Krylov subspace methods

Nicholas Hu · Last updated on 2025-03-24

Let $A \in \mathbb{R}^{n \times n}$ be invertible and $b \in \mathbb{R}^n$. A **Krylov subspace method** for solving $Ax = b$ is a projection method in which the search subspaces are **Krylov subspaces** – subspaces of the form $\mathcal{K}_k(A, v) := \operatorname{span}\{A^j v\}_{j=0}^{k-1}$ for some $k \geq 0$ and $v \in \mathbb{R}^n$.

Krylov subspaces

In this section, we regard A as a linear operator on a nontrivial finite-dimensional vector space V. Recall that the **minimal polynomial of** A is the *monic* polynomial μ_A of minimal degree such that $\mu_A(A) = 0$ and that the **minimal polynomial of** v **with respect to** A (also known as the A**-annihilator of** v) is the *monic* polynomial $\mu_{A,v}$ of minimal degree such that $\mu_{A,v}(A)v = 0$. In particular, any polynomial p with $p(A) = 0$ must be a multiple of μ_A, and similarly for $\mu_{A,v}$.

> $\deg(\mu_A) \leq \dim(V)$.

Proof. If $\dim(V) = 1$, this is trivial, so suppose that for some $n > 1$ it is true whenever $\dim(V) < n$. Let $v \in V \setminus \{0\}$ and $m := \deg(\mu_{A,v})$. Then $\{A^j v\}_{j=0}^{m-1}$ is linearly independent and annihilated by $\mu_{A,v}(A)$. Hence $W := \operatorname{im}(\mu_{A,v}(A))$ is an A-invariant subspace of V with $\deg(\mu_{A\restriction_W}) \leq n - m < n$, and $(\mu_{A\restriction_W}\mu_{A,v})(A) = 0$. ∎

Now suppose that $v \in V \setminus \{0\}$ and let $m := \deg(\mu_{A,v})$.

> $\{A^j v\}_{j=0}^{\min\{k,m\}-1}$ is a basis of $\mathcal{K}_k(A, v)$. In particular, $\dim(\mathcal{K}_k(A, v)) = \min\{k, m\}$.

Proof. If $x \in \mathcal{K}_k(A, v)$, then $x = p(A)v$ for some polynomial p with $\deg(p) \leq k - 1$. Dividing p by $\mu_{A,v}$, we obtain $p = q\mu_{A,v} + r$ for some polynomials q and r with $\deg(r) \leq m - 1$, so $x = r(A)v$ and hence $\{A^j v\}_{j=0}^{\min\{k,m\}-1}$ spans $\mathcal{K}_k(A, v)$. Moreover, if $\sum_{j=0}^{\min\{k,m\}-1} c_j A^j v = 0$, the c_j must be zero by the minimality of m. ∎

The A**-cyclic subspace generated by** v is $\mathcal{C}(A, v) := \operatorname{span}\{A^j v\}_{j=0}^{\infty}$ and is the smallest A-invariant subspace of V containing v. Clearly, $\mathcal{C}(A, v) = \mathcal{K}_m(A, v)$, so the cyclic subspace is also the largest Krylov subspace generated by v.

> The following are equivalent:
>
> - $k \geq m$
> - $\mathcal{K}_k(A, v)$ is A-invariant
> - $\mathcal{K}_k(A, v) = \mathcal{C}(A, v)$
>
> Moreover, if $v = r^{(0)} := b - Ax^{(0)}$, these are equivalent to $x^* := A^{-1}b \in x^{(0)} + \mathcal{K}_k(A, r^{(0)})$.

