Krylov method

$Ax = b$. $A \in \mathbb{R}^{n\times n}$. sym. positive definite.

equivalent to minimization problem.

$$\varphi(x) = \frac{1}{2} x^T A x - b^T x$$

$\nabla \varphi(x) = Ax - b = 0$. Hessian is $A \Rightarrow$ minimizer.

residual $\quad r = b - Ax = -\nabla \varphi(x)$ negative gradient.

$\langle u, v \rangle_A = u^T A v$. $\quad$ A-inner product.

$\langle u, u \rangle_A = u^T A u \geq 0$. $\quad \langle u, u \rangle_A = 0 \Rightarrow u = 0$.

$\| u \|_A^2 = \langle u, u \rangle_A$.

steepest descent

$$x_{k+1} = x_k + \alpha_k p_k, \quad p_k = r_k = b - A x_k, \quad \alpha_k \in \mathbb{R}.$$

$$\alpha_k = \arg\min_{\alpha} \varphi(x_k + \alpha_k p_k)$$

$$\varphi(x_k + \alpha_k p_k) = \frac{1}{2}(x_k + \alpha_k p_k)^T A (x_k + \alpha_k p_k) - b^T (x_k + \alpha_k p_k)$$

$$\frac{\partial \varphi(x_k + \alpha_k p_k)}{\partial \alpha_k} = (p_k^T A p_k)\alpha_k + p_k^T(\underbrace{A x_k - b}_{-r_k}) = 0$$

$$\alpha_k = \frac{p_k^T r_k}{p_k^T A p_k}.$$

Alg. for $k=1, \text{maxit}$

$$r_k = b - A x_k$$

if $\|r_k\| < \tau$. stop.

$$\alpha_k = \frac{r_k^T r_k}{r_k^T A r_k}$$

$$x_{k+1} = x_k + \alpha_k r_k$$

end.

A is applied twice.

Reformulate

$$r_{k+1} = b - A x_{k+1}$$

$$= b - A(x_k + \alpha_k r_k) = r_k - \alpha_k \underline{\underline{A r_k}}.$$

store this.

Alg. for $k = 1, \text{maxit}$

$w = A r_k$. if $\|r_k\| < \tau$ stop

$$\alpha_k = \frac{r_k^T r_k}{r_k^T w}$$

$x_{k+1} = x_k + \alpha_k r_k$, $r_{k+1} = r_k - \alpha_k w$

Convergence of SD.

<u>Thm</u>. A sym. pos. def. eigenvalues $0 < \lambda_1 \le \lambda_2 \le \cdots \le \lambda_N$

cond. number $K(A) = \dfrac{\lambda_N}{\lambda_1}$

$$\| x_k - x_* \|_A \le \left(\frac{K(A) - 1}{K(A) + 1} \right)^k \| x_0 - x_* \|_A .$$

1. SD converges. linear rate.

2. $\gamma = \dfrac{1 - \frac{1}{K(A)}}{1 + \frac{1}{K(A)}} \approx 1 - \dfrac{2}{K(A)}$

Pf: $\varphi(x) + \frac{1}{2} x_*^T A x_* = \frac{1}{2}(x - x_*)^T A (x - x_*)$

SD: $\varphi(x_k) \le \varphi(x_{k-1} + \alpha r_{k-1}) \quad \forall \alpha .$

opt. step length

$$\Rightarrow \cancel{\tfrac{1}{2}}(x_k - x_*)^T A (x_k - x_*) \le \cancel{\tfrac{1}{2}}(x_{k-1} + \alpha r_{k-1} - x^*)^T A (x_{k-1} + \alpha r_{k-1} - x^*)$$

$$\| x_k - x_* \|_A^2 = (x_k - x_*)^T A (x_k - x_*)$$

$$\le (x_{k-1} - x_*)^T (I - \alpha A) A \underbrace{(I - \alpha A)(x_{k-1} - x_*)}$$

$$\|P(A)(x_k - x_*)\|_A^2 \le \max_{1\le i\le n} |P(\lambda_i)|^2 \|x_{k-1}-x_*\|_A^2 \quad \Big| \quad P(A) = I - \alpha A$$

