A NOTE ON PRECONDITIONERS AND SCALAR PRODUCTS FOR KRYLOV METHODS IN HILBERT SPACE

ANDREAS GÜNNEL, ROLAND HERZOG, AND EKKEHARD SACHS

Abstract. Krylov subspace methods, viz. CG and MINRES, for the solution of linear systems $A\,x = b$ in a Hilbert space X are considered. The operator A is self-adjoint and it maps X into its dual X^*. This setting is natural for variational problems such as linear partial differential equations. The derivation of the two methods in Hilbert space shows that the choice of a preconditioner is equivalent to the choice of the scalar product in X.

In this note we consider Krylov subspace methods for the solution of self-adjoint linear equations

$$A\,x = b \tag{1}$$

in a Hilbert space X. The bounded linear operator A maps X to its dual X^*, i.e., $A \in \mathcal{L}(X, X^*)$, and the right hand side b belongs to X^*. This is the natural setting with variational formulations of, e.g., second-order partial differential equations in mind, as illustrated by the examples below.

We shall consider the solution of (1) by the conjugate gradient and minimal residual methods, depending on whether or not positive definiteness (coercivity) of A is assumed. These methods were originally introduced in [9, 12], respectively, for the case $X = \mathbb{R}^n$. Generalizations to infinite dimensional Hilbert spaces have been considered in [1, 3, 5–7, 11] and the references therein. In these publications, A was assumed to map X into itself. By contrast, the setting adopted here has a number of advantages. Besides being more natural in the context of variational problems, it also leads directly to a simple but important observation.

In a nutshell, this observation shows that selecting a preconditioner for problem (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 conjugate gradient and minimal residual methods, which avoids altogether the temporary use of Cholesky factors. Finally, our work also explains why the solution of self-adjoint indefinite problems in Minres requires positive definite preconditioners.

An essential role in the presentation is played by the Riesz map. As a matter of fact, Krylov methods in Hilbert space for (1) cannot be formulated without it. The significance of the Riesz map was recently also emphasized in [10], where its importance was substantiated by the part it plays in the proof of the Lax-Milgram lemma, which applies to the case of positive definite but possibly non-self-adjoint A. Also recently, [14] studied appropriate scalar products which can guide the choice of block-diagonal preconditioners for symmetric saddle-point problems.

The following two protypical examples illustrate that considering $A \in \mathcal{L}(X, X^*)$ is natural for variational problems. The first example is Poisson's equation $-\triangle u =$

Date: July 13, 2011.
Key words and phrases. Krylov methods, preconditioners, scalar products, Hilbert spaces, Riesz isomorphism.

f, endowed for simplicity with homogeneous Dirichlet boundary conditions. Its variational formulation is given by (1) with $X = H_0^1(\Omega)$ and

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

This problem gives rise to a positive definite operator so that (1) is amenable to the solution by the conjugate gradient method. The Stokes problem in fluid dynamics, by constrast, leads to an indefinite system which can be solved by Minres. Its variational formulation is given by $X = H_0^1(\Omega) \times L_0^2(\Omega)$ and

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

and b as above, see, for instance, [4, Section 5]. While we focus here on self-adjoint operators A, we mention that the considerations in this paper can be easily extended to Krylov methods for non-self-adjoint problems, such as Gmres [13], in Hilbert space.

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

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

The conjugate gradient method (CGM), developed by [9], can be conceived in one of 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 subsequent search directions.

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

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

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

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

where $r_k := b - A\, x_k \in X^*$ 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. It is now important to realize that the steepest descent direction depends upon the scalar product in X. Indeed, by definition,

$$d_k \text{ minimizes } \phi'(x_k)\, d = \langle A\, x_k - b,\, d\rangle = -\langle r_k,\, d\rangle \tag{2}$$

over all $d \in X$ of constant norm. To solve problem (2), we apply the Riesz map to obtain

$$\phi'(x_k)\, d = -(R\, r_k, d)_X,$$

which readily shows $d_k = R\, r_k$ 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.

