

Utrecht, 8 november 2017

Optimal Iterative Methods



<http://www.staff.science.uu.nl/~sleij101/>

Program Lecture 6

- Krylov basis & Hessenberg matrices
- Arnoldi's decomposition
- Linear systems and Arnoldi's decomposition
- GMRES and FOM
- Convergence
- Eigenvalues and Arnoldi's decomposition
- Arnoldi's method
- Convergence
- Stability issues in Arnoldi's decomposition
- Summary

$$\mathbf{Ax} = \mathbf{b} \quad \{\lambda \in \mathbb{C} \mid \mathbf{Av} = \lambda \mathbf{v}\} \subset \mathcal{E} \subset \mathbb{C}$$

Construct iteratively $\mathbf{r}_k = \mathbf{b} - \mathbf{Ax}_k$ with $\|\mathbf{r}_k\|_2$ small

\mathbf{x}_k gets a free ride: $\mathbf{r}_k = \mathbf{r}_{k-1} - \alpha_k \mathbf{A} \mathbf{u}_{k-1}$, $\mathbf{x}_k = \mathbf{x}_{k-1} + \alpha_k \mathbf{u}_{k-1}$

Richardson: $\alpha \in \mathbb{C}$, $\mathbf{r}_k = \mathbf{r}_{k-1} - \alpha \mathbf{A} \mathbf{r}_{k-1} = (\mathbf{I} - \alpha \mathbf{A})^k \mathbf{r}_0$

Krylov subspace: $\mathcal{K}_k(\mathbf{A}, \mathbf{r}_0) = \text{span}(\mathbf{r}_0, \mathbf{A} \mathbf{r}_0, \dots, \mathbf{A}^{k-1} \mathbf{r}_0)$
 $= \{p(\mathbf{A}) \mathbf{r}_0 \mid p \text{ pol. degree } < k\}$

Richardson: $\alpha_k \in \mathbb{C}$, $\mathbf{r}_k = \mathbf{r}_{k-1} - \alpha_k \mathbf{A} \mathbf{r}_{k-1} = p_k(\mathbf{A}) \mathbf{r}_0$

with $p_k(\lambda) = (1 - \alpha_1 \lambda) \dots (1 - \alpha_k \lambda)$

Selection α_k : **static**, **dynamic**

• $\alpha_k = \alpha_{k \bmod \ell}$: with $p(\lambda) = (1 - \alpha_1 \lambda) \dots (1 - \alpha_\ell \lambda)$, $p_{j\ell} = p^j$
 and $\max_{\zeta \in \mathcal{E}} |p(\zeta)|$ as small as possible.

• **Chebyshev iteration:** $p_k = \tilde{T}_k$, $\mathbf{r}_{k+1} = \tilde{\alpha}_k \mathbf{r}_k - \tilde{\beta}_k \mathbf{A} \mathbf{r}_k - \tilde{\gamma}_{k-1} \mathbf{r}_{k-1}$

• **Local Minimal Residual:** $\alpha_k = \text{argmin}_\alpha \|\mathbf{r}_{k-1} - \alpha \mathbf{A} \mathbf{r}_{k-1}\|_2$

• **Generalized Conjugate Residuals:** $\mathbf{R}_k \equiv [\mathbf{r}_0, \dots, \mathbf{r}_{k-1}]$
 $\mathbf{r}_k = \mathbf{r}_{k-1} - \mathbf{A} \mathbf{R}_k \vec{\alpha}_{k-1}$ with $\vec{\alpha}_{k-1} \equiv \text{minarg}_{\vec{\alpha} \in \mathbb{C}^k} \|\mathbf{r}_{k-1} - \mathbf{A} \mathbf{R}_k \vec{\alpha}\|_2$

Orthonormal basis $\mathcal{K}_k(\mathbf{A}, \mathbf{r}_0)$

Recall $\mathbf{V}_k \equiv [\mathbf{v}_1, \dots, \mathbf{v}_k]$.

Suppose $\mathbf{v}_1, \dots, \mathbf{v}_k$ is an orthonormal Krylov basis $\mathcal{K}_k(\mathbf{A}, \mathbf{r}_0)$.

