MA5233 Computational Mathematics

Lecture 9: Krylov Subspace Methods

Simon Etter

National University of Singapore

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.

Note that generally $V, \Lambda \in \mathbb{C}^{n \times n}$ are complex even if $A \in \mathbb{R}^{n \times n}$ is real!

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

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

Lemma: Consistency of matrix functions

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

Proof. See next slide.

Corollary

$p(A)$ for $p \in \mathcal{P}_m$ can be evaluated using only matrix sums and products.

Krylov Subspace Methods

Proof.

Constant: $$h(A) = Vh(\Lambda)V^{-1} = VIV^{-1} = I.$$

Identity: $$h(A) = Vh(\Lambda)V^{-1} = V\Lambda V^{-1} = A.$$

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 and constant result, we then have

$$f(A)\,g(A) = h(A) = I$$

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 the absolute values of its diagonal entries.

Krylov Subspace Methods

Discussion

The above results provide a rigorous 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

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

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

Polynomial approximation algorithm

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

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

- 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

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

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 $T_k(x)$,

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

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

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

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,

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

Proof. The above approximation problem is posed on an arbitrary interval $[c, d] \subset (0, \infty)$ 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

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

hence we conclude that

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

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

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

Proof.

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

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

Numerical demonstration

See `convergence()`.

Remark

If A is symmetric and positive definite, then its eigenvalues are also its singular values; hence if A is symmetric and $[c, d]$ is as tight as possible, i.e. $c = \min_k \lambda_k$ and $d = \max_k \lambda_k$, then

$$\kappa = \frac{d}{c} = \frac{\max_k \sigma_k}{\min_k \sigma_k} = \kappa_2(A)$$

is the condition number of A.

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.

Unfortunately, it is not at all obvious how we would determine such an S.

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

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

is not practical except if we are lucky enough to 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 can be obtained by demanding that the residual of the approximate solution $\tilde{x} = p(A)\,b$ should be as small as possible.

Def: Residual

$b - A\tilde{x}$ is called the *residual* of an approximate solution $\tilde{x}$ to $Ax = b$.

Algorithms based on this idea are known under the following name.

Def: Krylov subspace methods

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

Krylov Subspace Methods

Discussion

The rationale for minimising the residual is two-fold:

- On the one hand, we will see shortly that $\min_p \|A\,p(A)\,b - b\|$ can be solved reasonably efficiently.
- On the other hand, the following result shows that minimising the residual is in some sense equivalent to minimising the error.

Lemma: Equivalence of error and residual

$$\|A\|^{-1}\,\|A\tilde{x} - b\| \;\leq\; \|\tilde{x} - A^{-1}b\| \;\leq\; \|A^{-1}\|\,\|A\tilde{x} - b\|$$

Proof. Immediate consequence of

$$A\tilde{x} - b = A\,(\tilde{x} - A^{-1}b) \qquad \text{and} \qquad \tilde{x} - A^{-1}b = A^{-1}\,(A\tilde{x} - b).$$

Krylov Subspace Methods

Discussion

Note that the definition of Krylov subspace methods on slide 16 does not specify the norm in which to minimise the residual. This is because there are several Krylov subspace methods which are in part generated by choosing different norms for measuring the size of the residual.

If the residual is measured in the 2-norm, then we obtain a Krylov subspace method known under the following name.

Def: GMRES algorithm (Generalised Minimal RESidual)

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

Outlook

I will next show that GMRES converges with exactly the same rate as the polynomial approximation scheme from slide 9. The key to doing so is to reduce the above matrix-valued 2 norm minimisation problem to a scalar-valued supremum norm minimisation problem as shown next.

Krylov Subspace Methods

Thm: Error estimate for GMRES

Assume $A \in \mathbb{R}^{n\times n}$ has eigendecomposition $A = V \Lambda V^{-1}$, and assume all eigenvalues of A are contained in some set $S \subset \mathbb{C}$.

Then, the GMRES residual satisfies

$$\min_{p \in \mathcal{P}_m} \|A\,p(A)\,b - b\|_2 \quad \le \quad \kappa(V)\ \|b\|_2\ \min_{q \in \mathcal{P}_{m+1}} \tfrac{\|q\|_S}{|q(0)|}.$$

Proof. See next slide.

Remark

Note that this result reintroduces the eigenvalue set $S \subset \mathbb{C}$, which as discussed on slide 15 is generally not available in applications.

The key difference to what we had before is that Krylov subspace methods require S only to understand convergence, not to run the algorithm.

Krylov Subspace Methods

Thm: Error estimate for GMRES (copied from above)

$$\min_{p\in\mathcal{P}_m} \|A\,p(A)\,b - b\|_2 \quad \leq \quad \kappa(V)\ \|b\|_2\ \min_{q\in\mathcal{P}_{m+1}} \tfrac{\|q\|_S}{|q(0)|}.$$

Proof. Using the arguments outlined at the beginning of this lecture, we obtain

$$\min_{p\in\mathcal{P}_m} \left\|Ap(A)\,b - b\right\|_2 \leq \kappa(V)\,\|b\|_2\ \min_{p\in\mathcal{P}_m} \|x\,p(x) - 1\|_S.$$

It hence remains to show that

$$\min_{p\in\mathcal{P}_m} \|x\,p(x) - 1\|_S = \min_{q\in\mathcal{P}_{m+1}} \tfrac{\|q\|_S}{|q(0)|}.$$

This will be done on the next slide.

