A NOTE ON PRECONDITIONERS AND SCALAR PRODUCTS IN KRYLOV SUBSPACE METHODS FOR SELF-ADJOINT PROBLEMS IN HILBERT SPACE[*]

ANDREAS GÜNNEL[†], ROLAND HERZOG[†], AND EKKEHARD SACHS[‡]

Abstract. The conjugate gradient and minimal residual methods for the solution of linear systems $Ax = b$ are considered. The operator A is bounded and self-adjoint and maps a Hilbert space X into its dual X^*. This setting is natural for variational problems such as those involving linear partial differential equations. The derivation of the two methods in Hilbert spaces shows that the choice of a preconditioner is equivalent to the choice of the scalar product in X.

Key words. Krylov subspace methods, preconditioners, scalar products, Hilbert spaces, Riesz isomorphism

AMS subject classifications. 65F10, 65F08

1. Introduction. In this note we consider Krylov subspace methods for the solution of linear equations

$$(1.1) \qquad Ax = b \quad \text{in } X^*.$$

The unknown x belongs to a Hilbert space X, and X^* denotes its dual. The linear operator A maps X to X^*, and it is assumed to be bounded, i.e., $A \in \mathcal{L}(X, X^*)$, and self-adjoint. The right-hand side b belongs to X^*. This is a natural setting for a variational formulation of some boundary-value problems involving linear second-order elliptic partial differential equations as illustrated by the examples below.

We shall consider the solution of (1.1) by the conjugate gradient (CG) and minimal residual (MINRES) methods depending on whether or not positive definiteness (coercivity) of A is assumed. These methods were originally introduced in [14, 20] for the case $X = \mathbb{R}^n$. Generalizations to infinite-dimensional Hilbert spaces have been considered in [1, 5, 7, 8, 11, 13, 18, 19] and the references therein; see also the references in [12].

In most of the above publications, a problem $Bx = c$ is considered, where B is assumed to map X into itself. Clearly, this complies with the present setting when we set $B := RA$, where $R \in \mathcal{L}(X^*, X)$ denotes the Riesz isomorphism. This point of view is taken, at least implicitly, in [1, 11, 16] and in [2, Section 3]. In this presentation, we prefer the setting (1.1), and we keep the Riesz map explicit. This will make it easier to pinpoint the fact that the Riesz isomorphism precisely takes the role of a preconditioner. (Also in finite dimensions it is often preferred to keep the preconditioner explicit rather than to merge it with the system matrix.)

As a matter of fact, Krylov subspace methods in Hilbert spaces for (1.1) cannot be formulated without referring to the Riesz isomorphism/preconditioner since A^2 has no meaning due to non-matching mapping properties. Since in turn the Riesz map is defined by the scalar product in X (or X^*), we conclude that selecting a preconditioner for problem (1.1) is the same as selecting the scalar product in X (or X^*). If $X = \mathbb{R}^n$, an "unpreconditioned" method is one where implicitly the Euclidean scalar product is used.

In addition to this insight, we obtain an elegant derivation of the preconditioned CG and MINRES methods, which avoids altogether the temporary use of Cholesky factors. Finally, our work also explains why it is natural to have positive definite preconditioners when solving self-adjoint indefinite problems with MINRES.

The significance of the Riesz map or the scalar product, respectively, is pointed out in many publications to varying degrees. We mention [10, Chapter 16] and [11], where the CG method is described in a variational setting and the Riesz isomorphism appears implicitly in the conversion of the residual to the gradient. In [16], the importance of the Riesz map is substantiated by the role it plays in the proof of the Lax-Milgram lemma, which applies to the case of positive definite but possibly non-self-adjoint operators A. Also recently in [25], appropriate scalar products are studied, which can guide the choice of block-diagonal preconditioners for self-adjoint saddle-point problems.

We are thus aware of the fact that our observation of preconditioning and choosing the inner product in X being one and the same for CG and MINRES would be considered folklore by some. Nevertheless, it is our experience that this fact is not as widely known as one would expect, and it seems difficult to find references which pinpoint it. This note was written to serve as a reference.