Compute \mathbf{v}_{k+1} by **orthogonalising $\mathbf{A} \mathbf{v}_k$ against \mathbf{V}_k** :

- **Expand:** $\mathbf{w} = \mathbf{A} \mathbf{v}_k$,
- **Orthogonalize:** $\tilde{\mathbf{v}} = \mathbf{w} - \mathbf{V}_k \vec{h}'_k$ with $\vec{h}'_k = \mathbf{V}_k^* \mathbf{w}$,
- **Normalize:** $\mathbf{v}_{k+1} = \tilde{\mathbf{v}} / \nu_k$ with $\nu_k = \|\tilde{\mathbf{v}}\|_2$.

Note. With $\vec{h}_k \equiv (\vec{h}'_k, \nu_k)^\top$, we have

$$\mathbf{A} \mathbf{v}_k = \mathbf{w} = \mathbf{V}_k \vec{h}'_k + \nu_k \mathbf{v}_{k+1} = [\mathbf{V}_k, \mathbf{v}_{k+1}] \begin{bmatrix} \vec{h}'_k \\ \nu_k \end{bmatrix} = \mathbf{V}_{k+1} \vec{h}_k$$

$$\text{Assemble } \mathbf{A}[\mathbf{V}_{k-1}, \mathbf{v}_k] = [\mathbf{V}_k, \mathbf{v}_{k+1}] \begin{bmatrix} \underline{H}_{k-1} & \vec{h}'_k \\ 0 \dots 0 & \nu_k \end{bmatrix}$$

Theorem.

Orthogonalising $\mathbf{A}\mathbf{v}_j$ against \mathbf{V}_j for $j = 1, \dots, k$ leads to

$$\mathbf{A}\mathbf{V}_k = \mathbf{V}_{k+1} \underline{H}_k,$$

with \mathbf{V}_k orthonormal, spanning $\mathcal{K}_k(\mathbf{A}, \mathbf{v}_1)$,
 \underline{H}_k is $(k+1) \times k$ upper Hessenberg.

Note. The matrix \underline{H}_k comes for **free**
in the orthogonalisation process.

Application for solving $\mathbf{A}\mathbf{x} = \mathbf{r}_0$. Try $\mathbf{x}_k = \mathbf{V}_k \vec{y}_k$.

Find \vec{y}_k such that $\mathbf{A}\mathbf{V}_k \vec{y}_k \approx \mathbf{r}_0 = \rho_0 \mathbf{V}_k e_1$ with $\rho_0 \equiv \|\mathbf{r}_0\|_2$

\rightsquigarrow Find \vec{y}_k such that $\mathbf{V}_{k+1} \underline{H}_k \vec{y}_k \approx \rho_0 \mathbf{V}_{k+1} e_1$ if $\rho_0 \mathbf{v}_1 = \mathbf{r}_0$

\rightsquigarrow Find \vec{y}_k such that $\underline{H}_k \vec{y}_k \approx \rho_0 e_1$ if $\rho_0 \mathbf{v}_1 = \mathbf{r}_0$

Details later.

Orthogonalisation

Terminology.

If \mathbf{V} is an $n \times k$ orthonormal matrix and \mathbf{w} is an n vector,
then, with **orthonormalise \mathbf{w} against \mathbf{V}** , we mean:

construct an n -vector \mathbf{v} and a $(k+1)$ -vector \vec{h} such that

$$\mathbf{v} \perp \mathbf{V}, \quad \|\mathbf{v}\|_2 = 1, \quad \mathbf{w} = [\mathbf{V}, \mathbf{v}] \vec{h}$$

Notation. $[\mathbf{v}, \vec{h}] = \text{Orth}(\mathbf{V}, \mathbf{w})$

Use a stable variant of Gram–Schmidt.

Note that the last coordinate of \vec{h} is 0 if \mathbf{w} is in the span
of \mathbf{V} : in such a case (and if $k < n$), we select \mathbf{v} to be a
(random) normalized vector orthogonal to \mathbf{V} (we insist on
expanding to avoid stagnation in subsequential steps).

5

Hessenberg and Krylov

Hessenberg matrices and Krylov subspaces are intimately
related.

Theorem. Consider the relation $\mathbf{A}\mathbf{V}_k = \mathbf{V}_{k+1} \underline{H}_k$,

where $\mathbf{V}_{k+1} = [\mathbf{V}_k, \mathbf{v}_{k+1}]$ is $n \times (k+1)$,

and \underline{H}_k is $(k+1) \times k$.

