A convergence result for some Krylov-Tikhonov methods in Hilbert spaces

P. Novati,
Department of Mathematics and Geosciences
University of Trieste, Italy

Abstract

In this paper we present a convergence result for some Krylov projection methods when applied to the Tikhonov minimization problem in its general form. In particular we consider the method based on the Arnoldi algorithm and the one based on the Lanczos bidiagonalization process.

Keywords. Linear ill-posed problem, Compact operator, Arnoldi algorithm, Lanczos bidiagonalization.

AMS subject classifications. 65F10, 65F22, 65R32.

1 Introduction

When solving linear equations of the type

$$Au = f, \tag{1}$$

where u and f belong to a Hilbert space $\mathcal{H}$, and $A : \mathcal{H} \to \mathcal{H}$ is a compact linear operator, one is often forced to impose some regularization. Indeed, such operators possess the general property that the spectrum is either finite or countably infinite; in the latter case the sequence of eigenvalues $\{\lambda_n\}_{n\geq 1}$ (arranged in order of decreasing magnitude) converges to 0 (see e.g. [11, §1.8]). By consequence, the problem (1) is ill-posed since the operator does not possess a bounded inverse. The situation becomes even more difficult whenever f does not satisfy the Picard condition (because of errors, noise, see [7]), that is, when it does not belong to the range of A, $\mathcal{R}(A)$.

In this context, the well known Tikhonov regularization plays an important role. This method prescribes to replace (1) with the minimization problem

$$\min_{u\in\mathcal{H}} \left\{ \|Au - f\|^2 + \lambda^2 \|Lu\|^2 \right\}, \quad \lambda > 0, \tag{2}$$

where $L : \mathcal{H} \to \mathcal{H}$ is a linear regularization operator, that, depending on the magnitude of λ, should force the solution to have some characteristics (see [6] for a background).

As in many other areas of numerical analysis the minimization (2) can be approached by using projections on suitable finite dimensional spaces, that is, we can replace $\min_{u\in\mathcal{H}}$ by $\min_{u\in\mathcal{K}_m}$

in (2), where $\mathcal{K}_m \subset \mathcal{H}$ is an m-dimensional space. When $\mathcal{K}_m$ is a Krylov subspace the arising procedure is generally referred to as a Krylov-Tikhonov method (see for instance [3] for a recent overview). In this context, many Krylov based algorithms such as the Lanczos bidiagonalization and the Arnoldi algorithm are generally considered powerful tools. Nevertheless, to the best of our knowledge, no convergence result is available in the literature. Assuming that the compact operator A is not of finite rank and injective ($\mathcal{N}(A) = \{0\}$, where $\mathcal{N}(A)$ denotes the null space), the solution $u^\dagger$ of (2) is unique. Denoting by $\{u_m\}_{m\geq 0}$ the sequence of approximations arising from the constrained minimization $\min_{u\in\mathcal{K}_m}$, the convergence of this sequence to $u^\dagger$ is still to be proved.

In this paper we show that some Krylov-Tikhonov methods are in fact orthogonal projection methods for the linear operator equation

$$\left(A^*A + \lambda^2 L^*L\right) u = A^*f.$$

In this way we are able to show that there exists a norm in $\mathcal{H}$, $E(\cdot)$, that eventually may be a semi-norm, such that under suitable hypotheses $E(u_m - u^\dagger) \to 0$ and moreover that $\left\{E(u_m - u^\dagger)\right\}_{m\geq 0}$ is bounded by the remainder of a series corresponding to an ℓ_2 sequence. We restrict our analysis to the Krylov-Tikhonov methods based on the Lanczos bidiagonalization and the Arnoldi algorithm. We keep the analysis in the framework of a general infinite dimensional separable Hilbert space, since the properties of compact operators of infinite rank are not well replicated in the finite dimensional setting.

The paper is organized as follows. In Section 2 we recall the basic properties of the Arnoldi algorithm and GMRES. In section 3 we present some theoretical results about the extendibility of the Krylov subspaces. The convergence results for the Arnoldi-Tikhonov method are given in Section 4. In Section 5, the results are then extended to the Lanczos-Tikhonov method in which the Tikhonov minimization is solved by the Lanczos bidiagonalization.

2 The Arnoldi algorithm

Let $\mathcal{H}$ be a Hilbert space, with scalar product $<,>$ and norm $\|\cdot\|$ defined as

$$\|x\| = \langle x, x\rangle^{1/2}.$$

Throughout the paper we assume that $\mathcal{H}$ is separable, that is, it admits a countable orthonormal basis $\{\varphi_n\}_{n\in\mathbb{N}}$. We denote by A^* the adjoint of A, and by $I_\mathcal{H}$ the identity operator. We also denote by $\{\sigma_j(A)\}_{j\geq 1}$ the sequence of the singular values of A arranged in order of decreasing magnitude.

