Universität Karlsruhe
Rechenzentrum
Postfach 6980
76128 Karlsruhe
Germany

Interner Bericht Nr. 52/94

# Orthogonalization Methods

Abteilung “Numerikforschung für Supercomputer”
Rechenzentrum der Universität Karlsruhe
April 1994

Author: Rüdiger Weiss

# ORTHOGONALIZATION METHODS

RÜDIGER WEISS *

**Abstract.** The general concept of orthogonalization methods is introduced for the iterative solution of linear systems. A large scale of iterative methods is described by this concept. In particular popular and commonly used state-of-the-art techniques are covered including Krylov subspace methods such as generalized cg, error-minimizing and conjugate Krylov subspace methods.

Despite the generality of the concept convergence and geometric properties are proven giving a deep insight into the techniques. Relations between different approaches become apparent and the theoretical basis for new methods is given. Techniques to accelerate convergence like preconditioning can be interpreted consistently and used for unusual purposes. The aim of this paper is to give a unifying theory for iterative linear solvers and to stimulate research for new approaches.

**Key words.** conjugate gradients, convergence, iterative solvers, convergence, linear systems, Krylov methods.

**AMS subject classification.** 65F10, 65F50, 40A05

**Contents.**



* Numerikforschung für Supercomputer, Rechenzentrum der Universität Karlsruhe, Postfach 6980, D-76128 Karlsruhe, Germany, e-mail: weiss@rz.uni-karlsruhe.de

**1. Background.** We are interested in the solution of the linear system

$$Ax = b. \tag{1}$$

The matrix $A$ is a real, square matrix of dimension $n$, i. e. $A \in \mathbb{R}^{n\times n}$, and $x, b \in \mathbb{R}^n$. In general the matrix $A$ is non-symmetric and not positive definite.

We use the following notations: Let $Z$ be a symmetric, positive definite matrix, then the norm $\|y\|_Z$ of any vector $y \in \mathbb{R}^n$ is defined by $\|y\|_Z = \sqrt{y^T Z y}$. If $Z$ is nonsymmetric and not positive definite, then $\|y\|_Z^2$ is a mnemonic abbreviation for $y^T Z y$. $\|y\|_I$ is the Euclidean norm $\|y\|$. The symmetric part of the matrix $B \in \mathbb{R}^{n\times n}$ is $\frac{1}{2}(B + B^T)$ and the skew-symmetric part is $\frac{1}{2}(B - B^T)$.

**1.1. Iterates, residuals and errors.**

Starting from an initial guess $x_0$ an iterative solution method constructs a sequence of approximations $x_k$ so that

$$\lim_{k\to\infty} x_k \to x. \tag{2}$$

The recurrence can be very sophisticated and non-linear and in general

$$x_k = f_k(x_0, \ldots, x_{k-1}, A, b) \tag{3}$$

with $f_k : \mathbb{R}^{n\times(k+n+1)} \to \mathbb{R}^n$. The residuals are defined by

$$r_k = Ax_k - b, \tag{4}$$

and the errors by

$$e_k = x_k - x. \tag{5}$$

The residuals and errors are connected by

$$r_k = Ae_k. \tag{6}$$

The residuals and the errors are considered together in all further investigations.

We introduce $d_k \in \mathbb{R}^n$ as update direction, i. e.

$$x_k = x_{k-1} + d_k \tag{7}$$

for $k = 1, 2, \ldots$. The residuals and the errors can be expressed as

$$r_k = r_{k-1} + Ad_k, \tag{8}$$

$$e_k = e_{k-1} + d_k. \tag{9}$$

From (9) follows directly that the optimal update direction is

$$d_k = -e_{k-1} \tag{10}$$

because then $e_k = 0$ and the solution is obtained. (10) implies

$$Ad_k = -r_{k-1}, \tag{11}$$

i, e. the optimal update direction is the solution of the defect correction equation (11).

Note that (7), (8) and (9) cover in fact all iterative methods by setting:

$$d_k = f_k(x_0, \ldots, x_{k-1}, A, b) - x_{k-1}. \tag{12}$$

An iterative method is parameter-dependent if parameters have to be determined a priori for the calculation of the update directions $d_k$. Such methods are for example the SOR method [91] and the large family of semi-iterative methods [16, 17, 18, 49, 50] - to name only a few. Parameter-free methods calculate the update directions $d_k$ by using information of the iteration steps. Such methods are for example the the Jacobi, the Gauss-Seidel and the large families of generalized cg [4, 27, 37, 42, 72], error-minimizing [31, 55, 82], conjugate Krylov subspace and quasi-minimal residual methods [25, 26, 29, 30]. An excellent historical overview is given by Young [92] for established iterative solvers.

**1.2. Classical iterative methods.**

By means of an arbitrary non-singular matrix $B \in I\!R^{n\times n}$ an iterative method can be derived from

$$Bx_k + (A - B)x_{k-1} = b,$$

resulting in

$$x_k \;=\; x_{k-1} - B^{-1}r_{k-1} \tag{13}$$

with update direction $d_k = -B^{-1}r_{k-1}$. For the residuals and errors the following equations are satisfied:

$$r_k \;=\; (I - AB^{-1})r_{k-1}, \tag{14}$$

$$e_k \;=\; (I - B^{-1}A)e_{k-1}. \tag{15}$$

The methods are convergent if and only if the spectral radius of the iteration matrix

$$H = I - B^{-1}A \tag{16}$$

is less than 1, following directly from (15). The following three classical iterative methods are obtained by by splitting the matrix $A$ into

$$A = L + D + U, \tag{17}$$

where $L$ and $U$ are strictly lower and upper triangular matrices and $D$ is a diagonal matrix. If $B = D$, then the method is called Jacobi iteration [44]. For $B = L + D$ we obtain the Gauss-Seidel [67] method. For $B = L + \frac{1}{\omega}D$ we get the succesive relaxation method. The method is called succesive overrelaxation (SOR) method for $\omega > 1$. The method was theoretically investigated by Young [89, 90]. This method is a classical example for parameter-dependent iterative methods.

The method of steepest descent and the classical cg method are fundamental for many modern iterative techniques. The idea of the method of steepest descent, proposed by Temple [75], is to minimize the quadratic form

$$F(x) = \frac{1}{2}(Ax - b)^T A^{-1}(Ax - b). \tag{18}$$

If $A$ is symmetric and positive definite, such a minimum exists. The minimum of (18) is precisely the solution of (1). $x_k$ is determined by a one-dimensional minimization so that

$$F(x_k) = \min_{\delta_k} F(x_{k-1} + \delta_k r_{k-1}). \tag{19}$$

The gradient of (18) is $Ax - b$. Therefore, the update direction $\delta_k r_{k-1}$ is orthogonal to the isoline $F(x) = F(x_{k-1})$ and parallel to the isoline $F(x) = F(x_k)$. The method converges, if $A$ is positive real. Note that

$$2F(x_k) = r_k^T A^{-1} r_k = e_k^T A e_k = \|e_k\|_A \tag{20}$$

is called energy norm of the error because for systems arising from structural analysis this quantity is equivalent to the potential energy of the discretized problem.

The conjugate gradient (cg) method was developed by Hestenes and Stiefel [42] in the early 1950s for the solution of linear systems with a symmetric, positive definite matrix. The method can be considered as a direct method because the exact solution is obtained at least in the step $n$ in the absence of roundoff errors. Until the early 1970s the method was not widely used. With the increasing use of computers and the consideration as an iterative method by Reid [54] the method recovered importance. The idea of the classical conjugate gradient method is to minimize the quadratic form (18) by using a multi-dimensional minimization improving the steepest descent method. If the ellipses of the isolines of $F(x) = constant$ have very different semi-axes the convergence of the steepest descent method may be poor, while a more sophisticated update direction improves the convergence. For the cg method $x_k$ is determined from

$$\begin{aligned} F(x_k) &= \min_{\delta_1,\ldots,\delta_k} F(x_{k-1} + \sum_{i=1}^{k} \delta_i r_{i-1}) \\ &= \frac{1}{2} \min_{\delta_1,\ldots,\delta_k} \left\| e_{k-1} + \sum_{i=1}^{k} \delta_i r_{i-1} \right\|_A . \end{aligned} \tag{21}$$

The new iterate can be calculated by a simple three-term recurrence. There are various possible implementations of the algorithm. The method converges, if $A$ is symmetric and positive definite. The matrix $B$ in (13) is step-dependent.

Classical iterative methods are suited for matrices that are diagonally dominant or symmetric and positive definite. Of course the methods can also be applied to other systems, but the convergence is not guaranteed or may be very slow. Therefore these methods have been generalized in order to obtain converging techniques for general linear systems. A whole family of methods results from this generalization. For all generalized cg methods the iterates are in a shifted Krylov space

$$x_k \in x_0 + span(A^k r_0, A^{k-1} r_0, \ldots, A r_0, r_0). \tag{22}$$

Condition (21) is equivalent to the orthogonalities

$$r_k^T r_{k-i} = 0 \tag{23}$$

for $i = 1, \ldots, k$. This orthogonality relation has been the starting point for the following concept of orthogonalization methods.

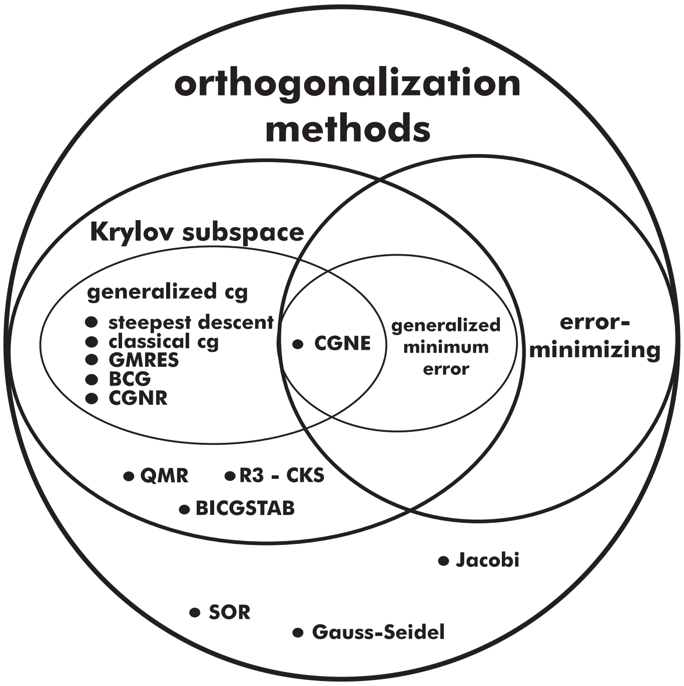


FIG. 1. *Subsets of orthogonalization methods*

## 2. Orthogonalization methods.

We start by giving a definition that includes most of the modern parameter-free iterative methods. In particular generalized cg methods [4, 19, 27, 37, 42, 43, 72, 81], error-minimizing methods [31, 82], conjugate Krylov subspace methods [83] and many semiiterative methods fit into this class.