Krylov Subspace Methods

Lemma

$$\min_{p \in \mathcal{P}_m} \|x\, p(x) - 1\|_S = \min_{q \in \mathcal{P}_{m+1}} \tfrac{\|q\|_S}{|q(0)|}$$

Proof. $\leq$: Assume $q \in \mathcal{P}_{m+1}$ is a minimiser of the right-hand side. Since $1 - q(0)/q(0) = 0$, the fundamental theorem of algebra guarantees that there exists a $p \in \mathcal{P}_m$ such that

$$1 - q(x)/q(0) = x\, p(x) \qquad \Longleftrightarrow \qquad x\, p(x) - 1 = -\tfrac{q(x)}{q(0)}.$$

This shows that $\min\limits_{p \in \mathcal{P}_m} \|x\, p(x) - 1\|_S \leq \min\limits_{q \in \mathcal{P}_{m+1}} \tfrac{\|q\|_S}{|q(0)|}$.

$\geq$: Assume $p \in \mathcal{P}_m$ is a minimiser of the left-hand side. We then have $q(x) = x\, p(x) - 1 \in \mathcal{P}_{m+1}$ and $q(0) = -1$, and hence

$$\min_{p \in \mathcal{P}_m} \|x\, p(x) - 1\|_S \geq \min_{q \in \mathcal{P}_{m+1}} \tfrac{\|q\|_S}{|q(0)|}.$$

Krylov Subspace Methods

Thm: Convergence of GMRES

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

$$\min_{q \in \mathcal{P}_m} \frac{\|q\|_{[c,d]}}{|q(0)|} \quad \leq \quad 2\left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^m .$$

Proof. Consider the polynomial

$$q(x) = T_m\big(\psi^{-1}(x)\big)$$

where $T_m(x) \in \mathcal{P}_m$ denotes the mth Chebyshev polynomial from Lecture 4, and

$$\psi^{-1}(x) = \frac{2}{d-c}\left(x - \frac{d+c}{2}\right)$$

is the unique linear function $\psi^{-1} : [c,d] \to [-1,1]$.

I will verify the claim by showing on the next slide that this particular choice of $q(x)$ satisfies

$$\|q\|_{[c,d]} = 1 \qquad \text{and} \qquad q(0) \geq \frac{1}{2}\left(\frac{\sqrt{\kappa}+1}{\sqrt{\kappa}-1}\right)^m .$$

Krylov Subspace Methods

Proof that $\|q\|_{[c,d]} = 1$.

$$\|q\|_{[c,d]} = \left\| T_m\big(\psi^{-1}(x)\big) \right\|_{[c,d]} = \| T_m(\hat{x}) \|_{[-1,1]} = 1.$$

We have shown the last step in Lecture 4.

Proof that $q(0) \geq \frac{1}{2}\left(\frac{\sqrt{\kappa}+1}{\sqrt{\kappa}-1}\right)^m$.

$$\begin{aligned} |q(0)| &= \left| T_m\!\left(\frac{d+c}{d-c}\right) \right| = \left| T_m\!\left(\frac{\kappa+1}{\kappa-1}\right) \right| \\ &= \frac{1}{2}\left| \phi_-^{-1}\!\left(\frac{\kappa+1}{\kappa-1}\right)^m + \phi_-^{-1}\!\left(\frac{\kappa+1}{\kappa-1}\right)^{-m} \right| \\ &= \frac{1}{2}\left| \left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^m + \left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^{-m} \right| \\ &\geq \frac{1}{2}\, \left(\frac{\sqrt{\kappa}+1}{\sqrt{\kappa}-1}\right)^m, \end{aligned}$$

where on the second line I used the definition of Chebyshev polynomials from Lecture 4, and on the third line I used the lemma from slide 14.

Krylov Subspace Methods

Remark

Using an equioscillation argument like in Lecture 4, one can show that, in fact,

$$T_m\big(\psi^{-1}(x)\big) \quad \in \quad \operatorname*{arg\,min}_{q \in \mathcal{P}_m} \frac{\|q\|_{[c,d]}}{|q(0)|}.$$

This is irrelevant in the above proof, but it indicates that the statement on slide 22 is sharp in the limit $m \to \infty$.

Krylov Subspace Methods

Corollary

Combining the results from slides 19 and 22, we conclude that

$$p = \operatorname*{arg\,min}_{p \in \mathcal{P}_m} \|A\,p(A)\,b - b\|_2 \quad \Longrightarrow \quad \|A\,p(A)\,b - b\|_2 \le 2\left(\tfrac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^m.$$

This shows that the GMRES solution $p(A)\,b$ converges to $A^{-1}b$ exponentially and with exactly the same rate as the polynomial approximation algorithm from slide 9.

Outlook

Recall that our motivation for considering the GMRES algorithm was that we wanted to be able to determine a polynomial $p(A)\,b \approx A^{-1}b$ efficiently.

The above result demonstrates that the GMRES polynomial satisfies $p(A)\,b \approx A^{-1}b$, and I will next show that the GMRES minimisation problem can indeed be solved efficiently.

