UNIVERSITÀ
DEGLI STUDI DI TRIESTE

Iterative Methods for Sparse Linear Systems

The Generalized Minimal Residual (GMRES) method (Part I)

Ángeles Martínez Calomardo
Department of Mathematics and Earth Science
University of Trieste
e-mail:amartinez@units.it

Iterative methods for general systems

- CG is expected to converge only for SPD matrices. Why?
- CG at iteration k minimizes the error on a subspace of dimension k and constructs an orthogonal basis using a short-term recurrence.

  For nonsymmetric matrices it is impossible to have at the same time orthogonality and a minimization property by means of a short term recurrence (Faber & Manteuffel, 1984).

- Extensions of CG for general nonsymmetric matrices:
  1. CG method applied to the (SPD) system $A^T A\boldsymbol{x} = A^T\boldsymbol{b}$ (Normal equations). This system is often ill-conditioned, in fact:

  $$\kappa_2(A^T A) = \sqrt{\frac{\lambda_{\max}(A^T A)^2}{\lambda_{\min}(A^T A)^2}} = \frac{\lambda_{\max}(A^T A)}{\lambda_{\min}(A^T A)} = \|A\|_2^2 \cdot \|A^{-1}\|_2^2 = \kappa(A)^2.$$

     Also difficult to find a preconditioner if $A^T A$ could not be explicitly formed.
  2. Methods that provide orthogonality + minimization by using a long-term recurrence (GMRES)
  3. Methods that provide (bi) orthogonality. Examples: BiCG, BiCGstab
  4. Methods that provide some minimization properties. Examples: QMR, TFQMR.

The GMRES method

The GMRES (Generalized Minimal RESidual) method finds the solution of the linear system

$$A\boldsymbol{x} = \boldsymbol{b}$$

by minimizing the norm of the residual $\boldsymbol{r}_k = \boldsymbol{b} - A\boldsymbol{x}_k$ over all the vectors $\boldsymbol{x}_k$ written as

$$\boldsymbol{x}_k = \boldsymbol{x}_0 + \boldsymbol{y}, \qquad \boldsymbol{y} \in \mathcal{K}_k(\boldsymbol{v}_1)$$

where $\boldsymbol{x}_0$ is an arbitrary initial vector and $\mathcal{K}_k$ is the Krylov subspace generated by the normalized initial residual ($\boldsymbol{v}_1 = \frac{\boldsymbol{r}_0}{\|\boldsymbol{r}_0\|}$).
First note that the basis

$$\{\boldsymbol{v}_1, A\boldsymbol{v}_1, A^2\boldsymbol{v}_1, \ldots, A^{k-1}\boldsymbol{v}_1\}$$

of $\mathcal{K}_k$ is “little linearly independent”.

Theorem

If the eigenvalues of A are such that $|\lambda_1| > |\lambda_2| \geq \cdots$ then $A^k\boldsymbol{r}_0 = \boldsymbol{u}_1 + O\left(\left(\frac{|\lambda_2|}{|\lambda_1|}\right)^k\right)$, with $\boldsymbol{u}_1$ the eigenvector corresponding to λ_1.

Orthogonalization of the Krylov basis: The Arnoldi method

To compute a really independent basis for $\mathcal{K}_m$ we have to orthonormalize such vectors using the Modified Gram-Schmidt procedure.

This algorithm is known as the **Arnoldi method**.