DEFINITION 1. *Let $x_0$ be any initial guess and $r_0 = Ax_0 - b$ the starting residual. The following recurrence is called an orthogonalization method. For $k \geq 1$ calculate the approximations so that*

$$x_k \in \tilde{x}_k + span(q_{k-\sigma_k,k}, \ldots, q_{k-1,k}) \tag{24}$$

*with $q_{k-i,k} \in I\!R^n$ and*

$$r_k^T Z_k q_{k-i,k} = 0 \tag{25}$$

*is satisfied for $i = 1, \ldots, \sigma_k$. $Z_k$ are auxiliary, non-singular matrices.*

*The method is called*

- *exact if $\sigma_k = k$, i. e. $k$ $q_{k-i,k}$ are used for the calculation of the new iterate,*
- *restarted if $\sigma_k = (k-1) \bmod \sigma_{res} + 1$, i. e. a restart is made periodically after $\sigma_{res}$ steps,*
- *truncated if $\sigma_k = min(k, \sigma_{max})$ with $\sigma_{max}$ fixed, i. e. only $\sigma_{max}$ $q_{k-i,k}$ are used for the calculation of the new iterate,*
- *combined if the truncated method is restarted,*
  *i. e. $\sigma_k = min\left((k-1) \bmod \sigma_{res} + 1, \sigma_{max}\right)$.*

In general $\tilde{x}_k = x_{k-\sigma_k}$ or $\tilde{x}_k = x_{k-1}$. Let us introduce the abbreviations:

$$\tilde{r}_k = A\tilde{x}_k - b, \tag{26}$$

$$\tilde{e}_k = \tilde{x}_k - x. \tag{27}$$

It is quite clear that the approximations can be represented as

$$x_k = \sum_{i=1}^{\sigma_k} \gamma_{i,k} q_{k-i,k} + \tilde{x}_k, \tag{28}$$

The residuals and errors can be expressed as follows

$$r_k = \sum_{i=1}^{\sigma_k} \gamma_{i,k} A q_{k-i,k} + \tilde{r}_k, \tag{29}$$

$$e_k = \sum_{i=1}^{\sigma_k} \gamma_{i,k} q_{k-i,k} + \tilde{e}_k. \tag{30}$$

From (29) and (25) follows directly that the $\gamma_{i,k}$ can be determined from the solution of the linear system of dimension $\sigma_k$

$$\sum_{i=1}^{\sigma_k} \gamma_{i,k} q_{k-i,k}^T A^T Z_k q_{k-j,k} = -\tilde{r}_k^T Z_k q_{k-j,k}, \tag{31}$$

for $j = 1, \ldots, \sigma_k$.

As the choice of the $q_{k-i,k}$ is arbitrary the definition covers in fact all commonly used iterative methods, see table 1 and figure 1. The methods distinguish by $q_{k-i,k}$, $Z_k$, $\sigma_k$ and $\tilde{x}_k$ and can be easily classified by indicating these quantities.

**2.1. Matrix splittings.** Let us consider the matrix splitting methods investigated in section 1.2, see (13),

$$\begin{aligned} r_k &= \left(I - AB^{-1}\right) r_{k-1}, \\ \tilde{r}_k &= r_{k-1}, \\ q_{k-1,k} &= -B^{-1} r_{k-1}. \end{aligned}$$

Let $S$ be any skew-symmetric matrix, then

$$Z_k = S(B - A) \tag{32}$$

fulfils the condition for orthogonalization methods:

$$\begin{aligned} r_k^T Z_k q_{k-1,k} &= -r_{k-1}^T \left(I - AB^{-1}\right)^T S(B - A) B^{-1} r_{k-1} \\ &= -r_{k-1}^T \left(I - AB^{-1}\right)^T S \left(I - AB^{-1}\right) r_{k-1} \\ &= 0, \end{aligned}$$

because $S$ is skew-symmetric.

**2.2. Generalized cg and cg applied to the normal equations.** The concept covers generalized cg methods by setting

$$q_{k-i,k} = r_{k-i} \tag{33}$$

and $Z_k = Z = const$. Of course the $q_{k-i,k}$ can be transformed to

$$q_{k-i,k} = A^{k-i} r_0, \tag{34}$$

so that it is obvious that a Krylov space is spanned. The concept describes in particular the steepest descent [75], classical cg [42], ORTHORES [91] and GMRES [61]. For the biconjugate gradients (BCG) [23, 47] the double system

$$\begin{pmatrix} A & 0 \\ 0 & A^T \end{pmatrix} \begin{pmatrix} x \\ x^* \end{pmatrix} = \begin{pmatrix} b \\ b^* \end{pmatrix}$$

is considered, where $b^*$ is arbitrary. The doubling has the advantage that the exact method can be formulated with a short recurrence.

If we apply classical cg to the normal equations we get the two methods CGNR [42] and CGNE [11, 42]. Both techniques can be implemented with a short recurrence. CGNR minimizes the residuals, while CGNE minimizes the errors.

**2.3. Conjugate Krylov subspace methods.** Conjugate Krylov subspace methods [83] are a very important subset of orthogonalization methods. The difference from generalized cg methods is that the matrix $Z_k$ varies in each iteration step. The R3-CKS method based on a rank-3 update of $Z_k$ was introduced in [83]. By this technique R3-CKS is exact and can be formulated as short recurrence at the same time. All other here discussed conjugate Krylov subspace methods have the same two desirable properties as well. CGS [70], minimal residual smoothing applied to the biconjugate gradients (MRS) [62, 63, 93] and BiCGSTAB approaches [36, 69, 79] fit into this class. Their classification is incomplete in table 1 because the determination of $Z_k$ is rather complex and would be beyond the scope of this paper. In particular, from the investigations of Barth and Manteuffel, see [6], follows for the QMR method by Freund and Nachtigal [28, 29]

$$Z_k = W_k \Omega_k^T \Omega_k W_k^T A. \tag{35}$$

Here, $W_k$ is the matrix consisting of the shadow vectors of the Lanczos process and $\Omega_k$ is a diagonal scaling matrix. The same holds for the generalizations of Tong [76] where $\Omega_k$ is an upper triangular matrix.

**2.4. Generalized minimum error methods.** Generalized minimum error methods [82] like Fridman's method [31] and the method proposed in [82], here denoted by GMERR, are orthogonalization methods.

## 3. Convergence and geometric properties.

The next lemma is the basis for all convergence investigations of orthogonalization methods.

LEMMA 2. *For any orthogonalization method*

$$r_k^T Z_k A^{-1} r_k = r_k^T Z_k A^{-1} \left( \sum_{i=1}^{\sigma_k} \eta_i A q_{k-i,k} + \tilde{r}_k \right) \tag{36}$$

| family of methods<br>particular method | $q_{k-i,k}$ | $Z_k$ | $\sigma_k$<br>type[1)] | $\tilde{x}_k$ |
|---|---|---|---|---|
| matrix splittings | $-B^{-1}r_{k-i}$ | as in (32) | 1 | $x_{k-1}$ |
| Jacobi | $-D^{-1}r_{k-i}$ | as in (32) with $B=D$ | 1 t | $x_{k-1}$ |
| Gauss-Seidel | $-(D+L)^{-1}r_{k-i}$ | as in (32) with $B=D+L$ | 1 t | $x_{k-1}$ |
| SOR | $-(D+\frac{1}{\omega}L)^{-1}r_{k-i}$ | as in (32) with $B=D+\frac{1}{\omega}L$ | 1 t | $x_{k-1}$ |
| generalized cg | $r_{k-i}$ | $Z$ | arbitrary | $x_{k-\sigma_k}$ |
| steepest descent | $r_{k-i}$ | $I$ | 1 t | $x_{k-1}$ |
| classical cg | $r_{k-i}$ | $I$ | 2 t[2)] | $x_{k-2}$ |
| ORTHORES | $r_{k-i}$ | $I$ | $k$ e | $x_0$ |
| GMRES | $r_{k-i}$ | $A$ | $k$ e | $x_0$ |
| BCG | $\begin{pmatrix} r_{k-i} \\ r^*_{k-i} \end{pmatrix}$ | $\begin{pmatrix} 0 & I \\ I & 0 \end{pmatrix}$ | $k$ s | $\begin{pmatrix} x_0 \\ x^*_0 \end{pmatrix}$ |
| cg applied to the normal equations | $A^T r_{k-i}$ | $Z$ | arbitrary | $x_{k-\sigma_k}$ |
| CGNR | $A^T r_{k-i}$ | $A$ | $k$ s | $x_0$ |
| CGNE | $A^T r_{k-i}$ | $A^{-T}$ | $k$ s | $x_0$ |
| conjugate Krylov subspace methods | $r_{k-i}$ | arbitrary | arbitrary | $x_{k-\sigma_k}$ |
| QMR | $r_{k-i}$ | $W_k\Omega_k^T\Omega_k W_k^T A$ | $k$ s | $x_0$ |
| R3-CKS | $r_{k-i}$ | $I+C_kD_k^T$<br>$C_k, D_k \in \mathbb{R}^{n\times 3}$ | $k$ s | $x_0$ |
| CGS | $r_{k-i}$ | 3) | $k$ s | $x_0$ |
| BiCGSTAB | $r_{k-i}$ | 3) | $k$ s | $x_0$ |
| MRS | $r_{k-i}$ | 3) | $k$ s | $x_0$ |
| generalized minimum error methods | $(A^T)^{k-i+1} y_0$<br>$y_0$ arbitrary | $A^{-T}$ | arbitrary | $x_{k-\sigma_k}$ |
| Fridman's method | $(A^T)^{k-i+1} r_0$ | $A^{-T}$ | 2 t[2)] | $x_{k-2}$ |
| GMERR | $(A^T)^{k-i+1} r_0$ | $A^{-T}$ | $k$ e | $x_0$ |
| Krylov subspace methods | $B^{k-i}z$<br>$B$, $z$ arbitrary | arbitrary | arbitrary | arbitrary |
| all methods in generalized cg, conjugate Krylov subspace and generalized minimum error methods | | | | |

TABLE 1

*Orthogonalization methods,* [1)] *e=exact, t=truncated, s=exact method that can be formulated with a short recurrence,* [2)] *exact for A symmetric, truncated else,* [3)] *see section 2.3* *is satisfied with $\eta_i$ arbitrary for $i = 1, \ldots, \sigma_k$. In particular*

$$r_k^T Z_k A^{-1}(r_k - \tilde{r}_k) = 0. \tag{37}$$

*Proof.*

$$\begin{aligned} r_k^T Z_k A^{-1} r_k &= r_k^T Z_k A^{-1} \left( \sum_{i=1}^{\sigma_k} \gamma_{i,k} A q_{k-i,k} + \tilde{r}_k \right) \\ &= r_k^T Z_k A^{-1} \left( \sum_{i=1}^{\sigma_k} \eta_i A q_{k-i,k} + \tilde{r}_k \right) \end{aligned}$$

with $\eta_i$ arbitrary for $i = 1, \ldots, \sigma_k$ because of the $Z_k$-orthogonalities (25). Thus (36) is proved. (37) follows from (36) directly by choosing all $\eta_i = 0$. $\square$

LEMMA 3. *If $M, R \in \mathbb{R}^{n\times n}$, $M$ is symmetric and positive definite and $R$ is skew-symmetric, then*

$$\|(I + M^{-1}R)\|_M^2 \leq 1 + \frac{\rho^2(R)}{\mu_m^2}, \tag{38}$$

*$\rho(R)$ is the spectral radius of $R$. $\mu_m$ is the minimum eigenvalue of $M$.*

*Proof.*

$$\begin{aligned} \|(I + M^{-1}R)\|_M^2 &= \max_{x\neq 0} \frac{x^T(I + R^T M^{-1})M(I + M^{-1}R)x}{x^T M x} \\ &= \max_{x\neq 0} \left(1 + \frac{x^T R^T M^{-1} R x}{x^T M x}\right) \\ & \quad as\ R\ is\ skew - symmetric \\ &\leq \max_{x\neq 0} \left(1 + \frac{\|M^{-1}\|\ \|Rx\|^2}{\mu_m \|x\|^2}\right) \\ &\leq 1 + \frac{\|R\|^2}{\mu_m^2} \\ &= 1 + \frac{\rho^2(R)}{\mu_m^2}. \end{aligned}$$