Let $\mathcal{K}_m(A, f) = \operatorname{span}\{f, Af, \ldots, A^{m-1}f\}$ be the m-dimensional Krylov subspace generated by A and f. Setting $N = \sup_n(\dim \mathcal{K}_n(A, f))$, the Arnoldi algorithm computes an orthonormal basis $\{w_1, ..., w_m\}$ of $\mathcal{K}_m(A, f)$ for each $m \leq N$. In particular, we have

$$w_1 = \frac{f}{\|f\|}, \quad w_{m+1} = \frac{(I_\mathcal{H} - P_m)Aw_m}{\|(I_\mathcal{H} - P_m)Aw_m\|},$$

where P_m is the orthogonal projection onto $\mathcal{K}_m(A, f)$. Let $W_m : \mathbb{C}^m \to \mathcal{K}_m(A, f) \subseteq \mathcal{H}$ be the isometry defined by

$$W_m y = \sum\nolimits_{j=1}^{m} y^{(j)} w_j, \quad y = \left(y^{(1)}, ..., y^{(m)}\right) \in \mathbb{C}^m, \tag{3}$$

so that $P_m = W_m W_m^*$. Let moreover $H_m \in \mathbb{C}^{(m+1)\times m}$ be the upper Hessenberg matrix such that

$$AW_m = W_{m+1} H_m. \tag{4}$$

It is known that the sequence of GMRES approximations, $\{u_m\}_{m\geq 0}$, $u_0 = 0$, is defined by solving at each step of the Arnoldi algorithm the least-squares problem

$$\min_{u\in\mathcal{K}_m(A,f)} \|Au - f\|,$$

whose solution is given by $u_m = W_m y_m$ where

$$y_m = \arg\min_{y\in\mathbb{C}^m} \|H_m y - \|f\| e_1\|, \tag{5}$$

$e_1 = (1, 0, ..., 0)^T \in \mathbb{C}^{m+1}$. Note that for $m \leq N$

$$u_m = W_m \left(H_m^* H_m\right)^{-1} H_m^* e_1 \|f\|. \tag{6}$$

We remark that the norm used in (5) is just the Euclidean norm in $\mathbb{C}^{m+1}$. We avoid the use of a different symbol since the meaning is clear from the context. The same holds for the adjoint symbol used in (6) to indicate the Hermitian transpose of a matrix.

3 Some theoretical results on Krylov subspaces

Theorem 1 *[11, Theorem 1.9.2] Let $A : \mathcal{H} \to \mathcal{H}$ be a compact normal operator. Let moreover $\{\lambda_n\}_{n\in\mathbf{S}}$ be the sequence (finite $\mathbf{S} = \{1, ..., d\}$ or countably infinite $\mathbf{S} = \mathbb{N}$) of non-zero eigenvalues counted according to their multiplicities and $\{\varphi_n\}_{n\in\mathbf{S}}$ the corresponding orthonormal sequence of eigenvectors. Then*

$$Ax = \sum\nolimits_{n\in\mathbf{S}} \lambda_n \langle x, \varphi_n\rangle \varphi_n, \quad x \in \mathcal{H}. \tag{7}$$

Moreover A is self-adjoint if and only if $\lambda_n \in \mathbb{R}$, $n \in \mathbf{S}$, and is positive if and only if $\lambda_n > 0$, $n \in \mathbf{S}$.

The closed linear span of the eigenvectors $\{\varphi_n\}_{n\in\mathbf{S}}$ is equal to $\mathcal{H}$ if and only if A is injective. Now, let $\mathcal{K}(A, f) = \cup_{m\in\mathbb{N}}\mathcal{K}_m(A, f)$ be the closed linear span of the vectors $f, Af, A^2 f, ...$ (eventually of finite dimension if $N < \infty$) and let P be the orthogonal projection onto $\mathcal{K}(A, f)$. Let moreover $\mathcal{K}(A, f)^\perp$ be the orthogonal complement of $\mathcal{K}(A, f)$. We have the following result.