Proof. The equivalence of the first three statements follows from the discussion above, and from the general theory of projection methods, we know that $x^* \in x^{(0)} + \mathcal{K}_k(A, r^{(0)})$ if $\mathcal{K}_k(A, r^{(0)})$ is A-invariant. On the other hand, if $x^* \in x^{(0)} + \mathcal{K}_k(A, r^{(0)})$, then $r^{(0)} = Ap(A)r^{(0)}$ for some polynomial p with $\deg(p) \leq k - 1$, so $q(t) := 1 - tp(t)$ is a polynomial such that $q(A)r^{(0)} = 0$. Hence $m \leq \deg(q) \leq k$. ∎

The Arnoldi iteration

To construct Krylov subspace methods, it is useful to generate well-conditioned bases of such subspaces. The **Arnoldi iteration** produces *orthonormal* bases of successive Krylov subspaces $\mathcal{K}_j(A, v)$ ($v \neq 0$) using (modified) Gram–Schmidt orthogonalization [1]:

$$
\begin{aligned}
&q_1 = v/\|v\| \\
&\\
&\texttt{for } j = 1 \texttt{ to } k\texttt{:} \\
&\quad v_j = Aq_j \\
&\quad \texttt{for } i = 1 \texttt{ to } j\texttt{:} \\
&\quad\quad h_{ij} = \langle v_j, q_i \rangle \\
&\quad\quad v_j = v_j - h_{ij} q_i \\
&\quad h_{j+1,\,j} = \|v_j\| \\
&\quad q_{j+1} = v_j / h_{j+1,\,j}
\end{aligned}
$$

Indeed, if $\{q_i\}_{i=1}^{j}$ is an orthonormal basis of $\mathcal{K}_j(A, v)$ (which it is for $j = 1$), then initially $v_j \in A\mathcal{K}_j(A, v) \subseteq \mathcal{K}_{j+1}(A, v)$. Subsequently, v_j is orthogonalized against $q_1, \ldots, q_j$ and normalized to form q_{j+1} (provided that $h_{j+1,\,j} \neq 0$), which implies that $\{q_i\}_{i=1}^{j+1}$ is an orthonormal set of vectors in $\mathcal{K}_{j+1}(A, v)$ and hence a basis thereof.

Thus, if $m = \deg(\mu_{A,v})$, the Arnoldi iteration will *break down in the m^{th} iteration* (in the sense that $h_{m+1,\,m} = 0$). For if it did not break down by the m^{th} iteration, we would have $\dim(\mathcal{K}_{m+1}(A, v)) = m + 1$; and if it breaks down (for the first time) in the j^{th} iteration, then $Aq_j - \sum_{i=1}^{j} h_{ij} q_i = 0$, which is to say that $p(A)v = 0$ for some polynomial p with $\deg(p) \leq j$, so $j \geq m$.

After completing the Arnoldi iteration, we obtain $Aq_j = \sum_{i=1}^{j+1} h_{ij} q_i$ for $1 \leq j \leq k$, which we can express in matrix form as

$$
A\underbrace{\begin{bmatrix} q_1 & \cdots & q_k \end{bmatrix}}_{=:\,Q_k} = \underbrace{\begin{bmatrix} q_1 & \cdots & q_{k+1} \end{bmatrix}}_{Q_{k+1}} \underbrace{\begin{bmatrix} h_{11} & \cdots & h_{1k} \\ h_{21} & \ddots & \vdots \\ & \ddots & h_{kk} \\ & & h_{k+1,\,k} \end{bmatrix}}_{=:\,\widetilde{H}_k}.
$$

We can also view this as a reduction of A to upper Hessenberg form: $Q_k^{\top} A Q_k = H_k$, where H_k denotes the upper $k \times k$ submatrix of $\widetilde{H}_k$.

GMRES

The **generalized minimal residual (GMRES)** method for solving $Ax = b$ is an iterative *residual projection method* whose k^{th} iterate $x^{(k)}$ lies in $x^{(0)} + \mathcal{K}_k(A, r^{(0)})$, where $r^{(k)} := b - Ax^{(k)}$. Thus, if we apply the Arnoldi iteration to $r^{(0)}$, we can write $x^{(k)} = x^{(0)} + Q_k y^{(k)}$ with Q_k as above and $y^{(k)} \in \mathbb{R}^k$ minimizing $\|r^{(k)}\| = \|r^{(0)} - AQ_k y^{(k)}\| = \|\beta_0 e_1^{(k)} - \widetilde{H}_k y^{(k)}\|$, where $\beta_0 := \|r^{(0)}\|$ and $e_1^{(k)} := \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}^{\top} \in \mathbb{R}^{k+1}$ (since Q_{k+1} has orthonormal columns).