As R is a skew-symmetric matrix, R is normal, and $\|R\| = \rho(R)$. $\square$

THEOREM 4. *If $Z_k A^{-1}$ is positive real, i. e. the symmetric part of $Z_k A^{-1}$ is positive definite, then*

$$\|r_k\|_{Z_k A^{-1}} \leq \sqrt{1 + \frac{\rho^2(R)}{\mu_m^2}} \min_{\eta_1, \ldots, \eta_{\sigma_k}} \left\| \sum_{i=1}^{\sigma_k} \eta_i A q_{k-i,k} + \tilde{r}_k \right\|_{Z_k A^{-1}} \tag{39}$$

$$\|e_k\|_{A^T Z_k} \leq \sqrt{1 + \frac{\rho^2(R)}{\mu_m^2}} \min_{\eta_1, \ldots, \eta_{\sigma_k}} \left\| \sum_{i=1}^{\sigma_k} \eta_i q_{k-i,k} + \tilde{e}_k \right\|_{A^T Z_k} \tag{40}$$

*holds for orthogonalization methods. $\rho(R)$ is the spectral radius of the skew-symmetric part $R$ of $Z_kA^{-1}$. $\mu_m$ is the minimum eigenvalue of $M$, the symmetric part of $Z_kA^{-1}$. In particular if $Z_kA^{-1}$ is symmetric, then*

$$\|r_k\|_{Z_kA^{-1}} = \min_{\eta_1,\ldots,\eta_{\sigma_k}} \left\| \sum_{i=1}^{\sigma_k} \eta_i A q_{k-i,k} + \tilde{r}_k \right\|_{Z_kA^{-1}}, \tag{41}$$

$$\|e_k\|_{A^TZ_k} = \min_{\eta_1,\ldots,\eta_{\sigma_k}} \left\| \sum_{i=1}^{\sigma_k} \eta_i q_{k-i,k} + \tilde{e}_k \right\|_{A^TZ_k}. \tag{42}$$

*Proof.* By Lemma 2, it follows with arbitrary $\eta_i$:

$$\begin{aligned} \|r_k\|^2_{Z_kA^{-1}} &= r_k^T Z_k A^{-1} \left( \sum_{i=1}^{\sigma_k} \eta_i A q_{k-i,k} + \tilde{r}_k \right) \\ &= r_k^T M(I + M^{-1}R) \left( \sum_{i=1}^{\sigma_k} \eta_i A q_{k-i,k} + \tilde{r}_k \right). \end{aligned}$$

Let be $\hat{r}_r$ the vector with

$$\|\hat{r}_k\|_{Z_kA^{-1}} = \min_{\eta_1,\ldots,\eta_{\sigma_k}} \left\| \sum_{i=1}^{\sigma_k} \eta_i A q_{k-i,k} + \tilde{r}_k \right\|_{Z_kA^{-1}},$$

then

$$\begin{aligned} \|r_k\|^2_{Z_kA^{-1}} &= r_k^T M(I + M^{-1}R)\hat{r}_k \\ &\leq \|r_k\|_M \ \|(I + M^{-1}R)\hat{r}_k\|_M. \\ &\leq \|r_k\|_M \ \|(I + M^{-1}R)\|_M \ \|\hat{r}_k\|_M \\ &\leq \|r_k\|_M \sqrt{1 + \frac{\rho^2(R)}{\mu_m^2}} \ \|\hat{r}_k\|_M \end{aligned}$$

by lemma 3. By dividing the last inequality by $\|r_k\|_M$ (39) is proved. (41) follows because $\rho(R) = 0$ if $Z_kA^{-1}$ is symmetric, i. e. the square root is equal to 1. The equality follows because $r_k$ can be represented as the item in the norm of the right-hand side following from (29). (40) and (42) follow by $r_k = Ae_k$. □

This theorem is a generalization of the convergence results for generalized cg methods [81]. A more detailed investigation is given in [71, 33, 73, 74]. From this theorem follow the convergence investigations for conjugate Krylov subspace methods [83] and generalized error-minimizing methods [82]. Moreover, for QMR [29] the minimization property of Barth and Manteuffel [6] follows directly from theorem 4

$$\begin{aligned} \|r_k\|_{W_k\Omega_k^T\Omega_kW_k^T} &= \min_{\eta_1,\ldots,\eta_k} \left\| \sum_{i=1}^{k} \eta_i A^i r_0 + r_0 \right\|_{W_k\Omega_k^T\Omega_kW_k^T} \\ &= \min_{z \in I\!R^k} \left\| \ \|r_0\|\Omega_k e_1 + \Omega_k H_k z \right\|, \end{aligned}$$

| condition | minimized item | method |
|---|---|---|
| $Z_kA^{-1}$ sym., pos. def. | $\|r_k\|_{Z_kA^{-1}}$<br>$\|e_k\|_{A^TZ_k}$ | minimum $Z_kA^{-1}$-residual |
| $Z_k = \bar{Z}_kA$ and<br>$\bar{Z}_k$ sym., pos. def. | $\|r_k\|_{\bar{Z}_k}$<br>$\|e_k\|_{A^T\bar{Z}_kA}$ | minimum $\bar{Z}_k$-residual |
| $Z_k = I$ and<br>$A$ sym., pos. def. | $\|r_k\|_{A^{-1}}$<br>$\|e_k\|_A$ | minimum energy norm |
| $Z_k = A$ | $\|r_k\|$<br>$\|e_k\|_{A^TA}$ | minimum residual |
| $Z_k = A^{-T}$ | $\|r_k\|_{A^{-T}A^{-1}}$<br>$\|e_k\|$ | minimum error |

Table 2
*Minimizing orthogonalization methods*

where $H_k$ is the tridiagonal matrix of the Lanczos process and $e_1$ is the first unit vector. The well known property of QMR,

$$\|r_k\| \quad \leq \quad \sqrt{k+1} \min_{z \in I\!R^k} \| \; \|r_0\|e_1 + H_k z \| \, ,$$

can be easily derived by means of

$$\begin{aligned} \|r_k\|^2 \quad &\leq \quad \frac{k+1}{\lambda_{max}(V_k^TV_k)}\|r_k\|^2 \leq (k+1)\lambda_{min}(W_kW_k^T)\|r_k\|^2 \\ &\leq \quad (k+1)\|W_k^Tr_k\|^2 = (k+1)\|r_k\|^2_{W_kW_k^T} \, , \end{aligned}$$

where $\Omega_k$ is omitted for simplicity and $V_k$ is the matrix consisting of the Lanczos vectors. In addition we also get convergence estimates for the classical Jacobi, Gauss-Seidel and SOR methods.

From Theorem 4 follows immediately that the exact solution of the linear system is obtained at least in the iteration step $n$ if the method is exact and the $q_{k-i,k}$ are linear independent because the whole space is spanned and the right-hand side of (40) is zero. Thus exact methods converge. The restarted version is more robust as regards the roundoff errors because only $\sigma_{res}$ of the $q_{k-i,k}$ have to be $Z_k$-orthogonal to $r_k$. But the convergence of the restarted version is not guaranteed even if $Z_kA^{-1}$ is positive real, because the right-hand side of the inequalities does not necessarily tend to zero. The same holds for truncated versions. However, for truncated versions more $q_{k-i,k}$ are used in average for minimization than for restarted versions.

A very important result of Theorem 4 is that the methods minimize different quantities if $Z_kA^{-1}$ is positive real. A method that calculates the iterates by (24), but determines the coefficients by a minimization procedure is thus a special case of orthogonalization methods. In Table 2 the different conditions for the minimized items are collected. Note that $\|e_k\|_{A^T} = \|e_k\|_A$; thus if $A$ is symmetric and positive definite, then the energy norm is minimized, i. e. the methods minimize the quadratic form $F(x_k)$; see (20).

As a result of (6) we get $\|r_k\| = \|Ae_k\| \leq \|A\| \cdot \|e_k\|$ and $\|e_k\| = \|A^{-1}r_k\| \leq$

$\|A^{-1}\| \cdot \|r_k\|$ and the following inequalities are valid:

$$\frac{1}{\kappa}\frac{\|r_k\|}{\|r_0\|} \leq \frac{\|e_k\|}{\|e_0\|} \leq \kappa\frac{\|r_k\|}{\|r_0\|} \leq \kappa^2\frac{\|e_k\|}{\|e_0\|}, \tag{43}$$

where $\kappa = \|A\| \cdot \|A^{-1}\|$ is the condition number of $A$. Iterative methods are usually controlled by the observation of the norm of the residual. In reality the error should be controlled, but the error is not available. If $\kappa$ is near to unity the norm of the relative residuals is strongly connected to the norm of the relative errors. Then the errors will be reduced if the residuals decrease sufficiently. For a large condition number the residuals can decrease while the norm of the errors does not decrease or even increases and vice-versa. For these systems iterative methods that minimize the residuals in the Euclidean norm or in the energy norm may require a large number of iterations to reduce the errors and the user may be misled to stop the iteration too early by small residuals. Therefore error-minimizing methods should be applied. The condition

$$Z_k = A^{-T} \tag{44}$$

for minimum error methods seems to be only of theoretical interest. If $A^{-T}$ and thus $A^{-1}$ would be available, there is no need to solve the system iteratively because the exact solution can be calculated directly by $x = A^{-1}b$. However, minimum error methods can be implemented without knowing $A^{-T}$ by a trick:

THEOREM 5. *Let $y_{k-i,k} \in I\!R^n$ be arbitrary for $i = 1, \ldots, \sigma_k$. If*

$$q_{k-i,k} = A^T y_{k-i,k} \tag{45}$$

*and the $\gamma_{i,k}$ are determined from the solution of the linear system*

$$\sum_{i=1}^{\sigma_k} \gamma_{i,k} q_{k-i,k}^T q_{k-j,k} = -\tilde{r}_k^T y_{k-j,k}, \tag{46}$$

*for $j = 1, \ldots, \sigma_k$, then*

$$x_k = \sum_{i=1}^{\sigma_k} \gamma_{i,k} q_{k-i,k} + \tilde{x}_k \tag{47}$$

*is equivalent to the iterate of an error-minimizing orthogonalization method, i. e.*

$$\|e_k\| = \min_{\beta_1,\ldots,\beta_{\sigma_k}} \left\| \sum_{i=1}^{\sigma_k} \beta_i q_{k-i,k} + \tilde{e}_k \right\|. \tag{48}$$

*Proof.* Following from Theorem 4 orthogonalization methods minimize the error in the Euclidean norm if $Z_k = A^{-T}$. The coefficients are determined from (31) being in this case

$$\sum_{i=1}^{\sigma_k} \gamma_{i,k} q_{k-i,k}^T A^T A^{-T} q_{k-j,k} = -\tilde{r}_k^T A^{-T} A^T y_{k-j,k}, \tag{49}$$

for $j = 1, \ldots, \sigma_k$. $\square$

Error-minimizing Krylov subspace methods were introduced in [31, 82].

The next investigations show some interesting facts concerning the geometric location of the residuals and the errors.