The following two prototypical examples illustrate that considering $A \in \mathcal{L}(X, X^*)$ is natural for variational problems. The first example is Poisson's equation $-\triangle u = f$ on a d-dimensional bounded domain Ω, endowed for simplicity with homogeneous Dirichlet boundary conditions. Its variational formulation is given by (1.1) with $X = H_0^1(\Omega)$ and

$$\langle Au, v \rangle = \int_\Omega \nabla u \cdot \nabla v \, \mathrm{d}x \quad \text{and} \quad \langle b, v \rangle = \int_\Omega fv \, \mathrm{d}x.$$

As usual, $H_0^1(\Omega)$ denotes the space of Sobolev functions which are square integrable up to their weak first-order derivatives and which have zero boundary values in the sense of traces. The Poisson problem gives rise to a positive definite operator (when X is endowed with the standard H^1-norm or H^1-seminorm) so that (1.1) is amenable to solution by the CG method.

The Stokes problem in fluid dynamics, by contrast, leads to an indefinite system which can be solved by MINRES. The associated variational formulation employs the Hilbert space $X = H_0^1(\Omega) \times \cdots \times H_0^1(\Omega) \times L_0^2(\Omega)$ and is given by

$$\langle A(u,p), (v,q) \rangle = \int_\Omega \nabla u : \nabla v \, \mathrm{d}x - \int_\Omega p \, \mathrm{div}\, v \, \mathrm{d}x - \int_\Omega q \, \mathrm{div}\, u \, \mathrm{d}x$$

and b as above; see, for instance, [6, Section 5]. Here $L_0^2(\Omega)$ is the space of square integrable functions with zero mean.

This paper is structured as follows. Section 2 addresses the CG method in Hilbert spaces for the self-adjoint and positive definite (coercive) case. Section 3 is devoted to the MINRES method in Hilbert spaces for self-adjoint and possibly indefinite problems.

Throughout, $\mathcal{L}(X, X^*)$ and $\mathcal{L}(X)$ denote the spaces of bounded linear operators $X \to X^*$ or $X \to X$, respectively. Moreover, $\langle \cdot, \cdot \rangle_{X^*, X}$ or, for short, $\langle \cdot, \cdot \rangle$ denotes the duality pairing of the Hilbert space X and its dual X^*. We shall denote by $R \in \mathcal{L}(X^*, X)$ the Riesz map. Given $b \in X^*$, its Riesz representation is defined by $(Rb, x)_X = \langle b, x \rangle$ for all $x \in X$, where $(\cdot, \cdot)_X$ is the scalar product in X. Clearly, the Riesz map R depends on the choice of the scalar product in X.

Let us mention that we do not address here the CG method for *indefinite* systems in so-called non-standard inner products as treated, for instance, in [4, 22, 23, 24].

2. The conjugate gradient method. In this section, $A \in \mathcal{L}(X, X^*)$ is assumed self-adjoint, i.e., $\langle Ax, y \rangle = \langle Ay, x \rangle$, and positive definite (coercive), i.e., $\langle Ax, x \rangle \geq \delta \|x\|_X^2$ for some $\delta > 0$ and all $x, y \in X$. This implies that A induces a norm, $\|x\|_A = \langle Ax, x \rangle^{1/2}$, which is equivalent to the norm $\|\cdot\|_X$. The unique solution of (1.1) is denoted by x_*.

The CG method, developed in [14], can be conceived in several ways. We follow here the derivation based on the one-dimensional minimization of the error in the A-norm for a

predetermined search direction and the update of these search directions using the direction of steepest descent while maintaining A-conjugacy of search directions.

Given an iterate $x_k \in X$ and a search direction $p_k \in X$, the CG method seeks to minimize the value of

$$\phi(x) = \frac{1}{2}\langle Ax, x \rangle - \langle b, x \rangle = \frac{1}{2}\|x - x_*\|_A^2 - \frac{1}{2}\|x_*\|_A^2$$

along the line $x_k + \alpha p_k$. This minimum is attained at

$$\alpha_k := \frac{\langle r_k, p_k \rangle}{\langle Ap_k, p_k \rangle},$$

where $r_k := b - Ax_k \in X^*$ denotes the residual. The search direction p_k is chosen as a linear combination of the previous direction p_{k-1} and the direction of steepest descent d_k for ϕ at x_k.