Then, $\mathbf{v}_1, \dots, \mathbf{v}_k$ form a **Krylov basis** for $\mathcal{K}_k(\mathbf{A}, \mathbf{v}_1)$

i.e., \mathbf{V}_j spans $\mathcal{K}_j(\mathbf{A}, \mathbf{r}_0)$ for all $j = 1, \dots, k$,

$\Leftrightarrow \underline{H}_k$ is Hessenberg.

In **Arnoldi's decomposition**, \mathbf{V}_k is selected to be ortho-
normal (to ease computations and to enhance stability).

Arnoldi's method:

orthonormalise $\mathbf{A}\mathbf{v}_k$ against \mathbf{V}_k to obtain \mathbf{v}_{k+1} all k .

6

Arnoldi's decomposition

$$\mathbf{A}\mathbf{V}_{k-1} = \mathbf{V}_k \underline{H}_{k-1},$$

with \mathbf{V}_k $n \times k$ orthonormal, \underline{H}_{k-1} $k \times (k-1)$ Hessenberg.

Expand the decomposition to $\mathbf{A}\mathbf{V}_k = \mathbf{V}_{k+1} \underline{H}_k$.

Notation. $[\mathbf{V}_{k+1}, \underline{H}_k] = \text{ArnStep}(\mathbf{A}, \mathbf{V}_k, \underline{H}_{k-1})$

$$\begin{aligned} \mathbf{w} &= \mathbf{A}\mathbf{v}_k \\ [\mathbf{V}_{k+1}, \vec{h}_k] &= \text{Orth}(\mathbf{V}_k, \mathbf{w}) \\ \mathbf{V}_{k+1} &= [\mathbf{V}_k, \mathbf{v}_{k+1}] \\ \underline{H}_k &= \begin{bmatrix} \underline{H}_{k-1} \\ \vec{0}_{k-1}^* \end{bmatrix}, \quad \underline{H}_k \leftarrow [\underline{H}_k, \vec{h}_k] \end{aligned}$$

7

8

Generalized Minimal Residuals

Proposition. With $\mathbf{x}_k \equiv \mathbf{V}_k \vec{y}_k$ and $\mathbf{r}_k \equiv \mathbf{b} - \mathbf{A}\mathbf{x}_k$, \vec{y}_k solves $\underline{H}_k \vec{y}_k = \rho_0 e_1$ in least square sense
 $\Leftrightarrow \|\mathbf{r}_k\|$ minimizes $\|\mathbf{b} - \mathbf{A}\tilde{\mathbf{x}}\|_2$ over all $\tilde{\mathbf{x}} \in \mathcal{K}_k(\mathbf{A}, \mathbf{b})$.

```

Select  $k_{\max}$  and tol
Set  $\rho_0 = \|\mathbf{b}\|_2$ ,  $\mathbf{V}_1 = [\mathbf{b}/\rho_0]$ ,  $\underline{H}_0 = []$ 
for  $k = 1, \dots, k_{\max}$  do
  Break if  $\rho_k < \text{tol}$ 
   $[\mathbf{V}_{k+1}, \underline{H}_k] = \text{ArnStep}(\mathbf{A}, \mathbf{V}_k, \underline{H}_{k-1})$ 
  Solve  $\underline{H}_k \vec{y}_k = \rho_0 e_1$  in least square sense
   $\rho_k = \|\rho_0 e_1 - \underline{H}_k \vec{y}_k\|_2$ 
end for
 $\mathbf{x} = \mathbf{V}_k \vec{y}_k$ .

```

9

GMRES & FOM

Notes.

- “Solve $\underline{H}_k \vec{y}_k = \rho_0 e_1$ in the least square sense”
and “solve $\underline{H}_k \vec{y}_k = \rho_0 e_1$ ”
are problems in k -space, where $k \ll n$: **costs are ‘negligible’**

- There is no need to compute the residuals \mathbf{r}_k :
computing residual-norm does not require computing \mathbf{r}_k

$$\rho_k = \|\mathbf{r}_k\|_2 = \|\rho_0 e_1 - \underline{H}_k \vec{y}_k\|_2$$

- Computation of the approximate solution only if residual accuracy has been reached.

