MA3227 Numerical Analysis II

Lecture 8: Symmetric Krylov Methods

Simon Etter

NUS
National University
of Singapore

2019/2020

Symmetric Krylov Methods

Introduction

Recap GMRES algorithm:

1. Run Arnoldi iteration to obtain Q_k, H_k.
2. Solve least squares problem $y_k = \arg\min \left\| H_k y_k - \|b\|_2 \, e_1 \right\|_2$
3. Set $x_k = Q_k \, y_k$.

We have seen that the Arnoldi iteration requires $\mathcal{O}(Nk^2)$ FLOP in addition to the matrix-vector products. This is expensive if we have to run GMRES for a large number of iterations k.

Additionally, GMRES requires us to store all of Q_k. This requires $\mathcal{O}(Nk)$ memory, which may be a problem for very large N and k.

It turns out that both of these issues disappear if A is symmetric.
The next slide provides the mathematical foundation for this.

Symmetric Krylov Methods

Lanczos theorem

Let $Q_k \in \mathbb{R}^{N\times k}$ and $H_k \in \mathbb{R}^{(k+1)\times k}$ be the sequence of matrices produced by the Arnoldi iteration. Then,

$$A \text{ is symmetric} \quad \Longrightarrow \quad H_k \text{ is tridiagonal} .$$

Warning: calling this result "Lanczos theorem" is my own invention.

You will not find a "Lanczos theorem" in the literature.

Remark

Tridiagonal in this context means

$$H_k[i,j] = 0 \text{ if } |i-j| > 1 \quad \Longleftrightarrow \quad H_k = \begin{pmatrix} x & x & & & & \\ x & x & x & & & \\ & x & x & x & & \\ & & x & x & x & \\ & & & x & x & x \\ & & & & x & x \\ & & & & & x \end{pmatrix}.$$

Symmetric Krylov Methods

Proof of Lanczos theorem.
Recall the Arnoldi relations $AQ_k = Q_{k+1}H_k$.
Multiplying with Q_k^T from the left yields

$$Q_k^T A Q_k = \begin{pmatrix} Q_k^T Q_k & Q_k^T Q_{k+1}[:, k+1] \end{pmatrix} H_k = \begin{pmatrix} I & 0 \end{pmatrix} H_k = \tilde{H}_k.$$

($\tilde{H}_k \in \mathbb{R}^{k \times k}$ is obtained from H_k by removing last row.)
Since $\tilde{H}_k$ is Hessenberg and symmetric, H_k must be tridiagonal.

Corollary of Lanczos theorem

Iterations $m = 1, \ldots, \ell - 2$ in the inner loop of the Arnoldi iteration effectively do the following:

```
for m = 1, ..., ℓ − 2 do
    H[m, ℓ] = 0
    q̃_{ℓ+1} = q̃_{ℓ+1} − 0
end for
```

Removing these "empty" loops from Arnoldi yields the Lanczos iteration show on the next slide.

Symmetric Krylov Methods

Algorithm 1 Lanczos iteration

$Q_1 = b/\|b\|_2$
for $\ell = 1, \dots, k$ **do**
  $\tilde{q}_{\ell+1} = AQ_\ell[:, \ell]$
  $H[\ell, \ell] = Q[:, \ell]^T \tilde{q}_{\ell+1}$
  if $\ell = 1$ **then**
    $\tilde{q}_{\ell+1} = \tilde{q}_{\ell+1} - Q[:, \ell]\, H[\ell, \ell]$
  else
    $\tilde{q}_{\ell+1} = \tilde{q}_{\ell+1} - Q[:, \ell]\, H[\ell, \ell] - Q[:, \ell-1]\, H[\ell-1, \ell]$
  end if
  $H[\ell+1, \ell] = H[\ell, \ell+1] = \|\tilde{q}_{\ell+1}\|_2$
  $Q_{\ell+1} = \begin{pmatrix} Q_\ell & \big| & \frac{\tilde{q}_{\ell+1}}{H[\ell+1,\ell]} \end{pmatrix}$
end for

In addition to removing the empty loop iterations, the above algorithm also exploits that

$$Q[:, \ell]^T \tilde{q}_{\ell+1} = H[\ell, \ell+1] = H[\ell+1, \ell] = \|\tilde{q}_\ell\|_2$$

which allows us to save one inner product.

Symmetric Krylov Methods

Terminology

GMRES applied to symmetric matrices and using Lanczos instead of Arnoldi is known as the MinRes (Minimal Residual) method.

It may seem silly that essentially the same method goes by two different names (same also holds for Arnoldi vs. Lanczos).

There are probably two reasons for that:

- MinRes was discovered first and later generalised to GMRES.
- The differences in runtime and memory requirements between the two algorithms is quite substantial.

Symmetric Krylov Methods

Implementation

See `lanczos()` and `minres()`.

This implementation requires only $\mathcal{O}(Nk)$ FLOP in addition to the matrix-vector products, but it still stores the full matrix Q which requires $\mathcal{O}(Nk)$ memory.

It is possible to interleave the Lanczos algorithm with solving the least squares problem such that only a small number of vectors need to be stored in memory and any given moment.

This is important for very large-scale computations where memory constraints are a concern.

Deriving the reduced-memory MinRes implementation requires tools which we have not developed in this class (Householder reflectors), so I will not explain further.

Loss of orthogonality

Lanczos' theorem guarantees that $Q[:,k]$ and $Q[:,\ell]$ are orthogonal for all k and ℓ even though we only explicitly orthogonalise for $|k-\ell| \leq 1$ in the Lanczos iteration.