Krylov Subspace Methods

Solving the GMRES optimisation problem

Making the ansatz $p(x) = \sum_{k=0}^{m} c_k \, x^k$, the GMRES minimisation problem

$$p = \underset{p \in \mathcal{P}_m}{\arg\min} \, \| A \, p(A) \, b - b \|_2$$

reduces to

$$c = \underset{c \in \mathbb{R}^{m+1}}{\arg\min} \left\| A \sum_{k=0}^{m} c_k \, A^k \, b - b \right\|_2 .$$

This problem can in turn be written as a standard least-squares problem,

$$c = \underset{c \in \mathbb{R}^{m+1}}{\arg\min} \, \| AVc - b \|_2, \qquad V = \left(b \;\middle|\; Ab \;\middle|\; \dots \;\middle|\; A^{m-1} b \;\middle|\; A^m b \right).$$

Recall from your undergraduate linear algebra course that least-squares problems can be solved using the QR factorisation. The following slide will recapitulate the fundamentals.

Krylov Subspace Methods

Thm: QR factorisation

Any matrix $A \in \mathbb{R}^{m \times n}$ can be written as $A = QR$ where

- $Q \in \mathbb{R}^{m \times n}$ is orthogonal ($Q^T Q = I$), and
- $R \in \mathbb{R}^{n \times n}$ is upper-triangular ($R[i, j] = 0$ if $i > j$).

Proof. Undergraduate material.

Thm: Least squares via QR

Assume $A \in \mathbb{R}^{m \times n}$ has QR factorisation $QR = A$, and let $b \in \mathbb{R}^m$. Then,

$$\arg\min_{x \in \mathbb{R}^n} \|Ax - b\|_2 = R^{-1} Q^T b.$$

Proof. Undergraduate material.

Thm: Runtime of QR factorisation

The QR factorisation $QR = A$ of a matrix $A \in \mathbb{R}^{m \times n}$ can be computed in $O(mn^2)$ floating-point operations.

Proof. Undergraduate material.

Krylov Subspace Methods

Least squares in Julia

Julia (and Matlab) provide the syntax `A\b` for solving the least-squares problem $\arg\min_{x \in \mathbb{R}^n} \|Ax - b\|_2$, assuming $A \in \mathbb{R}^{m \times n}$ is rectangular.

If x is underdetermined, i.e. if $\mathrm{rank}(A) < n$, then `A\b` returns the vector x which minimises both $\|Ax - b\|_2$ and $\|x\|_2$ simultaneously.

Remark

If $A \in \mathbb{R}^{n \times n}$ is square, then `A\b` solves the linear system $Ax = b$ instead.

This is (somewhat) consistent since if $\mathrm{rank}(A) = n$, then we indeed have

$$\arg\min_{x \in \mathbb{R}^n} \|Ax - b\|_2 = A^{-1} b.$$

Krylov Subspace Methods

Discussion

Given the above, we see that GMRES can be rewritten as follows.

GMRES, algorithmic form

1. Assemble $V = \begin{pmatrix} b & Ab & \dots & A^m b \end{pmatrix}$.
2. Solve $y = \arg\min \|AVy - b\|_2$ using the QR factorirsation
3. Set $\tilde{x} = Vy$.

Def: Krylov subspaces

A linear subspace $\mathcal{K}_m(A, b) = \operatorname{span}\{b, Ab, \dots, A^{m-1}b\} \subset \mathbb{R}^n$ where $A \in \mathbb{R}^{n \times n}$ and $b \in \mathbb{R}^n$ is called a *Krylov subspace*.

Etymology of "Krylov subspace methods"

The Krylov subspace methods from slide 16 can also be formulated in terms of Krylov subspaces as

$$\text{Compute } \tilde{x} \approx A^{-1}b \qquad \text{where} \qquad \tilde{x} = \operatorname*{arg\,min}_{\tilde{x} \in \mathcal{K}_m(A,b)} \|A\tilde{x} - b\|.$$

This explains why these methods are called Krylov subspace methods.

Krylov Subspace Methods

Implementation

See `gmres_unstable()` and `gmres_convergence()`.

Discussion

We observe that `gmres_unstable()` converges with the expected rate for $m \lesssim 10$ but diverges for larger m.

This shows that `gmres_unstable()` is an unstable numerical approximation to the mathematical function

$$(A, b, m) \quad \mapsto \quad p(A)\, b \qquad \text{where} \qquad p = \operatorname*{arg\,min}_{p \in \mathcal{P}_m} \|A\, p(A)\, b - b\|_2.$$

This numerical instability can be avoided by interleaving the processes of assembling the Krylov matrix V and computing the QR factorisation of AV in a clever way.

I will not go into details here but instead simply present the final result on the next slide.

Krylov Subspace Methods

Thm: Stable and fast GMRES

There exists an implementation of the GMRES function

$$(A, b, m) \quad \mapsto \quad p(A)\, b \qquad \text{where} \qquad p = \operatorname*{arg\,min}_{p \in \mathcal{P}_m} \|A\, p(A)\, b - b\|_2$$

which is numerically stable, requires $O(nm)$ memory and whose runtime is given by

$$m \text{ matrix-vector products and } O(nm^2) \text{ other operations.}$$

Proof. Omitted.