Theorem 2 *Let $A : \mathcal{H} \to \mathcal{H}$ be a compact normal operator. Assume moreover that A is injective. If $u \in \mathcal{H}$ is the solution of $Au = f$ then $u \in \mathcal{K}(A, f)$.*

Proof. Let $\mathcal{K} = \mathcal{K}(A, f)$ for simplicity. Assuming $u \notin \mathcal{K}$, then $u - Pu \perp \mathcal{K}$. Since $A\mathcal{K} \subseteq \mathcal{K}$ we have that $A^*(u - Pu) \in \mathcal{K}^\perp$. Moreover, $A(u - Pu) = b - APu \in \mathcal{K}$ and therefore

$$\langle A^*(u - Pu), A(u - Pu)\rangle = \left\langle (u - Pu), A^2(u - Pu)\right\rangle = 0.$$

Since A is injective $\{\varphi_n\}_{n\in\mathbb{N}}$ is a basis of $\mathcal{H}$ and hence, using (7), we obtain

$$\begin{aligned}\left\langle (u-Pu), A^2(u-Pu)\right\rangle &= \left\langle \sum\nolimits_{j\in\mathbb{N}} \langle u-Pu,\varphi_j\rangle \varphi_j, \sum\nolimits_{n\in\mathbb{N}} |\lambda_n|^2 \langle u-Pu,\varphi_n\rangle \varphi_n \right\rangle \\ &= \sum\nolimits_{n\in\mathbb{N}} |\lambda_n|^2 \left|\langle u-Pu,\varphi_n\rangle\right|^2 = 0,\end{aligned}$$

that is, $\langle u-Pu,\varphi_n\rangle = 0$ for $n\in\mathbb{N}$, and therefore $u = Pu$. ■

The above result can be generalized as follows.

Theorem 3 *Let $A:\mathcal{H}\to\mathcal{H}$ be a bounded linear operator. Assume moreover that A is injective. If $u\in\mathcal{H}$ is the solution of $Au=f$ then $u\in\mathcal{K}(A,f)$.*

Proof. As before let $\mathcal{K}=\mathcal{K}(A,f)$. Let moreover $u=u_{\mathcal{K}}+u_{\mathcal{K}}^{\perp}$, $u_{\mathcal{K}}\in\mathcal{K}$, $u_{\mathcal{K}}^{\perp}\in\mathcal{K}^{\perp}$. Observe that $Au_{\mathcal{K}}^{\perp}=Au-Au_{\mathcal{K}}=f-Au_{\mathcal{K}}\in\mathcal{K}$ and therefore

$$\left\langle Au_{\mathcal{K}}^{\perp}, v\right\rangle = \left\langle u_{\mathcal{K}}^{\perp}, A^*v\right\rangle = 0,$$

for each $v\in\mathcal{K}^{\perp}$. Since $\mathcal{K}^{\perp}$ is A^* invariant (because $\mathcal{K}$ is A invariant) we have that

$$\left\langle u_{\mathcal{K}}^{\perp}, z\right\rangle = 0 \text{ for each } z\in\mathcal{R}(A^*_{|\mathcal{K}^{\perp}}).$$

Since we can write $\mathcal{K}^{\perp}=\overline{\mathcal{R}(A^*_{|\mathcal{K}^{\perp}})}\oplus\mathcal{N}(A_{|\mathcal{K}^{\perp}})$ and A is injective we have that $\overline{R(A^*_{|\mathcal{K}^{\perp}})}=\mathcal{K}^{\perp}$ and therefore $u_{\mathcal{K}}^{\perp}=0$. ■

Clearly, if the solution belongs to $\mathcal{K}(A,f)$ then GMRES converges to it. We remark that in general $A\mathcal{K}(A,f)\subseteq\mathcal{K}(A,f)$, and the above theorem ensures that $A\mathcal{K}(A,f)=\mathcal{K}(A,f)$ if and only if $f\in\mathcal{R}(A)$. Let $W:\mathbb{C}^N\to\mathcal{K}(A,f)$ be the isometry defined as in (3) (note that it may be $N=\infty$ if the Arnoldi algorithm does not terminate in a finite number of steps). It is clear that $A\mathcal{K}(A,f)=\mathcal{K}(A,f)$ if and only if there exists a vector $c\in\mathbb{C}^N$ such that

$$AWc = f.$$

Using (4) with $H:\mathbb{C}^N\to\mathbb{C}^N$ such that $AW=WH$, and remembering that $f=\|f\|\,w_1=\|f\|\,We_1$, we have

$$WHc = \|f\|\,We_1,$$

and therefore c is the solution of $Hc=\|f\|\,e_1$. This means that $\|f\|\,e_1\in\mathbb{C}^N$ satisfies the Picard condition for the projected linear problem.

