MA5233 Computational Mathematics

Lecture 11: Krylov Subspace Methods: Theory

Simon Etter

2019/2020

Krylov Subspace Methods: Theory

Recap: Krylov subspace methods

Approximate solution to $Ax = b$ by

$$\tilde{x} := p_{n-1}(A)\, b \qquad \text{where} \qquad p_{n-1} := \operatorname*{arg\,min}_{p_{n-1} \in \mathcal{P}_{n-1}} \left\| \left(A p_{n-1}(A) - I \right) b \right\|.$$

Cost:

- n matrix-vector products,
- $\mathcal{O}(Nn^2)$ (GMRES) or $\mathcal{O}(Nn)$ (MinRes,CG) other FLOP.

Observations:

- Increasing n increases FLOP count but decreases error.
- $\tilde{x} = x$ for $n = N$ since then Krylov subspace has dimension N.
- Hope: $\tilde{x} \approx x$ already for $n \ll N$.
- To see whether hope is justified, we need a result which bounds error as a function of n.

Krylov Subspace Methods: Theory

Error measure

Natural error measure for GMRES and MinRes is $\|A\tilde{x} - b\|_2$ since this is the quantity which is minimised.

Bound on error $\|\tilde{x} - x\|_2$ results from

$$\|\tilde{x} - x\|_2 \leq \|A^{-1}\|_2 \, \|A\tilde{x} - b\|_2.$$

Similarly, the natural error measure for conjugate gradients is $\|\tilde{x} - x\|_A$.

We focus on GMRES case in the following, but all statements carry over to conjugate gradient.

Krylov Subspace Methods: Theory

Bound on GMRES residual

Assume A has eigendecomposition $A = V\Lambda V^{-1}$. Then,

$$\begin{aligned}
\|A\tilde{x} - b\|_2 &= \min_{p_{n-1}\in\mathcal{P}_{n-1}} \left\| \left(Ap_{n-1}(A) - I\right) b \right\|_2 \\
&\leq \min_{p_{n-1}\in\mathcal{P}_{n-1}} \|V\|_2 \left\|\Lambda p_{n-1}(\Lambda) - I\right\|_2 \|V^{-1}\|_2 \|b\|_2 \\
&\leq \kappa(V) \|b\|_2 \min_{p_{n-1}\in\mathcal{P}_{n-1}} \max_{\lambda_k} |\lambda_k\, p_{n-1}(\lambda_k) - 1|.
\end{aligned}$$

$\lambda_k = \Lambda_{kk}$ on last line are the eigenvalues of A.

Observations:

- Only the last factor depends on n.
- To get an asymptotic error estimate, we must study the behaviour of

$$\min_{p_{n-1}\in\mathcal{P}_{n-1}} \max_{\lambda_k} |\lambda_k\, p_{n-1}(\lambda_k) - 1|$$

  as a function of n.

Krylov Subspace Methods: Theory

The GMRES polynomial minimisation problem, observation 1

$$p_{n-1} \in \mathcal{P}_{n-1} \quad \Longrightarrow \quad q_n(x) := x\, p_{n-1}(x) - 1 \in \mathcal{P}_n,\ q(0) = -1$$

$$q_n \in \mathcal{P}_n,\ q_n(0) = -1 \quad \Longrightarrow \quad q_n(x) := x\, p_{n-1}(x) - 1 \text{ for some } p_{n-1} \in \mathcal{P}_{n-1}.$$

Proof. First implication is obvious.
Second implication: $p_{n-1}(x) = \frac{q_n(x)+1}{x}$ is a polynomial since $q_n(0) = -1$.

Corollary
The GMRES minimisation problem may equivalently be formulated as

$$\min_{q_n \in \mathcal{P}_n} \max_{\lambda_k} \frac{|q_n(\lambda_k)|}{|q_n(0)|}.$$

Hence, we want $q_n \in \mathcal{P}_n$ such that $|q_n(\lambda_k)|$ is small relative to $|q_n(0)|$.

Krylov Subspace Methods: Theory

The GMRES polynomial minimisation problem, observation 2

Eigenvalues λ_k typically cluster in certain sets $\mathcal{E} \subset \mathbb{C}$.
The GMRES minimisation problem

$$\min_{q_n \in \mathcal{P}_n} \max_{\lambda_k} \frac{|q_n(\lambda_k)|}{|q(0)|}$$

may then be replaced by

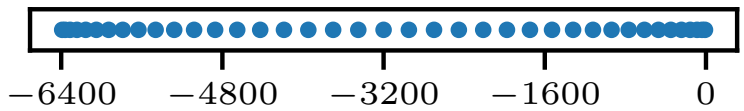
$$\min_{p_{n-1} \in \mathcal{P}_{n-1}} \max_{x \in \mathcal{E}} \frac{|q_n(x)|}{|q(0)|}$$

without losing much in sharpness.

