Chapter 5
Krylov Subspace Methods that are Based on the Minimization of the Residual

Remark 5.1. Goal. The goal of these methods consists in determining

$$\underline{x}^{(k)} \in \underline{x}^{(0)} + K_k\left(\underline{r}^{(0)}, A\right)$$

such that the Euclidean norm of the corresponding residual

$$\left\|\underline{r}^{(k)}\right\|_2 = \left\|\underline{b} - A\underline{x}^{(k)}\right\|_2$$

becomes minimal. Notice that in the Richardson iteration with the special choice (4.5), the norm of the residual will be minimized on the line $\underline{x}^{(k-1)} + \tau\underline{r}^{(k-1)}$. However, the minimum on this line is in general not the global minimum in $\underline{x}^{(0)} + K_k\left(\underline{r}^{(0)}, A\right)$. □

5.1 General Matrices

Remark 5.2. Construction of an orthonormal basis of the Krylov subspace. To perform the minimization of the Euclidean norm of the residual efficiently, i.e., to solve the corresponding least squares problem, an orthonormal basis of $K_m(\underline{q}_1, A)$ is needed. There are several ways to transform an arbitrary basis into an orthonormal one, e.g., the modified Gram[1]–Schmidt[2] method[3] or the Householder[4] algorithm. In the context of Krylov subspace methods, the computation of an orthonormal basis of $K_m(\underline{q}_1, A)$ is called Arnoldi's[5] method. □

Algorithm 5.3. Arnoldi method, modified Gram–Schmidt method. Given $A \in \mathbb{R}^{n\times n}$, $\underline{q}_1 \in \mathbb{R}^n$ with $\left\|\underline{q}_1\right\|_2 = 1$, and $m < n$.

---

[1] Jorgen Pedersen Gram (1850 – 1916)

[2] Erhard Schmidt (1876 – 1959)

[3] Given a set of orthonormal vectors $\{\underline{u}_1, \ldots, \underline{u}_{m-1}\}$ and a vector $\underline{v}_m$ that should be orthornormalized with this set. In the original Gram–Schmidt method, one computes all projections of $\underline{v}_m$ with respect to $\underline{u}_i$, $i = 1, \ldots, m-1$, and subtracts these projections. In the modified Gram–Schmidt method, one computes the projection of $\underline{v}_m$ to one of the vectors, say $\underline{u}_1$, and subtracts this projection. The result $\underline{v}_m^{(1)}$ is orthogonal to $\underline{u}_1$. Now the projection $\underline{v}_m^{(1)}$ with respect to a second vector, say $\underline{u}_2$ is computed and subtracted to give $\underline{v}_m^{(2)}$. This vector is orthogonal to $\underline{u}_1$ and $\underline{u}_2$. Continuing this procedure leads to the orthonomalization of $\underline{v}_m$ with respect to $\{\underline{u}_1, \ldots, \underline{u}_{m-1}\}$. In exact arithmetic, both versions are identical. From the numerical point of view, the original Gram–Schmidt method might be instable whereas the modified Gram–Schmidt method is stable.

[4] Alston Scott Householder (1904 – 1993)

[5] Walter Edwin Arnoldi (1917 – 1995)

1. for $j = 1 : m$
2. $\quad \underline{w}_j = A\underline{q}_j$
3. $\quad$ for $i = 1 : j$
4. $\qquad h_{ij} = (\underline{w}_j, \underline{q}_i)$
5. $\qquad \underline{w}_j = \underline{w}_j - h_{ij}\underline{q}_i$ % subtract projection
6. $\quad$ endfor
7. $\quad h_{j+1,j} = \|\underline{w}_j\|_2$
8. $\quad$ if $h_{j+1,j} == 0$
9. $\qquad$ stop
10. $\quad$ endif
11. $\quad \underline{q}_{j+1} = \underline{w}_j / h_{j+1,j}$ % normalize
12. endfor

□

Lemma 5.4. Computation of an orthonormal basis by Arnoldi's method. *If* $\dim K_m(\underline{q}_1, A) = m$, *then Arnoldi's method computes an orthonormal basis* $\{\underline{q}_1, \ldots, \underline{q}_m\}$ *of* $K_m(\underline{q}_1, A)$.

