Iterative Numerical Linear Algebra/Iterative Matrix Computations:
Krylov Subspace methods

Iterative Numerical Linear Algebra/Iterative Matrix Computations: Krylov Subspace methods

Suppose $A \in \mathbb{R}^{n \times n}$, n large $\Rightarrow O(n^3)$ complexity and memory required for the standard LU, QR factorizations is infeasible BUT Ar easy to compute for a general vector $r \in \mathbb{R}^n$ because

- A is structured eg. Toeplitz, Hankel, Circulant,$\ldots$
- A is sparse - typically $O(1)$ non-zero entries per row

then it is easy to compute $Ar, A(Ar), \ldots$ and so by taking linear combinations it is easy to compute vectors in the *Krylov Subspaces*

$$\mathcal{K}_k(A, r) = \mathit{span}\{r, Ar, A^2r, \ldots, A^{k-1}r\}, \quad k = 1, 2, \ldots$$

In particular if want to solve linear system

$$\boldsymbol{Ax = b}$$

Take initial guess $\boldsymbol{x_0}$, define residual $\boldsymbol{r_0 = b - Ax_0}$ then

$$\begin{aligned}
\mathcal{K}_{\boldsymbol{1}}(\boldsymbol{A, r_0}) &= \text{span}\{\boldsymbol{r_0}\}, \\
\mathcal{K}_{\boldsymbol{2}}(\boldsymbol{A, r_0}) &= \text{span}\{\boldsymbol{r_0, Ar_0}\} \\
\mathcal{K}_{\boldsymbol{3}}(\boldsymbol{A, r_0}) &= \text{span}\{\boldsymbol{r_0, Ar_0, A^2r_0}\} \\
&\vdots \\
\mathcal{K}_{\boldsymbol{k}}(\boldsymbol{A, r_0}) &= \text{span}\{\boldsymbol{r_0, Ar_0, A^2r_0, \ldots, A^{k-1}r_0}\}
\end{aligned}$$

are sequentially computed with 1 further matrix$\times$vector product.

Krylov subspace:

$$\mathcal{K}_k(A, r_0) := \text{span}\{r_0, Ar_0, A^2 r_0, \ldots, A^{k-1} r_0\}$$

$$\begin{aligned}
x_k \in x_0 + \mathcal{K}_k(A, r_0) \quad &\Leftrightarrow \\
x_k &= x_0 + q_{k-1}(A) r_0 \\
x - x_k &= x - x_0 - q_{k-1}(A) r_0 \\
A(x - x_k) &= A(x - x_0) - A q_{k-1}(A) r_0 \\
\underbrace{b - Ax_k}_{r_k} &= \underbrace{\underbrace{b - Ax_0}_{r_0} - A q_{k-1}(A) r_0}_{p_k(A) r_0}
\end{aligned}$$

so $\quad r_k = p_k(A) r_0$ with $p_k \in \Pi_k, p_k(0) = 1$,

$\Pi_k =$ real polynomials of degree $\leq k$.

Krylov subspace:

$$\mathcal{K}_k(A, r_0) := \text{span}\{r_0, Ar_0, A^2 r_0, \ldots, A^{k-1} r_0\}$$

$$\begin{aligned}
x_k \in x_0 + \mathcal{K}_k(A, r_0) \quad & \Leftrightarrow \\
x_k &= x_0 + q_{k-1}(A) r_0 \\
x - x_k &= x - x_0 - q_{k-1}(A) r_0 \\
A(x - x_k) &= A(x - x_0) - A q_{k-1}(A) r_0 \\
\underbrace{b - Ax_k}_{r_k} &= \underbrace{\underbrace{b - Ax_0}_{r_0} - A q_{k-1}(A) r_0}_{p_k(A) r_0}
\end{aligned}$$

so $\quad r_k = p_k(A) r_0$ with $p_k \in \Pi_k, p_k(0) = 1$,

$\Pi_k =$ real polynomials of degree $\leq k$.
For the same polynomial: $e_k = p_k(A) e_0, \;\; e_k = x - x_k$ as $Ae_k = r_k$

Krylov subspace methods for solving $\boldsymbol{Ax = b}$:
want $\boldsymbol{r_k}$ small $\Rightarrow$ want $\boldsymbol{p_k(A)r_0}$ small
eg. if $\boldsymbol{r_0 = \sum_{i=1}^{n} \alpha_i v_i,\ \ Av_i = \lambda_i v_i, i = 1, \ldots, n}$

$$\begin{aligned} \boldsymbol{r_k} &= \sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \boldsymbol{\alpha_i p_k(A) v_i} \\ &= \sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \boldsymbol{\alpha_i p_k(\lambda_i) v_i} \end{aligned}$$

so if $\boldsymbol{p_k}$ is small at the eigenvalues $\Rightarrow$ $\boldsymbol{r_k}$ small $\ldots$