$\beta = \|\boldsymbol{r}_0\|, \;\; \boldsymbol{v}_1 = \dfrac{\boldsymbol{r}_0}{\beta}$
for $k = 1 : m$ **do**
  $\boldsymbol{w}_{k+1} = A\boldsymbol{v}_k$
  for $j = 1 : k$ **do**
    $h_{jk} = \boldsymbol{w}_{k+1}^T \boldsymbol{v}_j$
    $\boldsymbol{w}_{k+1} := \boldsymbol{w}_{k+1} - h_{jk}\boldsymbol{v}_j$
  end for
  $h_{k+1,k} = \|\boldsymbol{w}_{k+1}\|$
  if $h_{k+1,k} \neq 0$ **then**
    $\boldsymbol{v}_{k+1} = \boldsymbol{w}_{k+1}/h_{k+1,k}$
  else
    STOP
  end if
end for

W. E. Arnoldi,
The Principle of Minimized Iteration in the Solution of the Matrix Eigenvalue Problem,
Quart. Appl. Math., 9, 1951

Orthogonalization of the Krylov basis: The Arnoldi method

Proposition

Assume that Arnoldi Algorithm does not stop before the m-th step. Then the vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_m$ *form an orthonormal basis of the Krylov subspace*

$$\mathcal{K}_m = span\{\boldsymbol{v}_1, A\boldsymbol{v}_1, A^2\boldsymbol{v}_1, \ldots, A^{m-1}\boldsymbol{v}_1\}$$

Prove it by exercise.

Hint: the vectors $\boldsymbol{v}_i$, $i = 1, 2, \ldots, m$, are orthonormal by construction. The only thing to prove is that they span $\mathcal{K}_m$. This can be done proving by induction on j that each vector $\boldsymbol{v}_j$ is of the form $q_{j-1}(A)\boldsymbol{v}_1$, where q_j is a polynomial of degree j.

Properties of the Arnoldi method

We now investigate under which conditions the Arnoldi method can *break down*.
We will consider in this section that a Krylov subspace $\mathcal{K}_m(A, \boldsymbol{v})$ can be not necessarily of dimension m.

Definition

The *grade* of a vector $\boldsymbol{v}$ (with respect to A) is the degree of the nonzero monic (i.e. with leading coefficient equal to 1) minimum degree polynomial satisfying $\boxed{p(A)\boldsymbol{v} = 0}$.

Proposition (A1)

Let r be the grade of $\boldsymbol{v}$. Then $\mathcal{K}_r(A, \boldsymbol{v})$ is invariant under A and $\mathcal{K}_m = \mathcal{K}_r, \forall m \geq r$.

Proof.

By hypothesis $p(A)\boldsymbol{v} = 0$ with p of degree r.

This is equivalent to $A^r\boldsymbol{v} + \displaystyle\sum_{k=0}^{r-1} \alpha_k A^k \boldsymbol{v} = 0$ which implies that $A^r\boldsymbol{v} = -\displaystyle\sum_{k=0}^{r-1} \alpha_k A^k \boldsymbol{v}$.

Now let $\boldsymbol{z} \in \mathcal{K}_r, \boldsymbol{z} = \displaystyle\sum_{k=0}^{r-1} \beta_k A^k \boldsymbol{v}$, we prove that $A\boldsymbol{z} \in \mathcal{K}_r$. In fact,

$$A\boldsymbol{z} = \beta_{r-1} A^r \boldsymbol{v} + \sum_{k=1}^{r-1} \beta_{k-1} A^k \boldsymbol{v} = -\beta_{r-1} \sum_{k=0}^{r-1} \alpha_k A^k \boldsymbol{v} + \sum_{k=1}^{r-1} \beta_{k-1} A^k \boldsymbol{v} \in \mathcal{K}_r.$$

□

Properties of the Arnoldi method

Proposition (A2)

The Krylov subspace $\mathcal{K}_m$ has dimension m if and only if grade$(\boldsymbol{v}) \geq m$.

Proof.

The vectors $\boldsymbol{v}, A\boldsymbol{v}, \dots, A^{m-1}\boldsymbol{v}$ form a basis of $\mathcal{K}_m$ if and only if for any set of m scalars $\alpha_j, j = 0, \dots, m-1$, not all zero, the linear combination $\displaystyle\sum_{i=0}^{m-1} \alpha_i A^i \boldsymbol{v} \neq \boldsymbol{0}$, which is equivalent to say that the only polynomial of degree $\leq m-1$ such that $p(A)\boldsymbol{v} = 0$ is the zero polynomial. $\square$

Summarizing:

- There exists a Krylov subspace of dimension m generated by $\boldsymbol{v}_1$ if and only the grade of $\boldsymbol{v}_1$ is $\geq m$.
- The grade of $\boldsymbol{v}_1$ is the maximum dimension of a Krylov subspace generated by $\boldsymbol{v}_1$.