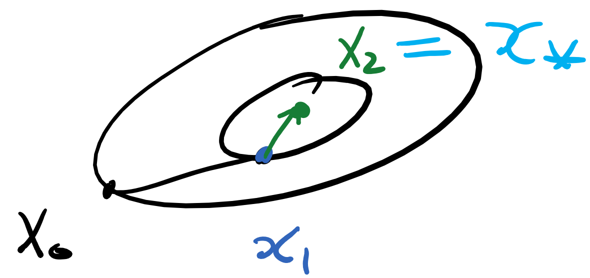
$$\le \underbrace{\min_{\alpha} \max_{\lambda_1 \le t \le \lambda_n}}_{} |1-\alpha t|^2 \|x_{k-1}-x_*\|_A^2$$

sol. $\alpha_* = \dfrac{2}{\lambda_1+\lambda_n}$

val: $|1-\alpha_* t| = \dfrac{\lambda_n-\lambda_1}{\lambda_n+\lambda_1}$. $\square$.

CG.

2D.



First step. SD.

$$\alpha_0 = \frac{r_0^T p_0}{p_0^T A p_0}$$

$$x_1 = x_0 + \alpha_0 p_0 . \quad p_0 = r_0 , \; r_0 = b - Ax_0 .$$

$$x_2 = x_1 + \alpha_1 p_1 \qquad \alpha_1 = \frac{r_1^T p_1}{p_1^T A p_1} \leftarrow \text{universal. opt.}$$

Choose $p_1^T A p_0 = 0$. A-ortho to p_0.

Compute $r_2 = b - A(x_1 + \alpha_1 p_1)$

$$= b - Ax_1 - \alpha_1 A p_1$$

$$= (I - \alpha_0 A) p_0 - \alpha_1 A p_1$$

$\{p_0, p_1\}$ form a basis in $\mathbb{R}^2$

$$p_0^T r_2 = (p_0^T p_0 - \underbrace{\alpha_0 p_0^T A p_0}_{=0}) - \alpha_1 \underbrace{p_0^T A p_1}_{=0} = 0$$

$$p_1^T r_2 = (p_1^T p_0 - \alpha_1 p_1^T A p_1) - \alpha_0 \underbrace{p_1^T A p_0}_{=0}$$

$$p_1^T p_0 - \alpha_1 p_1^T A p_1 = p_1^T p_0 - p_1^T r_0 = \alpha_0 p_1^T A p_0 = 0.$$

$$\Rightarrow r_2 = 0 \Rightarrow x_2 = x_*.$$

CG for $\mathbb{R}^n$.

$$p_1 = r_1 + \beta_1 p_0$$

$$p_0^T A p_1 = p_0^T A r_1 + \beta_1 p_0^T A p_0 \Rightarrow \beta_1 = -\frac{p_0^T A r_1}{p_0^T A p_0}$$

$$x_{k+1} = x_k + \alpha_k p_k \qquad \alpha_k = \frac{r_k^T p_k}{p_k^T A p_k} \quad \text{minimization}$$

$$r_{k+1} = b - A x_{k+1}$$

ortho.

$$p_{k+1} = r_{k+1} + \beta_k p_k \qquad \beta_k = -\frac{p_k^T A r_{k+1}}{p_k^T A p_k}.$$

conceptual implementation of CG.

2 applications of A each iteration.

practical implementation only requires

1.

In general.

Krylov subspace

$$\mathcal{K}_m(A, r_0) \equiv \mathcal{K}_m = \operatorname{span}\{r_0, Ar_0, \cdots, A^{m-1}r_0\}.$$

① For all m, $p_m \in \mathcal{K}_m$.

$$x_{m+1} \in x_0 + \mathcal{K}_m$$

② $p_{m+1} \in \mathcal{K}_{m+1}$, $p_{m+1} \perp_A \mathcal{K}_m$.

$\{P_0, P_1, \cdots, P_{m+1}\}$ mutually A-orthogonal.

$\longrightarrow$ short (3-term) recurrence.

③ x_{m+1} achieves the global minimum of $\varphi(x)$ within all $x_0 + K_m$

Convergence of CG.