2.1. The direction of steepest descent. It is now important to realize that the *steepest descent direction depends upon the scalar product* in X. Let us denote by $\phi'(x_k)$ the (Fréchet) derivative of ϕ at x_k, which is, by definition, an element of X^*. The directional derivative becomes $\phi'(x_k)d = \langle Ax_k - b, d \rangle$ for arbitrary directions $d \in X$. The direction of steepest descent d_k has, by its definition, the property that

(2.1) $$d_k \text{ minimizes } \phi'(x_k)d = \langle Ax_k - b, d \rangle = -\langle r_k, d \rangle$$

over all $d \in X$ of constant norm. To solve the problem (2.1), we apply the Riesz map to obtain the representation

$$\phi'(x_k)d = -(Rr_k, d)_X.$$

The Cauchy-Schwarz inequality now readily shows $d_k = Rr_k$ to be the direction of steepest descent for ϕ at x_k. Since the Riesz map depends on the scalar product in X, so does the direction of steepest descent d_k. Note that the direction of steepest descent of ϕ is also known as its *gradient*. We point out that even in $\mathbb{R}^n$ it is useful to distinguish the derivative of a function (which does not depend on the scalar product) from the gradient (which does).

For instance, let $A \in \mathbb{R}^{n \times n}$ be symmetric, and consider it as a linear map ($x \mapsto Ax$) from $X = \mathbb{R}^n$ to its dual. The directional derivative of $\phi(x) = \frac{1}{2}x^T Ax - b^T x$ in the direction $d \in \mathbb{R}^n$ is $\phi'(x)d = (Ax - b)^T d$. Then we distinguish the derivative $\phi'(x) = Ax - b \in X'$ from the gradient $\nabla\phi(x)$. The latter depends on the scalar product in X, which is represented by the symmetric positive definite matrix P. We write $(x, y)_X$ or $(x, y)_P$ for the scalar product of two elements $x, y \in X$. According to (2.1), the direction $d \in X$ is the gradient of ϕ at x if it minimizes

$$\phi'(x)d = (Ax - b)^T d = \left(P^{-1}(Ax - b), d\right)_P$$

over all $d \in X$ of constant norm. This shows that $\nabla\phi(x) = P^{-1}(Ax - b)$.

2.2. The conjugate gradient method finalized. Using a linear combination of d_k and the previous search direction p_{k-1}, i.e.,

(2.2) $$p_k = Rr_k + \beta_k p_{k-1},$$

the requirement $\langle Ap_k, p_{k-1} \rangle = 0$ of A-conjugacy of search directions leads to the choice

$$\beta_k := -\frac{\langle Ap_{k-1}, Rr_k \rangle}{\langle Ap_{k-1}, p_{k-1} \rangle}.$$

The procedure outlined above generates the following iterates and Krylov subspaces of dimension $k \geq 1$:

$$p_{k-1} \in \mathcal{K}_k(RA; Rr_0) = \mathrm{span}\{Rr_0, (RA)Rr_0, \ldots, (RA)^{k-1}Rr_0\} \subset X,$$
$$x_k \in x_0 + \mathcal{K}_k(RA; Rr_0),$$
$$r_k \in r_0 + AR\mathcal{K}_k(AR; r_0) = r_0 + \mathrm{span}\{(AR)r_0, \ldots, (AR)^k r_0\} \subset X^*.$$

Here, the primal and dual Krylov subspaces are related by $\mathcal{K}_k(RA; Rr_0) = R\mathcal{K}_k(AR; r_0)$.

It follows by standard arguments that the iterates also satisfy

$$\langle r_k, p_j \rangle = 0 \quad \text{for all } j = 0, 1, \ldots, k-1,$$
$$\langle r_k, Rr_j \rangle = 0 \quad \text{for all } j = 0, 1, \ldots, k-1,$$
$$\langle Ap_k, p_j \rangle = 0 \quad \text{for all } j = 0, 1, \ldots, k-1,$$

and that x_k minimizes ϕ over the entire affine space $x_0 + \mathcal{K}_k(RA; Rr_0)$. Using these relations, one arrives at the final form of the CG method in a Hilbert space; see Algorithm 2.1.

ALGORITHM 2.1 (CG Method for (1.1) in a Hilbert Space).