Numerical demonstration

See `gmres_convergence()` but set `unstable = false`.

Remark

The runtime estimate for GMRES can be interpreted as follows.

- m matrix-vector products: cost of evaluating $p(A)\, b$.
- $O(nm^2)$ other operations: cost of solving $\min_{p \in \mathcal{P}_m} \|A\, p(A)\, b - b\|_2$.

Krylov Subspace Methods

Discussion

For small degrees m, the $O(nm)$ memory estimate and the $O(nm^2)$ part of the runtime of GMRES are usually dominated by the memory footprint of A and the cost of evaluating m matrix-vector products, respectively, but they become dominant and often limiting for larger m.

Luckily, it turns out that we can artificially restrict the size of m using the following trick.

Def: GMRES with initial guess

Assume we have a rough guess $\tilde{x} \approx A^{-1}b$ for the solution to $Ax = b$.

Krylov subspace methods allow us to exploit this information by approximately solving

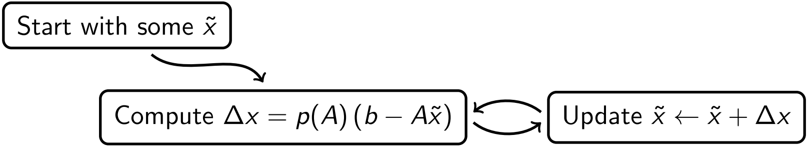
$$A\,\Delta x = b - A\tilde{x}$$

and returning an improved guess $\tilde{x} + \Delta x \approx A^{-1}b$.

Krylov Subspace Methods

Restarted GMRES

The "GMRES with initial guess" idea presented above allows us to gradually improve an initial guess $\tilde{x}$ by repeatedly running GMRES with a fixed degree m and using the output of the previous run as the input in the next run.



Start with some $\tilde{x}$

Compute $\Delta x = p(A)\,(b - A\tilde{x})$

Update $\tilde{x} \leftarrow \tilde{x} + \Delta x$

This algorithm is known as *restarted GMRES*.

The advantage of restarted GMRES is that it allows us to avoid the $O(m^2)$ scaling of standard GMRES.

The disadvantage is that it generally leads to slower convergence, and sometimes much slower convergence; see `restarted_gmres_good()` and `restarted_gmres_bad()`.

Krylov Subspace Methods

Termination criteria for GMRES

So far, I have assumed that the degree m of the GMRES polynomial $p \in \mathcal{P}_m$ is simply a parameter of the algorithm which must be specified by the user when we want to actually run GMRES.

This parameter m directly controls the runtime but provides no guarantees regarding the accuracy of the return value.

This is the exact opposite of what we usually want: we usually have an error tolerance to meet, and we want m to be large enough to meet this error tolerance but otherwise as small as possible.

Fortunately, the stable and fast GMRES algorithm from slide 31 can compute

$$m = \min\Big\{ m \in \mathbb{N} \;\Big|\; \min_{p \in \mathcal{P}_m} \|A\, p(A)\, b - b\|_2 \leq \tau \Big\}$$

at a cost which is negligible compared to directly running GMRES with the correct m.

Krylov Subspace Methods

Termination criteria for GMRES (continued)

Most GMRES implementations therefore allow the user to specify both a maximal degree m_{max} and a residual tolerance τ.

The return value is then a GMRES solution $p(A)\,b$ such that either the residual norm is less than τ, or $\text{degree}(p) = m_{\text{max}}$ if meeting the error tolerance is not possible with $\text{degree}(p) \le m_{\text{max}}$.

Note that imposing a largest admissible degree m_{max} means that your code must be able to handle the situation where the GMRES solution $p(A)\,b$ may be arbitrarily far from $A^{-1}b$.

Krylov Subspace Methods

Convergence theory for $S \neq [c, d] \subset (0, \infty)$

Looking back at the convergence theory developed on slides 19 onwards, we realise that the theory presented so far covers only the case where the eigenvalues are contained in some set

$$S = [c, d] \subset (0, \infty).$$

Our theory is hence fairly incomplete since eigenvalues can cluster in multiple intervals, on both the positive and negative real axis, and even anywhere in the complex plane assuming A is not symmetric.

Unfortunately, we cannot really develop a complete theory since

$$\min_{q \in \mathcal{P}_m} \frac{\|q\|_S}{|q(0)|} \qquad \text{for arbitrary } S \subset \mathbb{C}$$

is too complicated a problem to be solved using only pen and paper.

However, there is one more reasonably simple and useful result which I will present next.

Krylov Subspace Methods

Thm: Finite termination of GMRES

Assume $A \in \mathbb{R}^{n \times n}$ has n_{EV} distinct eigenvalues. Then,

$$\min_{p \in \mathcal{P}_m} \left\| A p(A)\, b - b \right\|_2 = 0 \qquad \text{for all } m \geq n_{\mathsf{EV}} - 1,$$

i.e. GMRES produces the exact solution if $m \geq n_{\mathsf{EV}} - 1$.