THEOREM 6. *The residuals $r_k$ and the errors $e_k$ of exact orthogonalization methods satisfy the following equations:*

$$\left\| r_k - \frac{\tilde{r}}{2} \right\|^2_{Z_k A^{-1}} = \frac{\|\tilde{r}\|^2_{Z_k A^{-1}}}{4}, \tag{50}$$

$$\left\| e_k - \frac{\tilde{e}}{2} \right\|^2_{A^T Z_k} = \frac{\|\tilde{e}\|^2_{A^T Z_k}}{4} \tag{51}$$

*for $j = 0, \ldots, k$ with*

$$\tilde{r} = 2 \ \left( Z_k A^{-1} + (Z_k A^{-1})^T \right)^{-1} Z_k A^{-1} \tilde{r}_k, \tag{52}$$

$$\tilde{e} = 2 \ \left( Z_k + (Z_k A^{-1})^T A \right)^{-1} Z_k \tilde{e}_k. \tag{53}$$

*In particular if $A^T Z_k$ is symmetric, then*

$$\tilde{r} = \tilde{r}_k, \tag{54}$$

$$\tilde{e} = \tilde{e}_k. \tag{55}$$

*Proof.*

$$r_k^T Z_k A^{-1} (r_k - \tilde{r}_k) = 0$$

by (37). Therefore,

$$r_k^T Z_k A^{-1} r_k - r_k^T Z_k A^{-1} \tilde{r}_k = 0.$$

From the definition (52) for $\tilde{r}$ follows:

$$r_k^T Z_k A^{-1} r_k - \frac{1}{2} r_k^T Z_k A^{-1} A Z_k^{-1} \left( Z_k A^{-1} + (Z_k A^{-1})^T \right) \tilde{r} = 0$$

and

$$\begin{aligned} r_k^T Z_k A^{-1} r_k \quad & - \quad \frac{1}{2} r_k^T \left( Z_k A^{-1} + (Z_k A^{-1})^T \right) \tilde{r} \\ & + \quad \frac{\|\tilde{r}\|^2_{Z_k A^{-1}}}{4} = \frac{\|\tilde{r}\|^2_{Z_k A^{-1}}}{4} \end{aligned}$$

equivalent to

$$\left\| r_k - \frac{\tilde{r}}{2} \right\|^2_{Z_k A^{-1}} = \frac{\|\tilde{r}\|^2_{Z_k A^{-1}}}{4}.$$

Eq. (51) follows by $r_j = A e_j$. $\square$

This theorem is a generalization of the results obtained for generalized cg methods [81] and Krylov subspace methods [83]. The equations (50) and (51) are quadratic forms and describe geometrical figures:

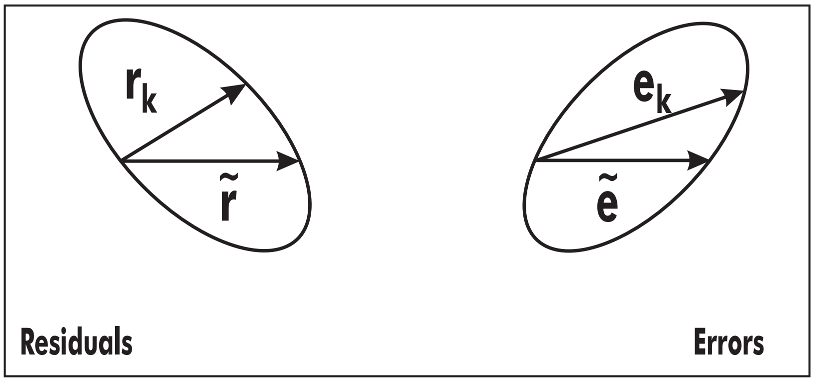


FIG. 2. *Residuals and errors if $A^T Z_k$ is definite*

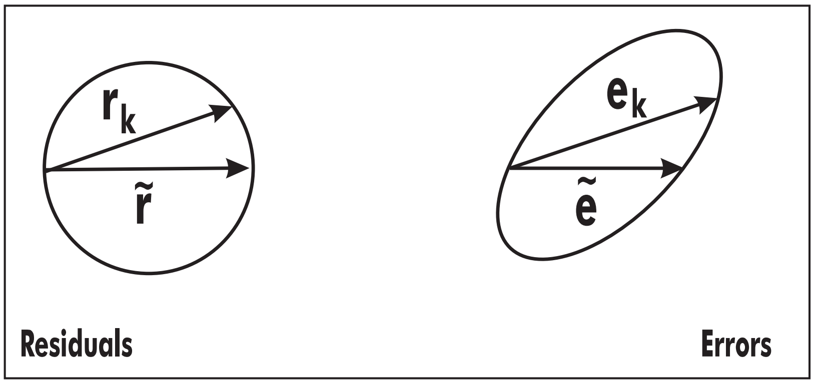


FIG. 3. *Residuals and errors if $Z_k = A$*

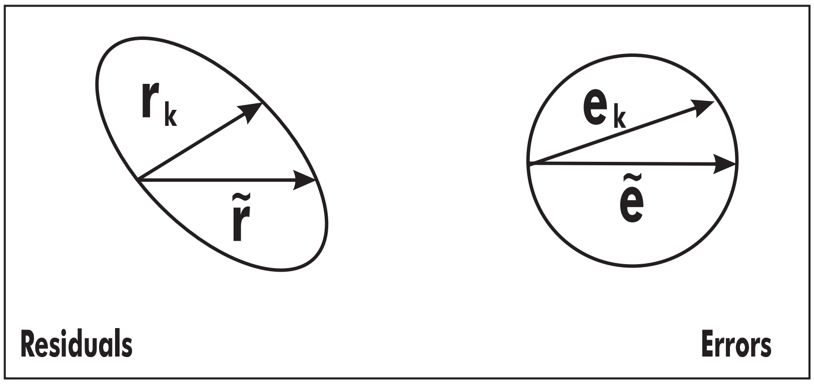


FIG. 4. *Residuals and errors if $Z_k = A^{-T}$*

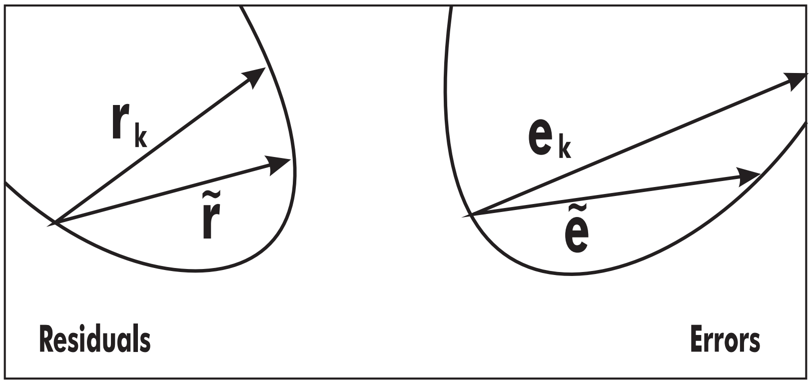


FIG. 5. *Residuals and errors if $A^T Z_k$ is indefinite*

- If $A^T Z_k$ ($Z_k A^{-1}$, respectively) is definite, then the residual $r_k$ and the error $e_k$ lie on hyperellipsoids. The lengths of the semiaxes of the hyperellipsoid for the residual are precisely the singular values of the matrix $AZ_k^{-1}$ multiplied by $\|\tilde{r}\|_{Z_k A^{-1}}$. The lengths of the semiaxes of the hyperellipsoid for the error are precisely the singular values of the matrix $Z_k^{-1} A^{-T}$ multiplied by $\|\tilde{e}\|_{A^T Z_k}$; see Fig. 2.
- If $Z_k = A$, then the residual $r_k$ lies on a n-dimensional sphere ; see Fig. 3. The norm of the residual is monotonically decreasing.
- If $Z_k = A^{-T}$, then the error $e_k$ lies on a n-dimensional sphere; see Fig. 4. The norm of the error is monotonically decreasing.
- If $A^T Z_k$ ($Z_k A^{-1}$, respectively) is indefinite, then the geometric figures are not closed and in general the method does not converge; see Fig. 5.

As a direct consequence Theorem 4 covers in fact all converging cases.

## 4. Matrix notation and short recurrences.

Let us rewrite exact orthogonalization methods in matrix form by introducing:

$$Q_k = (q_{0,k}, \ldots, q_{k-1,k}) \in I\!R^{n \times k}, \tag{56}$$
$$\gamma_k = (\gamma_{k,k}, \ldots, \gamma_{1,k})^T \in I\!R^k, \tag{57}$$

then (24), (29) and (30) are equivalent to

$$x_k = Q_k \gamma_k + \tilde{x}_k, \tag{58}$$
$$r_k = AQ_k \gamma_k + \tilde{r}_k, \tag{59}$$
$$e_k = Q_k \gamma_k + \tilde{e}_k. \tag{60}$$

(25) is equivalent to

$$r_k^T Z_k Q_k = 0, \tag{61}$$

and (31) to

$$Q_k^T Z_k^T A Q_k \gamma_k = -Q_k^T Z_k^T \tilde{r}_k. \tag{62}$$

For large systems exact methods are not feasible if they cannot be reformulated with a short recurrence because the required storage and computational effort increases with the iteration. In order to investigate the conditions for which the methods can be formulated as short recurrences we have to consider all possible transformations and formulations of the methods.

LEMMA 7. *Let $T_k \in \mathbb{R}^{k\times k}$ be a non-singular transformation matrix and*

$$U_k = Q_k T_k \in \mathbb{R}^{n\times k}, \tag{63}$$

*then the iterates can be calculated as*

$$x_k = U_k \delta_k + \tilde{x}_k., \tag{64}$$

*with $\delta_k \in \mathbb{R}^k$ and*

$$r_k^T Z_k U_k = 0, \tag{65}$$

*i. e. $\delta_k$ is the solution of*

$$U_k^T Z_k^T A U_k \delta_k = -U_k^T Z_k^T \tilde{r}_k. \tag{66}$$

*Proof.* From (58) follows

$$\begin{aligned} x_k &= Q_k \gamma_k + \tilde{x}_k \\ &= U_k T_k^{-1} \gamma_k + \tilde{x}_k. \end{aligned}$$

Set $\delta_k = T_k^{-1}\gamma_k$. From (61) follows

$$0 = r_k^T Z_k Q_k T_k = r_k^T Z_k U_k. \quad \square$$

The statement of lemma 7 is that orthogonalization methods are invariant to linear transformations of $Q_k$. But a linear tranformation may be useful to get short recurrences. The next lemma is the basis for the conditions that have to be fulfilled so that we will obtain short recurrences.

LEMMA 8. *If $B \in \mathbb{R}^{k\times k}$, $y, b \in \mathbb{R}^k$, and*

$$B = \begin{pmatrix} B_{1,1} & B_{1,2} \\ B_{2,1} & B_{2,2} \end{pmatrix}, y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}, b = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}, \tag{67}$$

*with $B_{1,1} \in \mathbb{R}^{k-l\times k-l}$ non-singular, $B_{1,2} \in \mathbb{R}^{k-l\times l}$, $B_{2,1} \in \mathbb{R}^{l\times k-l}$, $B_{2,2} \in \mathbb{R}^{l\times l}$ non-singular, $y_1 \in \mathbb{R}^{k-l}$, $y_2, b_2 \in \mathbb{R}^l$ and*

$$b_1 = B_{1,2} B_{2,2}^{-1} b_2, \tag{68}$$

*then the solution of the system*

$$By = b \tag{69}$$

*is*

$$y_1 = 0, \tag{70}$$

$$y_2 = B_{2,2}^{-1} b_2. \tag{71}$$

*Proof.* Trivial. □

In the following we will consider more general transformations than that of lemma 7. Let us assume that an orthogonalization method can be transformed to

$$x_k = U_k\delta_k + \tilde{x}_k^u, \tag{72}$$
$$r_k = AU_k\delta_k + \tilde{r}_k^u, \tag{73}$$
$$e_k = U_k\delta_k + \tilde{e}_k^u, \tag{74}$$

with

$$U_k = (u_{0,k}, \ldots, u_{k-1,k}) \in I\!R^{n\times k}, \tag{75}$$
$$\delta_k = (\delta_{k,k}, \ldots, \delta_{1,k})^T \in I\!R^k, \tag{76}$$

$\tilde{x}_k^u, \tilde{r}_k^u, \tilde{e}_k^u \in I\!R^n$. This means for $j = 0, \ldots, k-1$ that

$$u_{j,k} \in \tilde{x}_k + span(q_{0,k}, \ldots, q_{k-1,k}), \tag{77}$$
$$\tilde{x}_k^u \in \tilde{x}_k + span(q_{0,k}, \ldots, q_{k-1,k}). \tag{78}$$