11

Full Orthogonalisation Method

Proposition. With $\mathbf{x}_k \equiv \mathbf{V}_k \vec{y}_k$ and $\mathbf{r}_k \equiv \mathbf{b} - \mathbf{A}\mathbf{x}_k$, \vec{y}_k solves $\underline{H}_k \vec{y}_k = \rho_0 e_1 \Leftrightarrow \mathbf{r}_k \perp \mathcal{K}_k(\mathbf{A}, \mathbf{b})$.

```

Select  $k_{\max}$  and tol
Set  $\rho_0 = \|\mathbf{b}\|_2$ ,  $\mathbf{V}_1 = [\mathbf{b}/\rho_0]$ ,  $\underline{H}_0 = []$ 
for  $k = 1, \dots, k_{\max}$  do
  Break if  $\rho_k < \text{tol}$ 
   $[\mathbf{V}_{k+1}, \underline{H}_k] = \text{ArnStep}(\mathbf{A}, \mathbf{V}_k, \underline{H}_{k-1})$ 
  Solve  $\underline{H}_k \vec{y}_k = \rho_0 e_1$  for  $\vec{y}_k$ 
   $\rho_k = \|\rho_0 e_1 - \underline{H}_k \vec{y}_k\|_2$ 
end for
 $\mathbf{x} = \mathbf{V}_k \vec{y}_k$ .

```

10

GMRES & FOM

Let $\vec{\gamma}_{k+1} = (1, \gamma_2, \dots, \gamma_k, \gamma_{k+1})^T = (\vec{\gamma}_k^T, \gamma_{k+1})^T$ such that

$$\vec{\gamma}_{k+1}^* \underline{H}_k = \vec{0}_k^*$$

Note that $\vec{\gamma}_{k+1}$ can be computed by recursive updating.

Proposition.

$$\|\mathbf{r}_k^{\text{GMRES}}\|_2 = \frac{\rho_0}{\|\vec{\gamma}_{k+1}\|_2} \quad \|\mathbf{r}_k^{\text{FOM}}\|_2 = \frac{\rho_0}{|\gamma_{k+1}|}$$

Note.

[Sleijpen vd Eshof '04]

- Even the k -dimensional system has to be solved only (once) at reaching residual accuracy.

12

GMRES versus GCR

Both methods are mathematically equivalent, that is, in exact arithmetic, they have the same residuals (residual norms) at step k and the same approximate solutions.

GMRES is the most efficient method that gives the approximate solution from $\mathcal{K}_k(\mathbf{A}, \mathbf{b})$ with **smallest residual $\|\cdot\|_2$ -norm**. GCR needs \approx twice as many AXPYs as GMRES. *Assuming the same form of Gram-Schmidt is used.*

GCR trivially extends to a flexible variant (injecting an $\mathbf{u}_k \neq \mathbf{r}_k$ does not hamper convergence). A flexible variant of **GMRES lacks the efficiency advantage**.

Both methods suffer from **growing computational costs per step and growing memory requirements** with increasing step numbers.

13

Convergence

With the same \mathbf{r}_0 : $\|\mathbf{r}_k^{\text{GMRES}}\|_2 = \|\mathbf{r}_k^{\text{GCR}}\|_2 \leq \|\mathbf{r}_k^{\text{DOL}}\|_2$

'Basic convergence conditions' are:

- 1) the eigenvalues of \mathbf{A} cluster away from 0, and
- 2) the eigenvector basis is not very ill conditioned.

Proposition. If eigs \mathbf{A} in $[\lambda_-, \lambda_+] \subset (0, \infty)$, then

$$\|\mathbf{r}_k^{\text{GMRES}}\|_2 \leq \|\mathbf{r}_k^{\text{Cheb}}\|_2 \leq (2C_E \|\mathbf{r}_0\|_2) \exp\left(-\frac{2k}{\sqrt{C}}\right),$$

where $C \equiv \frac{\lambda_+}{\lambda_-}$ and C_E is the condition number of the eigenvectors (i.e., $C_E \equiv \|\mathbf{V}\|_2 \|\mathbf{V}^{-1}\|_2$ if $\mathbf{AV} = \mathbf{V}\Lambda$).

Note. It is easy to construct *sophisticated* examples with C_E extremely large (as 10^{100}). Actually, even for small $\frac{\lambda_+}{\lambda_-}$, convergence can be arbitrarily slow.

