A rational Arnoldi approach for ill-conditioned linear systems

C. Brezinski* P. Novati † M. Redivo–Zaglia ‡

August 8, 2011

Abstract

For the solution of full-rank ill-posed linear systems a new approach based on the Arnoldi algorithm is presented. Working with regularized systems, the method theoretically reconstructs the true solution by means of the computation of a suitable function of matrix. In this sense the method can be referred to as an iterative refinement process. Numerical experiments arising from integral equations and interpolation theory are presented. Finally, the method is extended to work in connection with the standard Tikhonov regularization with a right hand side contaminated by noise.

Keywords: Ill-conditioned linear systems. Arnoldi algorithm. Matrix function. Tikhonov regularization.

1 Introduction

In this paper we consider the solution of ill-conditioned linear systems

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

We mainly focus the attention on linear systems in which $A \in \mathbb{R}^{N\times N}$ is full rank with singular values that gradually decay to 0, as for instance in the case of the discretized Fredholm integral equations of the first kind. In order face this kind of problems one typically apply some regularization technique such as the well known Tikhonov regularization (see e.g. [18] for a wide background). The Tikhonov regularized system takes the form

$$(A^TA + \lambda H^TH)x_\lambda = A^Tb, \tag{2}$$

where $\lambda \in \mathbb{R}$ is a suitable parameter and H is the regularization matrix. The system (2) should have singular values bounded away from 0 in order to reduce the condition number and, at the same time, its solution x_λ should be closed to the solution of the original system.

*Laboratoire Paul Painlevé, UMR CNRS 8524, UFR de Mathématiques Pures et Appliquées, Université des Sciences et Technologies de Lille, 59655–Villeneuve d'Ascq cedex, France. E–mail: Claude.Brezinski@univ-lille1.fr

†Università degli Studi di Padova, Dipartimento di Matematica Pura ed Applicata, Via Trieste 63, 35121–Padova, Italy. E–mail: novati@math.unipd.it

‡Università degli Studi di Padova, Dipartimento di Matematica Pura ed Applicata, Via Trieste 63, 35121–Padova, Italy. E–mail: Michela.RedivoZaglia@unipd.it

For this kind of problem the method initially presented in this paper is based on the shift and invert transformation

$$Z_\lambda = (A + \lambda I)^{-1}, \tag{3}$$

where $\lambda > 0$ is a suitable parameter and I is the identity matrix. Provided that λ is large enough, if A is positive definite ($F(A) \subset \mathbb{C}^+ = \{z \in \mathbb{C} : \mathrm{Re}(z) > 0\}$, where $F(A)$ denotes the field of values) the shift $A + \lambda I$, that represents the most elementary example of regularization, has the immediate effect of moving the spectrum (that we denote by $\sigma(A)$) away from 0 so reducing the condition number. Moreover, since

$$x = A^{-1}b = f_\lambda(Z_\lambda)b,$$

where

$$f_\lambda(z) = \left(\frac{1}{z} - \lambda\right)^{-1} = (1 - \lambda z)^{-1} z, \tag{4}$$

the idea is to solve the system $Ax = b$ by computing $f_\lambda(Z_\lambda)b$. For the computation of $f_\lambda(Z_\lambda)b$, we use the standard Arnoldi method projecting the matrix Z_λ onto the Krylov subspaces generated by Z_λ and b, that is $K_m(Z_\lambda, b) = \mathrm{span}\{b, Z_\lambda b, ..., Z_\lambda^{m-1} b\}$. By definition of Z_λ the method is commonly referred to as the Restricted-Denominator (RD) rational Arnoldi method [12], [26].

Historically, a first attempt to reconstruct the solution from x_λ that solves

$$(A + \lambda I)\, x_\lambda = b, \tag{5}$$

was proposed by Riley in [29]. The algorithm is just based on the approximation of $f_\lambda(Z_\lambda)$ by means of its Taylor series. Indeed we have

$$A^{-1}b = \frac{1}{\lambda} \sum_{k=1}^{\infty} (\lambda Z_\lambda)^k b, \tag{6}$$

that leads to the recursion

$$x_{k+1} = y + \lambda Z_\lambda x_k, \quad x_0 = 0, \quad y = Z_\lambda b. \tag{7}$$