Then (62) is transformed to

$$Q_k^T Z_k^T AU_k\delta_k = -Q_k^T Z_k^T \tilde{r}_k^u. \tag{79}$$

We further introduce the matrix of the first $k-l$ and last $l$ $q_{k-i,k}$:

$${}^{k-l}Q_k = (q_{0,k}, \ldots, q_{k-l-1,k}) \in I\!R^{n\times(k-l)}, \tag{80}$$
$$Q_k^l = (q_{k-l,k}, \ldots, q_{k-1,k}) \in I\!R^{n\times l}, \tag{81}$$

the matrix of the first $k-l$ and last $l$ $u_{k-i,k}$:

$${}^{k-l}U_k = (u_{0,k}, \ldots, u_{k-l-1,k}) \in I\!R^{n\times(k-l)}, \tag{82}$$
$$U_k^l = (u_{k-l,k}, \ldots, u_{k-1,k}) \in I\!R^{n\times l}, \tag{83}$$

and the vector of the first $k-l$ and last $l$ components of $\delta_k$:

$${}^{k-l}\delta_k = (\delta_{k,k}, \ldots, \delta_{l+1,k})^T \in I\!R^{k-l}, \tag{84}$$
$$\delta_k^l = (\delta_{l,k}, \ldots, \delta_{1,k})^T \in I\!R^l, \tag{85}$$

then (79) can be written as:

$$\begin{pmatrix} {}^{k-l}Q_k^T Z_k^T A\ {}^{k-l}U_k & {}^{k-l}Q_k^T Z_k^T AU_k^l \\ \left[Q_k^l\right]^T Z_k^T A\ {}^{k-l}U_k & \left[Q_k^l\right]^T Z_k^T AU_k^l \end{pmatrix} \begin{pmatrix} {}^{k-l}\delta_k \\ \delta_k^l \end{pmatrix} = -\begin{pmatrix} {}^{k-l}Q_k^T Z_k^T \tilde{r}_k^u \\ \left[Q_k^l\right]^T Z_k^T \tilde{r}_k^u \end{pmatrix}. \tag{86}$$

Let us apply lemma 8 to orthogonalization methods:

THEOREM 9. *Exact orthogonalization methods can be formulated as l-term recurrence, i. e.*

$$x_k = U_k^l\delta_k^l + \tilde{x}_k^u \tag{87}$$

*if*

$$ {}^{k-l}Q_k^T Z_k^T A U_k^l \left( \left[Q_k^l\right]^T Z_k^T A U_k^l \right)^{-1} \left[Q_k^l\right]^T Z_k^T \tilde{r}_k^u = {}^{k-l} Q_k^T Z_k^T \tilde{r}_k^u . \tag{88} $$

*Proof.* Apply lemma 8 to (86). □

As special case of theorem 9 we can formulate:

COROLLARY 10. *Exact orthogonalization methods can be formulated as l-term recurrence, i. e.*

$$ x_k = U_k^l \delta_k^l + \tilde{x}_k^u \tag{89} $$

*if*

$$ {}^{k-l}Q_k^T Z_k^T A U_k^l = 0 \tag{90} $$

*and*

$$ {}^{k-l}Q_k^T Z_k^T \tilde{r}_k^u = 0. \tag{91} $$

*Proof.* By this choice the left- and right-hand side of (88) are zero. □

The conditions of Faber and Manteuffel [22] for short recurrences of exact ORTHODIR implementations, the results of Joubert and Young [46] for the simplification of ORTHOMIN and ORTHORES implementations can be derived from corollary 10. Well-known techniques like Craig's method [11], the biconjugate gradients (BCG) [23, 47], Sonnefeld's CGS [70], Freund's and Nachtigal's QMR method [29], the biconjugate gradients smoothed by Schönauer's minimal residual algorithm [62, 63, 93] and BiCGSTAB approaches [36, 69, 79] introduced by van der Vorst can be formulated with short recurrences. This can be also obtained from corollary 10. Moreover, corollary 10 is the basis for constructing short recurrences for conjugate Krylov subspace methods [83]. The more general theorem 9 offers the possibility to search for new methods with short recurrences.

## 5. Preconditioning.

The idea of preconditioning is to consider a related linear system with the same solution as the original problem, which can be easier solved. In fact today there is no efficient linear solver without any preconditioning. The problem is that each preconditioning technique is suited for a different type of problem. Up to now we know of no robust preconditioning technique that is successfully applicable for all or at least most type of problems.

### 5.1. Constant preconditioning.

Instead of solving the linear system (1) it may be convenient to solve the preconditioned system

$$ P_L A P_R y = P_L b. \tag{92} $$

$P_L \in I\!R^{n\times n}$ is a non-singular left-hand preconditioning matrix and $P_R \in I\!R^{n\times n}$ is a non-singular right-hand preconditioning matrix. The solution of (1) is obtained from the solution of (92) by

$$x = P_R y. \tag{93}$$

The left-hand preconditioning matrix $P_L$ acts on the equations, while the right-hand preconditioning matrix $P_L$ acts on the solution.

The optimal choice of $P_L, P_R$ would be

$$P_R P_L = A^{-1}, \tag{94}$$

because then the solution is obtained directly by

$$y = P_L b, \tag{95}$$
$$x = P_R P_L b = A^{-1} b. \tag{96}$$

Of course the solution of (94) is a more difficult task than the solution of (1). But all preconditioning techniques try to choose $P_L$ and $P_R$ so that $P_L P_R$ is close to $A^{-1}$ or that $P_L A P_R$ is closer to the unit matrix than the matrix $A$ itself.

An iterative method applied to the preconditioned system (92) produces iterates $y_k$, residuals $r_k^P$ and errors $e_k^P$:

$$y_k = y_{k-1} + t_k, \tag{97}$$
$$r_k^P = r_{k-1}^P + P_L A P_R t_k, \tag{98}$$
$$e_k^P = e_{k-1}^P + t_k, \tag{99}$$

where

$$t_k = f_k(y_0, \ldots, y_{k-1}, P_L A P_R, P_L b) - y_{k-1} \tag{100}$$

is the update direction of the preconditioned system.

The residuals and errors of the preconditioned system are connected by

$$r_k^P = P_L A P_R e_k^P. \tag{101}$$

Therefore

$$\frac{1}{\kappa_P}\frac{||r_k^P||}{||r_0^P||} \le \frac{||e_k^P||}{||e_0^P||} \le \kappa_P \frac{||r_k^P||}{||r_0^P||} \le \kappa_P^2 \frac{||e_k^P||}{||e_0^P||}, \tag{102}$$

where $\kappa_P = ||P_L A P_R|| \cdot ||(P_L A P_R)^{-1}||$ is the condition number of $P_L A P_R$. If

$$\kappa_P < \kappa, \tag{103}$$

then the relative errors and residuals of the preconditioned system are closer connected than those of the original system; see (43).

An iterative method for the solution of (92) induces an iterative method for the solution of (1) by

$$x_k = P_R y_k. \tag{104}$$

For the residuals of the preconditioned system

$$r_k^P = P_L A P_R y_k - P_L b = P_L r_k \tag{105}$$

is satisfied and for the errors the equation is valid:

$$e_k^P = y_k - y = P_R^{-1} e_k. \tag{106}$$

The sequence defined by (97), (98) and (99) can be formulated in terms of the original system, setting $d_k = P_R t_k$, by

$$\begin{aligned}
x_k &= P_R y_k = P_R(y_{k-1} + t_k) && (107)\\
&= x_{k-1} + d_k, \\
r_k &= P_L^{-1} r_k^P = P_L^{-1}(r_{k-1}^P + P_L A P_R t_k) = r_{k-1} + A P_R t_k && (108)\\
&= r_{k-1} + A d_k, \\
e_k &= P_R e_k^P = P_R(e_{k-1}^P + t_k) && (109)\\
&= e_{k-1} + d_k.
\end{aligned}$$

Thus a preconditioned iterative method only influences the update direction

$$\begin{aligned}
d_k &= P_R t_k && (110)\\
&= P_R \left( f_k(y_0, \ldots, y_{k-1}, P_L A P_R, P_L b) - y_{k-1} \right) \\
&= P_R f_k(P_R^{-1} x_0, \ldots, P_R^{-1} x_{k-1}, P_L A P_R, P_L b) - x_{k-1}.
\end{aligned}$$

We have a non-symmetry between left-hand and right-hand preconditioners: If left-hand preconditioners are used, a different linear system is solved in fact, but the preconditioning influences the formulas of the solution algorithm only so that $A$ has to be substituted by $P_L A$ and $b$ by $P_L b$. If right-hand preconditioning is used the preconditioning must be included in the algorithm if it is reformulated to the original system.

As stated in (94) the optimal preconditioning would fulfil $P_R = A^{-1} P_L^{-1}$. Therefore the optimal update direction would be

$$d_k = P_R t_k = A^{-1} P_L^{-1} t_k,$$

i. e. the solution of the linear system

$$P_L A d_k = t_k. \tag{111}$$

As direct consequence the matrix $P_R$ is not needed explicitly, it is sufficient to determine $d_k$ so that the solution of (111) is approximated.

### 5.2. Overview of constant preconditioners.

Preconditioning from the left-hand side is as usual as preconditioning from the right-hand side. For both techniques the same construction principles for the preconditioning matrices are used. In practice all three cases apply:

- preconditioning by $P_L$, while $P_R = I$,
- preconditioning by $P_R$, while $P_L = I$ and

- preconditioning by both: $P_L$ and $P_R$.

For right-hand preconditioning the preconditioning matrix $P_R$ must not be explicitly calculated. It is sufficient to know the update direction $d_k = P_R t_k$ that can be determined by approximating (111). Techniques that are based on this proceeding are the polynomial preconditioners [1, 2, 24, 32, 45, 57], the incomplete LU decomposition, see Wittum and Liebau [87] for theoretical investigations and [88] for efficient implementations on vector computers, and the incomplete Cholesky. The incomplete LU algorithm (ILU) has been modified in order to achieve more efficiency to MILU (modified ILU) by Gustafsson [35], to RILU (relaxed ILU) by Axelsson and Lindskog [4] and to ILUT (ILU threshold) [59] and ILUM (ILU multi-elimination) [58] by Saad. If $A$ is symmetric and positive definite, an incomplete Cholesky decomposition can be applied instead of the LU-decomposition. The advantage is that this algorithm is more stable and needs only half of the operations of ILU. The algorithm was investigated by Meijerling and van der Vorst [48, 77, 78]. Modifications have been investigated by Beauwens and Wilmet [7] and by Notay [51, 52].

If $P_R$ or $P_L$ is explicitly calculated it is considered to be an incomplete inverse for $A$. These techniques have the advantage that during the iteration there is only one additional matrix-vector multiplication per iteration step that can be optimally vectorized and parallelized, see e. g. [41, 62, 84]. Normalization, incomplete Gauss-Jordan, incomplete$^2$ Gauss-Jordan, incomplete Frobenius [34, 68, 10] and projection methods, a generalization of incomplete Frobenius preconditioning, are based on this approach. There are investigations on the structural properties of band matrices [12, 56] and their inverses leading to methods that prescribe a certain pattern for $P_R$. In order to get an overview, see also [3].

Parameter-dependent preconditioners can be constructed by choosing $P_R$ or $P_L$ as Chebyshev polynomial [20, 21] on an interval enclosing the eigenvalues if $A$ is symmetric. Eiermann, Li and Varga give an encompassing theoretical analysis [15] for more general cases.