Corollary 4 *Let $g\in\mathcal{R}(A)$ and let u be such that $Au=g$. If $g\in\mathcal{K}(A,f)$ then $u\in\mathcal{K}(A,f)$.*

Proof. Using the above theorem, $u\in\mathcal{K}(A,g)$. Since $g\in\mathcal{K}(A,f)$ we clearly have $\mathcal{K}(A,g)\subseteq\mathcal{K}(A,f)$, and hence $u\in\mathcal{K}(A,f)$. ■

The following theorem, extends to separable Hilbert spaces what is very well known in finite dimension. That is, working in $\mathbb{C}^n$, $n<\infty$, if the Arnoldi process does not terminate in $s<n$ steps, then the Arnoldi vectors form a basis of $\mathbb{C}^n$. This is not obvious in infinite dimension, because, in principle, $\mathcal{K}(A,f)$ might be a proper subspace of $\mathcal{H}$ even if $\dim\mathcal{K}(A,f)=\infty$. Before

starting we need an additional notation. Given a finite sequence of vectors $\{g_1, g_2, ..., g_n\}$ we denote by $\mathcal{K}(A, g_1, g_2, ..., g_n)$ the closed linear span of the set $\left\{A^k g_i, k = 0, 1, ..., i = 1, ..., n\right\}$. Clearly if $g \in \mathcal{K}(A, g_1, g_2, ..., g_n)$ then $\mathcal{K}(A, g) \subseteq \mathcal{K}(A, g_1, g_2, ..., g_n)$ so that the result of Corollary 4 still holds if we replace $\mathcal{K}(A, f)$ with $\mathcal{K}(A, g_1, g_2, ..., g_n)$.

Definition 5 *Let $A : \mathcal{H} \to \mathcal{H}$ be a bounded linear operator. A is* cyclic *if there is a vector v in $\mathcal{H}$ such that $\mathcal{K}(A, v) = \mathcal{H}$. In this case, v is a* cyclic vector *of A.*

Theorem 6 *Let $A : \mathcal{H} \to \mathcal{H}$ be a bounded linear operator. Assume moreover that A is injective. If all eigenvalues of A are simple then there exists a subset $\mathcal{M} \subset \mathcal{H}$ such that $\overline{\mathcal{M}} = \mathcal{H}$ and $\mathcal{K}(A, v) = \mathcal{H}$ for each $v \in \mathcal{M}$.*

Proof. Let $g_1 \in \mathcal{R}(A)$. By Theorem 3 we know that $A\mathcal{K}(A, g_1) = \mathcal{K}(A, g_1)$. Denoting by $P^{(1)}$ the orthogonal projection onto $\mathcal{K}(A, g_1)$, for each $u \in \mathcal{H}$, $u - P^{(1)}u \perp A\mathcal{K}(A, g_1)$, that is, $A^*(u - P^{(1)}u) \perp \mathcal{K}(A, g_1)$. This implies $P^{(1)}A^*(u - P^{(1)}u) = 0$ and hence $P^{(1)}A^* = P^{(1)}A^*P^{(1)}$, which finally leads to $AP^{(1)} = P^{(1)}AP^{(1)}$. Assuming that $\mathcal{K}(A, g_1) \subsetneq \mathcal{H}$, take $g_2 \in \mathcal{R}(A) \cap \mathcal{K}(A, g_1)^{\perp}$. Now observe that $Ag_2 \notin \mathcal{K}(A, g_1)$. Indeed, by Corollary 4, if $Ag_2 \in \mathcal{K}(A, g_1)$ then $g_2 \in \mathcal{K}(A, g_1)$, that contradicts our choice. Using the same argument we thus have $A^k g_2 \notin \mathcal{K}(A, g_1)$ for each $k \geq 0$. Denoting by $P^{(2)}$ the orthogonal projection onto $\mathcal{K}(A, g_2)$, as before we have $AP^{(2)} = P^{(2)}AP^{(2)}$. Moreover since $g_2 \in \mathcal{R}(A)$ by Theorem 3 we still have $A\mathcal{K}(A, g_2) = \mathcal{K}(A, g_2)$. Assuming $\mathcal{K}(A, g_1, g_2) \subsetneq \mathcal{H}$ we can take $g_3 \in \mathcal{R}(A) \cap \mathcal{K}(A, g_1, g_2)^{\perp}$ and work as before. By induction, since $\overline{\mathcal{R}(A)} = \mathcal{H}$ we are able to construct a finite or countably infinite number of Krylov subspaces $\mathcal{K}(A, g_n)$ with corresponding orthogonal projections $P^{(n)}$ such that the operator A can be written in the following upper triangular form

$$\begin{bmatrix} A_1 & & & * \\ & A_2 & & \\ & & \ddots & \\ 0 & & & \end{bmatrix}, \tag{8}$$

because $AP^{(n)} = P^{(n)}AP^{(n)}$ for each n, and where $A_n = P^{(n)}AP^{(n)}$. Now using [13, Proposition 1.3], we have that if the A_n's are all cyclic with mutually disjoint spectra then A is cyclic. Because of our hypotheses on the spectrum of A, in order to prove that A is cyclic we just need to prove that the A_n's are all cyclic. Using the Arnoldi decomposition we have that each A_n is a quasiaffine transform (through the isometry $W^{(n)}$ defined by the orthonormal basis of $\mathcal{K}(A, g_n)$) of a Hessenberg matrix $H^{(n)}$ whose subdiagonal entries are all strictly positive. Therefore $H^{(n)}$ is cyclic with respect to e_1 ([5, Problem 167]), and hence A_n is cyclic with respect to $g_n = \|g_n\| W^{(n)}e_1$, since quasiaffine transforms preserve cyclicity. Denoting by $\mathcal{M}$ the set of all cyclic vectors, it has been proved in [5, Problem 166] that $\mathcal{M}$ is either empty or dense in $\mathcal{H}$. ■