Proof. Let $\left(\lambda_k\right)_{k=1}^{n_{\mathsf{EV}}}$ be the n_{EV} distinct eigenvalues, and consider the polynomial

$$q(x) = \prod_{k=1}^{n_{\mathsf{EV}}} \left(x - \lambda_k\right) \in \mathcal{P}_{n_{\mathsf{EV}}}.$$

According to the GMRES error estimate from slide 19, we then have

$$\min_{p \in \mathcal{P}_m} \|A\, p(A)\, b - b\|_2 \quad \leq \quad \kappa(V)\ \|b\|_2 \min_{q \in \mathcal{P}_{m+1}} \frac{\|q\|_{\{\lambda_1, \dots, \lambda_{n_{\mathsf{EV}}}\}}}{|q(0)|} \quad = \quad 0.$$

Krylov Subspace Methods

Corollary

GMRES with $m = n - 1$ is an exact linear system solver comparable to LU factorisation.

Proof. Combine the above result with the fact that a matrix $A \in \mathbb{R}^{n \times n}$ has at most n distinct eigenvalues.

Discussion

While interesting, the above corollary is rarely immediately useful because for $m = n - 1$, the runtime of GMRES becomes

$$m = n \text{ matrix-vector products and } O(nm^2) = O(n^3) \text{ other operations,}$$

which is as expensive as a dense LU factorisation.

However, the proof of the corollary suggests the rule-of-thumb shown on the next slide, which is very helpful for understanding the convergence of GMRES on a qualitative level.

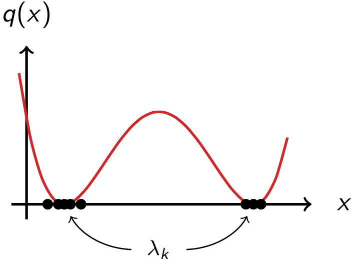
Krylov Subspace Methods

Rule of thumb

GMRES converges faster the more the eigenvalues cluster around a few points x_k far away from 0.

"Proof." Given the assumptions, we can obtain a good upper bound on

$$\min_{q \in \mathcal{P}_m} \frac{\|q\|_S}{|q(0)|} \qquad \text{by choosing} \qquad q(x) = \prod_k (x - x_k)^{m_k}.$$



Krylov Subspace Methods

Discussion

It turns out that Krylov subspace methods can be accelerated significantly if we assume A to be symmetric. In fact, the speedup is so significant that "GMRES with symmetric A" has been given its own name.

Def: MinRes

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

assuming A is symmetric.

Thm: Fast MinRes algorithm

The MinRes function can be evaluated using $O(n)$ memory,

m matrix-vector products and $O(nm)$ other operations.

Discussion. (A complete proof is beyond our scope.)

Recall that solving $\min_{p \in \mathcal{P}_m} \|A\,p(A)\,b - b\|_2$ requires us to compute the QR factorisation of AV where V is a matrix such that its columns span the Krylov subspace $\mathcal{K}_m$.

Krylov Subspace Methods

Discussion (continued)

In general, the QR factorisation of AV must be computed according to the following algorithm.

Algorithm GMRES version of $QR = AV$

```
for $k = 1, \dots, m$ do
    Orthogonalise $AV[:, k]$ against $Q[:, 1{:}k-1]$.
    Write the result into $Q[:, k]$.
end for
```

It turns out that if A is symmetric, then the following is sufficient to guarantee orthogonality of all columns of Q.

Algorithm MinRes version of $QR = AV$

```
for $k = 1, \dots, m$ do
    Orthogonalise $AV[:, k]$ against $Q[:, k-3{:}k-1]$.
    Write the result into $Q[:, k]$.
end for
```

Krylov Subspace Methods

Discussion (continued)

The above modification has two consequences.

- The cost of computing $QR = AV$ reduces from $O(nm^2)$ to $O(n)$.
- By cleverly interleaving the computations for $QR = AV$, $c = R^{-1}\,Q^T b$ and $x = Vc$, we can avoid storing all of Q and thereby reduce the memory requirements from $O(nm)$ to $O(n)$.

Unfortunately, the above claim that orthogonalising $AV[:,k]$ against only $Q[:,k{-}3{:}k{-}1]$ is sufficient to guarantee the orthogonality of Q holds true only in exact arithmetic.

In floating-point arithmetic, the two columns $Q[:,k]$ and $Q[:,\ell]$ computed by the MinRes version of $QR = AV$ will gradually lose their orthogonality for increasing $|k-\ell|$ due to rounding errors.

In practice, this means that MinRes usually converges somewhat more slowly than GMRES, see `gmres_vs_minres()`, but this slower convergence is offset by the lower cost per iteration.

Krylov Subspace Methods

Convergence theory for MinRes

Since MinRes and GMRES evaluate the same mathematical function

$$(A, b, m) \;\mapsto\; p(A)\, b \approx A^{-1} b \quad \text{where} \quad p = \operatorname*{arg\,min}_{p \in \mathcal{P}_m} \|A\, p(A)\, b - b\|_2,$$

the MinRes residuals satisfy exactly the same error bound as the GMRES residuals, namely

$$\min_{p \in \mathcal{P}_m} \|A\, p(A)\, b - b\|_2 \quad \le \quad \kappa(V)\; \|b\|_2 \; \min_{q \in \mathcal{P}_{m+1}} \tfrac{\|q\|_S}{|q(0)|}.$$

However, in the case of MinRes this bound can be somewhat simplified using the following observations.