Properties of the Arnoldi method

Proposition

The Arnoldi algorithm breaks down at step k (i.e. $h_{k+1,k} = 0$) if and only if the grade of $\boldsymbol{v}_1$ is k. In this case $\mathcal{K}_k$ is invariant under A ($A\mathcal{K}_k \subset \mathcal{K}_k$).

Proof.

($\Longleftarrow$)
If the grade of $\boldsymbol{v}_1$ is k then there exists a polynomial p of degree k such that $p(A)\boldsymbol{v}_1 = 0$.

Let us suppose by contradiction that $h_{k+1,k} \neq 0 \implies \boldsymbol{w}_{k+1} \neq 0$.

In this case we could compute $\boldsymbol{v}_{k+1}$ therefore constructing $\mathcal{K}_{k+1}(A, \boldsymbol{v}_1)$.

This contradicts Proposition (A2), since there can not be a Krylov subspace generated by $\boldsymbol{v}_1$ with dimension greater than the grade of $\boldsymbol{v}_1$.

($\Longrightarrow$)
Assume now that $\boldsymbol{w}_{k+1} = 0$. Recall that $\boldsymbol{w}_{k+1} \in \langle \boldsymbol{v}_1, A\boldsymbol{v}_1, \ldots, A^k\boldsymbol{v}_1 \rangle$, so that there is a suitable polynomial p of degree $r \leq k$ such that $\boldsymbol{w}_{k+1} = p(A)\boldsymbol{v}_1$.

It remains to prove that the degree of p is exactly k. In fact if $r < k$, then $\boldsymbol{w}_{r+1}$ would be zero for the proof of ($\Longleftarrow$) and therefore $h_{r+1,r} = 0$, and the algorithm would have stopped at step r.

The second statement comes from Proposition (A1). $\square$

Properties of the Arnoldi method

Theorem

The new vectors $\boldsymbol{v}_k$ satisfy: $\quad A\boldsymbol{v}_k = \sum_{j=1}^{k+1} h_{jk}\boldsymbol{v}_j, \qquad k = 1, \ldots, m$

This result is easy to prove.

From steps 4–7 of the previous Arnoldi algorithm we have

$$\boldsymbol{w}_{k+1} = A\boldsymbol{v}_k - \sum_{j=1}^{k} h_{jk}\boldsymbol{v}_j$$

Now using 8. we can substitute $\boldsymbol{w}_{k+1} = h_{k+1,k}\boldsymbol{v}_{k+1}$ to obtain

$$h_{k+1,k}\boldsymbol{v}_{k+1} = A\boldsymbol{v}_k - \sum_{j=1}^{k} h_{jk}\boldsymbol{v}_j$$

so that the thesis holds.

Matrix relations of the Arnoldi procedure

The relation

$$A\boldsymbol{v}_k = \sum_{j=1}^{k} h_{jk}\boldsymbol{v}_j + h_{k+1,k}\boldsymbol{v}_{k+1} = \sum_{j=1}^{k+1} h_{jk}\boldsymbol{v}_j$$