We remark that in [5, Problem 167] it is also proved that an operator is cyclic if and only if it can be represented by an Hessenberg matrix with all subdiagonal elements different from zero. The density of the set of cyclic vectors therefore tell us that there is a dense subset of starting vectors for the Arnoldi process for which the Arnoldi orthonormal system spans $\mathcal{H}$ and then forms an orthonormal basis.

We conclude this section by showing that for normal operators the representation (8) is in fact diagonal since $AP^{(n)} = P^{(n)}A$. We just need to prove that in this situation $\mathcal{K}(A, v)$ is A^* invariant.

Proposition 7 *Let $A : \mathcal{H} \to \mathcal{H}$ be a compact normal operator. Then for any given $v \in \mathcal{H}$, $\mathcal{K}(A, v)$ is A^* invariant.*

Proof. By (7), we can write $A = \Phi\Lambda\Phi^*$ where Λ is diagonal and Φ is an isometry. Therefore for each polynomial q we have $q(A) = \Phi q(\Lambda)\Phi^*$. Taking q such that $q(\lambda_n) = \lambda_n^*$ we have $q(A) = \Phi\Lambda^*\Phi^* = A^*$. Therefore for each $w \in \mathcal{K}(A, v)$, $A^*w = q(A)w \in \mathcal{K}(A, v)$. ■

4 The Arnoldi-Tikhonov method

The Arnoldi-Tikhonov method, used for the first time in [2], consists in approximating the solution of (2) by means of

$$\begin{aligned} & \min_{u_m \in \mathcal{K}_m(A,f)} \left\{ \|Au_m - f\|^2 + \lambda^2 \|Lu_m\|^2 \right\} \qquad (9) \\ = & \min_{y_m \in \mathbb{C}^m} \left\{ \|AW_m y_m - f\|^2 + \lambda^2 \|LW_m y_m\|^2 \right\} \\ = & \min_{y_m \in \mathbb{C}^m} \left\{ \|W_{m+1}H_m y_m - W_{m+1}e_1 \|f\|\|^2 + \lambda^2 \|Q_m R_m y_m\|^2 \right\} \\ = & \min_{y_m \in \mathbb{C}^m} \left\{ \|H_m y_m - e_1 \|f\|\|^2 + \lambda^2 \|R_m y_m\|^2 \right\}, \qquad (10) \end{aligned}$$

where $Q_m : \mathbb{C}^m \to L\mathcal{K}_m(A, f) \subseteq \mathcal{H}$ is an isometry and $R_m \in \mathbb{C}^{m\times m}$ is upper triangular, such that $Q_m R_m = LW_m$. In this way we obtain

$$u_m = W_m \left(H_m^* H_m + \lambda^2 R_m^* R_m\right)^{-1} H_m^* e_1 \|f\|.$$

For the standard Tikhonov minimization, where L is the identity operator the above formulation still works with $Q_m = W_m$ and $R_m = I_m$, the identity in $\mathbb{C}^m$.

Theorem 8 *For each $m \leq N$ let $u_m = W_m y_m$, where y_m solves (10). Let moreover $u^\dagger$ be the solution of (2). If $\mathcal{N}(A) \cap \mathcal{N}(L) = \{0\}$ then there exists a norm $E(\cdot)$ in $\mathcal{H}$ such that*

$$E(u_m - u^\dagger) \leq E(P_m u^\dagger - u^\dagger). \qquad (11)$$

Proof. Let $r_m = Au_m - f$. Then by (10) and since $f = W_{m+1}e_1 \|f\|$,

$$\begin{aligned} r_m &= AW_m \left[\left(H_m^* H_m + \lambda^2 R_m^* R_m\right)^{-1} H_m^* e_1 \|f\|\right] - W_{m+1}e_1 \|f\| \\ &= W_{m+1}H_m \left[\left(H_m^* H_m + \lambda^2 R_m^* R_m\right)^{-1} H_m^* e_1 \|f\|\right] - W_{m+1}e_1 \|f\| \\ &= W_{m+1} \left[H_m \left(H_m^* H_m + \lambda^2 R_m^* R_m\right)^{-1} H_m^* - I_{m+1}\right] e_1 \|f\|. \end{aligned}$$