It is easy to see that the method is equivalent to the *iterative improvement*

$$\begin{aligned} (A + \lambda I)\, e_k &= b - A x_k \\ x_{k+1} &= x_k + e_k \end{aligned}$$

generally referred to as *iterated Tikhonov regularization* or *preconditioned Landweber iteration* (see e.g. [15], [20], [22], [23], [27]). The main problem concerning this kind of algorithms is that they can be extremely slow because the spectrum of Z_λ accumulates at $1/\lambda$ (cf. (3), (6)). This, of course, for large values of λ, that is, when $A + \lambda I$ is well conditioned. From the point of view of the computation of function of matrices this is a well known problem, i.e., the the computation by means of the Taylor series generally provides poor results unless the spectrum of the matrix is close to the expansion point. Indeed, from well known results of complex approximation, the rate of convergence of a polynomial method for the computation of a function of matrix depends on the position of the singularity of the function, with respect to the location of the spectrum of the matrix.

We also point out that, in [7], the authors construct an improved approximation via extrapolation with respect to the regularization parameter, using the singular values representation of

the solution. Extrapolation techniques can also be applied to accelerate (7), as suggested in [6] and also indicated by Fasshauer in [13].

For problems in which the right hand side is affected by noise, instead of working with the transformation (3) or implicitly with systems of type (5), we shall work with the standard regularization (2) and hence on the transformation

$$Z_\lambda = (A^T A + \lambda H^T H)^{-1}.$$

As we shall see, the subsequent Arnoldi-based algorithm for the reconstruction of the exact solution will be almost identical to the one based on (3), but the use of a regularization matrix H different from the identity allows to define methods less sensitive to perturbations on the right hand side.

The paper is organized as follows. In Section 2, we describe the Arnoldi method for the computation of $f_\lambda(Z_\lambda)b$ and, in Section 3, we present a theoretical a-priori error analysis. In Section 4, we show an a-posteriori representation of the error. In Section 5, we analyze the choice of the parameter λ. Some numerical experiments taken out from Hansen's Matlab toolbox on regularization [17, 19], and from the theory of interpolation with radial basis functions are presented in Section 6. Finally, in Section 7, we extend our method to the Tikhonov regularization in its general form (2) showing also some tests with data affected by noise. In this section we also consider a simmetric alternative of the method that allows to reduce the computational costs working with the Lanczos algorithm.

2 The Arnoldi method for $f_\lambda(Z_\lambda)b$.

For the construction of the subspaces $K_m(Z_\lambda, b)$, the Arnoldi algorithm generates an orthonormal sequence $\{v_j\}_{j\geq 0}$, with $v_1 = b/\|b\|$, such that $K_m(Z_\lambda, b) = \text{span}\{v_1, v_2, ..., v_m\}$ (here and below the norm used is always the Euclidean norm). For every m we have

$$Z_\lambda V_m = V_m H_m + h_{m+1,m} v_{m+1} e_m^T, \tag{8}$$

where $V_m = [v_1, v_2, ..., v_m]$, H_m is an upper Hessenberg matrix with entries $h_{i,j} = v_i^T Z_\lambda v_j$ and e_j is the j-th vector of the canonical basis of $\mathbb{R}^m$. Formula (8) is just the matrix formulation of the algorithm.

The m-th Arnoldi approximation to $x = f_\lambda(Z_\lambda)b$ is defined as

$$x_m = \|b\| V_m f_\lambda(H_m) e_1.$$

Regarding the computation $f_\lambda(H_m)$, since the method is expected to produce a good approximation of the solution in a relatively small number of iterations, that is for $m \ll N$, one typically considers a certain rational approximation to f_λ, or the Schur-Parlett algorithm (see e.g. [16, Chapter 11] or [21]).

Denoting by Π_{m-1} the vector space of polynomials of degree at most $m-1$, it can be seen that

$$x_m = \overline{p}_{m-1}(Z_\lambda)b, \tag{9}$$

where $\overline{p}_{m-1} \in \Pi_{m-1}$ interpolates, in the Hermite sense, the function f_λ at the eigenvalues of H_m [30].