can be written in matrix form as

$$\underbrace{A\begin{pmatrix} & & & \downarrow \\ \boldsymbol{v}_1 & \boldsymbol{v}_2 & \dots & \boldsymbol{v}_k \\ & & & \end{pmatrix}}_{V_k} = \underbrace{\begin{pmatrix} \\ \boldsymbol{v}_1 & \boldsymbol{v}_2 & \dots & \boldsymbol{v}_k & \boldsymbol{v}_{k+1} \\ \\ \end{pmatrix}}_{V_{k+1}} \underbrace{\begin{pmatrix} h_{11} & h_{12} & \dots & \dots & h_{1k} \\ h_{21} & h_{22} & h_{23} & \dots & h_{2k} \\ 0 & h_{32} & h_{33} & \dots & \dots \\ 0 & 0 & h_{43} & \dots & \dots \\ 0 & 0 & 0 & \dots & \dots \\ 0 & 0 & 0 & \dots & h_{kk} \\ 0 & 0 & 0 & 0 & h_{k+1,k} \end{pmatrix}}_{H_{k+1,k}}$$

Now define

$$V_m = [\boldsymbol{v}_1, \boldsymbol{v}_2, \dots \boldsymbol{v}_m]$$

and $H_{m+1,m} = (h_{jk}), j = 1, \dots m+1,\ \ k = 1, \dots, m.$

Matrix relations of the Arnoldi procedure

At the end of the Arnoldi process we have

$$AV_m = V_{m+1} H_{m+1,m}.$$

Denoting as H_m the matrix obtained from $H_{m+1,m}$ by deleting its last row we have also the relation

$$V_m^T A V_m = V_m^T V_{m+1} H_{m+1,m} = H_m.$$

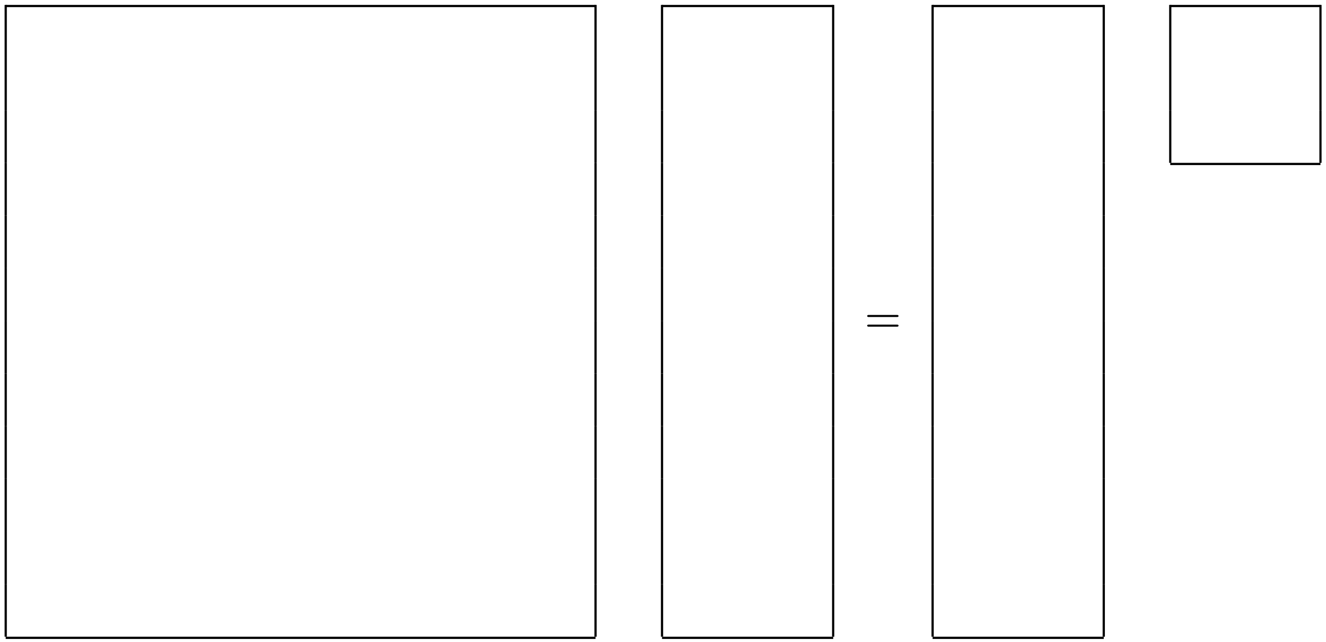
In fact, since

$$V_m^T V_{m+1} = V_m^T \begin{bmatrix} V_m & \boldsymbol{v}_{m+1} \end{bmatrix} = \begin{bmatrix} I_m & \mathbf{0} \end{bmatrix},$$