Proof. The vectors $\underline{q}_1, \ldots, \underline{q}_m$ are orthonormal by construction: orthogonal by line 3 – 6 and normalized by line 7 and 11. One has to show that they belong all to $K_m(\underline{q}_1, A)$. This statement will follow from the fact that each vector $\underline{q}_j$ is of the form $p_{j-1}(A)\,\underline{q}_1$, where p_{j-1} is a polynomial of degree $j-1$. The proof is done by induction. For $j = 1$, one has $\underline{q}_1 = p_0(A)\,\underline{q}_1$ such that $p_0(t) = 1$. Assume, the statement is true for $j \le k$. One has by using first line 11, which implies $h_{k+1,k} \neq 0$, then lines 2 – 6, and finally the assumption of the induction

$$h_{k+1,k}\underline{q}_{k+1} = \underline{w}_k = A\underline{q}_k - \sum_{i=1}^{k} h_{ik}\underline{q}_i = A\underline{q}_k - \sum_{i=1}^{k}\left(A\underline{q}_k, \underline{q}_i\right)\underline{q}_i = Ap_{k-1}(A)\underline{q}_1 - \sum_{i=1}^{k} h_{ik}p_{i-1}(A)\underline{q}_1 = p_k(A)\underline{q}_1. \tag{5.1}$$

Division by $h_{k+1,k}$ finishes the proof. ■

Remark 5.5. Factorization of the system matrix. Denote

$$Q_m = \left(\underline{q}_1, \underline{q}_2, \ldots, \underline{q}_m\right) \in \mathbb{R}^{n\times m}, \quad H_m = \begin{pmatrix} h_{11} & h_{12} & \cdots & h_{1,m-1} & h_{1m} \\ h_{21} & h_{22} & \cdots & h_{2,m-1} & h_{2m} \\ 0 & h_{32} & \cdots & h_{3,m-1} & h_{3m} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & h_{m,m-1} & h_{mm} \\ 0 & 0 & \cdots & 0 & h_{m+1,m} \end{pmatrix} \in \mathbb{R}^{(m+1)\times m}.$$

A matrix of this form, i.e., $h_{ij} = 0$ for $i > j+1$, is called (upper) Hessenberg[6] matrix. It follows from (5.1) that

$$A\underline{q}_k = \underline{q}_{k+1}h_{k+1,k} + \sum_{i=1}^{k}\underline{q}_i h_{ik} = \sum_{i=1}^{k+1}\underline{q}_i h_{ik}, \quad k = 1, \ldots, m, \tag{5.2}$$

such that one gets for Arnoldi's method the compact representation

$$AQ_m = Q_{m+1}H_m. \tag{5.3}$$

□

Remark 5.6. Initial vector in Krylov subspace methods. In the Krylov subspace methods, $\underline{r}^{(0)}/\left\|\underline{r}^{(0)}\right\|_2$ plays the role of $\underline{q}_1$ in Arnoldi's method. □

[6] Karl Hessenberg (1904 – 1959)

Remark 5.7. Principle approach to minimize the residual. The goal of the methods presented in this section is to minimize the Euclidean norm of the residual. One has

$$\left\|\underline{r}^{(k)}\right\|_2 = \left\|\underline{b} - A\underline{x}^{(k)}\right\|_2 = \left\|\underline{r}^{(0)} + A\underline{x}^{(0)} - A\underline{x}^{(k)}\right\|_2 = \left\|\underline{r}^{(0)} - A\left(\underline{x}^{(k)} - \underline{x}^{(0)}\right)\right\|_2$$

with $\underline{x}^{(k)} - \underline{x}^{(0)} \in K_k\left(\underline{r}^{(0)}, A\right)$, see Remark 4.9. Since the vectors $\{\underline{q}_1, \ldots, \underline{q}_k\}$ computed with Arnoldi's method form a generating system of $K_k\left(\underline{r}^{(0)}, A\right)$, it is

$$\underline{x}^{(k)} - \underline{x}^{(0)} = \sum_{i=1}^{k} z_i \underline{q}_i = Q_k \underline{z} \tag{5.4}$$