15

GMRES versus GCR

In contrast to GCR, GMRES exploits the Krylov–Hessenberg structure.

Advantage. More efficiency.

If for mathematical reasons an inner product is zero, then there is no need to compute it: thus, saving flops.

Disadvantage. Sensitive to perturbations that affect the Krylov structure.

Due to perturbations, the actual value of such an inner product may not be zero

Example. If at step k in GCR " $\mathbf{u}_k = \mathbf{r}_k$ " is replaced by "Select a random \mathbf{u}_k ", then convergence is delayed by 1 step. If at step k in GMRES $\mathbf{A}\mathbf{v}_k$ is replaced by a random vector then GMRES stagnates forever.

Perturbations may come from inexact MVs, variable pre-conditioners, etc.

14

Convergence

With the same \mathbf{r}_0 : $\|\mathbf{r}_k^{\text{GMRES}}\|_2 = \|\mathbf{r}_k^{\text{GCR}}\|_2 \leq \|\mathbf{r}_k^{\text{DOL}}\|_2$

'Basic convergence conditions' are:

- 1) the eigenvalues of \mathbf{A} cluster away from 0, and
- 2) the eigenvector basis is not very ill conditioned.

A modification of the above examples shows that any monotonic convergence curve is possible with unitary matrices and also with any eigenvalue distribution:

Proposition. For $\rho_0 \geq \rho_1 \geq \dots \geq \rho_{n-1} \geq \rho_n = 0$, with $\mathbf{x}_0 = \mathbf{0}$ and \mathbf{A} is $n \times n$, consider the statement

(*) for a \mathbf{b} and $\mathbf{Ax} = \mathbf{b}$, we have $\|\mathbf{r}_k^{\text{GMRES}}\|_2 = \rho_k$ all k .

Then a) (*) holds for some unitary matrix \mathbf{A} , and

b) given $\lambda_1, \dots, \lambda_n$ in \mathbb{C} , (*) holds for some matrix \mathbf{A} with eigenvalues $\lambda_1, \dots, \lambda_n$.

16

Eigenvalues and Arnoldi's decomposition

$$\mathbf{Ax} = \lambda \mathbf{x}.$$

Find a normalized $\mathbf{x}_k \in \mathcal{K}_k(\mathbf{A}, \mathbf{r}_0)$ such that

$$\mathbf{r}_k = \lambda^{(k)} \mathbf{x}_k - \mathbf{Ax}_k \quad \text{with} \quad \lambda^{(k)} \equiv \mathbf{x}_k^* \mathbf{Ax}_k$$

is small in some sense and $\lambda^{(k)}$ 'almost' has the desired properties

Arnoldi's decomposition: $\mathbf{AV}_k = \mathbf{V}_{k+1} \underline{H}_k$.

$$\mathbf{r}_k = \mathbf{V}_{k+1} (\lambda^{(k)} \vec{y}_k - \underline{H}_k \vec{y}_k), \quad \lambda^{(k)} = \vec{y}_k^* \underline{H}_k \vec{y}_k.$$

Note

$$\|\mathbf{r}_k\|_2 = \|\lambda^{(k)} \vec{y}_k - \underline{H}_k \vec{y}_k\|_2$$

The computation of $\|\mathbf{r}_k\|_2$ and $\lambda^{(k)}$ is in k -space!

17

(Harmonic) Ritz values and (GMRES) FOM residuals

Consider the linear system $\mathbf{Ax} = \mathbf{r}$. Take $\mathbf{v}_1 = \mathbf{r}_0 / \|\mathbf{r}_0\|_2$.

On the next transparencies a one-one relation is given between Ritz values and FOM residuals and between harmonic Ritz values and GMRES residuals. This relation provides **theoretical insight**: it allows to relate convergence of (harmonic) Ritz values towards eigenvalues to convergence of FOM (GMRES). More details on consequences for the convergence will be discussed next lecture.