Using a linear combination of d_k and the previous search direction p_{k-1}, i.e.,

$$p_k = R\,r_k + \beta_k\,p_{k-1}, \tag{3}$$

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

$$\beta_k := -\frac{\langle A\,p_{k-1},\,R\,r_k\rangle}{\langle A\,p_{k-1},\,p_{k-1}\rangle}.$$

The procedure outlined above generates the following iterates and Krylov spaces:

$$\begin{aligned}
p_k &\in \mathcal{K}_k(RA;R\,r_0) = \operatorname{span}\{R\,r_0, (RA)R\,r_0, \ldots, (RA)^k R\,r_0\} \subset X,\\
x_k &\in \mathcal{K}_k(RA;R\,r_0) + x_0,\\
r_k &\in \mathcal{K}_k(AR;r_0) = \operatorname{span}\{r_0, (AR)\,r_0, \ldots, (AR)^k\,r_0\} \subset X^*.
\end{aligned}$$

We point out that the primal and dual Krylov spaces are related by $\mathcal{K}_k(RA;R\,r_0) = R\,\mathcal{K}_k(AR;r_0)$.

It follows by standard arguments that the iterates also satisfy

$$\begin{aligned}
\langle A\,p_k,\,p_j\rangle &= 0 \quad \text{for all } j = 0,1,\ldots,k-1,\\
\langle r_k,\,p_j\rangle &= 0 \quad \text{for all } j = 0,1,\ldots,k-1,
\end{aligned}$$

and that x_k minimizes ϕ over the entire affine space $x_0 + \mathcal{K}_k(RA;R\,r_0)$. Using these relations one arrives at the final form of the CGM in Hilbert space, see Algorithm 1.

Algorithm 1 Conjugate gradient method for (1) in Hilbert space

Set $r_0 := b - A\,x_0 \in X^*$

Set $p_0 := R\,r_0 \in X$

Set $k := 0$

while not converged **do**

  Set $\alpha_k := \dfrac{\langle r_k,\,R\,r_k\rangle}{\langle A\,p_k,\,p_k\rangle}$

  Set $x_{k+1} := x_k + \alpha_k\,p_k$

  Set $r_{k+1} := r_k - \alpha_k A\,p_k$

  Set $\beta_{k+1} := \dfrac{\langle r_{k+1},\,R\,r_{k+1}\rangle}{\langle r_k,\,R\,r_k\rangle}$

  Set $p_{k+1} := R\,r_{k+1} + \beta_{k+1}\,p_k$

  Set $k := k+1$

end while

Comparing Algorithm 1 to standard forms of the CGM in the literature, it stands out that the Riesz map R takes precisely the role of a preconditioner P^{-1}. Recalling that the Riesz map depends on the scalar product, we conclude that the choice of a preconditioner in the traditional preconditioned CGM 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 CGM are one and the same, they merely differ in the choice of the scalar product in X. The unpreconditioned CGM corresponds to the implicit choice of the Euclidean scalar product.

In infinite dimensions, the use of the Riesz map is essential in formulating the CGM for the solution of (1). Without it, the residual $r_k \in X^*$ cannot be used in (3) to update the search direction $p_k \in X$ since the two of them belong to different spaces.

The convergence properties of Algorithm 1 in Hilbert space depend on the eigenvalues of the operator pair (A, R^{-1}), i.e., the solutions of

$$A\,x = \lambda\,R^{-1}x \text{ in } X^*, \quad \text{or equivalently,} \quad RA\,x = \lambda\,x \text{ in } X.$$

This eigenvalue problem is analogous to the eigenvalue problem for $P^{-1}A$ appearing in the classical convergence analysis of the preconditioned CGM. The significance of (the inverse of) R as a preconditioner was recently also pointed out in [10]. 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 finite element discretizations of (1). More details will be given in a forthcoming publication [8].