Example

Recall eigenvalues $\lambda_\ell := (n+1)^2\left(2\cos\left(\pi \frac{\ell}{n+1}\right) - 2\right)$ of discrete Laplacian.
These eigenvalues cluster in the interval $\mathcal{E} = [-4\,(n+1)^2, 0]$.

$$n = 39 \quad \longrightarrow \quad \lambda_\ell =$$



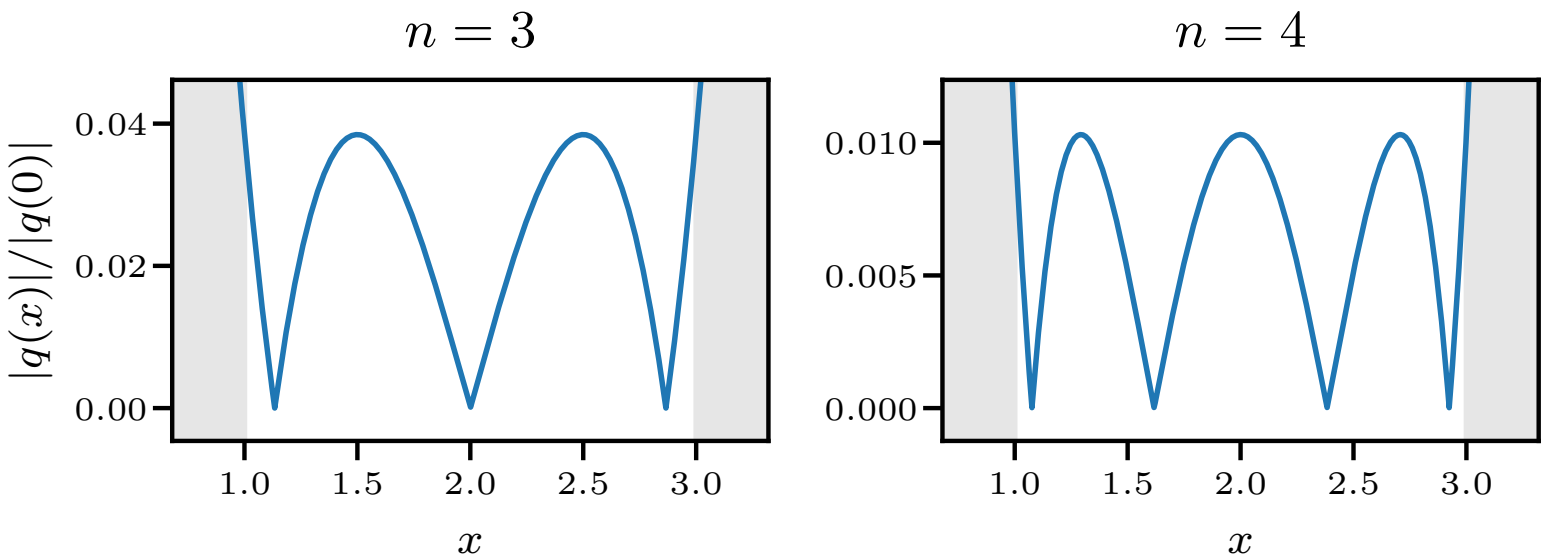
Krylov Subspace Methods: Theory

The GMRES polynomial minimisation problem, final formulation

$$\min_{q_n \in \mathcal{P}_n} \max_{x \in \mathcal{E}} \frac{|q_n(x)|}{|q_n(0)|}.$$

Idea: choose $q_n(x) = \prod_{k=1}^{n}(x - x_k)$ with x_k distributed over $\mathcal{E}$.

Example: $\mathcal{E} = [1, 3]$.



Krylov Subspace Methods: Theory

Properties of $\mathcal{E}$ which make GMRES converge fast

- $\mathcal{E}$ is small: few x_k are enough to ensure $q_n(x)$ is small on $\mathcal{E}$.
- $\mathcal{E}$ is far away from 0: $|q(0)| \geq \left(\min_{x\in\mathcal{E}} |x|\right)^n$.

The above two points are equivalent after scaling.

Example:

- $\mathcal{E} = [\frac{1}{n}, 1]$ is small but close to 0
- $\mathcal{E} = [1, n]$ is large but far away from 0.

Convergence is the same in both cases (see next slide).

Properties of $\mathcal{E}$ which make GMRES converge slowly

- $\mathcal{E}$ surrounds 0, e.g. $\mathcal{E} = [-2, -1] \cup [1, 2]$: GMRES will converge, but convergence will be slower than for $\mathcal{E} = [1, 2]$.

Krylov Subspace Methods: Theory

Proof that GMRES convergence is independent of scaling.
Assume there exists $\lambda \in \mathbb{R} \setminus \{0\}$ such that

$$\min_{q_1 \in \mathcal{P}_n} \max_{x \in \mathcal{E}} \frac{|q_1(x)|}{|q_1(0)|} \leq \min_{q_\lambda \in \mathcal{P}_n} \max_{x \in \lambda\mathcal{E}} \frac{|q_\lambda(x)|}{|q_\lambda(0)|}, \tag{1}$$

and let $q_1(x) = \prod_{k=1}^{n}(x - x_k)$ be the minimiser from the left-hand side.