Therefore

$$\begin{aligned} (AW_m)^* r_m &= (W_{m+1}H_m)^* r_m \\ &= \left[H_m^* H_m \left(H_m^* H_m + \lambda^2 R_m^* R_m\right)^{-1} H_m^* - H_m^*\right] e_1 \|f\|. \end{aligned}$$

Using the identity

$$H_m^* H_m \left(H_m^* H_m + \lambda^2 R_m^* R_m\right)^{-1} = I_m - \lambda^2 R_m^* R_m \left(H_m^* H_m + \lambda^2 R_m^* R_m\right)^{-1},$$

and the relation $Q_m R_m = L W_m$, we obtain

$$\begin{aligned} (AW_m)^* r_m &= -\lambda^2 R_m^* R_m \left(H_m^* H_m + \lambda^2 R_m^* R_m\right)^{-1} H_m^* e_1 \|f\| \\ &= -\lambda^2 R_m^* R_m y_m \\ &= -\lambda^2 W_m^* L^* L W_m W_m^* u_m. \end{aligned}$$

Since $W_m W_m^* u_m = u_m$ we finally have

$$\begin{aligned} 0 &= W_m^* \left(A^* r_m + \lambda^2 L^* L u_m\right) \\ &= W_m^* \left[\left(A^* A + \lambda^2 L^* L\right) u_m - A^* f\right]. \end{aligned}$$

In other words, u_m is the result of an orthogonal projection method for the linear problem

$$\left(A^* A + \lambda^2 L^* L\right) u = A^* f,$$

since

$$u_m \in \mathcal{K}_m, \quad \left[\left(A^* A + \lambda^2 L^* L\right) u_m - A^* f\right] \perp \mathcal{K}_m.$$

Because the operator $A^* A + \lambda^2 L^* L$ is positive and injective by the hypotheses on A and L, the functional

$$E(z) = \left\langle \left(A^* A + \lambda^2 L^* L\right) z, z \right\rangle^{1/2}, \quad z \in \mathcal{H}, \tag{12}$$

defines a norm in $\mathcal{H}$. By [12, Proposition 5.2] we know that u_m is the result of an orthogonal projection method if and only if

$$E(u_m - u^\dagger) \leq E(\widetilde{u} - u^\dagger), \quad \text{for each } \widetilde{u} \in \mathcal{K}_m.$$

Taking $\widetilde{u} = P_m u^\dagger$ we obtain the result. ■

We remark that the hypothesis $\mathcal{N}(A) \cap \mathcal{N}(L) = \{0\}$ is fulfilled if A or L are injective. In the contrary case $E(\cdot)$ is just a seminorm.

Corollary 9 *Let $A : \mathcal{H} \to \mathcal{H}$ be a compact operator. Assume moreover that A is injective. If f in (9) is cyclic then the Arnoldi-Tikhonov method converges. Moreover if L is bounded then there exists a nonnegative sequence $\{a_j\}_{j\geq 1} \in \ell_2$ such that*

$$E(u_m - u^\dagger)^2 \leq \sum\nolimits_{j>m} a_j^2.$$

Proof. By Definition 5 the Krylov vectors $\{w_j\}_{j\in\mathbb{N}}$ form a basis of $\mathcal{H}$ and hence

$$u^\dagger - P_m u^\dagger = \sum\nolimits_{j>m} \left\langle u^\dagger, w_j \right\rangle w_j.$$

Then, by Parseval's identity,

$$\left\|u^\dagger - P_m u^\dagger\right\|^2 = \sum\nolimits_{j>m} \left|\left\langle u^\dagger, w_j \right\rangle\right|^2,$$

and $\left\|u^\dagger - P_m u^\dagger\right\| \to 0$. Moreover, by (12) we have that

$$E(z)^2 \leq \left\|A^*A + \lambda^2 L^*L\right\| \left\|z\right\|^2,$$

and therefore

$$\begin{aligned} E(u_m - u^\dagger)^2 &\leq \left\|A^*A + \lambda^2 L^*L\right\| \left\|u^\dagger - P_m u^\dagger\right\|^2 \\ &\leq \left\|A^*A + \lambda^2 L^*L\right\| \sum\nolimits_{j>m} \left|\left\langle u^\dagger, w_j \right\rangle\right|^2. \end{aligned}$$

■

We remark that if A is injective and the Arnoldi algorithm terminates in $N < \infty$ steps then GMRES finds the exact solution of $Au = f$ (cf. [9, Section 2]). This means that $f \in \mathcal{R}(A)$ (that is, it satisfies the Picard condition) and hence the problem does not require regularization. More generally if $f \in \mathcal{R}(A)$ then $u \in \mathcal{K}(A, f)$ by Theorem 3 and then GMRES converges. On the other side, if $f \notin \mathcal{R}(A)$ then the GMRES residual stagnates, and the corresponding approximation explodes. Indeed the Picard condition cannot hold, asymptotically, for the projected least squares problem (5)

$$\min_{y \in \mathbb{C}^m} \left\|H_m y - \left\|f\right\| e_1\right\|,$$

because otherwise $A\mathcal{K}(A, f) = \mathcal{K}(A, f)$, that is, $f \in \mathcal{R}(A)$, as pointed out after the proof of Theorem 3.