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 R-scalar product, i.e.,

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

The minimal residual method, introduced by [12], uses the same Krylov spaces as the CGM, but it seeks to minimize the R-norm of the residual $r_k \in b - A\,x_k \in X^*$, i.e.,

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

over the dual Krylov space $\mathcal{K}_k(AR; r_0) \subset X^*$. To carry out this minimization, Minres builds an orthonormal basis of this space. This growing basis at iteration k 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., $\langle v_i,\, R\,v_j\rangle = \delta_{ij}$, is obtained via the Lanczos recursion

$$A\,R\,V_k = V_k\,T_k + \gamma_{k+1}\,v_{k+1}\,\vec{e}_k^{\top} = V_{k+1}\,\widehat{T}_k, \tag{4}$$

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

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

The matrix $T_k \in \mathbb{R}^{k\times k}$ in (4) equals $\widehat{T}_k$ without the last row.

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

$$x_k = x_0 + R\,V_k\,\vec{y} \quad \text{for some } \vec{y} \in \mathbb{R}^k.$$

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

$$\begin{aligned}
&\|b - A\,x_0 - A\,R\,V_k\,\vec{y}\|_R \\
&\qquad = \big\|r_0 - A\,R\,V_k\,\vec{y}\big\|_R \\
&\qquad = \big\|r_0 - V_{k+1}\widehat{T}_k\,\vec{y}\big\|_R && \text{by (4)} \\
&\qquad = \big\|\|r_0\|_R\,v_1 - V_{k+1}\widehat{T}_k\,\vec{y}\big\|_R && \text{since } v_1 = r_0/\|r_0\|_R \\
&\qquad = \big\|V_{k+1}\big(\|r_0\|_R\,\vec{e}_1 - \widehat{T}_k\,\vec{y}\big)\big\|_R && \text{where } \vec{e}_1 = (1, 0, \ldots, 0)^\top \in \mathbb{R}^{k+1} \\
&\qquad = \big\|\|r_0\|_R\,\vec{e}_1 - \widehat{T}_k\,\vec{y}\big\|_{\mathbb{R}^{k+1}} && \text{by orthonormality.}
\end{aligned}$$

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. Therefore, from here, the derivation of the Hilbert space 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 2. It coincides with the (preconditioned) implementation given in [4, Algorithm 6.1] except that in Algorithm 2, we scale both quantities v_k and z_k such that $\|v_k\|_R = 1$ and $z_k = R\,v_k$ are maintained throughout. 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 Krylov spaces.

Algorithm 2 Minimal residual method for (1) in Hilbert space

Set $v_0 := 0 \in X^*$ and $w_0 := w_1 := 0 \in X$
Set $v_1 := b - A\,x_0 \in X^*$
Set $z_1 := R\,v_1$
Set $\gamma_1 := \langle v_1, z_1 \rangle^{1/2}$
Set $z_1 := z_1/\gamma_1$ and $v_1 := v_1/\gamma_1$
Set $\eta_0 := \gamma_1$, $s_0 := s_1 := 0$, $c_0 := c_1 := 1$
Set $k := 1$
while not converged **do**
  Set $\delta_k := \langle z_k, z_k \rangle$
  Set $v_{k+1} := A\,z_k - \delta_k\,v_k - \gamma_k\,v_{k-1}$
  Set $z_{k+1} := R\,v_{k+1}$
  Set $\gamma_{k+1} := \langle v_{k+1}, z_{k+1} \rangle^{1/2}$
  Set $z_{k+1} := z_{k+1}/\gamma_{k+1}$ and $v_{k+1} := v_{k+1}/\gamma_{k+1}$
  Set $\alpha_0 := c_k\,\delta_k - c_{k-1}\,s_k\,\gamma_k$ and $\alpha_1 := (\alpha_0^2 + \gamma_{k+1}^2)^{1/2}$
  Set $\alpha_2 := s_k\,\delta_k + c_{k-1}\,c_k\,\gamma_k$ and $\alpha_3 := s_{k-1}\,\gamma_k$
  Set $c_{k+1} := \alpha_0/\alpha_1$ and $s_{k+1} := \gamma_{k+1}/\alpha_1$
  Set $w_{k+1} := (1/\alpha_1)\big[z_k - \alpha_3\,w_{k-1} - \alpha_2\,w_k\big]$
  Set $x_k := x_{k-1} + c_{k+1}\,\eta_{k-1}\,w_{k+1}$
  Set $\eta_k := -s_{k+1}\,\eta_{k-1}$
  Set $k := k+1$
end while