As already mentioned, this kind of approach is commonly referred to as the RD rational Arnoldi method since it is based on the use of single pole rational forms of the type

$$R_{m-1}(x) = \frac{q_{m-1}(x)}{(x+a)^{m-1}}, \quad a \in \mathbb{R}, \quad q_{m-1} \in \Pi_{m-1}, \quad m \geq 1,$$

introduced and studied by Nørsett in [28] for the approximation of the exponential function. In other words, with respect to A, formula (9) is actually a rational approximation.

It is worth noting that, at each step of the Arnoldi algorithm, we have to compute the vectors $w_j = Z_\lambda v_j$, $j \geq 1$, which leads to solve the systems

$$(A + \lambda I)w_j = v_j, \quad j \geq 1.$$

Since $v_1 = b/\|b\|$, the corresponding w_1 is just the scaled solution of a regularized system (with the rough regularization $A \to A + \lambda I$). In this sense if λ arises from the standard techniques that seek for the optimal regularization parameter λ_{opt} (L-curve, Generalized Cross Validation, etc.) this procedure can be employed as a tool to improve the quality of the approximation $w_1\|b\|$. Anyway we shall see that, using the Arnoldi algorithm, larger values for λ are more reliable.

3 Error analysis

The error $E_m := x - x_m$ can be expressed and bounded in many ways (see e.g. the recent paper [1] and the references therein). In any case, however, the sharpness of the bound essentially depends on the amount of information about the location of the field of values of Z_λ, defined by

$$F(Z_\lambda) := \left\{ \frac{x^H Z_\lambda x}{x^H x}, x \in \mathbb{C}^N \setminus \{0\} \right\}.$$

The bound we propose is based on the use of Faber polynomials. We need some definitions and we refer to [31] or [32] for a wide background of what follows.

Let Ω be a compact set of the complex plane with simply connected complement. By the Riemann mapping theorem there exists a conformal surjection

$$\psi : \overline{\mathbb{C}} \setminus \{w : |w| \leq 1\} \to \overline{\mathbb{C}} \setminus \Omega, \quad \psi(\infty) = \infty, \quad \psi'(\infty) = \gamma, \tag{10}$$

that has a Laurent expansion of the type

$$\psi(w) = \gamma w + c_0 + \frac{c_1}{w} + \frac{c_2}{w^2} + \cdots$$

The constant γ is the capacity of Ω. If Ω is an ellipse or a line segment then $c_i = 0$ for $i \geq 2$. Let us denote by $\|\cdot\|_\Omega$ the uniform norm on Ω. Given a function g analytic in Ω, it is known that defining p_{m-1} as the truncated Faber series of exact degree $m-1$ with respect to g and ψ, then p_{m-1} provides an asymptotically optimal uniform approximation to g in Ω, that is

$$\lim_{m\to\infty} \sup \|p_{m-1} - g\|_\Omega^{1/m} = \lim_{m\to\infty} \sup \left\|p^*_{m-1} - g\right\|_\Omega^{1/m}, \tag{11}$$

$\left\{p^*_{m-1}(z)\right\}_{m\geq 1}$ being the sequence of polynomials of best uniform approximation to g in Ω. Property (11) is also called *maximal convergence*. Let moreover $\phi : \overline{\mathbb{C}} \setminus \Omega \to \overline{\mathbb{C}} \setminus \{w : |w| \leq 1\}$ be the inverse of ψ. For any $r > 1$, let Γ_r be the equipotential curve

$$\Gamma_r := \{z : |\phi(z)| = r\},$$

and let us denote by Ω_r the bounded domain with boundary Γ_r. Let $\widehat{r} > 1$ be the largest number such that g is analytic in Ω_r for each $\gamma < r < \widehat{r}$ and has a singularity on $\Gamma_{\widehat{r}}$. Then, it is known that the rate of convergence of the sequence $\{p_{m-1}(z)\}_{m\geq 1}$ is given by

$$\lim_{m\to\infty} \sup \|p_{m-1} - g\|_\Omega^{1/m} = \frac{1}{\widehat{r}}. \tag{12}$$

For this reason we know that superlinear convergence is only attainable for entire functions, where asymptotically one can set $\widehat{r} := m$. In order to derive error bounds for the computation of $f_\lambda(Z_\lambda)b$ we need the following classical result