- Eigenvectors of a symmetric matrix are orthogonal; hence $\kappa_2(V) = 1$.
- The eigenvalues of a symmetric matrix are real; hence we can assume $S \subset \mathbb{R}$ without loss of generality.

Thus, the error bound in the case of MinRes becomes

$$\min_{p \in \mathcal{P}_m} \|A\, p(A)\, b - b\|_2 \quad \le \quad \|b\|_2 \; \min_{q \in \mathcal{P}_{m+1}} \tfrac{\|q\|_S}{|q(0)|}.$$

Krylov Subspace Methods

Discussion

If A is both symmetric and positive definite, then we can obtain yet another Krylov subspace method by minimising the residual in the following norm.

Thm: Energy norm

Assume A is symmetric and positive definite. Then,

$$\|v\|_A = \sqrt{v^T A v} \quad \text{is a norm on } \mathbb{R}^n.$$

This norm is called the A-norm or energy norm.

Proof. Omitted but straightforward.

Def: Conjugate gradients

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

assuming A is symmetric and positive definite.

Krylov Subspace Methods

Discussion

Conjugate gradients is by far the most well-known and most widely used Krylov subspace algorithm. There are several reasons why this is the case.

Conjugate gradients is simple.

The algorithm described on the previous slide can be translated into just 8 lines of basic linear algebra operations which are very easy to implement in any programming environment.

Algorithm Conjugate gradients

```
$x_0 = 0, r_0 = b, p_0 = r_0$
for $k = 1, \dots, m$ do
    $\alpha_k = (r_{k-1}^T r_{k-1})/(p_{k-1}^T A p_{k-1})$
    $x_k = x_{k-1} + \alpha_k \, p_{k-1}$
    $r_k = r_{k-1} - \alpha_k \, A p_{k-1}$
    $\beta_k = (r_k^T r_k)/(r_{k-1}^T r_{k-1})$
    $p_k = r_k + \beta_k \, p_{k-1}$
end for
```

Krylov Subspace Methods

Conjugate gradients is fast.

It follows immediately from the algorithm on the previous slide that conjugate gradients can be evaluated using $O(n)$ memory and

m matrix-vector products and $O(nm)$ other operations.

This is the same runtime as MinRes, but the prefactor hidden by the big O notation is slightly better for conjugate gradients.

Symmetric and positive definite matrices are ubiquitous.

Out of the three Krylov subspace methods presented in this lecture, conjugate gradients imposes the most stringent assumptions on the coefficient matrix A (namely A must be symmetric and positive definite).

However, many matrices occurring in applied mathematics do satisfy these extra assumptions; hence the lack of generality of conjugate gradients on paper is often not a problem in practice.

Krylov Subspace Methods

Conjugate gradients controls the error in the A-norm.

We have seen that GMRES allows us to cheaply determine the smallest degree m such that

$$\min_{p \in \mathcal{P}_m} \|A\, p(A)\, b - b\|_2 \leq \tau.$$

Conjugate gradients, on the other hand, allows us to cheaply determine the smallest degree m such that

$$\min_{p \in \mathcal{P}_m} \|A\, p(A)\, b - b\|_{A^{-1}} \leq \tau.$$

The A^{-1}-norm of the residual is in fact the same as the A-norm of the error,

$$\begin{aligned} \|A\tilde{x} - b\|_{A^{-1}}^2 &= (A\tilde{x} - b)^T A^{-1} (A\tilde{x} - b) \\ &= (\tilde{x} - A^{-1}b)^T A^T A^{-1} A (\tilde{x} - A^{-1}b) \\ &= (\tilde{x} - A^{-1}b)^T A (\tilde{x} - A^{-1}b) \\ &= \|\tilde{x} - A^{-1}b\|_A^2, \end{aligned}$$

and it turns out that $\|\tilde{x} - A^{-1}b\|_A$ is often a more meaningful error measure than $\|A\tilde{x} - b\|_2$. In such applications, conjugate gradients thus enables us to more easily determine exactly the right degree m at which to terminate the algorithm.

Krylov Subspace Methods

Discussion

Recall that conjugate gradients evaluates a slightly different function compared to GMRES; hence the GMRES convergence theory does not immediately carry over to conjugate gradients.

However, it turns out that only minor adjustments are needed to adapt the GMRES convergence theory to conjugate gradients. This slide and the following provide the details.

Lemma

Assume $A \in \mathbb{R}^{n \times n}$ is symmetric, positive definite and has eigenvalues $(\lambda_k)_{k=1}^n$. Then,

$$\|f(A)\|_A \quad \le \quad \max_k |f(\lambda_k)|.$$

Proof (not examinable). See next slide.

Krylov Subspace Methods

Proof (not examinable).

