MA5233 Computational Mathematics

Lecture 9: Krylov Subspace Methods

Simon Etter

Semester I, AY 2020/2021

Krylov Subspace Methods

Problem statement

$$\text{Given } A \in \mathbb{R}^{n \times n} \text{ and } b \in \mathbb{R}^n,$$
$$\text{determine } p \in \mathcal{P}_m \text{ such that } p(A)\, b \approx A^{-1}\, b.$$

Why polynomial approximation?

We have seen in Lecture 7 that finite differences reduces PDEs to large and sparse linear systems, and in Lecture 8 that LU factorisation cannot solve these linear systems with optimal $O(n)$ runtime if the PDE is posed on a $(d > 1)$-dimensional domain.

By contrast, evaluating

$$p(A)\, b = \sum_{k=0}^{m} c_k\, A^k\, b$$

requires only m matrix-vector products $A\,(A^{k-1}\, b)$ and m vector sums. The overall runtime $p(A)\, b$ is hence $O(mn)$ if A has $O(n)$ nonzero entries, which is optimal as long as $m = O(1)$ for $n \to \infty$.

Krylov Subspace Methods

Outlook

Tackling the above problem requires some background in the theory of eigenvalues and matrix functions. This slide and the following will review the basics.

Def: Eigendecomposition

Let $A \in \mathbb{R}^{n \times n}$. Its eigendecomposition is a pair of matrices $V, \Lambda \in \mathbb{C}^{n \times n}$ such that $A = V \Lambda V^{-1}$ and Λ is diagonal.

Rem: Eigendecomposition vs. eigenpairs

Rewriting $A = V \Lambda V^{-1}$ in the form $AV = V\Lambda$ and considering the kth column in this identity, we obtain

$$A\, V[:, k] = V[:, k]\, \Lambda[k, k].$$

This shows that the columns of V are the eigenvectors and the diagonal entries of Λ are the eigenvalues of A.

Notation

I will frequently use the abbreviation $\lambda_k = \Lambda[k, k]$ in the following.

Krylov Subspace Methods

Def: Function of a matrix

Assume $A \in \mathbb{R}^{n \times n}$ has eigendecomposition $A = V \Lambda V^{-1}$, and let $f : \mathbb{C} \to \mathbb{C}$. The matrix function $f(A)$ is then given by

$$f(A) = V\, f(\Lambda)\, V^{-1} \qquad \text{where} \qquad f(\Lambda) = \operatorname{diag}\big(f(\lambda_1), \ldots, f(\lambda_n)\big).$$

Lemma: Consistency of matrix functions

The above definition of matrix functions is consistent with matrix sums, products and inverses, i.e. we have

$$\begin{aligned} h(x) = f(x) + g(x) \quad &\Longleftrightarrow \quad h(A) = f(A) + g(A) \\ h(x) = f(x)\, g(x) \quad &\Longleftrightarrow \quad h(A) = f(A)\, g(A) \\ h(x) = 1/f(x) \quad &\Longleftrightarrow \quad h(A) = f(A)^{-1}. \end{aligned}$$

Proof. See next slide.

Krylov Subspace Methods

Proof.

Addition:

$$\begin{aligned} h(A) &= V\, h(\Lambda)\, V^{-1} = V\, \big(f(\Lambda) + g(\Lambda)\big)\, V^{-1} \\ &= V\, f(\Lambda)\, V^{-1} + V\, g(\Lambda)\, V^{-1} = f(A) + g(A). \end{aligned}$$

Multiplication:

$$\begin{aligned} h(A) &= V\, h(\Lambda)\, V^{-1} = V\, f(\Lambda)\, g(\Lambda)\, V^{-1} \\ &= V\, f(\Lambda)\, V^{-1}\, V\, g(\Lambda)\, V^{-1} = f(A)\, g(A) \end{aligned}$$

Inverse:

Consider $g(x) = 1/f(x)$ and $h(x) = f(x)\, g(x) = f(x)/g(x) = 1$.
According to the multiplication result, we then have

$$f(A)\, g(x) = h(A) = V\, I\, V^{-1} = I$$

and hence $g(A) = f(A)^{-1}$.