These least squares problems can be solved by incrementally triangularizing the upper Hessenberg matrices $\widetilde{H}_k$. More precisely, if Ω_k is an orthogonal matrix such that $\Omega_k \widetilde{H}_k =: \begin{bmatrix} R_k \\ 0_{1\times k} \end{bmatrix} \in \mathbb{R}^{(k+1)\times k}$ is upper triangular, then

$$\begin{bmatrix} \Omega_k & \\ & 1 \end{bmatrix} \underbrace{\begin{bmatrix} \widetilde{H}_k & * \\ & h_{k+2,\,k+1} \end{bmatrix}}_{\widetilde{H}_{k+1}} = \begin{bmatrix} R_k & * \\ 0 & * \\ & h_{k+2,\,k+1} \end{bmatrix}.$$

Hence this matrix can in turn be triangularized using a single Givens rotation G_{k+1} (and Ω_1 itself can be chosen to be a Givens rotation):

$$\underbrace{G_{k+1} \begin{bmatrix} \Omega_k & \\ & 1 \end{bmatrix}}_{=:\,\Omega_{k+1}} \underbrace{\begin{bmatrix} \widetilde{H}_k & * \\ & h_{k+2,\,k+1} \end{bmatrix}}_{\widetilde{H}_{k+1}} = \begin{bmatrix} R_k & * \\ 0 & * \\ & 0 \end{bmatrix} =: \begin{bmatrix} R_{k+1} \\ 0_{1\times(k+1)} \end{bmatrix}.$$

Furthermore, if $\Omega_k(\beta_0 e_1^{(k)}) =: \begin{bmatrix} b_k \\ \beta_k \end{bmatrix} \in \mathbb{R}^{k+1}$, then $\|r^{(k)}\| = \left\| \begin{bmatrix} b_k \\ \beta_k \end{bmatrix} - \begin{bmatrix} R_k \\ 0_{1\times k} \end{bmatrix} y^{(k)} \right\|$ is minimized when $y^{(k)} = R_k^{-1} b_k$ and $\|r^{(k)}\| = |\beta_k|$, and we have

$$\underbrace{G_{k+1} \begin{bmatrix} \Omega_k & \\ & 1 \end{bmatrix}}_{\Omega_{k+1}} \underbrace{\begin{bmatrix} \beta_0 e_1^{(k)} \\ 0 \end{bmatrix}}_{\beta_0 e_1^{(k+1)}} = G_{k+1} \begin{bmatrix} b_k \\ \beta_k \\ 0 \end{bmatrix} = \begin{bmatrix} b_k \\ * \\ * \end{bmatrix} =: \begin{bmatrix} b_{k+1} \\ \beta_{k+1} \end{bmatrix}.$$

We note that GMRES breaks down precisely when the underlying Arnoldi iteration does, which in view of the discussion above is equivalent to the approximate solution being exact.

Convergence

By definition, the residuals in GMRES satisfy

$$\|r^{(k)}\| = \min_{p_{k-1} \in P_{k-1}} \|(I - A p_{k-1}(A)) r^{(0)}\| = \min_{p_k \in P_k,\, p_k(0)=1} \|p_k(A) r^{(0)}\|,$$

where P_k denotes the vector space of polynomials with degree at most k. This immediately yields the following estimate for diagonalizable matrices.