$\forall\, x \in x_0 + K_m$.

$$x_* - x = x_* - (x_0 + c_1 r_0 + \cdots + c_m A^{m-1} r_0)$$

$$= A^{-1}(r_0 + c_1 A r_0 + \cdots + c_m A^m r_0)$$

$$= A^{-1} P_m(A)\, r_0 .$$

$P_m(z) = 1 + c_1 z + \cdots + c_m z^m$. $P_m(0) = 1$

$$\|x_* - x_{m+1}\|_A \le \left(\max_{\lambda_i} |P_m(\lambda_i)|\right) \underbrace{\|A^{-1} r_0\|_A}_{\substack{\| \\ \|x_0 - x_*\|_A}} .$$

$$\le \left(\min_{\substack{p \in \mathcal{P}_m \\ p(0)=1}} \max_{\lambda_1 \le t \le \lambda_n} |p(t)|\right) \|x_0 - x_*\|_A .$$

$\downarrow$

sol. Chebyshev poly.

$$\le 2\left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^{m+1} \|x_0 - x_*\|_A$$

SD : κ. CG : $\sqrt{\kappa}$.

General non-sym matrices.

$$K_m = \text{span}\{r_0, Ar_0, \cdots, A^{m-1} r_0\}$$

Orthonormalize $\Longrightarrow$ Arnoldi process.

$\beta v_1 = r_0$, $\beta = \|r_0\|$, $\|v_1\| = 1$.

$$A v_1 = h_{21} v_2 + h_{11} v_1$$

$v_2^T v_1 = 0$, $v_2^T v_2 = 1$

$\Rightarrow h_{11} = v_1^T A v_1$

normalization. orthogonal. $\Rightarrow h_{21} = \| A v_1 - h_{11} v_1 \|$

$$A v_1 = [v_1 \ \ v_2] \begin{bmatrix} h_{11} \\ h_{21} \end{bmatrix}$$

$$A[v_1 \ \ v_2] = [v_1 \ \ v_2 \ \ v_3] \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \\ 0 & h_{32} \end{bmatrix}$$

orthogonalization.

normalization

$$\vdots$$

$$A[v_1, \dots, v_m] = [v_1, \dots, v_{m+1}] \underbrace{\begin{bmatrix} h_{11} & \cdots & h_{1m} \\ h_{21} & \cdot & h_{2m} \\ 0 & \ddots & \\ \vdots & & h_{mm} \\ \hline 0 & & h_{m+1,m} \end{bmatrix}}_{m} \to \begin{bmatrix} H_m \\ \hline 0 \cdots 0 \ \ h_{m+1,m} \end{bmatrix}$$

m+1

Arnoldi:

$$A\,[v_1 \cdots v_m] = [v_1 \cdots v_m]\,\underbrace{H_m}_{\text{upper Hessenberg}} + \underbrace{v_{m+1}\, e_{m+1}^T\, h_{m+1,m}}_{\text{"error"}}\,.$$

Alg. Arnoldi process.

Pick v_1, $\|v_1\| = 1$.

for $k = 1, \cdots, m$.

$w = A v_k$

for $j = 1, \cdots, k$

$h_{j,k} = v_j^T w$

$w = w - h_{jk} v_j$

end

$h_{k+1,k} = \|w\|$

if $|h_{k+1,k}| < \tau$. stop

$v_{k+1} = w / h_{k+1,k}$

end.

Arnoldi for linear sys.

start from $r_0 = b - Ax_0$.

Arnoldi. $V_m = [v_1, \cdots, v_m]$

$$AV_m = V_m H_m + v_{m+1} e_{m+1}^T h_{m+1,m}.$$

Project to the space V_m

$$V_m^T A V_m = H_m ..$$

Find $y_m \in \mathbb{R}^m$

$$x_m = x_0 + V_m y_m .$$

$$r_m \perp K_m . \Rightarrow V_m^* (b - A x_m) = 0$$

$$\Updownarrow$$

$$V_m^* (b - A x_0 - A V_m y_m) = 0 .$$

$\beta e_1 = H_m y_m \longrightarrow$ $m \times m$ eq. can solve .

Full orthogonalization method (FOM)

Problems of FOM

① Error can be very large. in the middle.

② Keep entire V_m in memory.

technical. $\longrightarrow$ restarting.

$\downarrow$

Arnoldi for fixed # steps.