If A is of finite rank clearly the Arnoldi process terminates in a finite number of steps N. In this situation, under suitable hypotheses on A (for instance if it is normal, cf. [9, Section 2]) GMRES again finds the exact solution of $Au = f$ if $f \in \mathcal{R}(A)$. If $f \notin \mathcal{R}(A)$ then $Au = f$ does not have a solution in $\mathcal{H}$ and the Arnoldi-Tikhonov may fail to converge unless $P_m u^\dagger = u^\dagger$ for $m \leq N$. In this case the only measure of accuracy is given by (11) where $E(\cdot)$ may be a seminorm if L is not injective.

In order to monitor the decay of $\left\|u_m - u^\dagger\right\|$ one can use the so-called complementary condition [8, p.21] which states the existence of a constant $\gamma > 0$ such that

$$\left\|u - v\right\|^2 \leq \gamma^2 \left(\left\|Au - Av\right\|^2 + \left\|Lu - Lv\right\|^2\right), \tag{13}$$

for any $u, v \in \mathcal{H}$. Indeed in [8, Theorem 4] it is proved that (13) is a necessary condition for the existence of a unique solution for (2). Nevertheless, under the hypothesis that L is invertible we can state the following bound.

Proposition 10 *If the operator L is invertible then $\left\|u_m - u^\dagger\right\| \leq \lambda \sigma_{\min}(L) E(u_m - u^\dagger)$.*

Proof. Denoting by $(L^*L)^{1/2}$ the positive square root of the positive operator L^*L, we have

$$\begin{aligned} \left\langle \left(A^*A + \lambda^2 L^*L\right) z, z \right\rangle &= \left\langle \left(A^*A\left(L^*L\right)^{-1} + \lambda^2 I_{\mathcal{H}}\right) L^*Lz, z \right\rangle \\ &= \left\langle \left(L^*L\right)^{1/2} \left(L^*L\right)^{-1/2} \left(A^*A\left(L^*L\right)^{-1} + \lambda^2 I_{\mathcal{H}}\right) \left(L^*L\right)^{1/2} \left(L^*L\right)^{1/2} z, z \right\rangle \\ &= \left\langle \left(L^*L\right)^{-1/2} \left(A^*A\left(L^*L\right)^{-1} + \lambda^2 I_{\mathcal{H}}\right) \left(L^*L\right)^{1/2} x, x \right\rangle, \end{aligned} \tag{14}$$

where $x = (L^*L)^{1/2} z$. Since $(L^*L)^{-1/2} \left(A^*A (L^*L)^{-1} + \lambda^2 I_{\mathcal{H}}\right) (L^*L)^{1/2}$ and $\left(A^*A (L^*L)^{-1} + \lambda^2 I_{\mathcal{H}}\right)$ have the same eigenvalues $\{\lambda_j\}_{j\in\mathbb{N}}$, and $\lambda_j \geq \lambda$ for each j, by (14) we have

$$\begin{aligned} \left\langle \left(A^*A + \lambda^2 L^*L\right) z, z\right\rangle &\geq \lambda^2 \langle x, x\rangle \\ &= \lambda^2 \langle L^*Lz, z\rangle \\ &\geq \lambda^2 \sigma_{\min}(L)^2 \langle z, z\rangle . \end{aligned}$$

■

The above proposition states that if the operator L is invertible (as for the standard Tikhonov regularization where $L = I_{\mathcal{H}}$) then we can monitor the convergence using $\|\cdot\|$. On the other side, if the operator L is compact (this includes the common case of L bounded and of finite rank) then $A^*A + \lambda^2 L^*L$ is also compact (see [1, §2.4]) and therefore $A^*A + \lambda^2 L^*L$ possesses arbitrarily small eigenvalues so that the result of Proposition 10 does not hold anymore.