19

Arnoldi's method

Proposition. With $\mathbf{x}_k \equiv \mathbf{V}_k \vec{y}_k$ and $\mathbf{r}_k \equiv \vartheta_k \mathbf{x}_k - \mathbf{Ax}_k$, \vec{y}_k solves $\underline{H}_k \vec{y}_k = \vartheta_k \vec{y}_k \Leftrightarrow \mathbf{r}_k \perp \mathcal{K}_k(\mathbf{A}, \mathbf{b})$

```

Select  $k_{\max}$  and tol
Set  $\rho_0 = 1$ ,  $\mathbf{V}_1 = [\mathbf{b} / \|\mathbf{b}\|_2]$ ,  $\underline{H}_0 = []$ 
for  $k = 1, \dots, k_{\max}$  do
  Break if  $\rho_k < \text{tol}$ 
   $[\mathbf{V}_{k+1}, \underline{H}_k] = \text{ArnStep}(\mathbf{A}, \mathbf{V}_k, \underline{H}_{k-1})$ 
  Solve  $\underline{H}_k \vec{y}_k = \vartheta_k \vec{y}_k$  for  $k$  eigenpairs  $(\vartheta_k^{(i)}, \vec{y}_k^{(i)})$ .
  Select a pair, say,  $(\vartheta_k, \vec{y}_k)$ ,  $\vec{y}_k \leftarrow \vec{y}_k / \|\vec{y}_k\|_2$ 
   $\rho_k = |h_{k+1,k}| |e_k^* \vec{y}_k|$ 
end for
 $\mathbf{x} = \mathbf{V}_k \vec{y}_k$ ,  $\lambda = \vartheta_k$ .

```

18

Ritz values and FOM residuals

Consider the linear system $\mathbf{Ax} = \mathbf{r}$. Take $\mathbf{v}_1 = \mathbf{r}_0 / \|\mathbf{r}_0\|_2$.

If \mathbf{r}_k is the FOM residual,

then $\mathbf{r}_k = p_k(\mathbf{A}) \mathbf{r}_0$ for some polynomial p_k of degree k : p_k is the so-called k th residual **FOM polynomial**.

Theorem. For a $\vartheta \in \mathbb{C}$ we have that

$$\vartheta \text{ is a Ritz value} \Leftrightarrow p_k(\vartheta) = 0.$$

Proof. If $p_k(\vartheta) = 0$, then $p_k(\lambda) = (\lambda - \vartheta)q(\lambda)$ for some polynomial q of degree $< k$.

$$\mathbf{r}_k = (\mathbf{A} - \vartheta \mathbf{I})q(\mathbf{A})\mathbf{r}_0 = (\mathbf{A} - \vartheta \mathbf{I})\mathbf{u}_k \perp \mathbf{V}_k$$

where $\mathbf{u}_k \equiv q(\mathbf{A})\mathbf{r}_0 \in \text{span}(\mathbf{V}_k)$. Hence, (ϑ, \mathbf{u}) is a Ritz pair.

A counting argument completes the proof (there are k Ritz values and p_k has k zeros).

20

Convergence

The space $\mathcal{K}_k(\mathbf{A}, \mathbf{b}) = \text{span}(\mathbf{V}_k)$ contains all vectors that can be computed with $k - 1$ steps of (shifted) power method, and also the vectors computed with a $k - 1$ -degree polynomial filter.

⇒ faster convergence than any polynomial filter method.

A shifted power method, with appropriate shift, is effective in computing eigenpairs with 'extremal', 'isolated' eigenvalues. Arnoldi (without shift) is even more successful in detecting such eigenpairs.

Achieving better convergence also depends on how the approximate eigenpairs are **extracted** from the **search subspaces** $\text{span}(\mathbf{V}_k)$.

Using Ritz-Galerkin, for extremal eigenvalues
(selecting extremal Ritz values)

Arnoldi:shifted power ~ GMRES:Richardson

21

Gram–Schmidt orthogonalisation

$[\mathbf{v}, \vec{h}] = \text{Orth}(\mathbf{V}, \mathbf{w})$, then $\mathbf{v} \perp \mathbf{V}$, $\|\mathbf{v}\|_2 = 1$, $\mathbf{w} = [\mathbf{V}, \mathbf{v}]\vec{h}$

Classical Gram–Schmidt

$$\vec{h} = \mathbf{V}^* \mathbf{w}, \quad \mathbf{v} = \mathbf{w} - \mathbf{V} \vec{h}$$

$$\nu = \|\mathbf{v}\|_2, \quad \vec{h} \leftarrow (\vec{h}^T, \nu)^T, \quad \mathbf{v} \leftarrow \mathbf{v}/\nu$$

Loss of stability.