Consider an arbitrary vector $b \in \mathbb{R}^n$ and let $V \Lambda V^T = A$ be the eigendecomposition of A (recall from your undergraduate studies that if A is symmetric, then the eigendecomposition exists and V is orthogonal). We then have

$$\begin{aligned}
\|f(A)\, b\|_A^2 &= b^T f(A)^T A\, f(A)\, b \;=\; b^T\, V\, \Lambda\, f(\Lambda)^2\, V^T\, b \\
&= \sum_{k=1}^{n} \Lambda\, f(\Lambda)^2\, (V^T\, b)[k]^2 \\
&\le \left(\max_k |f(\lambda_k)|^2 \right) \sum_{k=1}^{n} \lambda_k\, (V^T\, b)[k]^2 \\
&= \left(\max_k |f(\lambda_k)|^2 \right) b^T V \Lambda V^T b \\
&= \left(\max_k |f(\lambda_k)|^2 \right) \|b\|_A^2
\end{aligned}$$

where on the third line I used that λ_k and $(V^T b)[k]^2$ are both positive.

Krylov Subspace Methods

Thm: Error estimate for conjugate gradients

Assume $A \in \mathbb{R}^{n\times n}$ is symmetric and positive definite, and all its eigenvalues are contained in some set $S \subset \mathbb{R}$.

Then, the conjugate gradients error satisfies

$$\min_{p\in\mathcal{P}_m} \|p(A)\,b - A^{-1}b\|_A \quad \leq \quad \|A^{-1}b\|_A \min_{q\in\mathcal{P}_{m+1}} \tfrac{\|q\|_S}{|q(0)|}.$$

Proof. Exactly the same as on slide 20 and following, but use the bound for $\|f(A)\|_A$ rather than $\|f(A)\|_2$.for

Discussion

The above result shows that the convergence of conjugate gradients is determined by the ratio

$$\min_{q\in\mathcal{P}_{m+1}} \tfrac{\|q\|_S}{|q(0)|}$$

just like GMRES and MinRes. Combined with our earlier observations, we therefore conclude that the only relevant differences between MinRes and conjugate gradients are the ones listed on slides 45 and following.

Krylov Subspace Methods

Preconditioning

It is sometimes possible to accelerate the convergence of Krylov subspace methods by replacing the original linear system

$$Ax = b \qquad \text{with} \qquad PAx = Pb$$

for some suitably chosen matrix P.

This technique is known as *preconditioning* and the matrix P is called a *preconditioner*.

Krylov Subspace Methods

The ideal preconditioner

Perhaps to best way to develop some intuition for the preconditioning idea is to consider the so-called ideal preconditioner $P = A^{-1}$.

For this choice of P, we have $PA = I$; hence the only distinct eigenvalue of PA is 1 and Krylov subspace methods applied to $PAx = Pb$ would produce the exact result already for $m = 0$.

However, choosing $P = A^{-1}$ does not make sense: evaluating the $Pb = A^{-1}b$ part of $p(PA)\,Pb$ would require solving $Ax = b$, and if we had an algorithm for doing this efficiently then we would not be looking into Krylov subspace methods.

Features of a good preconditioner

The above indicates that a good preconditioner should balance the following two conflicting goals.

- P should be close to A^{-1} in some sense so Krylov subspace methods applied to $PAx = Pb$ converge quickly.
- P should allow for a cheap matrix-vector product so $p(PA)\,Pb$ can be evaluated efficiently.

Krylov Subspace Methods

Determining preconditioners

There are two general strategies for determining good preconditioners.

- *Analytic approach.*

  Sometimes we understand the matrix A well enough that we can rigorously prove that a particular preconditioner will lead to better performance.

- *Empirical approach.*

  There are several well-known heuristics for constructing preconditioners, and we can always just try and see if they help.

Both of these approaches are fairly problem-specific. I will not discuss them further.

Krylov Subspace Methods

Conjugate gradients for solving $-\Delta_n^{(d)}\, u_n = f$

We have now completed the discussion of Krylov subspace methods for general linear systems $Ax = b$.

In the remainder of this lecture, I will discuss how Krylov subspace methods compare against LU factorisation when applied to the finite-difference-discretised Poisson equation $-\Delta_n^{(d)}\, u_n = f$.

Krylov Subspace Methods

Conjugate gradients for solving $-\Delta_n^{(d)}\, u_n = f$ (continued)

To this end, I observe the followig.

- We can solve $-\Delta_n^{(d)} u_n = f$ using the conjugate gradient algorithm since $-\Delta_n^{(d)}$ is symmetric and positive definite.

- Conjugate gradients applied to $-\Delta_n^{(d)} u_n = f$ requires $O(Nm)$ operations since the general runtime estimate is

  m matrix vector products and $O(Nm)$ other operations.

  and matvecs with $-\Delta_n^{(d)}$ can be evaluated in $O(N)$ operations.

  $N = n^d$ denotes the number of unknowns as usual.

- The conjugate gradients solution satisfies

$$\min_{p \in \mathcal{P}_m} \|p(A)\, b - A^{-1}\, b\|_A \ \leq \ C r^m$$

  where

$$C = 2\, \|A^{-1}\, b\|_A\,, \qquad r = \tfrac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}, \qquad \kappa = \tfrac{4}{\pi^2}\,(n+1)^2$$

  since we know from Lecture 7 that all eigenvalues of $-\Delta_n^{(d)}$ are contained in $\left[\pi^2\, d, 4d\,(n+1)^2\right]$.