we obtain

$$V_m^T V_{m+1} H_{m+1,m} = \begin{bmatrix} I_m & \mathbf{0} \end{bmatrix} \begin{bmatrix} H_m \\ h_{m+1,m} \boldsymbol{e}_m^T \end{bmatrix} = H_m.$$

Size of the matrices involved



$$A \quad V_m \quad = \quad V_{m+1} \quad H_{m+1,m}$$

The Full Orthogonalization Method (FOM)

After the Arnoldi process has been used to compute an orthonormal basis of $\mathcal{K}(A, \boldsymbol{v}_1)$, different selections of the subspace L_m yield different methods.
Consider e.g. the choice $L_m = \mathcal{K}_m$.

This gives raise to the the Full Orthogonalizaton Method which search for an approximation $\boldsymbol{x}_m \in \boldsymbol{x}_0 + \mathcal{K}(A, \boldsymbol{v}_1) = \boldsymbol{x}_0 + V_m \boldsymbol{y}$ for some vector $\boldsymbol{y}$, by imposing the orthogonality condition $\mathcal{K}_m \perp \boldsymbol{r}_m$:

$$0 = V_m^T \boldsymbol{r}_m = V_m^T (\boldsymbol{r}_0 - A V_m \boldsymbol{y}) = V_m^T \boldsymbol{r}_0 - V_m^T A V_m \boldsymbol{y}.$$

which gives

$$\boldsymbol{y} = (V_m^T A V_m)^{-1} V_m^T \boldsymbol{r}_0 = H_m^{-1} V_m^T \boldsymbol{r}_0,$$

where H_m is the (square) matrix obtained by dropping the last row in $H_{m+1,m}$.

So that the m-th approximation computed by FOM is:

$$\boldsymbol{x}_m = \boldsymbol{x}_0 + V_m H_m^{-1} V_m^T \boldsymbol{r}_0.$$

For the algorithm and variants of FOM see Saad's book (pag.153).

GMRES: Minimization of the residual

The GMRES method is a projection method onto $\mathcal{K}_m$ where the subspace $\mathcal{L}_m = A\mathcal{K}_m$, being $\mathcal{K}_m$ the m-th dimensional Krylov subspace generated by $\boldsymbol{v}_1 = \boldsymbol{r}_0/\|\boldsymbol{r}_0\|$.

It minimizes $\|\boldsymbol{r}_m\|_2$ among all vectors $\boldsymbol{x}_m$ of the form:

$$\boldsymbol{x}_m = \boldsymbol{x}_0 + \sum_{j=1}^{m} y_i \boldsymbol{v}_i = \boldsymbol{x}_0 + V_m \boldsymbol{y}, \qquad \boldsymbol{y} = (y_1, \ldots, y_m)^T$$

Now

$$\begin{aligned}
\min_{\boldsymbol{x} \in \boldsymbol{x}_0 + \mathcal{K}_m(A, \boldsymbol{r}_0)} \|\boldsymbol{r}_m\|_2 &= \min_{\boldsymbol{x} \in \boldsymbol{x}_0 + \mathcal{K}_m(A, \boldsymbol{r}_0)} \|\boldsymbol{b} - A\boldsymbol{x}_m\|_2 = \\
&= \min_{\boldsymbol{y}} \|\boldsymbol{b} - A\left(\boldsymbol{x}_0 + V_m \boldsymbol{y}\right)\|_2 = \\
&= \min_{\boldsymbol{y}} \|\boldsymbol{r}_0 - AV_m \boldsymbol{y}\|_2 = (\text{recalling that } \boldsymbol{v}_1 = \frac{\boldsymbol{r}_0}{\|\boldsymbol{r}_0\|} = \frac{\boldsymbol{r}_0}{\beta}) \\
&= \min_{\boldsymbol{y}} \|\beta \boldsymbol{v}_1 - V_{m+1} H_{m+1,m} \boldsymbol{y}\|_2 = (\text{writing } \boldsymbol{v}_1 \text{ as } V_{m+1}\boldsymbol{e}_1) \\
&= \min_{\boldsymbol{y}} \|V_{m+1}\left(\beta \boldsymbol{e}_1 - H_{m+1,m}\boldsymbol{y}\right)\|_2
\end{aligned}$$