with $\underline{z} = (z_1, \ldots, z_k)^T$, $Q_k = \left(\underline{q}_1, \ldots, \underline{q}_k\right)$. Using (5.4), (5.3), $Q_{k+1}\underline{e}_1 = \underline{q}_1 = \underline{r}^{(0)}/\left\|\underline{r}^{(0)}\right\|_2$, and the fact that the Euclidean norm is invariant under orthonormal transformations, one obtains

$$\begin{aligned}
\left\|\underline{r}^{(k)}\right\|_2^2 &= \left\|\underline{r}^{(0)} - AQ_k\underline{z}\right\|_2^2 = \left\|\underline{r}^{(0)} - Q_{k+1}H_k\underline{z}\right\|_2^2 \\
&= \left\|\left\|\underline{r}^{(0)}\right\|_2 Q_{k+1}\underline{e}_1 - Q_{k+1}H_k\underline{z}\right\|_2^2 = \left\|Q_{k+1}\left(\left\|\underline{r}^{(0)}\right\|_2 \underline{e}_1 - H_k\underline{z}\right)\right\|_2^2 = \left\|\left\|\underline{r}^{(0)}\right\|_2 \underline{e}_1 - H_k\underline{z}\right\|_2^2.
\end{aligned}$$

The minimizer of the residual is obtained by solving the least squares problem

$$\min_{\underline{z}\in\mathbb{R}^k} \left\|\left\|\underline{r}^{(0)}\right\|_2 \underline{e}_1 - H_k\underline{z}\right\|_2^2. \tag{5.5}$$

This problem possesses k unknowns $z_1, \ldots, z_k$ and the vector that has to be minimized has $k+1$ components. It can be solved, e.g., with a QR factorization of H_k, see lecture notes Numerical Mathematics I. Let $\underline{z}^{(k)}$ be a solution of this problem, then the next iterate of the Krylov subspace method is, compare (5.4)

$$\underline{x}^{(k)} = \underline{x}^{(0)} + Q_k\underline{z}^{(k)}. \tag{5.6}$$

This algorithm is called GMRES (generalized minimal residual). It has been proposed the first time in Saad & Schultz (1986). □

Theorem 5.8. Properties of GMRES.

i) In the case that Arnoldi's method has an early breakdown, i.e., $h_{l+1,l} = 0$, then $\dim K_k\left(\underline{r}^{(0)}, A\right) = l < k$ *and* $\underline{r}^{(l)} = \underline{0}$*. Hence* $A\underline{x}^{(l)} = \underline{b}$*.*

ii) The iterate $\underline{x}^{(k)} = \underline{x}^{(0)} + Q_k\underline{z}^{(k)}$ *is uniquely determined.*

iii) It holds

$$\left\|\underline{r}^{(k)}\right\|_2 \le \left\|\underline{r}^{(k-1)}\right\|_2, \quad k = 1, 2, 3, \ldots.$$

Proof. *i)* The breakdown of Arnoldi's method after l steps, $h_{l+1,l} = 0$, is equivalent to $\underline{w}_l = \underline{0}$, see line 8 of Algorithm 5.3. It follows from (5.1) that

$$A\underline{q}_l = \sum_{i=1}^{l} h_{il}\underline{q}_i,$$

where $\underline{q}_1 = \underline{r}^{(0)}/\left\|\underline{r}^{(0)}\right\|_2$ in the case of GMRES. Since, in contrast to (5.2) (upper limit of the summation), $A\underline{q}_l$ is a linear combination of the already computed l basis vectors, one has $\dim K_{l+1}\left(\underline{r}^{(0)}, A\right) = \dim K_l\left(\underline{r}^{(0)}, A\right)$. One obtains by induction

$$\dim K_k\left(\underline{r}^{(0)}, A\right) = \dim K_l\left(\underline{r}^{(0)}, A\right) \text{ for } k \ge l.$$