5 The Lanczos bidiagonalization

The Lanczos (or Golub-Kahan) bidiagonalization process [4] computes two orthonormal bases $\{w_1, ..., w_m\}$ and $\{z_1, ..., z_m\}$ for the Krylov subspaces $\mathcal{K}_m(A^*A, A^*f)$ and $\mathcal{K}_m(AA^*, f)$ respectively, for $m \leq N = \min\{\sup_n (\dim \mathcal{K}_n(A^*A, A^*f)), \sup_n (\dim \mathcal{K}_n(AA^*, f))\}$. Since $\mathcal{K}_m(A^*A, A^*f) = A^*\mathcal{K}_m(AA^*, f)$ for $m \leq N$, If A is injective then $N = \sup_n (\dim \mathcal{K}_n(A^*A, A^*f))$. Defining the isometries W_m, Z_m, corresponding to the two bases as in (3), the following decomposition holds

$$AW_m = Z_{m+1}B_m, \tag{15}$$

where $B_m \in \mathbb{C}^{(m+1)\times m}$ is lower bidiagonal. LSQR method [10] is defined by solving at each step

$$\min_{u\in\mathcal{K}_m(A^*A,A^*f)} \|Au - f\| ,$$

whose solution is given by $u_m = W_m y_m$, where

$$y_m = \arg\min_{y\in\mathbb{C}^m} \|B_m y - \|f\| e_1\| .$$

Observe that for the Lanczos bidiagonalization we have $f = Z_{m+1}e_1 \|f\|$. Then, using (15) and working as in (9)-(10) the Lanczos-Tikhonov method is based on the solution of

$$\min_{u_m\in\mathcal{K}_m(A^*A,A^*f)} \left\{\|Au_m - f\|^2 + \lambda^2 \|Lu_m\|^2\right\} \tag{16}$$

$$= \min_{y_m\in\mathbb{C}^m} \left\{\|B_m y_m - e_1 \|f\|\|^2 + \lambda^2 \|R_m y_m\|^2\right\} . \tag{17}$$

Theorem 11 *The results of Theorem 8 and Corollary 9 hold also for the Lanczos-Tikhonov method.*

Proof. As for Theorem 8, the proof is almost identical. We just need to replace W_{m+1} and H_m by Z_{m+1} and B_m respectively. P_m is now the orthogonal projection onto $\mathcal{K}_m(A^*A, A^*f)$. The decomposition (15) is used in place of (4). The proof of the result of Corollary 9 remains unaltered provided that A^*f is a cyclic vector for A^*A. ■

The case of finite rank compact operators is simpler when working with the Lanczos bidiagonalization. Indeed, without hypothesis on A, if $f \in \mathcal{R}(A)$ then LSQR finds the exact solution of $Au = f$ in a finite number of steps and hence the regularization is not necessary. Like the Arnoldi-Tikhonov, if $f \notin \mathcal{R}(A)$ then $Au = f$ does not have a solution in $\mathcal{H}$ and the Lanczos-Tikhonov may fail to converge unless $P_m u^\dagger = u^\dagger$ for $m \leq N$. The main advantage of the Lanczos bidiagonalization is that if $L = I_{\mathcal{H}}$ the method converges even if A is of finite rank since $u^\dagger = (A^*A + \lambda^2 I_{\mathcal{H}})^{-1} A^* f$ so that $u^\dagger \in \mathcal{K}_m(A^*A, A^*f)$ for some $m < \infty$.

References

[1] Y. A. Abramovich, C. D. Aliprantis, Problems in operator theory, AMS Graduate Studies in Mathematics, Volume 51, 2002.

[2] D. Calvetti, S. Morigi, L. Reichel, F.Sgallari, Tikhonov regularization and the L-curve for large discrete ill-posed problems, J. Comput. Appl. Math. 123 (2000), pp. 423–446.

[3] S. Gazzola, P. Novati, M. R. Russo, On Krylov projection methods and Tikhonov regularization, Electron. Trans. Numer. Anal. 44 (2015), pp. 83–123.

[4] G. Golub, W. Kahan, Calculating the singular values and pseudo-inverse of a matrix, J. Soc. Indust. Appl. Math. Ser. B Numer. Anal. 2 (1965), pp. 205–224.

[5] P.R. Halmos, A Hilbert space problem book, 2nd ed., Springer-Verlag, New York, 1982.

[6] P. C. Hansen, Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion. SIAM, Philadelphia, 1998.

[7] P. C. Hansen, The discrete Picard condition for discrete ill-posed problems, BIT 30 (1990), pp. 658–672.

[8] V. A. Morozov. Regularization methods for ill-posed problems. CRC press, 1993.

[9] P. Novati, Some properties of the Arnoldi based methods for linear ill-posed problems, SIAM J. Numer. Anal. 55 (2017), pp. 1437-1455.

[10] C. C. Paige, M. A. Saunders, LSQR: an algorithm for sparse linear equations and and sparse least squares, ACM Trans. Math. Software 8 (1982), pp. 43–71.

[11] J. R. Ringrose, Compact non-self-adjoint operators. Van Nostrand Reinhold Company London, 1971.

[12] Y. Saad, Iterative Methods for Sparse Linear Systems, 2nd ed. SIAM, Philadelphia, 2003.

[13] P. Y. Wu, Sums and products of cyclic operators, Proc. Amer. Math. Soc. 122 (1994), pp. 1053–1063.