Krylov Subspace Methods

Conjugate gradients for solving $-\Delta_n^{(d)}\,u_n = f$ (continued)

It follows from the last point that if we want to guarantee

$$\min_{p\in\mathcal{P}_m} \|p(A)\,b - A^{-1}\,b\|_A \;\leq\; \tau,$$

then we must choose m such that $Cr^m \leq \tau$, or equivalently,

$$\begin{aligned} m &= \tfrac{\log(\tau/C)}{\log(r)} = O\big(\log(\tau)\,(1-r)^{-1}\big) = O\Big(\log(\tau)\,\big(1-\tfrac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\big)^{-1}\Big) \\ &= O\Big(\log(\tau)\,\big(\tfrac{\sqrt{\kappa}+1-\sqrt{\kappa}+1}{\sqrt{\kappa}+1}\big)^{-1}\Big) = O\Big(\log(\tau)\,)(\sqrt{\kappa}+1)\Big) \\ &= O\big(\log(\tau)\,n\big), \end{aligned}$$

where in the second step I used that $\log(r)^{-1} = O\big((1-r)^{-1}\big)$ since

$$\lim_{r\to 1} \tfrac{\log(r)^{-1}}{(1-r)^{-1}} = \lim_{r\to 1} \tfrac{1-r}{\log(r)} = \lim_{r\to 1} \tfrac{-1}{1/r} = -1$$

according to L'Hôpital's rule.

Krylov Subspace Methods

Conjugate gradients for solving $-\Delta_n^{(d)}\, u_n = f$ (continued)

Combining the above estimate for the degree m with the $O(Nm)$ runtime estimate for conjugate gradients, we obtain the following table.

	Runtime		**Memory**	
	LU	CG	LU	CG
$d = 1$	$O\left(N\right)$	$O\left(N^2\right)$	$O\left(N\right)$	$O\left(N\right)$
$d = 2$	$O\left(N^{3/2}\right)$	$O\left(N^{3/2}\right)$	$O\left(N \log(N)\right)$	$O\left(N\right)$
$d = 3$	$O\left(N^2\right)$	$O\left(N^{4/3}\right)$	$O\left(N^{4/3}\right)$	$O\left(N\right)$

$N = n^d$ denotes the number of unknowns.

This shows that Krylov subspace methods are mainly advantageous for partial differential equations on three-dimensional domains.

Numerical demonstration

See `cg_poisson_1d()` and `cg_poisson_2d()`.

Krylov Subspace Methods

Final words on PDEs and linear systems

We have now reached the end of a major topic of this module. Hence, let us summarise the main insights.

- Lecture 7: Solving partial differential equations amounts to solving very large but sparse linear systems.
- Lecture 8: LU factorisation can solve these linear systems with optimal $O(N)$ runtime if the PDE is posed on a one-dimensional domain, but its runtime increases to $O(N^{3/2})$ in two dimensions and $O(N^2)$ in three dimensions.
- Lecture 9: Krylov subspace methods can solve three-dimensional PDEs in $O(N^{4/3})$ runtime. Furthermore, they are able to exploit additional information about the problem in the form of initial guesses and preconditioners.

Krylov Subspace Methods

Summary

Krylov subspace methods:

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

Concrete Krylov subspace methods:

	GMRES	**MinRes**	**Conjugate Gradients**
Residual norm	2-norm	2-norm	A^{-1} norm
Assumptions on A	none	symmetric	symmetric and positive definite
Runtime	m matvecs $+\ O(nm^2)$	m matvecs $+\ O(nm)$	m matvecs $+\ O(nm)$
Memory	$O(nm)$	$O(n)$	$O(n)$

Krylov Subspace Methods

Summary (continued)

Algorithmic details:

- $\min_{p \in \mathcal{P}_m} \|A\, p(A)\, b - b\|$ can be solved using QR factorisation.
- Polynomial degree m can be chosen adaptively at no extra cost.
- GMRES may require restarts to be feasible.
- MinRes and conjugate gradients suffer from loss of orthogonality.

Convergence theory:

- $p(A)\, b$ is exact if $\text{degree}(p) \geq [\#\text{ distinct eigenvalues}] - 1$.
- Krylov subspace methods converge faster the more the eigenvalues cluster around a few points far away from 0.
- We can prove exponential convergence if the eigenvalues cluster in some interval $[c, d] \subset (0, \infty)$.
- We can sometimes improve the performance of Krylov subspace methods by replacing

  $$Ax = b \qquad \text{with} \qquad PAx = Pb$$

  where P should be close to A^{-1} and allow for a cheap matrix-vector product. This trick is known as *preconditioning*.

Krylov Subspace Methods

Summary (continued)

Conjugate gradients vs LU factorisation for $-\Delta_n^{(d)} u_n = f$:

	Runtime		**Memory**	
	LU	CG	LU	CG
$d = 1$	$O\big(N\big)$	$O(N^2)$	$O\big(N\big)$	$O(N)$
$d = 2$	$O\big(N^{3/2}\big)$	$O(N^{3/2})$	$O\big(N \log(N)\big)$	$O(N)$
$d = 3$	$O\big(N^2\big)$	$O(N^{4/3})$	$O\big(N^{4/3}\big)$	$O(N)$

$N = n^d$ denotes the number of unknowns.