Krylov subspace methods for solving $Ax = b$:
want r_k small $\Rightarrow$ want $p_k(A)r_0$ small
eg. if $r_0 = \sum_{i=1}^{n} \alpha_i v_i, \;\; Av_i = \lambda_i v_i, i = 1, \ldots, n$

$$\begin{aligned} r_k &= \sum_{i=1}^{n} \alpha_i p_k(A) v_i \\ &= \sum_{i=1}^{n} \alpha_i p_k(\lambda_i) v_i \end{aligned}$$

so if p_k is small at the eigenvalues $\Rightarrow$ r_k small $\ldots$

Krylov subspace methods for $Ax = b$ define/compute

$$x_k \in x_0 + \mathcal{K}_k(A, r_0), \quad k = 1, 2, \ldots$$

residual $r_k = b - Ax_k$ so that x_k is best approximation of x in some sence $\leftrightarrow$ r_k is minimal.

eg. Conjugate Gradient method for A SPD:

$$\|x - x_k\|_A = \|r_k\|_{A^{-1}} \quad \text{minimal}$$

$(\|y\|_A^2 = y^T A y = \langle y, Ay \rangle)$

eg. minimum residual method:

$$\|r_k\|_I = \|x - x_k\|_{A^T A} \quad \text{minimal}$$

for

$$x_k \in x_0 + \mathcal{K}_k(A, r_0) \iff r_k = p_k(A) r_0$$

A symmetric ($\Rightarrow$ orthonormal eigenvectors), Minimum Residuals:
$\|r_k\|_I$ minimal for $x_k \in x_0 + \mathcal{K}_k(A, r_0), \quad k = 1, 2, \ldots$

$$\begin{aligned} \Rightarrow \qquad & \|\sum_{i=1}^{n} \alpha_i p_k(\lambda_i) v_i\|_I^2 \\ = \; & \sum_i \sum_j \alpha_i \alpha_j p_k(\lambda_i) p_k(\lambda_j) v_i^T v_j \\ = \; & \sum_{i=1}^{n} \alpha_i^2 p_k^2(\lambda_i) \end{aligned}$$

minimal
Krylov subspace methods for $Ax = \lambda x$:
p_k small at the eigenvalues of $A \rightarrow$ roots of p_k approximate eigenvalues.

Even though
$\mathcal{K}_k(A, r) = \text{span}\{r, Ar, A^2r, \dots, A^{k-1}r\}, \quad k = 1, 2, \dots,$
the basis

$$\{r, Ar, A^2r, \dots, A^{k-1}r\}$$

is for practical purposes a poor basis for these vector subspaces because $A^k r$ tends to the eigenvector corresponding to the dominant eigenvalue as k gets large:
$r = \sum_{i=1}^{n} \alpha_i v_i, \;\; Av_i = \lambda_i v_i, i = 1, \dots, n$ with $|\lambda_1|$ largest

$$\begin{aligned} \Rightarrow A^k r &= \sum_{i=1}^{n} \alpha_i \lambda_i^k v_i \\ &= \lambda_1^k \left[\alpha_1 v_1 + \sum_{i=2}^{n} \alpha_i \underbrace{\left(\frac{\lambda_i}{\lambda_1}\right)^k}_{\to 0 \text{ as } k \to \infty} v_i \right] \end{aligned}$$

Arnoldi's Algorithm

```
Set  v_1 = r/||r||
for k = 1, 2 ... n
    w = A v_k
    for j = 1, ..., k
        h_{j,k} = v_j^T w
        w = w - h_{j,k} v_j
    end    h_{k+1,k} = ||w||
    v_{k+1} = w / h_{k+1,k}
end
```

computes an orthonormal basis $\{v_1, v_2, \ldots, v_k\}$ for $\mathcal{K}_k(A, r_0)$ for each k

Conjugate Gradient Algorithm

choose x_0 , $r_0 = b - Ax_0 = p_0$ and for $k = 0, 1, 2, \ldots$

$$\begin{aligned} \alpha_k &= p_k^T r_k / p_k^T A p_k \\ x_{k+1} &= x_k + \alpha_k p_k \\ r_{k+1} &= b - A x_{k+1} \\ \beta_k &= -p_k^T A r_{k+1} / p_k^T A p_k \\ p_{k+1} &= r_{k+1} + \beta_k p_k \end{aligned}$$

computes iterates $\{x_k\}$, corresponding residuals $\{r_k\}$ and search directions $\{p_k\}$ so that as long as $x_k \neq x$ we have

$$\begin{aligned} r_k^T p_j &= r_k^T r_j = 0 \ , \ j < k \qquad (1) \\ p_k^T A p_j &= 0 \ , \ j < k \ , \ (p_k^T A p_k \neq 0) \ (2) \\ \text{span}\{r_0, r_1, \ldots, r_{k-1}\} &= \text{span}\{p_0, p_1, \ldots, p_{k-1}\} \\ &= \mathcal{K}_k(A, r_0) \qquad (3) \end{aligned}$$