Krylov Subspace Methods

Lemma: 2-norm of matrix functions

Assume $A \in \mathbb{R}^{n \times n}$ has eigendecomposition $A = V \Lambda V^{-1}$. Then,

$$\|f(A)\|_2 \quad \leq \quad \kappa_2(V) \quad \max_k |f(\lambda_k)|.$$

Proof. We have $\|AB\| \leq \|A\| \, \|B\|$ for any induced matrix norm; hence

$$\|f(A)\|_2 \leq \|V\|_2 \, \|f(\Lambda)\|_2 \, \|V^{-1}\|_2 = \kappa_2(V) \, \|f(\Lambda)\|_2.$$

The claim follows by noting that the 2-norm of a matrix is equal to its largest singular value and that the singular values of a diagonal matrix are given by its diagonal entries.

Krylov Subspace Methods

Discussion

The above results provide a rigorous mathematical justification for why it makes sense to attempt to approximate $A^{-1}b$ using $p(A)\,b$:

The lemma from slide 4 shows that A^{-1} is just $f(x) = 1/x$ applied to A, and the lemma from slide 6 shows that if we replace $f(A)$ with a polynomial approximation $p(A)$, then we introduce an error which we can bound in terms of the supremum norm difference between $p(x)$ and $f(x)$.

The corollary on the next slide summarises and completes this line of thought.

Krylov Subspace Methods

Corollary: Polynomial approximation for solving linear systems

Assume $A \in \mathbb{R}^{n \times n}$ is a matrix with eigendecomposition $A = V \Lambda V^{-1}$ such that all eigenvalues λ_k are contained in some set $S \subset \mathbb{C}$, and let $b \in \mathbb{R}^n$ and $p \in \mathcal{P}_m$.

We then have

$$\left\| p(A)\, b - A^{-1} b \right\|_2 \quad \leq \quad \kappa(V) \, \left\| p(x) - 1/x \right\|_S \, \|b\|_2.$$

Proof. Let $f(x) = 1/x$. We then have

$$\begin{aligned} \left\| p(A)\, b - A^{-1} b \right\|_2 &\leq \left\| p(A) - f(A) \right\|_2 \, \|b\|_2 \\ &\leq \kappa(V) \, \left\| p(x) - 1/x \right\|_S \, \|b\|_2, \end{aligned}$$

where on the first line I used the matrix function lemma from slide 4, the distributivity of matrix-vector products and the fact that $\|Ab\| \leq \|A\| \, \|b\|$ for any induced matrix norm, and on the second line I used the 2-norm lemma from slide 6 and the fact that $\lambda_k \in S$ for all $k \in \{1, \ldots, n\}$.

Krylov Subspace Methods

Discussion

The above corollary shows that we can solve $Ax = b$ to arbitrary accuracy using the following three-step program.

- Determine a set $S \subset \mathbb{C}$ containing all eigenvalues of A.
- Determine a polynomial p such that

$$\left\| p(x) - 1/x \right\|_S \quad \leq \quad \frac{\text{error tolerance}}{\kappa(V) \left\| b \right\|_2}.$$

- Evaluate $\tilde{x} \approx p(A)\, b$.

Implementation of this scheme is straightforward except that we must be careful to evaluate $p(A)\, b$ using an algorithm which is both efficient and stable. The next slide will elaborate further.

Krylov Subspace Methods

Evaluating matrix polynomials

Perhaps the most obvious way to evaluate $p(A)\,b$ would be to use the formula

$$p(A)\,b = \sum_{k=0}^{n} c_k\,A^k\,b,$$

but one can show that the resulting algorithm is not numerically stable.

We do not have the mathematical tools to understand why the above formula is unstable while the one below is stable. You just have to believe me that the stability properties are as I claim them to be.