Then, we obtain for $\tilde{q}_\lambda(x) = \prod_{k=1}^{n}(x - \lambda\, x_k) \in \mathcal{P}_n$ that

$$\max_{x \in \lambda\mathcal{E}} \frac{|\tilde{q}_\lambda(x)|}{|\tilde{q}_\lambda(0)|} = \max_{x \in \mathcal{E}} \frac{\prod_{k=1}^{n} |\lambda x - \lambda\, x_k|}{\prod_{k=1}^{n} |\lambda x_k|} = \max_{x \in \mathcal{E}} \frac{|q_1(x)|}{|q_1(0)|}. \tag{2}$$

Equation (2) contradicts the assumption (1).

Krylov Subspace Methods: Theory

Quantitative convergence estimate

Let $\mathcal{E} = [a, b]$ with $0 < a < b$, and set $\kappa := \frac{b}{a}$. Then,

$$\min_{q_n \in \mathcal{P}_n} \max_{x \in \mathcal{E}} \frac{|q_n(x)|}{|q_n(0)|} \leq 2 \left(\frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1} \right)^n .$$

We will prove this result in the lecture on approximation theory.

Summary of GMRES convergence theory

$$\|A\tilde{x} - b\|_2 \leq \kappa(V) \|b\|_2 \min_{q_n \in \mathcal{P}_n} \max_{\lambda_k} \frac{|q_n(\lambda_k)|}{|q(0)|}.$$

Previous slides discussed choice of q_n.
Following slides will look at b and λ_k.
Surprisingly, it turns out we can modify b and λ_k to improve convergence of GMRES.

Krylov Subspace Methods: Theory

GMRES with initial guess (modifying b)

Assume we already have a good initial guess x_0 for solution to $Ax = b$. Instead of applying GMRES to $Ax = b$, we can use GMRES to solve

$$A(x_0 + \Delta x) = b$$

for Δx, i.e. we apply GMRES to

$$A\,\Delta x = b - Ax_0.$$

In this case, we obtain using the GMRES error estimate that

$$\begin{aligned} \|A(x_0 + \Delta x) - b\|_2 &= \|A\,\Delta x - (b - Ax_0)\|_2 \\ &\leq \kappa(V)\,\|b - Ax_0\|_2 \min_{q_n \in \mathcal{P}_n} \max_{\lambda_k} \frac{|q_n(\lambda_k)|}{|q(0)|}. \end{aligned}$$

If x_0 is close to x, then $\|b - Ax_0\|_2$ is small; hence $\|A(x_0 + \Delta x) - b\|_2$ is small as well.

Krylov Subspace Methods: Theory

Preconditioning (modifying λ_k)

Instead of $Ax = b$, we can solve

$$P^{-1}Ax = P^{-1}b$$

for some invertible matrix P.

Idea: use P to move eigenvalues such that GMRES converges faster.

Example

Assume eigenvalues of A cluster in $\mathcal{E} = [a, b]$.
P is a good preconditioner if

- eigenvalues of $P^{-1}A$ cluster in $\mathcal{E}' = [a', b']$ with $\frac{b'}{a'} < \frac{b}{a}$, and
- $P^{-1}v$ is cheap to evaluate.

Remark

Preconditioner for MinRes and conjugate gradients must be symmetric positive definite.

Krylov Subspace Methods: Theory

Advantages of Krylov subspace methods compared LU

- Require only basic linear algebra operations like Av, $w^T v$, $v + \alpha w$.
  Assume you can represent your matrices and/or vectors in a compressed format. Krylov methods allow you to solve linear systems as long as your format supports the above operations.
- Corollary of previous point: can exploit fast matrix-vector product.
  Primary example: sparse matrices. LU factorisation typically requires $\mathcal{O}(N^\alpha)$ FLOP with $\alpha > 1$. Krylov methods may require only $\mathcal{O}(N)$ FLOP.
- Can exploit initial guesses.
  Assume you need to solve $Ax_1 = b$ and $(A + \Delta A)x_2 = b$ for some small ΔA. LU factorisation must be recomputed from scratch for $A + \Delta A$. Krylov methods can use x_1 as initial guess for x_2.
- Can be faster than LU factorisation if matrix is well-conditioned.

Krylov Subspace Methods: Theory

Disadvantages of Krylov subspace methods compared LU

- Errors (rounding errors, truncation tolerance) are hard to control.
- Almost always require preconditioning to be effective.

Summary

- Krylov methods are much harder to use then LU factorisation.
- Krylov subspace methods allow us to solve problems which we cannot solve using LU.

Advice:

- Use LU factorisation as long as you can.
- Only use Krylov methods if you have to.

Krylov Subspace Methods: Theory

References and further reading

Recommended since closest to presentation above:

- L. N. Trefethen and D. Bau. *Numerical Linear Algebra.* Society for Industrial and Applied Mathematics (1997),

Other references:

- G. H. Golub and C. F. Van Loan. *Matrix Computations.* Johns Hopkins University Press (1996),
- J. W. Demmel. *Applied Numerical Linear Algebra.* Society for Industrial and Applied Mathematics (1997), doi:10.1137/1.9781611971446
- N. J. Higham. *Accuracy and Stability of Numerical Algorithms.* Society for Industrial and Applied Mathematics (2002), doi:10.1137/1.9780898718027