Theorem 1 *[11] Let Ω be a compact and convex subset such that g is analytic in Ω. For $1 < r < \widehat{r}$ the following bound holds*

$$\|p_{m-1} - g\|_\Omega \leq 2\,\|g\|_{\Gamma_r} \frac{\left(\frac{1}{r}\right)^m}{1 - \frac{1}{r}}. \tag{13}$$

Using the above theorem, for our function $f_\lambda(z) = z/(1-\lambda z)$, singular at $1/\lambda$, we can state the

Proposition 2 *Assume that Ω is an ellipse of the complex plane, symmetric with respect to the real axis with associated conformal mapping $\psi(w) = \gamma w + c_0 + c_1/w$. Assume that $\psi(1) < 1/\lambda$ and let $\widehat{r}$ be such that $\psi(\widehat{r}) = 1/\lambda$. Let moreover $\overline{m}$ be the smallest integer such that*

$$\frac{\widehat{r}}{\overline{m}+1} < \widehat{r} - 1.$$

Then for $m \geq \overline{m}$

$$\|p_{m-1} - f_\lambda\|_\Omega \leq \frac{2\,e\,\overline{m}\,\widehat{r}}{\overline{m}(\widehat{r}-1) - 1} \frac{1}{\lambda^2 \psi'(\widehat{r})} \frac{m+1}{\widehat{r}^m}, \tag{14}$$

and for $m < \overline{m}$

$$\|p_{m-1} - f_\lambda\|_\Omega \leq \frac{4}{\lambda^2\,(\widehat{r}-1)\,\psi'(\widehat{r})} \left(\frac{2}{\widehat{r}+1}\right)^m \frac{\widehat{r}+1}{\widehat{r}-1}. \tag{15}$$

Proof. Let $r = \widehat{r} - \varepsilon$, with $0 < \varepsilon < \widehat{r} - 1$. By the properties of Ω, we have

$$\|f_\lambda\|_{\Gamma_r} = \frac{\psi(r)}{1 - \lambda\psi(r)},$$

and, by direct computation

$$\psi(r) = \psi(\widehat{r}) - \gamma\varepsilon + \frac{c_1\varepsilon}{(\widehat{r}-\varepsilon)\widehat{r}}.$$

Hence using $\psi(\widehat{r}) = 1/\lambda$ we find

$$\begin{aligned} \|f_\lambda\|_{\Gamma_r} &\leq \frac{\psi(\widehat{r})}{1 - \lambda\left(\psi(\widehat{r}) - \gamma\varepsilon + \frac{c_1\varepsilon}{(\widehat{r}-\varepsilon)\widehat{r}}\right)}, \\ &= \frac{1}{\lambda^2\varepsilon\left(\gamma - \frac{c_1}{(\widehat{r}-\varepsilon)\widehat{r}}\right)}, \\ &\leq \frac{1}{\lambda^2\varepsilon\psi'(\widehat{r})}. \end{aligned}$$

By (13), we thus obtain

$$\|p_{m-1} - f_\lambda\|_\Omega \leq \frac{2}{\lambda^2\varepsilon\psi'(\widehat{r})} \frac{1}{(\widehat{r}-\varepsilon)^m} \frac{1}{1 - \frac{1}{\widehat{r}-\varepsilon}}. \tag{16}$$

Now setting

$$\varepsilon = \frac{\widehat{r}}{m+1}, \tag{17}$$

since this value minimizes

$$\frac{1}{\varepsilon\left(\widehat{r}-\varepsilon\right)^{m}},$$

let $\overline{m}$ be the smallest positive integer such that

$$\frac{\widehat{r}}{\overline{m}+1} < \widehat{r} - 1.$$

By inserting (17) into (16) and using

$$\frac{1}{1-\dfrac{1}{\widehat{r}-\varepsilon}} \leq \frac{\overline{m}\widehat{r}}{\overline{m}(\widehat{r}-1)-1},$$

we find (14). For $m < \overline{m}$ we can take for instance

$$\varepsilon = \frac{\widehat{r}-1}{2}. \tag{18}$$

Substituting (18) into (16) we obtain (15). ■

Remark 3 *Note that the assumption $\psi(1) < 1/\lambda$ in Proposition 2 just means that the ellipse is strictly on the left of the singularity of f_λ.*

