Lecture 20:

Observation: Let A = symmetric and eigenvalues:

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

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

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

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

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

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

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

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

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

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

converges to $\parallel$ eigenvector $\vec{q}_n$

Initial matrix for simultaneous iteration:

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

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 $\mu^{(k)}$: Rayleign quotient:

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

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:

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

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

We can easily show:

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

Gradient descent method

Goal: Look for an iterative scheme:

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

such that:

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

From calculus, we can check:

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

Hessian of f $\stackrel{\text{def}}{=} \left(\frac{\partial^2 f}{\partial \eta_i \partial \eta_j} \right) = A.$

"

$f''(\vec{\eta})$

$n \times n$ matrix

Taylor's expansion:

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

$\vec{\zeta}^{k+1}$ ''

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

If α_k is small enough, we choose

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

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

Then: $1 - \alpha\lambda_1, 1 - \alpha\lambda_2, \dots, 1 - \alpha\lambda_n$ 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 $\alpha < \frac{2}{\lambda_{max}}$ $\quad$ $\lambda_{max} = \max\{\lambda_1, \lambda_2, \dots, \lambda_n\}$