where $\boldsymbol{e}_1 = [1, 0, \ldots, 0]^T$.

Youcef Saad and M. H. Schultz,
GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems
SIAM Journal on Scientific and Statistical Computing, 1986

Minimizing the residual

Recalling that $V_{m+1}^T V_{m+1} = I_{m+1}$:

$$\begin{aligned}
\|\boldsymbol{r}_m\|_2 &= \sqrt{\boldsymbol{r}_m^T \boldsymbol{r}_m} = \\
&= \sqrt{(\beta \boldsymbol{e}_1 - H_{m+1,m}\boldsymbol{y})^T \, V_{m+1}^T V_{m+1} \, (\beta \boldsymbol{e}_1 - H_{m+1,m}\boldsymbol{y})} = \\
&= \sqrt{(\beta \boldsymbol{e}_1 - H_{m+1,m}\boldsymbol{y})^T \, (\beta \boldsymbol{e}_1 - H_{m+1,m}\boldsymbol{y})} = \\
&= \| \, (\beta \boldsymbol{e}_1 - H_{m+1,m}\boldsymbol{y}) \, \|_2
\end{aligned}$$

and hence $\boldsymbol{y} = \text{argmin}\|\beta \boldsymbol{e}_1 - H_{m+1,m}\boldsymbol{y}\|_2$.

Comments:

- $H_{m+1,m}$ is rectangular then the linear system $H_{m+1,m}\boldsymbol{y} = \beta \boldsymbol{e}_1$ has (in general) no solutions.
- Minimization problem $\boldsymbol{y} = \text{argmin}\|\beta \boldsymbol{e}_1 - H_{m+1,m}\boldsymbol{y}\|_2$ is very **small** $m = 20, 50, 100$ as compared to the original size n. Computational solution of this problem will be very cheap.

Least square minimization

A standard method for solving least squares problems like

$$\min_{\boldsymbol{y}} \|\beta \boldsymbol{e}_1 - H_{m+1,m}\boldsymbol{y}\|_2$$

is to factorize the coefficient matrix as $H_{m+1,m} = QR$, with Q orthogonal $(m+1) \times (m+1)$ and R $(m+1) \times m$ with the $m \times m$ submatrix upper triangular and the last row equal to zero.

Then the least squares problem can be solved by the following steps:

$$H_{m+1,m}\boldsymbol{y} = \beta \boldsymbol{e}_1 \implies QR\boldsymbol{y} = \beta \boldsymbol{e}_1 \implies Q^T QR\boldsymbol{y} = \beta Q^T \boldsymbol{e}_1 \implies R\boldsymbol{y} = \beta Q^T \boldsymbol{e}_1.$$

After computing the QR factorization: $H_{m+1,m} = QR$ (computational cost $O(m^3)$), we have

$$\min \|\beta \boldsymbol{e}_1 - H_{m+1,m}\boldsymbol{y}\|_2 = \min \|\beta Q^T \boldsymbol{e}_1 - R\boldsymbol{y}\|_2 = \min \|\boldsymbol{g} - R\boldsymbol{y}\|_2$$

where $\boldsymbol{g} = \beta Q^T \boldsymbol{e}_1$.