Regarding the field of values of Z_λ, $F(Z_\lambda)$, it is well known that it is convex, that $\sigma(Z_\lambda) \subset F(Z_\lambda)$, and that $F(H_m) \subseteq F(Z_\lambda)$ (where H_m is defined in Section 2). Moreover we need the following

Proposition 4 *If $F(A) \subset \mathbb{C}^+$ (A is positive definite) then $F(Z_\lambda) \subset B_\lambda := \{z \in \mathbb{C} : 0 < \mathrm{Re}(z) < 1/\lambda\}$.*

Proof. Let $\chi(a) = (a+\lambda)^{-1}$ for $\mathrm{Re}(a) \geq 0$, $\lambda > 0$. The function χ maps the imaginary axis onto the circumference of the disk D_λ centered at $1/(2\lambda)$ and with radius $1/(2\lambda)$. If $\mathrm{Re}(a) > 0$ then $\chi(a) \in int(D_\lambda)$, the interior part of D_λ. Obviously $\sigma(Z_\lambda) = \chi(\sigma(A))$, so $F(Z_\lambda)$ cannot lie entirely outside $D_\lambda \subset B_\lambda$. We want prove the proposition showing that the intersection between $F(Z_\lambda)$ and the boundary of B_λ is empty.

Assume there exists $s \in \mathbb{R}$ such that $1/\lambda + is \in F(Z_\lambda)$. Hence, there exists $y \in \mathbb{C}^M$, $\|y\| = 1$, such that

$$y^H \left(A+\lambda I\right)^{-1} y = \frac{1}{\lambda} + is. \tag{19}$$

Defining $x := (A+\lambda I)^{-1} y$ we easily obtain

$$x^H \left(A^T + \lambda I\right) x = \frac{1}{\lambda} + is,$$

and hence

$$\frac{x^H A^T x}{x^H x} + \lambda = \left(\frac{1}{\lambda} + is\right) \frac{1}{\|x\|^2}. \tag{20}$$

By (19) we have

$$\|x\| \geq \left| \frac{1}{\lambda} + is \right|. \tag{21}$$

Now let us define $a := \frac{x^H A^T x}{x^H x} \in F(A)$. By (20) we have

$$\|x\|^2 = \left(\frac{1}{\lambda} + is \right) (a + \lambda)^{-1}, \tag{22}$$

and since by hypothesis $(a+\lambda)^{-1} \in int(D_\lambda)$ we clearly have

$$\left|(a+\lambda)^{-1}\right| < 1/\lambda \leq \left| \frac{1}{\lambda} + is \right|,$$

so that by (22)

$$\|x\| < \left| \frac{1}{\lambda} + is \right|,$$

that contradicts (21).

Now assume there exists $s \in \mathbb{R}$ such that $is \in F(Z_\lambda)$. Hence, there exists $y \in \mathbb{C}^M$, $\|y\| = 1$, such that

$$y^H (A + \lambda I)^{-1} y = is.$$

Defining as before $x := (A + \lambda I)^{-1} y$ we have

$$x^H \left(A^T + \lambda I \right) x = is,$$

and hence

$$\frac{x^H A^T x}{x^H x} + \lambda = \frac{is}{\|x\|^2},$$

that contraddicts the hypothesis.

Since the field of values is connected the proof is complete. ■

We are now on the point to demonstrate the following

Theorem 5 *Assume that $F(A) \subset \mathbb{C}^+$. Let $\Omega \subset int(B_\lambda)$ be an ellipse (with associated conformal mapping ψ, and inverse ϕ) symmetric with respect to the real axis and such that $F(Z_\lambda) \subseteq \Omega$. Then, for m large enough, we have*

$$\|E_m\| \leq 4\, e\, C \frac{\widehat{r}}{\widehat{r}-1} \frac{1}{\psi'(\widehat{r})}\, K\, \frac{m+1}{\widehat{r}^m}, \tag{23}$$

where $K = 1/\lambda^2$, $\widehat{r} = \phi(1/\lambda)$, and $C = 2 + 2/\sqrt{3}$ ($C = 1$ if A is symmetric).