### 5.3. Step-dependent preconditioning.

In analogy to the previous classical preconditioning technique we define a step-dependent preconditioning. The idea is to improve the preconditioning matrices in each iteration step together with the approximations. We consider the sequence of linear systems

$$P_{L,k} A P_{R,k} y^k = P_{L,k} b. \tag{112}$$

$P_{L,k} \in I\!R^{n\times n}$ are non-singular left-hand, $P_{R,k} \in I\!R^{n\times n}$ are non-singular right-hand preconditioning matrices. The solution of (1) is obtained from each system of (112) by

$$x = P_{R,k} y^k. \tag{113}$$

The solution of the system in the step $k$ can be obtained from the solution of the system in the step $k-1$ by

$$y^k = P_{R,k}^{-1} P_{R,k-1} y^{k-1}. \tag{114}$$

An iterative method applied to the systems (112) produces a sequence of approximations $y_k$ with

$$y_k \rightarrow y^k. \tag{115}$$

The iterates are defined by

$$y_k = P_{R,k}^{-1} P_{R,k-1} y_{k-1} + t_k, \tag{116}$$

with the update direction

$$\begin{aligned} t_k = {} & f_k(P_{R,k}^{-1} P_{R,0} y_0, \ldots, P_{R,k}^{-1} P_{R,k-1} y_{k-1}, P_{L,k} A P_{R,k}, P_{L,k} b) \\ & - P_{R,k}^{-1} P_{R,k-1} y_{k-1}. \end{aligned} \tag{117}$$

Each $y_k$ is an approximation for the solution of a different linear system. Let

$$r_k^{P_k} = P_{L,k} A P_{R,k} y_k - P_{L,k} b$$

be the residual for the approximation $y_k$ for the linear system of the step $k$,

$$r_{k-1}^{P_{k-1}} = P_{L,k-1} A P_{R,k-1} y_{k-1} - P_{L,k-1} b$$

be the residual for the approximation $y_{k-1}$ for the linear system of the step $k-1$, then

$$\begin{aligned} r_k^{P_k} &= P_{L,k} A P_{R,k} y_k - P_{L,k} b \\ &= P_{L,k} A P_{R,k-1} y_{k-1} - P_{L,k} b + P_{L,k} A P_{R,k} t_k \\ & \quad by\ (116) \\ &= P_{L,k} P_{L,k-1}^{-1} r_{k-1}^{P_{k-1}} + P_{L,k} A P_{R,k} t_k. \end{aligned} \tag{118}$$

Let

$$e_k^{P_k} = y_k - y^k$$

be the error for the approximation $y_k$ for the linear system of the step $k$,

$$e_{k-1}^{P_{k-1}} = y_{k-1} - y^{k-1}$$

be the error for the approximation $y_{k-1}$ for the linear system of the step $k-1$, then

$$\begin{aligned} e_k^{P_k} &= y_k - y^k \\ &= P_{R,k}^{-1} P_{R,k-1}(y_{k-1} - y^{k-1}) + t_k \\ & \quad by\ (116)\ and\ (114) \\ &= P_{R,k}^{-1} P_{R,k-1} e_{k-1}^{P_{k-1}} + t_k. \end{aligned} \tag{119}$$

An iterative method for the original system is induced by

$$x_k = P_{R,k} y_k. \tag{120}$$

The sequence can be formulated in terms of the original system with the update direction

$$
\begin{aligned}
(121)\, d_k &= P_{R,k} t_k \\
&= P_{R,k} f_k(P_{R,k}^{-1} P_{R,0} y_0, \ldots, P_{R,k}^{-1} P_{R,k-1} y_{k-1}, P_{L,k} A P_{R,k}, P_{L,k} b) - x_{k-1} \\
&= P_{R,k} f_k(P_{R,k}^{-1} x_0, \ldots, P_{R,k}^{-1} x_{k-1}, P_{L,k} A P_{R,k}, P_{L,k} b) - x_{k-1}
\end{aligned}
$$

Note the equivalence of (121) with (110). Thus

$$
\begin{aligned}
(122) \qquad x_k &= P_{R,k} y_k \\
&= P_{R,k}(P_{R,k}^{-1} P_{R,k-1} y_{k-1} + t_k) \\
&= P_{R,k-1} y_{k-1} + d_k \\
&= x_{k-1} + d_k, \\
(123) \qquad r_k &= P_{L,k}^{-1} r_k^P \\
&= P_{L,k}^{-1} P_{L,k} P_{L,k-1}^{-1} r_{k-1}^{P_{k-1}} + A P_{R,k} t_k \\
&= P_{L,k-1}^{-1} r_{k-1}^{P_{k-1}} + A P_{R,k} t_k \\
&= r_{k-1} + A d_k, \\
(124) \qquad e_k &= P_{R,k} e_k^P \\
&= P_{R,k-1} e_{k-1}^{P-1} + d_k \\
&= e_{k-1} + d_k.
\end{aligned}
$$

As for constant preconditioners follows directly that the optimal update direction is the solution of the linear system

$$
P_{L,k} A d_k = t_k. \tag{125}
$$

Thus the matrices $P_{R,k}$ are not needed explicitly, it is sufficient to determine $d_k$ so that the solution of (125) is approximated.

**5.4. Overview of step-dependent preconditioners.**
As it is sufficient to determine the update direction from (125) any iterative method can be used as preconditioner. It is also possible to precondition the preconditioner and so on, so that we arrive at a cascade of iterative methods. For practical experience with this type of preconditioning see [60, 64, 80].

Multilevel methods have become a powerful tool for the solution of partial differential equations. The technique can be used as preconditioner. For details we refer to [5, 8, 9, 13, 14, 38, 39, 40, 53]. A special multilevel technique tuned for preconditioning cg methods is DARE [65, 66, 85, 86].

**5.5. Preconditioned orthogonalization methods.**
In this section we focus on step-dependent preconditioners because constant preconditioners are obtained as special case. A preconditioned orthogonalization method calculates the iterates according to

$$
y_k = \sum_{i=1}^{\sigma_k} \gamma_{i,k} q_{k-i,k}^P + \tilde{y}_k \tag{126}
$$

with $q^P_{k-i,k} \in I\!R^n$. The $\gamma_{i,k}$ are determined from

$$\left(r_k^P\right)^T Z_k q^P_{k-i,k} = 0 \tag{127}$$

for $i = 1, \ldots, \sigma_k$, i. e. from the solution of the linear system

$$\sum_{i=1}^{\sigma_k} \gamma_{i,k} \left(q^P_{k-i,k}\right)^T P^T_{R,k} A^T P^T_{L,k} Z_k q^P_{k-j,k} = -\left(\tilde{r}_k^P\right)^T Z_k q^P_{k-j,k}, \tag{128}$$

for $j = 1, \ldots, \sigma_k$ with

$$\tilde{r}_k = P_{L,k} A P_{R,k} \tilde{y}_k - P_{L,k} b. \tag{129}$$

As stated in the chapter on preconditioning methods they can be reformulated to the original system. From (122) follows

$$\begin{aligned} x_k &= \sum_{i=1}^{\sigma_k} \gamma_{i,k} P_{R,k} q^P_{k-i,k} + P_{R,k} \tilde{y}_k \\ &= \sum_{i=1}^{\sigma_k} \gamma_{i,k} P_{R,k} q^P_{k-i,k} + \tilde{x}_k. \end{aligned} \tag{130}$$

From Theorem 4 follows:

Corollary 11. *If $P_{L,k} A P_{R,k} Z_k^{-1}$ is positive real, i. e. the symmetric part of $P_{L,k} A P_{R,k} Z_k^{-1}$ is positive definite, then*

$$\|r_k\|_{P^T_{L,k} Z_k P^{-1}_{R,k} A^{-1}} \leq \sqrt{1 + \frac{\rho^2(R)}{\mu_m^2}} \min_{\eta_1, \ldots, \eta_{\sigma_k}} \left\| \sum_{i=1}^{\sigma_k} \eta_i A P_{R,k} q^P_{k-i,k} + \tilde{r}_k \right\|_{P^T_{L,k} Z_k P^{-1}_{R,k} A^{-1}}, \tag{131}$$

$$\|e_k\|_{A^T P^T_{L,k} Z_k P^{-1}_{R,k}} \leq \sqrt{1 + \frac{\rho^2(R)}{\mu_m^2}} \min_{\eta_1, \ldots, \eta_{\sigma_k}} \left\| \sum_{i=1}^{\sigma_k} \eta_i P_{R,k} q^P_{k-i,k} + \tilde{e}_k \right\|_{A^T P^T_{L,k} Z_k P^{-1}_{R,k}} \tag{132}$$

*holds for orthogonalization methods. $\rho(R)$ is the spectral radius of the skew-symmetric part $R$ of $Z_k P^{-1}_{R,k} A^{-1} P^{-1}_{L,k}$. $\mu_m$ is the minimum eigenvalue of $M$, the symmetric part of $Z_k P^{-1}_{R,k} A^{-1} P^{-1}_{L,k}$. In particular if $Z_k P^{-1}_{R,k} A^{-1} P^{-1}_{L,k}$ is symmetric, then*

$$\|r_k\|_{P^T_{L,k} Z_k P^{-1}_{R,k} A^{-1}} = \min_{\eta_1, \ldots, \eta_{\sigma_k}} \left\| \sum_{i=1}^{\sigma_k} \eta_i A P_{R,k} q^P_{k-i,k} + \tilde{r}_k \right\|_{P^T_{L,k} Z_k P^{-1}_{R,k} A^{-1}}, \tag{133}$$

$$\|e_k\|_{A^T P^T_{L,k} Z_k P^{-1}_{R,k}} = \min_{\eta_1, \ldots, \eta_{\sigma_k}} \left\| \sum_{i=1}^{\sigma_k} \eta_i P_{R,k} q^P_{k-i,k} + \tilde{e}_k \right\|_{A^T P^T_{L,k} Z_k P^{-1}_{R,k}}. \tag{134}$$

*Proof.* Apply (123) and (124) to Theorem 4. □

Two particular cases resulting from corollary 11 are of special interest:

- The residuals are minimized in the Euclidean norm if

$$P_{L,k}^T Z_k = A P_{R,k}. \tag{135}$$

  This can be realized by different choices of $P_{L,k}$, $P_{R,k}$ and $Z_k$. If $Z_k$ is determined from (135) for fixed preconditioning matrices, then $P_{L,k}$ has to be the unit matrix to avoid the inversion and to get a feasible method resulting in

$$Z_k = A P_{R,k}. \tag{136}$$

  For the residuals

$$\|r_k\| = \min_{\eta_1,\dots,\eta_{\sigma_k}} \left\| \sum_{i=1}^{\sigma_k} \eta_i A P_{R,k} q_{k-i,k}^P + \tilde{r}_k \right\| \tag{137}$$

  is satisfied.
- The errors are minimized in the Euclidean norm if

$$P_{R,k} = A^T P_{L,k}^T Z_k. \tag{138}$$

  This can be realized by choosing $P_{R,k}$ accordingly. For the errors

$$\|e_k\| = \min_{\eta_1,\dots,\eta_{\sigma_k}} \left\| \sum_{i=1}^{\sigma_k} \eta_i A^T P_{L,k} Z_k q_{k-i,k}^P + \tilde{e}_k \right\| \tag{139}$$

  is satisfied. For fixed preconditioning matrices (138) seems to be unfeasible without matrix inversion. However, from Theorem 5 follows that for $P_{R,k} = I$ we can realize

$$Z_k = P_{L,k}^{-T} A^{-T}, \tag{140}$$

  thus fulfilling (138) without matrix inversion, by introducing $y_{k-i,k} \in I\!R^n$. If

$$q_{k-i,k} = A^T P_{L,k}^T y_{k-i,k}, \tag{141}$$

  then the coefficients $\gamma_{i,k}$ can be calculated by the solution of the linear system

$$\sum_{i=1}^{\sigma_k} \gamma_{i,k} \left(q_{k-i,k}^P\right)^T q_{k-j,k}^P = -\left(\tilde{r}_k^P\right)^T y_{k-j,k}^P, \tag{142}$$

  for $j = 1, \dots, \sigma_k$. For the errors

$$\|e_k\| = \min_{\eta_1,\dots,\eta_{\sigma_k}} \left\| \sum_{i=1}^{\sigma_k} \eta_i q_{k-i,k}^P + \tilde{e}_k \right\| \tag{143}$$

| minimum residual | | | |
|---|---|---|---|
| in general | $P_{L,k}Z_k = AP_{R,k}$ | | |
| in particular | $Z_k = AP_{R,k}$ | $P_{L,k} = I$ | |
| minimum error | | | |
| in general | $P_{R,k} = A^T P_{L,k}^T Z_k$ | | |
| in particular | $Z_k = P_{L,k}^{-T} A^{-T}$ | $P_{R,k} = I$ | $q_{k-i,k} = A^T P_{L,k} y_{k-i,k}$ |

TABLE 3

*Residual- and error-minimizing preconditioned orthogonalization methods*

is satisfied. Note that the two different choices for error-minimizing methods lead to different methods as can be seen by comparing (139) with (143). From (139) results CGNE [11], while from (143) results Fridman's method [31] or GMERR [55, 82].