End of minimization

The system to be solved is

$$\left(\begin{array}{ccccc} r_{11} & r_{12} & \dots & & \\ 0 & r_{22} & r_{23} & \dots & \\ & \dots & \dots & \dots & \\ & & & 0 & r_{mm} \\ \hline & & & & 0 \end{array}\right) \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_m \end{pmatrix} = \beta Q^T \boldsymbol{e}_1 = \beta \begin{pmatrix} q_{1,1} \\ q_{1,2} \\ \dots \\ q_{1,m} \\ q_{1,m+1} \end{pmatrix} \equiv \left(\begin{array}{c} g_1 \\ g_2 \\ \dots \\ g_m \\ \hline g_{m+1} \end{array}\right)$$

- The solution to $\min \|\boldsymbol{g} - R\boldsymbol{y}\|_2$ is simply accomplished by solving $\tilde{R}\boldsymbol{y} = \tilde{\boldsymbol{g}}$ where $\tilde{R}$ is obtained from R by dropping the last row and $\tilde{\boldsymbol{g}}$ the first m components of $\boldsymbol{g}$.
- This last system being square, small, and upper triangular, is easily and cheaply solved.
- Finally note that

$$\begin{aligned} \min \|\boldsymbol{r}_m\|_2 &= \min_{\boldsymbol{y}} \|\beta \boldsymbol{e}_1 - H_{m+1,m}\boldsymbol{y}\|_2 = \\ &= \min_{\boldsymbol{y}} \|R\boldsymbol{y} - \beta Q^T \boldsymbol{e}_1\|_2 = \text{(since the first } m \text{ equations are solved exactly)} \\ &= \| \begin{bmatrix} 0 & \dots & 0 & \beta q_{1,m+1} \end{bmatrix}^T \|_2 \\ &= |\beta q_{1,m+1}| = |g_{m+1}|. \end{aligned}$$

"Lucky" breakdown of GMRES

If $h_{k+1,k} = 0$ the Arnoldi algorithms breaks down as it has found an invariant Krylov subspace.

This is called *lucky breakdown* for the following reason:

The last row of $H_{k+1,k}$ is zero so that the relation

$$AV_k = V_{k+1}H_{k+1,k}$$

is replaced by (note that the $k+1$-th column of V_{k+1} does not exist)

$$AV_k = V_kH_k.$$

Hence

$$\begin{aligned}
\min_{\boldsymbol{x}\in\boldsymbol{x}_0+\mathcal{K}_k(A,\boldsymbol{r}_0)} \|\boldsymbol{b} - A\boldsymbol{x}_k\|_2 &= \min_{\boldsymbol{y}} \|\boldsymbol{r}_0 - AV_k\boldsymbol{y}\|_2 = \\
&= \min_{\boldsymbol{y}} \|\beta\boldsymbol{v}_1 - V_kH_k\boldsymbol{y}\|_2 = \quad (\text{writing now } \boldsymbol{v}_1 = V_k\boldsymbol{e}_1,\ \boldsymbol{e}_1 \in \mathbb{R}^k) \\
&= \min_{\boldsymbol{y}} \|V_k\left(\beta\boldsymbol{e}_1 - H_k\boldsymbol{y}\right)\|_2 = \\
&= \min_{\boldsymbol{y}} \|\beta\boldsymbol{e}_1 - H_k\boldsymbol{y}\|_2 = 0
\end{aligned}$$

since the minimum is obtained for $\boldsymbol{y}$ which is the solution of the square linear system $H_k\boldsymbol{y} = \beta\boldsymbol{e}_1$.

The norm of the residual is zero, GMRES has found the exact solution!