A numerically stable way to evaluate $p(A)\,b$ is to express $p(A)\,b$ in terms of the Chebyshev polynomials $T_k(x)$,

$$p(A)\,b = \sum_{k=0}^{n} c_k\,T_k(A)\,b,$$

and use the Chebyshev recurrence formula

$$T_{k+1}(A)\,b = 2A\,T_k(A)\,b - T_{k-1}(A)\,b$$

to evaluate $T_k(A)\,b$ efficiently.

Krylov Subspace Methods

Numerical implementation

See `solve()` for an implementation of the polynomial approximation scheme from slide 9.

Discussion

We have seen on slide 8 that

$$\left\| p(A)\, b - A^{-1} b \right\|_2 \quad \leq \quad \kappa(V)\, \left\| p(x) - 1/x \right\|_S\, \|b\|_2.$$

We can therefore estimate the rate of convergence of the scheme from slide 9 using polynomial approximation.

The result on the next slide fills in the details.

Krylov Subspace Methods

Thm: Convergence of polynomial approximation to $\frac{1}{x}$

Assume $d > c > 0$, and set $\kappa = \frac{d}{c}$. Then,

$$\min_{p \in \mathcal{P}_m} \|p(x) - 1/x\|_{[c,d]} \leq \left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^m$$

Proof. The above approximation problem is posed on an arbitrary interval $[c, d]$, while most of the theory from Lecture 4 was formulated for the particular interval $[-1, 1]$.

In order to apply the theory from Lecture 4, we hence need to map the interval $[c, d]$ to $[-1, 1]$ using the fact that

$$\min_{p \in \mathcal{P}_m} \|p(x) - 1/x\|_{[c,d]} \quad = \quad \min_{p \in \mathcal{P}_m} \|p(\hat{x}) - 1/\psi(\hat{x})\|_{[-1,1]},$$

where

$$\psi(\hat{x}) = \tfrac{d+c}{2} + \tfrac{d-c}{2}\,\hat{x}, \qquad \psi^{-1}(x) = \tfrac{2}{d-c}\left(x - \tfrac{d+c}{2}\right)$$

is the unique linear polynomial $\psi \in \mathcal{P}_1$ such that $\psi([-1, 1]) = [c, d]$.

See Lecture 4 for proofs of the various claims in the above sentence.

Krylov Subspace Methods

Proof (continued).

Since $\psi(\hat{x}) \in \mathcal{P}_1$ is analytic, $1/\psi(\hat{x})$ is analytic except if

$$\psi(\hat{x}) = 0 \quad \iff \quad \hat{x} = \psi^{-1}(0) = \tfrac{d+c}{d-c};$$

hence we conclude that

$$\min_{p \in \mathcal{P}_m} \|p(x) - 1/x\|_{[c,d]} \;\leq\; \left|\phi_-^{-1}\left(\tfrac{d+c}{d-c}\right)\right|^m \tag{1}$$

where

$$\phi_-^{-1}(x) = x - \sqrt{x^2 - 1}$$

is the branch of the inverse Joukowsky map $\phi_\pm^{-1}(x)$ such that $|\phi_-^{-1}(x)| < 1$ for positive arguments $x = \frac{d+c}{d-c} > 0$.

Equation (1) follows from the "Chebyshev interpolation error for analytic functions" theorem from Lecture 4.

It remains to show that

$$\left|\phi_-^{-1}\left(\tfrac{d+c}{d-c}\right)\right| = \left|\phi_-^{-1}\left(\tfrac{\kappa+1}{\kappa-1}\right)\right| = \tfrac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}.$$

This will be tackled in the lemma on the next slide.