- Sensitive to perturbations on \mathbf{V} .
- DOTs and AXPYs introduce roundig errors.
- Scaling by ν amplifies rounding errors if $\tan(\angle(\text{span}(\mathbf{V}), \mathbf{w})) = \nu/\|\vec{h}\|_2 \ll 1$.

Note. Costs of computing $\|\vec{h}\|_2$ are negligible (wrt costs computing $\|\mathbf{v}\|_2$).

22

Gram–Schmidt orthogonalisation

$[\mathbf{v}, \vec{h}] = \text{Orth}(\mathbf{V}, \mathbf{w})$, then $\mathbf{v} \perp \mathbf{V}$, $\|\mathbf{v}\|_2 = 1$, $\mathbf{w} = [\mathbf{V}, \mathbf{v}]\vec{h}$

Modified Gram–Schmidt

$$\mathbf{v} = \mathbf{w}$$

$$\text{for } j = 1, \dots, k \text{ do}$$

$$h_j = \mathbf{v}_j^* \mathbf{v}, \quad \mathbf{v} \leftarrow \mathbf{v} - \mathbf{v}_j h_j$$

$$\text{end for}$$

$$\nu = \|\mathbf{v}\|_2, \quad \vec{h} = (h_1, h_2, \dots, h_k, \nu)^T, \quad \mathbf{v} \leftarrow \mathbf{v}/\nu$$

Loss of stability.

- Sensitive to perturbations on \mathbf{V}
- **Smaller** rounding errors from AXPYs.
- Scaling by ν amplifies rounding errors if $\nu/\|\vec{h}\|_2 \ll 1$

+ More stable. - Harder to parallelise.

23

Gram–Schmidt orthogonalisation

$[\mathbf{v}, \vec{h}] = \text{Orth}(\mathbf{V}, \mathbf{w})$, then $\mathbf{v} \perp \mathbf{V}$, $\|\mathbf{v}\|_2 = 1$, $\mathbf{w} = [\mathbf{V}, \mathbf{v}]\vec{h}$

Repeated Gram–Schmidt with DGKS criterion

$$\begin{aligned} \vec{h} &= \mathbf{V}^* \mathbf{w}, \quad \mathbf{v} = \mathbf{w} - \mathbf{V} \vec{h} \\ \nu &= \|\mathbf{v}\|_2, \quad \mu = \|\vec{h}\|_2 \\ \text{while } \nu &\leq \tau \mu \\ \quad \vec{g} &= \mathbf{V}^* \mathbf{v}, \quad \mathbf{v} \leftarrow \mathbf{v} - \mathbf{V} \vec{g} \\ \quad \nu &= \|\mathbf{v}\|_2, \quad \mu = \|\vec{g}\|_2, \quad \vec{h} \leftarrow \vec{h} + \vec{g} \\ \text{end while} \\ \vec{h} &\leftarrow (\vec{h}^\top, \nu)^\top, \quad \mathbf{v} \leftarrow \mathbf{v} / \nu \end{aligned}$$

Loss of stability.

- **Not** sensitive to perturbations on \mathbf{V}
- **Smaller** rounding errors from AXPYs.
- Scaling by ν amplifies rounding errors if $\nu / \|\vec{h}\|_2 \ll 1$

24

Gram–Schmidt and Arnoldi

Theorem. Modified Gram–Schmidt is sufficiently stable for solving linear systems.

Proof. In Arnoldi, the $n \times (k+1)$ matrix \mathbf{W} is

$$\mathbf{W} = [\mathbf{v}_1, \mathbf{A}\mathbf{v}_1, \dots, \mathbf{A}\mathbf{v}_k] \quad \text{and} \quad R = \begin{bmatrix} 1 & & \\ & H_k & \\ \mathbf{0} & & \end{bmatrix}.$$

Hence, when solving $\mathbf{A}\mathbf{x} = \mathbf{r}_0$ with $\mathbf{x}_k = \mathbf{V}_k \vec{y}_k$, we have

$$\min \frac{\|\mathbf{W}z\|_2}{\|z\|_2} \leq \frac{\|\mathbf{r}_k\|_2}{\|\mathbf{r}_0\|_2} \quad (\text{take } z = (-\|\mathbf{r}_0\|_2, \vec{y}_k^\top)^\top).$$

Therefore, we have the (sharp) estimate

$$c_2(\mathbf{W}) \gtrsim \|\mathbf{A}\|_2 \frac{\|\mathbf{r}_0\|_2}{\|\mathbf{r}_k\|_2}.$$

Prop. Eigenvalue computations requires more stability.