Best known fix for ①

Generalized Minimal Residual Method (GMRES)

[Saad. Schultz. 1986].

$$H_m = \begin{pmatrix} h_{11} & \cdots & & - \\ h_{21} & \ddots & & \vdots \\ 0 & & \ddots & \vdots \\ & & h_{m,m-1} & h_{m,m} \end{pmatrix} \qquad \tilde{H}_m = \begin{pmatrix} H_m \\ \hline 0 \; \cdots \; h_{m+1,m} \end{pmatrix}$$

Find $x_m = x_0 + V_m y_m$. s.t.

$$\min_y \| b - A x_m \|_2$$

$$\|b - A x_m\| = \|b - A x_0 - A V_m y_m\|$$

$$= \|\beta v_1 - V_{m+1} \tilde{H}_m y_m\| \quad \xrightarrow[\text{exact}]{\text{Project to } V_{m+1}}$$

$$= \|\beta \tilde{e}_1 - \tilde{H}_m y_m\|$$

$$\begin{array}{cc} & m \\ (m+1) & \begin{bmatrix} & & \\ & \tilde{H}_m & \\ & & \end{bmatrix} \end{array} y_m = \beta \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} (m+1)$$

least squares.

solved via QR decomp.

$$A = Q \quad R \qquad\qquad Q^T Q = I_n.$$

$$\begin{array}{cc} & n \\ m & \begin{bmatrix} \quad & \quad \\ \quad & \quad \\ \quad & \quad \end{bmatrix} \end{array} = \begin{array}{cc} & n \\ m & \begin{bmatrix} \quad & \quad \\ \quad & \quad \\ \quad & \quad \end{bmatrix} \end{array} \begin{array}{cc} & n \\ n & \begin{bmatrix} \diagdown & * & * \\ & & * \end{bmatrix} \end{array}$$

ortho

upper triangular.

$$\min \| \beta \tilde{e}_1 - \tilde{H}_m y_m \|. \qquad \tilde{H}_m = QR.$$

$$\Updownarrow$$

$$\min \| \beta Q^T \tilde{e}_1 - R y_m \|.$$

If $\tilde{H}_m$ has full col rank.

$$R y_m = \beta Q^T \tilde{e}_1 \Rightarrow y_m = R^{-1} (\beta Q^T \tilde{e}_1)$$

$$x_{m+1} = x_0 + \bar{V}_m y_m.$$

Doing GMRES more efficiently.

Remove redundant operations in QR decomposition.

<u>Givens rotation.</u>

$$\min \| Q_m^T \cdots Q_2^T Q_1^T \tilde{H}_m y_m - Q_m^T \cdots Q_1^T \beta e_1 \|$$ e.g. size $(m+1)\times(m+1)$

$$\Downarrow$$

$$\min \left\| \begin{bmatrix} R \\ \hline 0 \cdots 0 \end{bmatrix} y_m - \begin{bmatrix} r_1 \\ \hline s \end{bmatrix} \right\| = |s|. \text{ when } R y_m = r_1$$

Efficient because

① More efficient QR. Reuse previous info

② Automatically keep track of error

w.o. forming x_m.

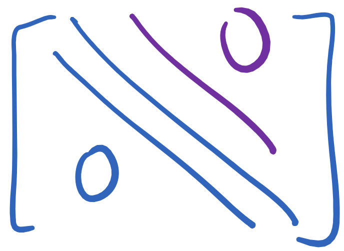
Go back to Pos. Def. matrices. $A \in \mathbb{R}^{n \times n}$.

Assume do Arnoldi.

$$AV = VH$$ $\to$ upper Hess. $\quad V^* V = I$.

$$V^* A V = H$$

sym



Tridiagonal. 3-term recurrence.

Efficient Arnoldi + Sym $\longrightarrow$ Lanczos.

Alg. (Lanczos. conceptual. no reorthogonalization)

Pick v_1, $\|v_1\|=1$, $\beta_1=0$, $v_0=0$.

for $k=1,\dots,m$.

$\tilde{v}_{k+1} = Av_k - \beta_k v_{k-1}$ ← ortho. w. v_{k-1}

$\alpha_k = (v_k, \tilde{v}_{k+1})$

$\tilde{v}_{k+1} = \tilde{v}_{k+1} - \alpha_k v_k$ ← orth. w. v_k

$\beta_{k+1} = \|\tilde{v}_{k+1}\|$, $v_{k+1} = \tilde{v}_{k+1} / \beta_{k+1}$ ← normal.

end

$$AV = VT, \quad T = \begin{pmatrix} \alpha_1 & \beta_1 & & 0 \\ \beta_1 & \ddots & \ddots & \\ & \ddots & \ddots & \beta_{N-1} \\ 0 & & \beta_{N-1} & \alpha_N \end{pmatrix}$$

Lanczos + FOM $\longrightarrow$ CG.