Unfortunately, this result only holds in exact arithmetic. In the presence of rounding errors, the orthogonality gradually deteriorates for increasing $|k-\ell|$, see `loss_of_orthogonality()`.

One consequence of this that $x_N = A^{-1}b$ may no longer hold for large N. For illustration, change `N = 2:9` to `N = 2:20` in `test()`.

Fortunately, this finite termination property is not important in practice since we are only interested in Krylov methods if they converge for $k \ll N$.

Loss of orthogonality effectively makes MinRes similar to GMRES with restarts: x_k may be interpreted as the exact MinRes solution after Δk steps starting from $x_{k-\Delta k}$. MinRes iterations $k' = 1, \ldots, k - \Delta k$ serve to provide a good initial guess $x_{k-\Delta k}$.

Symmetric Krylov Methods

Recap Krylov subspace algorithms

Recall: GMRES and MinRes both solve

$$x_k = p_{k-1}(A)\, b \qquad \text{where} \qquad p_{k-1} = \underset{p_{k-1} \in \mathcal{P}_{k-1}}{\arg\min} \left\| \left(A p_{k-1}(A) - I \right) b \right\|_2 .$$

The only difference between the two algorithms is that MinRes exploits the special properties resulting from the symmetry of A.

Genuinely new Krylov subspace algorithms can be defined by changing the norm $\|\cdot\|$ in which we minimise the residual $\|Ax_k - b\|$.

Our next aim is to do precisely this, but for that purpose we must introduce the norm first.

Symmetric Krylov Methods

Def: Energy norm

Let $A \in \mathbb{R}^{N \times N}$ be symmetric positive definite (spd, see below).
The energy norm associated with A is given by

$$\|v\|_A = \sqrt{v^T A v}.$$

Def: Positive definiteness

A symmetric matrix $A \in \mathbb{R}^{N \times N}$ is called positive definite if $v^T A v > 0$ for all $v \in \mathbb{R}^N \setminus \{0\}$.

Equivalently, A is positive definite if all its eigenvalues are positive.

Some useful results:

- Positive definiteness of A is a necessary and sufficient condition for $\| \cdot \|_A$ to be a norm.
- If A is positive definite, then so is A^{-1}.
- $-\Delta_n^{(d)}$ is positive definite.

Showing these claims is easy, but you may accept them as facts.

Symmetric Krylov Methods

Def: Conjugate gradients

Let A be a symmetric positive definite matrix.

The kth conjugate gradient approximation to $x = A^{-1}b$ is then given by

$$x_k = p_{k-1}(A)\, b \qquad \text{where} \qquad p_{k-1} = \operatorname*{arg\,min}_{p_{k-1} \in \mathcal{P}_{k-1}} \big\| \big(A p_{k-1}(A) - I\big)\, b \big\|_{A^{-1}}.$$

Discussion

Conjugate gradients is the most well-known Krylov subspace algorithm.

There are probably four reasons for this:

- It was discovered before MinRes and GMRES.
- It is slightly more efficient than MinRes in terms of runtime and memory (but same as MinRes in $\mathcal{O}$-sense).
- It is very easy to implement, see next slide.
- Depending on the application, $\|Ax_k - b\|_{A^{-1}}$ may have physical meaning, see slide after next slide.

Symmetric Krylov Methods

Algorithm 2 Conjugate gradients

1: $x_0 = 0, r_0 = b, p_0 = r_0$
2: **for** $\ell = 1, \dots, k$ **do**
3: $\quad \alpha_\ell = (r_{\ell-1}^T r_{\ell-1})/(p_{\ell-1}^T A p_{\ell-1})$
4: $\quad x_\ell = x_{\ell-1} + \alpha_\ell \, p_{\ell-1}$
5: $\quad r_\ell = r_{\ell-1} - \alpha_\ell \, A p_{\ell-1}$
6: $\quad \beta_\ell = (r_\ell^T r_\ell)/(r_{\ell-1}^T r_{\ell-1})$
7: $\quad p_\ell = r_\ell + \beta_\ell \, p_{\ell-1}$
8: **end for**

- $r_k = b - Ax_k$ is the residual of the kth iterate.
- p_k is often called the search direction since it is the direction in which we update x_k, see line 4.
- α_k is often called the step size for similar reasons.

Symmetric Krylov Methods

Interpretation of $\|Ax_k - b\|_{A^{-1}}$

Straightforward computations reveal

$$\begin{aligned}\|Ax_k - b\|_{A^{-1}} &= (Ax_k - b)^T A^{-1} (Ax_k - b) \\ &= (x_k - x)^T AA^{-1}A (x_k - x) \\ &= (x_k - x)^T A (x_k - x) \\ &= \|x_k - x\|_A .\end{aligned}$$

Conjugate gradients therefore not only minimises the residual, but it also minimises the error $x_k - x$.

The catch is that it minimises the error in the A-norm, but in some applications this is a meaningful quantity.

More precisely, $\|x - x_k\|_A$ can sometimes be interpreted as the "energy" of the error; hence the name "energy norm".

Symmetric Krylov Methods

Summary

- GMRES simplifies to MinRes when applied to a symmetric matrix A.
- If A is symmetric and positive definite, then we may also consider the conjugate gradients method. This method minimises $\|Ax_k - b\|_{A^{-1}} = \|x_k - x\|_A$ rather than $\|Ax_k - b\|_2$.
- Both MinRes and conjugate gradients require k matrix-vector products, $\mathcal{O}(Nk)$ other operations and $\mathcal{O}(N)$ memory.