As we observed for the CGM, 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 [14]. Finally, the exposition above explains why indefinite systems, which are to be solved by MINRES require self-adjoint positive definite preconditioners, compare, e.g., [2, Table 9.1].

Acknowledgments. This work was supported by a DAAD travel grant 50743976 *Preconditioners in PDE-Constrained Optimization*. The authors would like to thank Andy Wathen for fruitful discussions and the Numerical Analysis group at Oxford University for their hospitality. They also thank Martin Stoll for his comments on an earlier version of the manuscript.

References

[1] O. Axelsson and J. Karátson. On the rate of convergence of the conjugate gradient method for linear operators in Hilbert space. *Numerical Functional Analysis and Optimizationn*, 23(3-4):285–302, 2002.

[2] M. Benzi, G. Golub, and J. Liesen. Numerical solution of saddle point problems. *Acta Numerica*, 14:1–137, 2005.
[3] James W. Daniel. The conjugate gradient method for linear and nonlinear operator equations. *SIAM Journal on Numerical Analysis*, 4:10–26, 1967.
[4] Howard C. Elman, David J. Silvester, and Andrew J. Wathen. *Finite elements and fast iterative solvers: with applications in incompressible fluid dynamics*. Numerical Mathematics and Scientific Computation. Oxford University Press, New York, 2005.
[5] 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.
[6] Zenon Fortuna. Some convergence properties of the conjugate gradient method in Hilbert space. *SIAM Journal on Numerical Analysis*, 16(3):380–384, 1979.
[7] 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*, Sci. Comput., pages 223–245. Springer, Berlin, 2004.
[8] A. Günnel, R. Herzog, and E. Sachs. Convergence analysis for Krylov subspace methods in Hilbert space. Technical report, in preparation.
[9] Magnus R. Hestenes and Eduard Stiefel. Methods of conjugate gradients for solving linear systems. *Journal of Research of the National Bureau of Standards*, 49:409–436 (1953), 1952.
[10] R. Kirby. From functional analysis to iterative methods. *SIAM Review*, 52(2):269–293, 2010.
[11] Olavi Nevanlinna. *Convergence of iterations for linear equations*. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, 1993.
[12] C. Paige and M. Saunders. Solution of sparse indefinite systems of linear equations. *SIAM Journal on Numerical Analysis*, 12(4), 1975.
[13] Youcef Saad and Martin H. Schultz. GMRES: a generalized minimal residual algorithm for solving nonsymmetric linear systems. *Society for Industrial and Applied Mathematics. Journal on Scientific and Statistical Computing*, 7(3):856–869, 1986.
[14] W. Zulehner. Non-standard norms and robust estimates for saddle point problems. *SIAM Journal on Matrix Analysis and Applications*, 32:536–560, 2011.

Chemnitz University of Technology, Faculty of Mathematics, D–09107 Chemnitz, Germany

E-mail address: andreas.guennel@mathematik.tu-chemnitz.de

Chemnitz University of Technology, Faculty of Mathematics, D–09107 Chemnitz, Germany

E-mail address: roland.herzog@mathematik.tu-chemnitz.de

University of Trier, Department of Mathematics, D–54286 Trier, Germany

E-mail address: sachs@uni-trier.de