Proof. Observe first that f_λ is analytic in Ω. Using the properties of the Arnoldi algorithm, we know that for every $p_{m-1} \in \Pi_{m-1}$,

$$V_m p_{m-1}(H_m) e_1 = p_{m-1}(Z_\lambda) b. \tag{24}$$

Hence, from (24), it follows that, for $m \geq 1$ and for every $p_{m-1} \in \Pi_{m-1}$,

$$E_m = x - x_m = f_\lambda(Z_\lambda) b - p_{m-1}(Z_\lambda) b - V_m (f_\lambda(H_m) - p_{m-1}(H_m)) e_1. \tag{25}$$

Since $\|V_m\| = 1$ we have (see [2])

$$\|E_m\| \leq 2C \left\| p_{m-1} - f_\lambda \right\|_{F(Z_\lambda)}. \tag{26}$$

Therefore taking p_{m-1} as the $(m-1)$-th truncated Faber (Chebyshev) series, the result follows from Proposition 2 since $F(Z_\lambda) \subseteq \Omega$. ■

Remark 6 *Using the theory developed in [1] based on the use of the Faber transform we have that for* $1 < r < \widehat{r}$

$$\|E_m\| \leq 4 \|f_\lambda\|_{\Gamma_r} \frac{\left(\frac{1}{r}\right)^m}{1 - \frac{1}{r}}.$$

Using the bound for $\|f_\lambda\|_{\Gamma_r}$ *given in Proposition 2, we obtain (23) with* $C = 2$ *(cf. [1] Theorem 3.2 and Remark 3.3).*

Theorem 5 is surely important from a theoretical point of view since it states that the Arnoldi algorithm produces asymptotically optimal approximations. However, if we consider for simplicity the symmetric case, we can also understand that it cannot be used to suggest the choice of λ.

Indeed, let $\lambda_1 \gtrsim 0$ and λ_N be respectively the smallest and the largest eigenvalues A. Then $F(A) = [\lambda_1, \lambda_N]$ and

$$F(Z_\lambda) = \left[\frac{1}{\lambda_N + \lambda}, \frac{1}{\lambda_1 + \lambda}\right] =: I_\lambda.$$

In this case, by (26) we have

$$\|E_m\| \leq 2 \max_{I_\lambda} |f_\lambda(z) - p_{m-1}(z)|.$$

As already mentioned, the conformal mapping ψ associated to I_λ takes the form

$$\psi(w) = \gamma w + c_0 + \frac{c_1}{w} \tag{27}$$

where

$$\begin{aligned} \gamma &= \frac{1}{4}\left(\frac{1}{\lambda_1 + \lambda} - \frac{1}{\lambda_N + \lambda}\right) = \frac{1}{4}\frac{\lambda_N - \lambda_1}{(\lambda_1 + \lambda)(\lambda_N + \lambda)}, \\ c_0 &= \frac{1}{2}\left(\frac{1}{\lambda_1 + \lambda} + \frac{1}{\lambda_N + \lambda}\right) = \frac{1}{2}\frac{\lambda_N + \lambda_1 + 2\lambda}{(\lambda_1 + \lambda)(\lambda_N + \lambda)}, \\ c_1 &= \gamma. \end{aligned} \tag{28}$$

For $r > 1$, Ω_r is the confocal ellipse (foci in $\dfrac{1}{\lambda_N + \lambda}$ and $\dfrac{1}{\lambda_1 + \lambda}$) described by $\psi(re^{i\theta})$, $0 \leq \theta < 2\pi$. Since $f_\lambda(z)$ is singular at $1/\lambda$, $\widehat{r}$ is the solution (> 1) of

$$\gamma \widehat{r} + c_0 + \frac{\gamma}{\widehat{r}} = \frac{1}{\lambda} \tag{29}$$

that is

$$\widehat{r} = u + \sqrt{u^2 - 1}, \tag{30}$$

where

$$u = \frac{2\lambda_1\lambda_N}{\lambda(\lambda_N - \lambda_1)} + \frac{\lambda_N + \lambda_1}{\lambda_N - \lambda_1}. \tag{31}$$

Thus, $\widehat{r}$ monotonically decreases with respect to λ and $\widehat{r} \to \infty$ for $\lambda \to 0$.