26

Stability of the Gram–Schmidt variants

Orthogonalisation recursively applied to the columns of an $n \times k$ matrix \mathbf{W} leads to computed $\widehat{\mathbf{V}}$ and \widehat{R} such that

$$\mathbf{W} + \Delta = \widehat{\mathbf{V}} \widehat{R}$$

for some $n \times k$ perturbation matrix Δ with

- \widehat{R} is $k \times k$ upper triangular,
- $\|\Delta\|_F \leq 4k^2 u \|\mathbf{W}\|_F$,
- Loss of orthogonality: $\|\widehat{\mathbf{V}}^* \widehat{\mathbf{V}} - I_k\|_2 \leq \kappa u (c_2(\mathbf{W}))^\ell$

with κ, ℓ depend on orth. method:

ClassGS: κ of order \sqrt{kn} , $\ell = 2$ (conjecture).

ModGS: κ of order \sqrt{kn} , $\ell = 1$.

RepGS: κ may depend on $\frac{1}{\tau^k}$ (rarely), $\ell = 0$.

Householder QR: $\kappa = \mathcal{O}(\sqrt{kn})$, $\ell = 0$.

25

Arnoldi's decomposition based methods

1) **Expansion.** Use recursive expansion for building a Krylov basis \mathbf{V}_k (involves high dimensional operations)

2) **Extraction.** For theoretical analysis, consider a projected problem as:

$$\begin{aligned} \mathbf{b} - \mathbf{A}\mathbf{V}_k \vec{y}_k &\perp \mathbf{V}_k, & \mathbf{b} - \mathbf{A}\mathbf{V}_k \vec{y}_k &\perp \mathbf{A}\mathbf{V}_k \\ \vartheta \mathbf{V}_k \vec{y}_k - \mathbf{A}\mathbf{V}_k \vec{y}_k &\perp \mathbf{V}_k, & \text{or } \vartheta \mathbf{V}_k \vec{y}_k - \mathbf{A}\mathbf{V}_k \vec{y}_k &\perp \mathbf{A}\mathbf{V}_k \end{aligned}$$

For practical computations,

2.a) Form a projected matrix, as $H_k = \mathbf{V}_k^* \mathbf{A}\mathbf{V}_k$. (high dim)

2.b) Use the projected matrix to solve the projected problem for \vec{y}_k in k -space (only k -dimensional operations)

2.c) Assemble $\mathbf{x}_k = \mathbf{V}_k \vec{y}_k$. (high dim)

Note. When recursively using Gram–Schmidt to compute the component of $\mathbf{A}\mathbf{v}_k$ that is orthonormal to \mathbf{V}_k , the projected matrix H_k comes for free.

27

Krylov subspace methods

Krylov subspace methods search for approximate solutions in a Krylov subspace: the search subspace is a Krylov subspace.

Stages.

- **Expansion.**
Expand a Krylov basis $\mathbf{v}_1, \dots, \mathbf{v}_k$ recursively
- **Extraction.**
Extract an approximate solution from $\text{span}(\mathbf{V}_k)$
- If space becomes too large
Shrinking. (Restart)
For some $\ell < k$, select a Krylov basis $\tilde{\mathbf{v}}_1, \dots, \tilde{\mathbf{v}}_\ell$ in the space $\text{span}(\mathbf{V}_k)$ such that $\text{span}(\tilde{\mathbf{V}}_\ell)$ contains promising approximations.

28

Krylov subspace methods

Why searching for approximations in Krylov subspaces?

- 1) Convergence based on polynomial approximation theory (better than Richardson, Power method, etc.)
- 2) Krylov structure can be exploited to enhance efficiency.

For instance,

- with Arnoldi's method, the Hessenberg matrix (projected matrix) comes for free.
- if \mathbf{A} is Hermitian then expansion vectors can efficiently be computed (as in CR, CG, ...)

29

Subspace methods

Why searching for approximations in general subspaces?

To allow detection of more effective expansion vectors (as in Flexible GCR, and Flexible GMRES).

30