1: *Set $r_0 := b - Ax_0 \in X^*$*
2: *Set $p_0 := Rr_0 \in X$*
3: *Set $k := 0$*
4: **while** *not converged* **do**
5:     *Set $\alpha_k := \dfrac{\langle r_k, Rr_k \rangle}{\langle Ap_k, p_k \rangle}$*
6:     *Set $x_{k+1} := x_k + \alpha_k p_k$*
7:     *Set $r_{k+1} := r_k - \alpha_k Ap_k$*
8:     *Set $\beta_{k+1} := \dfrac{\langle r_{k+1}, Rr_{k+1} \rangle}{\langle r_k, Rr_k \rangle}$*
9:     *Set $p_{k+1} := Rr_{k+1} + \beta_{k+1} p_k$*
10:     *Set $k := k + 1$*
11: **end while**

Comparing Algorithm 2.1 to the standard forms of the CG method in the literature, it stands out that the Riesz map R takes precisely the role of the application of a preconditioner P, i.e., the evaluation of P^{-1} times a vector. Recalling that the Riesz map depends on the scalar product, we conclude that the choice of a preconditioner in the traditional preconditioned CG method is actually equivalent to the choice of the scalar product in X. Even if $X = \mathbb{R}^n$, the preconditioned and unpreconditioned variants of the CG method are one and the same; they merely differ in the choice of the scalar product in X. The unpreconditioned CG method corresponds to the implicit choice of the Euclidean scalar product.

In infinite dimensions, the use of the Riesz map is essential in formulating the CG method for the solution of (1.1). Without it, the residual $r_k \in X^*$ cannot be used in (2.2) to update the search direction $p_k \in X$ since these two vectors belong to different spaces.

The convergence properties of Algorithm 2.1 in a Hilbert space depend on the spectrum of the preconditioned operator $RA \in \mathcal{L}(X)$, which is self-adjoint, i.e.,

$$(RAx, y)_X = (x, RAy)_X.$$

The spectrum is the complement of $\{\mu \in \mathbb{R} : (RA - \mu I)^{-1} \text{ exists in } \mathcal{L}(X)\}$ in $\mathbb{R}$. In the finite-dimensional situation, the spectrum consists of the eigenvalues of the generalized eigenvalue problem $Ax = \lambda R^{-1}x$. Using the condition number, i.e., the quotient of the

extremal points in the spectrum of RA, one can reproduce the linear convergence bound of the finite-dimensional preconditioned CG method; see, e.g., [6, Section 2.2] and [5, Section 1.2]. We emphasize, however, that linear bounds based on the condition number are not, in general, descriptive for the practical convergence behaviour of the CG method; see [17, Section 5.6] or [9, 15].

The significance of (the inverse of) R as a preconditioner was recently also pointed out in [16] and [18]. Good preconditioners (Riesz maps) are those which are induced by scalar products in X close to the A-scalar product. This statement continues to hold when X is replaced by one of its finite-dimensional subspaces as one does, e.g., in conforming finite element discretizations of (1.1). We thus infer that practical preconditioners/scalar products for discretized versions of the operator equation $Ax = b$ can be derived by studying the undiscretized problem.

3. The minimal residual method. In this section, $A \in \mathcal{L}(X, X^*)$ is assumed self-adjoint but it may be indefinite (non-coercive). Note that X^* is naturally endowed with the scalar product

$$(3.1) \qquad\qquad (\cdot, \cdot)_{X^*} = \langle \cdot, R \cdot \rangle,$$

in which the Riesz map pulls one of the factors from X^* back into X. We sometimes refer to the induced norm $\|r\|_{X^*} = \langle r, Rr \rangle^{1/2} = \|r\|_R$ as the R-norm.

The MINRES method, introduced in [20], uses the same Krylov subspaces as the CG method, but it seeks to minimize the R-norm of the residual $r_k \in b - Ax_k \in X^*$, i.e.,

$$\|r_k\|_R = \langle r_k, Rr_k \rangle^{1/2},$$

over the shifted dual Krylov subspaces $r_0 + A\mathcal{K}_k(RA; Rr_0) = r_0 + AR\mathcal{K}_k(AR; r_0) \subset X^*$. To carry out this minimization, MINRES builds an orthonormal basis (with respect to the inner product (3.1)) of $\mathcal{K}_k(AR; r_0) \subset X^*$, which is denoted by $V_k \in \mathcal{L}(\mathbb{R}^k, X^*)$ with "columns" $v_i \in X^*$, $i = 1, \ldots, k$.