Obtaining GMRES by the orthogonality condition

The GMRES method can be developed in an alternative way by imposing the orthogonality condition $L_m \equiv A\mathcal{K}_m \perp \boldsymbol{r}_m$ which writes:

$$V_m^T A^T (\boldsymbol{r}_0 - AV_m\boldsymbol{y}) = 0.$$

This gives

$$\begin{aligned}
\boldsymbol{y} &= (V_m^T A^T A V_m)^{-1} V_m^T A^T \boldsymbol{r}_0 \\
&= (H_{m+1,m}^T V_{m+1}^T V_{m+1} H_{m+1,m})^{-1} H_{m+1,m}^T V_{m+1}^T \boldsymbol{r}_0 \\
&= (H_{m+1,m}^T H_{m+1,m})^{-1} H_{m+1,m}^T V_{m+1}^T \boldsymbol{r}_0 \\
&= (H_{m+1,m}^T H_{m+1,m})^{-1} H_{m+1,m}^T \beta \boldsymbol{e}_1, \qquad (1)
\end{aligned}$$

so vector $\boldsymbol{y}$ is the solution of the normal equations system which is known to be equivalent to the least square solution of

$$H_{m+1,m}\boldsymbol{y} = \beta\boldsymbol{e}_1.$$

In fact by multiplying the previous system by $H_{m+1,m}^T$ we obtain

$$H_{m+1,m}^T H_{m+1,m}\boldsymbol{y} = H_{m+1,m}^T \beta\boldsymbol{e}_1.$$

whose solution is (1).

Algorithm

ALGORITHM: GMRES

Input: x_0, A, b, $k_{\max}$, tol

- $\boldsymbol{r}_0 = \boldsymbol{b} - A\boldsymbol{x}_0,\ k = 0,\ \rho_0 = \|\boldsymbol{r}_0\|_2,\ \beta = \rho_0,\ \boldsymbol{v}_1 = \dfrac{\boldsymbol{r}_0}{\beta}$
- WHILE $\rho_k >$ tol $\|b\|_2$ AND $k < k_{\max}$ DO
  1. $k = k + 1$
  2. $\boldsymbol{v}_{k+1} = A\boldsymbol{v}_k$
  3. FOR $j = 1, k$

     $h_{jk} = \boldsymbol{v}_{k+1}^T \boldsymbol{v}_j$

     $\boldsymbol{v}_{k+1} = \boldsymbol{v}_{k+1} - h_{jk}\boldsymbol{v}_j$

     END FOR
  4. $h_{k+1,k} = \|\boldsymbol{v}_{k+1}\|_2$
  5. $\boldsymbol{v}_{k+1} = \boldsymbol{v}_{k+1}/h_{k+1,k}$
  6. Compute the QR factorization of $H_{k+1,k}$: $H_{k+1,k} = QR$.
  7. $\rho_k = |\beta q_{1,k+1}|$
- END WHILE
- $\boldsymbol{y}_k = \text{argmin}\|\beta\boldsymbol{e}_1 - H_{k+1,k}\boldsymbol{y}\|_2$ by using the QR factorization of $H_{k+1,k}$
- $\boldsymbol{x}_k = \boldsymbol{x}_0 + V_k\boldsymbol{y}_k$

Exercise

GMRES implementation.

- Implement the GMRES method as a Matlab function following the Algorithm in the previous slide.

  The syntax of the function should be the following:

  ```
  function [x,iter,resvec,flag] = mygmres(A,b,tol,maxit,x0)
  ```

  where the output parameters are
  - `x` is the solution vector
  - `iter` the number of iterations employed
  - `resvec` the vector with the norm of the residuals
  - `flag` a variable signaling breakdown ($= 0$: canonical termination, $= -1$: breakdown has occurred).

  The input parameters are the coefficient matrix, the right hand side, the tolerance, the maximum number of iterations and the initial guess vector.

- Upload (from Moodle) the matrix `mat13041.rig`, which is stored in coordinate format. Solve the linear system $A\boldsymbol{x} = \boldsymbol{b}$ with $\boldsymbol{b}$ corresponding to an exact solution with components $x_i = \frac{1}{\sqrt{i}}$. Use your GMRES implementation with `tol` $= 10^{-10}$, `itmax` $= 550$, and `x0` the all zero vector.
- Plot the residual norm vs the number of iterations in a `semilogy` profile.