.

$\therefore$ $\alpha\lambda_j < 2$ for all j.

$\therefore$ Choose : α such that $\alpha < \dfrac{2}{\lambda_{max}}$ $\quad$ $\lambda_{max} = \max\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$

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

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

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!!

$A\vec{x} = \vec{b} \xrightarrow{\text{MATLAB}} A \backslash \vec{b}$

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 $\vec{b} \in \mathbb{R}^n$.

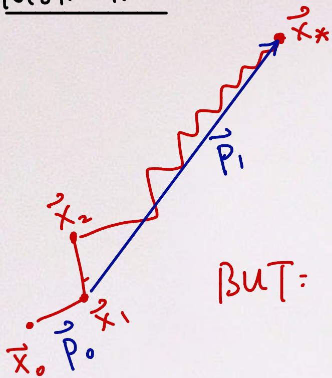
Recall: $\nabla \varphi(\vec{x}) = A\vec{x} - \vec{b}$ and $\underbrace{\varphi''(\vec{x})}_{\text{Hessian}} = A$ "$\left(\frac{\partial^2 \varphi}{\partial x_i \partial x_j}\right)$"

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

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 $\vec{x}_*$.

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

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

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

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

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

Get convergence in JUST 2 steps!!

Summary:
(Goal:)

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