> Let $\sigma(A)$ denote the spectrum of A. If $A = V \Lambda V^{-1}$ for some diagonal matrix Λ, then
>
> $$\|r^{(k)}\| \leq \kappa(V) \cdot \min_{p_k \in P_k,\, p_k(0)=1} \max_{\lambda \in \sigma(A)} |p_k(\lambda)| \cdot \|r^{(0)}\|.$$

The Lanczos iteration

When A is *symmetric*, the Arnoldi iteration reduces to what is known as the **Lanczos iteration**. Since $H_k = Q_k^\top A Q_k$, the upper Hessenberg matrix H_k must also be symmetric and therefore *symmetric tridiagonal*. For this reason, we denote it by T_k and define $\alpha_j := t_{jj}$ and $\beta_j := t_{j,\,j+1} = t_{j+1,\,j}$. With this notation, the algorithm is as follows.

$\beta_0 = 0,\ q_0 = 0$
$q_1 = v/\|v\|$

for $j = 1$ **to** k**:**
  $v_j = Aq_j$
  $\alpha_j = \langle v_j, q_j \rangle$
  $v_j = v_j - \beta_{j-1} q_{j-1} - \alpha_j q_j$
  $\beta_j = \|v_j\|$
  $q_{j+1} = v_j / \beta_j$

CG

The **conjugate gradient (CG)** method for solving $Ax = b$ when A is *symmetric positive definite* is an iterative *error projection method* whose k^{th} iterate $x^{(k)}$ lies in $x^{(0)} + \mathcal{K}_k(A, r^{(0)})$, where $r^{(k)} := b - Ax^{(k)}$. Although it is possible to derive CG from the Lanczos iteration just as GMRES was derived from the Arnoldi iteration, a simpler and more direct derivation is given in the notes on the conjugate gradient method.

Convergence

By definition, the errors in CG satisfy

$$\|e^{(k)}\|_A = \min_{p_{k-1} \in P_{k-1}} \|(I - Ap_{k-1}(A))e^{(0)}\|_A = \min_{p_k \in P_k,\, p_k(0)=1} \|p_k(A)e^{(0)}\|_A,$$

where P_k denotes the vector space of polynomials with degree at most k. Using the fact that $A = Q\Lambda Q^{-1}$ for some orthogonal matrix Q and some diagonal matrix Λ, and that $\|Qp_k(\Lambda)Q^{-1}\|_A = \|p_k(\Lambda)\|_2$, we obtain an estimate analogous to the one for GMRES.

> Let $\sigma(A)$ denote the spectrum of A. Then
>
> $$\|e^{(k)}\|_A \leq \min_{p_k \in P_k,\, p_k(0)=1} \max_{\lambda \in \sigma(A)} |p_k(\lambda)| \cdot \|e^{(0)}\|_A.$$

Now suppose that the eigenvalues of A are $\lambda_1 \geq \cdots \geq \lambda_n > 0$ with $\lambda_1 \neq \lambda_n$. We know from approximation theory that the polynomial $p_k \in P_k$ with $p_k(0) = 1$ that minimizes $\max_{\lambda \in [\lambda_n, \lambda_1]} |p_k(\lambda)|$ is $p_k = \frac{T_k \circ \alpha}{(T_k \circ \alpha)(0)}$, where T_k is the k^{th} Chebyshev polynomial and $\alpha(t) := -1 + \frac{1-(-1)}{\lambda_1 - \lambda_n}(t - \lambda_n)$ maps $[\lambda_n, \lambda_1]$ affinely to $[-1, 1]$. Since $|T_k| \leq 1$ on $[-1, 1]$ and $\alpha(0) = -\frac{\kappa+1}{\kappa-1}$, where $\kappa := \kappa_2(A)$, upon evaluating T_k at $\alpha(0)$ we obtain

> $$\|e^{(k)}\|_A \leq \frac{2}{\left(\frac{\sqrt{\kappa}+1}{\sqrt{\kappa}-1}\right)^k + \left(\frac{\sqrt{\kappa}+1}{\sqrt{\kappa}-1}\right)^{-k}} \|e^{(0)}\|_A \leq 2\left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^k \|e^{(0)}\|_A.$$

---

1. In the algorithm below, standard Gram–Schmidt orthogonalization would set $v_j = v_j - \sum_{i=1}^{j} h_{ij} q_i$ after computing the h_{ij} instead of updating it inside the loop. It is also possible to use Householder reflections. ↩