Using matrix notations, Arnoldi's method gives in this case

$$AQ_l = Q_l \tilde{H}_l \quad \text{with} \quad \tilde{H}_l = \begin{pmatrix} h_{11} & h_{12} & \cdots & h_{1l} \\ h_{21} & h_{22} & \cdots & h_{2l} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & h_{l,l-1} & h_{ll} \end{pmatrix} \in \mathbb{R}^{l\times l}, \quad Q_l \in \mathbb{R}^{n\times l}.$$

Since A is non-singular and $\operatorname{rank}(Q_l) = l$, one has $\operatorname{rank}(AQ_l) = l$. Consequently, it is $\operatorname{rank}\left(Q_l\tilde{H}_l\right) = l$ and $\operatorname{rank}\left(\tilde{H}_l\right) = l$ and $\tilde{H}_l$ is invertible. In the same way as in Remark 5.7, one obtains

$$\left\|\underline{r}^{(l)}\right\|_2^2 = \min_{\underline{z}\in\mathbb{R}^l} \left\|\left\|\underline{r}^{(0)}\right\|_2 \underline{e}_1 - \tilde{H}_l\underline{z}\right\|_2^2.$$

The minimizer is given by $\underline{z}^{(l)} = \tilde{H}_l^{-1}\left(\left\|\underline{r}^{(0)}\right\|_2 \underline{e}_1\right)$ which implies $\left\|\underline{r}^{(l)}\right\|_2 = 0$.

ii) If $\operatorname{rank}(H_k) = k$, the minimizer of (5.5) is unique (theory of least squares problems, see Numerical Mathematics I). If $\operatorname{rank}(H_k) < k$, then $\underline{x}^{(k)} = \underline{x}$, see i).

iii) The set in which the minimizer is computed becomes larger since the inclusion $K_k\left(\underline{r}^{(0)}, A\right) \supseteq K_{k-1}\left(\underline{r}^{(0)}, A\right)$ holds. ■

Remark 5.9. Implementational issues.

- The GMRES process consists in principle of two steps:
  1. computing the orthonormal basis of $K_k\left(\underline{r}^{(0)}, A\right)$,
  2. solving the least squares problem (5.5) to find the minimizer of the residual with a standard method.

  In the practical use of GMRES, Step 2 is performed only at the end of the iteration. Thus, the iterate $\underline{x}^{(k)}$ is not directly available. It is computed in a post-processing step. However, there is an elegant and inexpensive way to compute $\left\|\underline{r}^{(k)}\right\|_2$ without having access to $\underline{x}^{(k)}$, see Saad (2003). With $\left\|\underline{r}^{(k)}\right\|_2$, one can control the iterative process.

  For concrete ways to implement GMRES, it is refered to Saad & Schultz (1986); Saad (2003).
- Each step of GMRES requires one matrix-vector multiplication, line 2 of Arnoldi's method.
- In exact arithmetic, GMRES terminates with the solution in at most n steps. This property is, however, of no practical use for large n.
- From the practical point of view, the greatest problem of GMRES is that one has to store the basis $\{\underline{q}_1, \ldots, \underline{q}_k\}$ of $K_k\left(\underline{r}^{(0)}, A\right)$, see lines 3 – 6 of Arnoldi's method. Thus, with every new iteration, one has to store an additional vector. This situation is called long recurrence.

  In practice, one prescribes a maximal dimension m of the Krylov subspace. After m iterations, GMRES is stopped with the iterate $\underline{x}^{(m)}$. If $\underline{x}^{(m)}$ is not yet sufficiently close to the solution, GMRES is started from the beginning with $\underline{x}^{(0)} = \underline{x}^{(m)}$. This approach is called GMRES(m) (with restart). An optimal choice of m is in general an unresolved problem. Often $m \in [5, 50]$ is used.

  GMRES(m) might also fail to converge, see Saad & Schultz (1986) for the simple example

  $$A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad \underline{b} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \underline{x}^{(0)} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

  GMRES converges in two steps whereas GMRES(1) computes the stationary sequence $\underline{x}^{(1)} = \underline{x}^{(0)}$, $\underline{x}^{(2)} = \underline{x}^{(1)} = \underline{x}^{(0)}$ and so on. Despite of the possibility to fail, GMRES(m) is one of the most popular and best performing iterative methods for solving linear systems of equations with non-symmetric matrix.

□