Orthonormality, i.e., $\langle v_i, Rv_j \rangle = \delta_{ij}$, is obtained via the Lanczos recursion

$$(3.2) \qquad\qquad ARV_k = V_k T_k + \gamma_{k+1} v_{k+1} \vec{e}_k^T = V_{k+1} \widehat{T}_k,$$

with $\vec{e}_k = (0, \ldots, 0, 1)^T \in \mathbb{R}^k$ and a coefficient matrix of the form

$$\widehat{T}_k = \begin{bmatrix} \delta_1 & \gamma_2 & & \\ \gamma_2 & \delta_2 & \diagdown & \\ & \diagdown & \diagdown & \gamma_k \\ & & \gamma_k & \delta_k \\ & & 0 & \gamma_{k+1} \end{bmatrix} \in \mathbb{R}^{(k+1) \times k}.$$

The matrix $T_k \in \mathbb{R}^{k \times k}$ in (3.2) equals $\widehat{T}_k$ without the last row.

Using the basis V_k, the iterates $x_k \in x_0 + \mathcal{K}_k(RA; Rr_0) = x_0 + R\mathcal{K}_k(AR; r_0)$ and residuals r_k can be written as

$$x_k = x_0 + RV_k \vec{y}_k,$$
$$r_k = r_0 - ARV_k \vec{y}_k$$

for some $\vec{y}_k \in \mathbb{R}^k$. The objective of minimizing the residual in the R-norm can thus be expressed as the minimization over all $\vec{y} \in \mathbb{R}^k$ of

$$
\begin{aligned}
\|r_k\|_R &= \|b - Ax_0 - ARV_k\vec{y}_k\|_R \\
&= \|r_0 - ARV_k\vec{y}\|_R \\
&= \|r_0 - V_{k+1}\widehat{T}_k\vec{y}\|_R && \text{by (3.2)} \\
&= \|\,\|r_0\|_R v_1 - V_{k+1}\widehat{T}_k\vec{y}\|_R && \text{since } v_1 = r_0/\|r_0\|_R \\
&= \|V_{k+1}\big(\|r_0\|_R\vec{e}_1 - \widehat{T}_k\vec{y}\big)\|_R && \text{where } \vec{e}_1 = (1,0,\ldots,0)^T \in \mathbb{R}^{k+1} \\
&= \|\,\|r_0\|_R\vec{e}_1 - \widehat{T}_k\vec{y}\|_{\mathbb{R}^{k+1}} && \text{by orthonormality.}
\end{aligned}
$$

We conclude that the minimization of the residual in the R-norm leads to a least-squares problem in $\mathbb{R}^{k+1}$ with respect to the Euclidean norm regardless of the space in which the original problem $Ax = b$ is posed. Therefore, from here, the derivation of the Hilbert space version of MINRES parallels the derivation of the classical finite-dimensional method. We only mention that the least-squares problem is solved by maintaining a QR factorization of the matrices $\widehat{T}_k$ by means of Givens rotations. For convenience, we state the complete algorithm as Algorithm 3.1. It coincides with the (preconditioned) implementation given in [6, Algorithm 6.1] except that in Algorithm 3.1 we scale both quantities v_k and z_k such that $\|v_k\|_R = 1$ and $z_k = Rv_k$ are maintained throughout.

ALGORITHM 3.1 (MINRES Method for (1.1) in a Hilbert Space).