The related conditions for residual- and error-minimizing methods are listed in Table 3.

From Theorem 6 follows for the geometric properties of preconditioned methods:

COROLLARY 12. *The residuals $r_k$ and the errors $e_k$ of preconditioned exact orthogonalization methods satisfy the following equations:*

$$\|r_k - \frac{\tilde{r}}{2}\|^2_{P_{L,k}^T Z_k P_{R,k}^{-1} A^{-1}} = \frac{\|\tilde{r}\|^2_{P_{L,k}^T Z_k P_{R,k}^{-1} A^{-1}}}{4}, \tag{144}$$

$$\|e_k - \frac{\tilde{e}}{2}\|^2_{A^T P_{L,k}^T Z_k P_{R,k}^{-1}} = \frac{\|\tilde{e}\|^2_{A^T P_{L,k}^T Z_k P_{R,k}^{-1}}}{4} \tag{145}$$

*for $j = 0, \ldots, k$ with*

$$\tilde{r} = 2\ \left(Z_k P_{R,k}^{-1} A^{-1} + (Z_k P_{R,k}^{-1} A^{-1} P_{L,k}^{-1})^T P_{L,k}\right)^{-1} Z_k P_{R,k} A^{-1} \tilde{r}_k\ , \tag{146}$$

$$\tilde{e} = 2\ \left(Z_k P_{R,k}^{-1} + (Z_k P_{R,k}^{-1} A^{-1} P_{L,k}^{-1})^T P_{L,k} A\right)^{-1} Z_k P_{R,k} \tilde{e}_k. \tag{147}$$

*In particular if $P_{R,k}^T A^T P_{L,k}^T Z_k$ is symmetric, then*

$$\tilde{r} = \tilde{r}_k, \tag{148}$$

$$\tilde{e} = \tilde{e}_k. \tag{149}$$

*Proof.* Apply (123) and (124) to Theorem 6. □

Corollary 11 and 12 allow the interpretation that preconditioning transforms orthogonalization methods by changing the $q^P_{k-i,k}$ and the matrix $Z_k$ according to

$$q^P_{k-i,k} \quad \rightarrow \quad P_{R,k} q^P_{k-i,k}, \tag{150}$$

$$Z_k \quad \rightarrow \quad P_{L,k}^T Z_k P_{R,k}^{-1}. \tag{151}$$

If we choose a matrix $\bar{Z}$ with

$$\bar{Z} P_{R,k} = P_{L,k}^T Z_k, \tag{152}$$

then we get rid off the step-dependent norms in corollary 11, i. e.:

$$\|r_k\|_{\bar{Z}A^{-1}} \leq \sqrt{1+\frac{\rho^2(R)}{\mu_m^2}} \min_{\eta_1,\ldots,\eta_{\sigma_k}} \left\| \sum_{i=1}^{\sigma_k} \eta_i A P_{R,k} q_{k-i,k}^P + \tilde{r}_k \right\|_{\bar{Z}A^{-1}}, \tag{153}$$

$$\|e_k\|_{A^T\bar{Z}} \leq \sqrt{1+\frac{\rho^2(R)}{\mu_m^2}} \min_{\eta_1,\ldots,\eta_{\sigma_k}} \left\| \sum_{i=1}^{\sigma_k} \eta_i P_{R,k} q_{k-i,k}^P + \tilde{e}_k \right\|_{A^T\bar{Z}} \tag{154}$$

holds for exact orthogonalization methods. Moreover if we choose $Z_k$ according to (135) then we get residual-minimizing methods. If we choose $Z_k$ according to (138) then we get error-minimizing methods. Thus preconditioning can also be used to get methods that have properties that the original methods do not have. To summarize: Preconditioning was introduced

- to get a problem that is easier to solve.

However, preconditioning can be used

- to get step-independent norms,
- to get residual-minimizing methods,
- to get error-minimizing methods.

## 6. Conclusion.

The general concept of orthogonalization methods has been introduced describing many iterative methods. Convergence results, geometric properties and conditions for short recurrences have been derived. Well-known results for commonly used methods follow as special cases. Preconditioning has been explained consistently. Many recently developed methods like QMR, BiCGSTAB and minimal residual smoothing are beyond the concept of generalized cg methods. They can be easily classified by applying the concept of orthogonalization methods.

This theoretical framework offers the possibility to search for new iterative approaches where desired properties can be immediately obtained from the theoretical investigations. Moreover, a classification of the qualitative behavior of the methods and a different viewpoint of preconditioning has been introduced. The here introduced concept could be the basis for new methods. Research for efficient and robust techniques should be stimulated.

## REFERENCES


[1] S.F. Ashby. Minimax polynomial preconditioning for Hermitian linear systems. *SIAM J. Matrix Anal. Appl.*, 12(4):766–789, October 1991.

[2] S.F. Ashby, T.A. Manteuffel, and J.S. Otto. A comparison of adaptive Chebyshev and least squares polynomial preconditioning for Hermitian positive definite linear systems. Research Report UCRL-JC-106726, Law. Livermore Nat. Lab., Num. Math. Group, March 1991.

[3] O. Axelsson. A survey of preconditioned iterative methods for linear systems of algebraic equations. *BIT*, 1:166–187, 1985.

[4] O. Axelsson and G. Lindskog. On the rate of convergence of the preconditioned conjugate gradient method. *Numer. Math.*, 48:499–523, 1986.

[5] O. Axelsson and P.S. Vassilevski. A survey of multilevel preconditioned iterative methods. *BIT*, 29:769–793, 1989.

[6] T. Barth and T. Manteuffel. Estimating the spectrum of A using the roots of the polynomials associated with the QMR iteration. University of Colorado at Denver, 1994.

[7] R. Beauwens and R. Wilmet. Conditioning analysis of positive definite matrices by approximate factorizations. *J. Comput. Appl. Math.*, 26:257–269, 1989.

[8] A. Brandt. Multi-level adaptive solutions to boundary-value problems. *Mathematics of Computation*, 31(138):333–390, April 1977.

[9] W.L. Briggs. *A Multigrid Tutorial*. Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania, 1987.

[10] J.D.F. Cosgrove, J.C. Diaz, and A. Griewank. Approximate inverse preconditionings for sparse linear systems. *Intern. J. Computer Math.*, 44:91–110, 1992.

[11] E.J. Craig. The n-step iteration procedures. *Math. Phys.* , 34:64–73, 1955.

[12] S. Demko, W.F. Moss, and P.W. Smith. Decay rates for inverses of band matrices. *Mathematics of Computation*, 43(168):491–499, October 1984.

[13] J. E. Dendy, Jr. Black box multigrid for nonsymmetric problems. *Applied Mathematics and Computation*, 13:261–283, 1983.

[14] J.E. Dendy, Jr. Black box multigrid. *J. Comput. Phys.*, 48(3):366–386, December 1982.

[15] M. Eiermann, X. Li, and R.S. Varga. On hybrid semi-iterative methods. *SIAM J. Numer. Anal.*, 26(1):152–168, February 1989.

[16] M. Eiermann and W. Niethammer. On the construction of semiiterative methods. *SIAM J. Numer. Anal.*, 20(6):1153–1160, 1983.

[17] M. Eiermann, W. Niethammer, and R.S. Varga. A study of semiiterative methods for nonsymmetric systems of linear equations. *Numer. Math.*, 43:505–533, 1985.

[18] M. Eiermann, R.S. Varga, and W. Niethammer. Iterationsverfahren für nichtsymmetrische Gleichungssysteme und Approximationsmethoden im Komplexen. *Jber. d. Dt. Math.-Verein*, 89:1–32, 1987.

[19] H.C. Elman. *Iterative Methods for Large Sparse Nonsymmetric Systems of Linear Equations.* PhD thesis, Department of Computer Science, Yale University, 1982.

[20] G. Faber. Über polynomische Entwickelungen. *Math. An.*, 57:389–408, 1903.

[21] G. Faber. Über Tschebyscheffsche Polynome. *J. Reine Angew. Math.*, 150:79–106, 1920.

[22] V. Faber and T. Manteuffel. Necessary and sufficient conditions for the existence of a conjugate gradient method. *SIAM J. Numer. Anal.*, 21:352–362, 1984.

[23] R. Fletcher. Conjugate gradient methods for indefinite systems. In G.A. Watson, editor, *Proc. Dundee Biennial Conf. on Num. Anal.*, pages 73–89. Springer-Verlag, 1975.

[24] R.W. Freund. On conjugate gradient type methods and polynomial preconditioners for a class of non-Hermitian matrices. *Numer. Math.*, 57:285–312, 1990.

[25] R.W. Freund. Quasi-kernel polynomials and convergence results for quasi-minimal residual iterations. RIACS Technical Report 92/07, Research Institute for Advanced Computer Science, NASA Ames Research Center, March 1992. To appear in: Numerical Methods of Approximation Theory.

[26] R.W. Freund. A transpose-free quasi-minimal residual algorithm for non-Hermitian linear systems. *SIAM J. Sci. Comput.*, 14(2):470–482, March 1993.

[27] R.W. Freund, G.H. Golub, and N.M. Nachtigal. Iterative solution of linear systems. *Acta Numerica*, pages 1–44, 1991.

[28] R.W. Freund and M. Hochbruck. On the use of two QMR algorithms for solving singular systems and applications in Markov Chain modeling. to appear.

[29] R.W. Freund and N.M. Nachtigal. QMR: a quasi-minimal residual method for non-Hermitian linear systems. *Numer. Math.*, 60:315–339, 1991.

[30] R.W. Freund and T. Szeto. A quasi-minimal residual squared algorithm for non-hermitian linear systems. RIACS Technical Report 91/26, Research Institute for Advanced Computer Science, NASA Ames Research Center, December 1991. To appear in: Proceedings of the Copper Mountain Conference on Iterative Methods, 1992.

[31] V.M. Fridman. The method of minimum iterations with minimum errors for a system of linear algebraic equations with a symmetrical matrix. *USSR Comput. Math. and Math. Phys.* , 2:362–363, 1963.

[32] A. Greenbaum. Comparison of splittings used with the conjugate gradient algorithm. *Numer. Math.*, 33:181–194, 1979.

[33] A. Greenbaum and Z. Strakoš. Predicting the behavior of finite precision Lanczos and conjugate gradient computations. *SIAM J. Matrix Anal. Appl.*, 13(1):121–137, January 1992.

[34] M. Grote and H.D. Simon. Parallel preconditioning and approximate inverses on the connection machine. Research Report RNR-92-020, NAS Systems Division, NASA Ames Research Center, May 1992. To appear in: Proceedings of the Scalable High Performance Computer Conference 1992, IEEE Computer Society Press, 1992, pp. 76-83.