Krylov Subspace Methods

Lemma

$$\left|\phi_-^{-1}\left(\tfrac{\kappa+1}{\kappa-1}\right)\right| = \tfrac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}$$

Proof. $$\phi_-^{-1}\left(\tfrac{\kappa+1}{\kappa-1}\right) = \tfrac{\kappa+1}{\kappa-1} - \sqrt{\left(\tfrac{\kappa+1}{\kappa-1}\right)^2 - 1} = \tfrac{\kappa+1-\sqrt{(\kappa+1)^2-(\kappa-1)^2}}{\kappa-1}$$
$$= \tfrac{\kappa-2\sqrt{\kappa}+1}{\kappa-1} = \tfrac{(\sqrt{\kappa}-1)^2}{(\sqrt{\kappa}+1)(\sqrt{\kappa}-1)} = \tfrac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}.$$

Remark

If A is symmetric, then its eigenvalues are also its singular values; hence if

$$c = \min_k \lambda_k, \qquad d = \max_k \lambda_k,$$

then

$$\kappa = \frac{d}{c} = \frac{\max_k \sigma_k}{\min_k \sigma_k} = \kappa(A)$$

is the condition number of A.

Numerical demonstration

See `convergence()`.

Krylov Subspace Methods

Discussion

We have now seen how we can use polynomial approximation to solve $Ax = b$, assuming we know a set $S \subset \mathbb{C}$ containing all eigenvalues of A.

This is a very big assumption: it is not at all obvious how we can determine such a set S cheaply.

For example, computing all eigenvalues of $A \in \mathbb{R}^{n \times n}$ requires $O(n^3)$ operations and is hence as expensive as computing a dense LU factorisation.

On the other hand, if we settle for only approximate knowledge of the eigenvalues, then we must trade off the following risks.

- If we choose S too large, then we waste matrix-vector products to approximate $x \mapsto \frac{1}{x}$ in regions where this is not needed.
- If we choose S too small, then $p(A)\, b$ diverges since

$$p \in \mathcal{P}_m \qquad \Longrightarrow \qquad |p(x)| = O(|x|^m) \text{ for } |x| \to \infty.$$

See `polynomial()` for numerical demonstration.

Krylov Subspace Methods

Discussion (continued)

We therefore conclude that solving $Ax = b$ using

$$\tilde{x} = p(A)\, b \qquad \text{where} \qquad p = \min_{p \in \mathcal{P}_m} \|p(x) - 1/x\|_S$$

is not practical except in special circumstances where we have a good guess for S.

However, this does not invalidate the overall idea of finding $p(x) \approx 1/x$ first and then evaluating $p(A)\, b$; we just need to find a rule for choosing p which does not require any knowledge about the eigenvalues of A.

One such rule is given by

$$p = \arg\min_{p \in \mathcal{P}_m} \|A\tilde{x} - b\| \quad \text{where} \quad \tilde{x} = p(A)\, b.$$

This rule does not make any reference to the eigenvalues of A, and we will see in a moment that the resulting optimisation problem can also be solved without any such knowledge.

I will give details soon, but first let us establish some terminology.

Krylov Subspace Methods

Def: Krylov subspace method

$$\text{Compute } p(A)\,b \approx A^{-1}b \qquad \text{where} \qquad p = \arg\min_{p \in \mathcal{P}_m} \|A\,p(A)\,b - b\|.$$

Note that this definition does specify the norm in which we minimise $A\,p(A)\,b - b$. The above definition hence describes a family of algorithms which are in part generated by choosing different norms in which to minimise $A\,p(A)\,b - b$.

Terminology: Residual

The vector $A\tilde{x} - b$ minimised in Krylov subspace methods is called the *residual* of the approximate solution $\tilde{x} \approx A^{-1}b$.

Def: GMRES algorithm (Generalised Minimal RESidual)

$$\text{Compute } p(A)\,b \approx A^{-1}b \qquad \text{where} \qquad p = \arg\min_{p \in \mathcal{P}_m} \|A\,p(A)\,b - b\|_2.$$

Put differently, GMRES is the Krylov subspace method where the residual is minimised in the 2-norm.

Krylov Subspace Methods

Solving the GMRES optimisation problem

[To be added.]