1: *Set $v_0 := 0 \in X^*$ and $w_0 := w_1 := 0 \in X$*
2: *Set $v_1 := b - Ax_0 \in X^*$*
3: *Set $z_1 := Rv_1$*
4: *Set $\gamma_1 := \langle v_1, z_1 \rangle^{1/2}$*
5: *Set $z_1 := z_1/\gamma_1$ and $v_1 := v_1/\gamma_1$*
6: *Set $\eta_0 := \gamma_1$, $s_0 := s_1 := 0$, $c_0 := c_1 := 1$*
7: *Set $k := 1$*
8: **while** *not converged* **do**
9:     *Set $\delta_k := \langle Az_k, z_k \rangle$*
10:     *Set $v_{k+1} := Az_k - \delta_k v_k - \gamma_k v_{k-1}$*
11:     *Set $z_{k+1} := Rv_{k+1}$*
12:     *Set $\gamma_{k+1} := \langle v_{k+1}, z_{k+1} \rangle^{1/2}$*
13:     *Set $z_{k+1} := z_{k+1}/\gamma_{k+1}$ and $v_{k+1} := v_{k+1}/\gamma_{k+1}$*
14:     *Set $\alpha_0 := c_k\delta_k - c_{k-1}s_k\gamma_k$ and $\alpha_1 := (\alpha_0^2 + \gamma_{k+1}^2)^{1/2}$*
15:     *Set $\alpha_2 := s_k\delta_k + c_{k-1}c_k\gamma_k$ and $\alpha_3 := s_{k-1}\gamma_k$*
16:     *Set $c_{k+1} := \alpha_0/\alpha_1$ and $s_{k+1} := \gamma_{k+1}/\alpha_1$*
17:     *Set $w_{k+1} := (1/\alpha_1)\big[z_k - \alpha_3 w_{k-1} - \alpha_2 w_k\big]$*
18:     *Set $x_k := x_{k-1} + c_{k+1}\eta_{k-1}w_{k+1}$*
19:     *Set $\eta_k := -s_{k+1}\eta_{k-1}$*
20:     *Set $k := k + 1$*
21: **end while**

We mention that the quantity

$$|\eta_k| = \|r_k\|_R$$

gives access to the R-norm of the residual, which is minimized over the sequence of growing shifted Krylov subspaces.

As we have observed for the CG method, we conclude that the choice of the preconditioner is equivalent to choosing the Riesz map, i.e., equivalent to choosing the scalar product in X. This observation also recently guided the study of appropriate scalar products for symmetric saddle-point problems in [25]. Finally, the exposition above explains why indefinite systems, which are to be solved by MINRES, require self-adjoint positive definite preconditioners; compare, e.g., [3, Table 9.1].

REMARK 3.2. For non-symmetric problems in finite dimensions, the replacement of the symmetric Lanczos by the Arnoldi process leads from MINRES to the generalized minimal residual method (GMRES) introduced in [21]. The coefficient matrix $\widetilde{T}_k$ will be upper Hessenberg; see, for instance, [6, Section 4.1.1] or [17, Section 2.5.5]. Precisely the same modifications will lead to GMRES for non-self-adjoint problems of type (1.1) in a Hilbert space setting.

4. Conclusions. In this paper we consider the CG and MINRES methods for self-adjoint operator equations $Ax = b$ with $A \in \mathcal{L}(X, X^*)$, where X is a Hilbert space. We present a comprehensive derivation of these methods, which shows that both CG and MINRES inevitably depend on the scalar product one chooses in X. We also see that the "preconditioned" and "unpreconditioned" versions of these algorithms, which are often presented as distinct methods, are actually one and the same. The "unpreconditioned" versions simply correspond to the implicit choice of the Euclidean scalar product, which, of course, is only possible if $X \cong \mathbb{R}^n$. We emphasize that the choice of a preconditioner for CG and MINRES is equivalent to the choice of the scalar product in X. This choice, even for discretized versions of the problem, can be guided by an analysis of the spectrum carried out for the *undiscretized* problem, e.g., at the level of the partial differential equation to be solved. This is the central idea behind *operator preconditioning* techniques developed, for instance, in [2, 15, 18]. Condition numbers bounded independently of the mesh size can be achieved in this way, which are, however, not in one-to-one correspondence with the fast convergence of Krylov subspace methods.

In a forthcoming publication we will continue this work and address convergence results for CG and MINRES including a new superlinear convergence result for MINRES.

REFERENCES

[1] O. AXELSSON AND J. KARÁTSON, *On the rate of convergence of the conjugate gradient method for linear operators in Hilbert space*, Numer. Funct. Anal. Optim., 23 (2002), pp. 285–302.

[2] ———, *Equivalent operator preconditioning for elliptic problems*, Numer. Algorithms, 50 (2009), pp. 297–380.

[3] M. BENZI, G. GOLUB, AND J. LIESEN, *Numerical solution of saddle point problems*, Acta Numer., 14 (2005), pp. 1–137.

[4] J. BRAMBLE AND J. PASCIAK, *A preconditioning technique for indefinite systems resulting from mixed approximations of elliptic problems*, Math. Comp., 50 (1988), pp. 1–17.