[35] I. Gustafsson. A class of first order factorization methods. *BIT*, 18(1):142–156, 1978.

[36] M.H. Gutknecht. Variants of BICGSTAB for matrices with complex spectrum. Research Report 91-14, Eidgenössische Technische Hochschule Zürich, Interdisziplinäres Projektzentrum für Supercomputing, ETH-Zentrum, CH-8092 Zürich, August 1991.

[37] M.H. Gutknecht. A completed theory of the unsymmetric Lanczos process and related algorithms, part 1. *SIAM J. Matrix. Anal.*, 13(2):594–639, April 1992.

[38] W. Hackbusch. *Multigrid Methods and Applications*. Springer-Verlag, Berlin, Heidelberg, New York, 1985.

[39] W. Hackbusch and U. Trottenberg. *Multigrid Methods, Proceedings of the Conference held at Köln-Porz, 1981*, volume 960 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, Heidelberg, New York, 1982.

[40] W. Hackbusch and U. Trottenberg, editors. *Multigrid Methods III*, volume 98 of *International Series of Numerical Mathematics*. Birkhäuser, Basel, Boston, Berlin, 1991.

[41] M. Hanke, M. Hochbruck, and W. Niethammer. Experiments with Krylov subspace methods on a massively parallel computer. Research Report 92-16, Eidgenössische Technische Hochschule Zürich, Interdisziplinäres Projektzentrum für Supercomputing, ETH-Zentrum, CH-8092 Zürich, September 1992.

[42] M.R. Hestenes and E. Stiefel. Methods of conjugate gradients for solving linear systems. *J. Res. Nat. Bur. Standards*, 49:409–435, 1952.

[43] M. Hochbruck. *Lanczos- und Krylov-Verfahren für nicht-Hermitische lineare Systeme*. PhD thesis, Universität Karlsruhe, Germany, April 1992.

[44] C.G.J. Jacobi. Über eine neue Auflösungsart der bei der Methode der kleinsten Quadrate vorkommenden linearen Gleichungen. *Astr. Nachr.*, 22(523):297–306, 1845. also in 'C.G.J. Jacobi's Gesammelte Werke, Dritter Band', p.:469–478, ed. by K. Weierstrass, Berlin, 1884.

[45] O.G. Johnson, C.A. Micchelli, and G. Paul. Polynomial preconditioners for conjugate gradient calculations. *SIAM J. Numer. Anal.*, 20(2):362–376, 1983.

[46] W.D. Joubert and D. Young. Necessary and sufficient conditions for the simplification of generalized conjugate-gradient algorithms. *Lin. Alg. and its Appl.*, 88/89:449–485, 1987.

[47] C. Lanczos. Solution of systems of linear equations by minimized iterations. *J. Res. Nat. Bur. Standards*, 49:33–53, 1952.

[48] J.A. Meijerink and H.A. van der Vorst. An iterative solution method for linear systems of which the coefficient matrix is a symmetric M-matrix. *Math. Comp.*, 31:148–162, 1977.

[49] W. Niethammer and W. Schempp. On the construction of iteration methods for linear equations in Banach spaces by summation methods. *Aequat. Math.*, 5:273–284, 1970.

[50] W. Niethammer and R.S. Varga. The analysis of k-step iterative methods for linear systems from summability theory. *Numer. Math.*, 41:177–206, 1983.

[51] Y. Notay. A dynamic version of the RIC method, July 1992. to appear.

[52] Y. Notay. Ordering methods for approximate factorization preconditioning, January 1993. to appear.

[53] C. Popa. Preconditioning for the fulfilment of the approximation assumption in the algebraic

multigrid method. Submitted to Computing J.

[54] J.K. Reid. On the method of conjugate gradients for the solution of large sparse systems of linear equations. In J.K. Reid, editor, *Large Sparse Sets of Linear Equations*, pages 231–254. Academic Press, New York, 1971.

[55] M. Rozložník. Variants of the generalized minimal error method. In *Proceedings of Mathematical Modelling and Computational Methods "Modelling 94", Prague*, 1994. submitted.

[56] P. Rózsa, R. Bevilacqua, F. Romani, and P. Favati. On band matrices and their inverses. *Linear Algebra and its Applications*, 150:287–295, 1991.

[57] Y. Saad. Practical use of polynomial preconditionings for the conjugate gradient method. Research Report YALEU/DCS/RR-282, Yale University, Department of Computer Science, August 1983.

[58] Y. Saad. ILUM: A parallel multi-elimination ILU preconditioner for general sparse matrices. Research Report UMSI 92/241, University of Minnesota, Supercomputer Institute, 1200 Washington Avenue South, Minneapolis, Minnesota 55415, November 1992.

[59] Y. Saad. ILUT: A dual threshold incomplete LU factorization. Research Report UMSI 92/38, University of Minnesota, Supercomputer Institute, 1200 Washington Avenue South, Minneapolis, Minnesota 55415, March 1992.

[60] Y. Saad. A flexible inner-outer preconditioned GMRES algorithm. *SIAM J. Sci. Comput.*, 14(2):461–469, March 1993.

[61] Y. Saad and M.H. Schultz. GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems. *SIAM J. Sci. Stat. Comp.*, 7(3):856–869, 1986.

[62] W. Schönauer. *Scientific Computing on Vector Computers.* North-Holland, Amsterdam, New York, Oxford, Tokyo, 1987.

[63] W. Schönauer, H. Müller, and E. Schnepf. Pseudo-residual type methods for the iterative solution of large linear systems on vector computers. In M. Feilmeier, J. Joubert, and U. Schendel, editors, *Parallel Computing 85*, pages 193–198. Elsevier Science Publishers B.V., North-Holland, 1986.

[64] W. Schönauer, M. Schlichte, and R. Weiss. Wrong ways and promising ways towards robust and efficient iterative linear solvers. In *Advances in Computer Methods for Partial Differential Equations VI*, pages 7–14. Publ. IMACS, 1987.

[65] W. Schönauer, M. Schlichte, and R. Weiss. Numerical experiments with data reduction (DARE) methods for the solution of large linear systems on vector computers. In R. Vichnevetsky, P. Borne, and J. Vignes, editors, *IMACS 1988, 12th World Congress on Scientific Computation, Proceedings, Paris*, pages 227–229, 1988.

[66] W. Schönauer, R. Weiss, and M. Schlichte. The basic ideas of the data reduction (DARE) method for the solution of large linear systems on vector computers. *ZAMM*, 69:T 184–T 185, 1989.

[67] L. Seidel. Über die Grenzwerte eines unendlichen Potenz-Ausdruckes. *Bayrische Akad. Wiss., München, Math. Phys. Kl.*, 3(11), 1874.

[68] H.D. Simon. Parallel preconditioners for the connection machine. In J. Dongarra, P. Messina, D.C. Sorensen, and R.G. Voigt, editors, *Proceedings on the Fourth SIAM Conference on Parallel Processing for Scientific Computing*, pages 389–390, 1990.

[69] G.L.G. Sleijpen and D.R. Fokkema. BiCGStab(l) for linear equations involving unsymmetric matrices with complex spectrum. *Electronic Transactions on Numer. Anal. (ETNA)*, 1:11–32, September 1993.

[70] P. Sonneveld. CGS, a fast Lanczos-type solver for nonsymmetric linear systems. *SIAM J. Sci. Stat. Comp.*, 10(1):36–52, 1989.

[71] G. Starke. Subspace orthogonalization for substructuring preconditioners for nonselfadjoint elliptic problems. *SIAM J. Sci. Comput.*, 1994. to appear.

[72] J. Stoer and R.W. Freund. On the solution of large indefinite systems of linear equations by conjugate gradient algorithms. In *Proceedings of the International Symposium on Computing Methods in Appl. Sc. and Engrg., Dec. 14.-18., Versailles, France*, pages 35–53. North-Holland Pub. Comp., 1981.

[73] Z. Strakoš. A note on the rate of convergence of the conjugate gradient method and the convergence of Ritz values. Preprint 6 of contributed papers, Universita' Degli Studi di Bergamo, NATO ASI on Computer Algorithms for Solving Linear Equations: the State of Art, Via Salvecchio n. 19 - Bergamo (Italy), September 1990.

[74] Z. Strakoš. On the real convergence rate of the conjugate gradient method. *Linear Algebra and its Applications*, 154-156/91:535–549, 1991.

[75] G. Temple. The general theory of relaxation methods applied to linear systems. *Proc. Roy. Soc. (London)*, A169(A939):476–500, 1939.

[76] C. H. Tong. A family of quasi-minimal residual methods for nonsymmetric linear systems. *SIAM J. Sci. Comput.*, 15(1):89–108, 1994.

[77] H.A. van der Vorst. Guidelines for the usage of incomplete decompositions in solving sets of linear equations as they occur in practical problems. *J. Comput. Phys.*, 44:134–155, 1981.

[78] H.A. van der Vorst. *Preconditioning by Incomplete Decompositions.* PhD thesis, University of Utrecht, The Netherlands, 1982. ACCU Series 32.

[79] H.A. van der Vorst. BI-CGSTAB: A fast and smoothly converging variant of BI-CG for the solution of nonsymmetric linear systems. *SIAM J. Sci. Stat. Comp.*, 13(2):631–644, March 1992.

[80] C. Vuik. Further experiences with GMRESR. *Supercomputer*, 55:13–27, 1993.

[81] R. Weiss. Convergence behavior of generalized conjugate gradient methods. Internal report 43/90, University of Karlsruhe, Computing Center, 1990.

[82] R. Weiss. Error-minimizing Krylov subspace methods. *SIAM J. Sci. Comput.*, 15(3):511–527, 1994.

[83] R. Weiss. Minimization properties and short recurrences for Krylov subspace methods. *Electronic Transactions on Numerical Analysis (ETNA)*, 2:57–75, 1994.

[84] R. Weiss, H. Häfner, and W. Schönauer. Tuning the matrix-vector multiplication in diagonal form. In G.R. Joubert, D.J. Evans, and F.J. Peters, editors, *Proceedings of Parallel Computing 89, Leiden*, pages 93–98. Elsevier Science Publishers B. V., North-Holland, 1990.

[85] R. Weiss and W. Schönauer. Data reduction (DARE) preconditioning for generalized conjugate gradient methods. In *Preconditioned Conjugate Gradient Methods, Lecture Notes in Mathematics, 1457*, pages 137–153. Springer-Verlag, Berlin, Heidelberg, 1990.

[86] R. Weiss and W. Schönauer. Solution of partial differential equations by means of data reduction (DARE). In *Multigrid Methods: Special Topics and Applications II, GMD-Studien Nr. 189*, pages 361–372. GMD, St. Augustin, 1991.

[87] G. Wittum and F. Liebau. On truncated incomplete decompositions. *BIT*, 29:719–740, 1989.

[88] F. Yamamoto and S. Takahashi. Vectorized LUdecomposition algorithms for large-scale circuit simulation. *IEEE Transactions on CAD*, CAD-4(3):232–239, July 1985.

[89] D.M. Young. *Iterative Methods for Solving Partial Difference Equations of Elliptic Type.* PhD thesis, Havard University, Cambridge, Massachusetts, 1950.

[90] D.M. Young. Iterative methods for solving partial difference equations of elliptic type. *Trans. Amer. Math. Soc.*, 76:92–111, 1954.

[91] D.M. Young. *Iterative Solution of Large linear Systems.* Academic Press, New York, San Francisco, London, 1971.

[92] D.M. Young. A historical overview of iterative methods. *Computer Physics Communications*, 53:1–17, 1989.

[93] L. Zhou and H.F. Walker. Residual smoothing techniques for iterative methods. *SIAM J. Sci. Comput.*, 15(2):297–312, 1994.