[5] J. W. DANIEL, *The conjugate gradient method for linear and nonlinear operator equations*, SIAM J. Numer. Anal., 4 (1967), pp. 10–26.

[6] H. ELMAN, D. SILVESTER, AND A. WATHEN, *Finite Elements and Fast Iterative Solvers: with Applications in Incompressible Fluid Dynamics*, Oxford University Press, New York, 2005.

[7] O. ERNST, *Minimal and Orthogonal Residual Methods and their Generalizations for Solving Linear Operator Equations*. Habilitation Thesis, Faculty of Mathematics and Computer Sciences, TU Bergakademie Freiberg, 2000.

[8] Z. FORTUNA, *Some convergence properties of the conjugate gradient method in Hilbert space*, SIAM J. Numer. Anal., 16 (1979), pp. 380–384.

[9] T. GERGELITS AND Z. STRAKOŠ, *Composite convergence bounds based on Chebyshev polynomials and finite precision conjugate gradient computations*, Numer. Algorithms, in press, published online June 2, 2013, doi 10.1007/s11075-013-9713-z.

[10] R. GLOWINSKI, *Finite element methods for incompressible viscous flow*, in Handbook of Numerical Analysis, Vol. IX. Numerical Methods for Fluids, P. G. Ciarlet and J. L. Lions, eds., North-Holland, Amsterdam, 2003, pp. 3–1176.

[11] R. GLOWINSKI AND S. LAPIN, *Iterative solution of linear variational problems in Hilbert spaces: some conjugate gradients success stories*, in Conjugate Gradient Algorithms and Finite Element Methods, M. Křížek, P. Neittaanmäki, R. Glowinski, and S. Korotov, eds., Springer, Berlin, 2004, pp. 223–245.

[12] G. H. GOLUB AND D. P. O'LEARY, *Some history of the conjugate gradient and Lanczos algorithms: 1948–1976*, SIAM Rev., 31 (1989), pp. 50–102.

[13] R. M. HAYES, *Iterative methods of solving linear problems on Hilbert space*, National Bureau of Standards Applied Mathematics Series No. 39 (1954), pp. 71–103.

[14] M. R. HESTENES AND E. STIEFEL, *Methods of conjugate gradients for solving linear systems*, J. Research Nat. Bur. Standards, 49 (1952), pp. 409–436.

[15] R. HIPTMAIR, *Operator preconditioning*, Comput. Math. Appl., 52 (2006), pp. 699–706.

[16] R. KIRBY, *From functional analysis to iterative methods*, SIAM Rev., 52 (2010), pp. 269–293.

[17] J. LIESEN AND Z. STRAKOŠ, *Krylov Subspace Methods*, Oxford University Press, Oxford, 2013.

[18] K.-A. MARDAL AND R. WINTHER, *Preconditioning discretizations of systems of partial differential equations*, Numer. Linear Algebra Appl., 18 (2011), pp. 1–40.

[19] O. NEVANLINNA, *Convergence of Iterations for Linear Equations*, Birkhäuser, Basel, 1993.

[20] C. PAIGE AND M. SAUNDERS, *Solution of sparse indefinite systems of linear equations*, SIAM J. Numer. Anal., 12 (1975), pp. 617–629.

[21] Y. SAAD AND M. H. SCHULTZ, *GMRES: a generalized minimal residual algorithm for solving nonsymmetric linear systems*, SIAM J. Sci. Statist. Comput., 7 (1986), pp. 856–869.

[22] J. SCHÖBERL AND W. ZULEHNER, *Symmetric indefinite preconditioners for saddle point problems with applications to PDE-constrained optimization*, SIAM J. Matrix Anal. Appl., 29 (2007), pp. 752–773.

[23] M. STOLL AND A. WATHEN, *Combination preconditioning and the Bramble-Pasciak$^+$ preconditioner*, SIAM J. Matrix Anal. Appl., 30 (2008), pp. 582–608.

[24] M. TANI AND V. SIMONCINI, *Refined spectral estimates for preconditioned saddle point linear systems in a non-standard inner product*, ANZIAM J. Electron. Suppl., 54 (2012), pp. C291–C308.

[25] W. ZULEHNER, *Non-standard norms and robust estimates for saddle point problems*, SIAM J. Matrix Anal. Appl., 32 (2011), pp. 536–560.