The above arguments simply show that the error analysis does not take into account of the computational problems in the inversion of $A + \lambda I$ for $\lambda \approx 0$. The method is very fast for $\lambda \approx 0$ because, at each step, we are inverting something very close to the original operator A. In order to derive a more useful estimate one should modify the above analysis imposing in some way the requirement $\lambda \gg \lambda_1$. In some sense this will be done in Section 5 where we consider the conditioning in the computation of $f_\lambda(Z_\lambda)b$ that is obviously closely related to the rate of convergence of any iterative method.

4 A-posteriori error representation

By a result on Padé–type approximation proved in [4], we know that the Hermite interpolation polynomial of the function

$$g(s) = \frac{1}{1 - st}$$

at the zeros of any polynomial ν_m of exact degree m in s is given by

$$R_{m-1}(s) = \frac{1}{1 - st}\left(1 - \frac{\nu_m(s)}{\nu_m(t^{-1})}\right).$$

Setting $\lambda = t^{-1}$, we have that

$$f_\lambda(\xi) = \frac{1}{\xi^{-1} - \lambda} = -\lambda^{-1} g\left(\xi^{-1}\right),$$

and so

$$-\lambda^{-1} R_{m-1}(\xi^{-1}) = \frac{1}{1 - \xi^{-1}\lambda^{-1}}\left(1 - \frac{\nu_m(\xi^{-1})}{\nu_m(\lambda)}\right) \tag{32}$$

interpolates $f_\lambda(\xi)$. By (9) let $\overline{p}_{m-1} \in \Pi_{m-1}$ be the polynomial that interpolates, in the Hermite sense, the function $f_\lambda(z)$ at the eigenvalues of H_m, $\xi_1, ..., \xi_{m'}$, $m' \leq m$, with multiplicity k_i, $i = 1, ..., m'$.Then

$$\overline{p}_{m-1}^{(j)}(\xi_i) = -\lambda^{-1} R_{m-1}^{(j)}(\xi_i^{-1}) = f_\lambda^{(j)}(\xi_i), \quad 1 \leq i \leq m',\ 0 \leq j \leq k_i - 1.$$

By (32) and using the above relation is it easy to see that $\nu_m(s) = \det(sI - H_m^{-1})$. In this way, by direct computation,

$$\begin{aligned} x_m &= \overline{p}_{m-1}(Z_\lambda) b, \\ &= A^{-1} b - A^{-1}\left(\frac{\nu_m(Z_\lambda^{-1})}{\nu_m(\lambda)}\right) b. \end{aligned} \tag{33}$$

Since, of course, A^{-1} and Z_λ^{-1} commute, we find

$$\frac{\|x_m - x\|}{\|x\|} \leq \frac{\|\nu_m(A + \lambda I)\|}{|\nu_m(\lambda)|}.$$

A posteriori error estimate can be derived in this way. Since

$$\begin{aligned} \nu_m(s) &= \det(sI - H_m^{-1}), \\ &= \frac{s^m \det(H_m - s^{-1} I)}{\det H_m}, \end{aligned}$$

defining $q_m(\xi) = \det(H_m - \xi I)$, we have

$$\frac{\|x_m - x\|}{\|x\|} \leq \frac{\|(A + \lambda I)^m q_m(Z_\lambda)\|}{\lambda^m \left|q_m(\lambda^{-1})\right|}. \tag{34}$$

It is worth noting that, using the relation

$$q_m(Z_\lambda) b = \left(\prod_{j=1}^m h_{j+1,j}\right) v_{m+1},$$

(see [26]), we obtain from (33)

$$\|x_m - x\| = \frac{\left(\prod_{j=1}^m h_{j+1,j}\right)}{\lambda^m \left|q_m(\lambda^{-1})\right|} \left\|A^{-1}(A + \lambda I)^m v_{m+1}\right\|,$$

which proves the convergence in a finite number $m^* \leq N$ of steps of the method in exact arithmetics. Note that by (33) the corresponding ν_{m^*} is the minimal polynomial of $A + \lambda I$ for the vector b.

5 The choice of λ

As already mentioned, the arguments of Section 3 reveal that the standalone error analysis of the computation of $f_\lambda(Z_\lambda)b$ is not reliable to suggest the choice of λ, since $\kappa(Z_\lambda) \to \kappa(A)$ as $\lambda \to 0$ ($\kappa(\cdot)$ denoting the standard condition number of a matrix). In other words, it does not take into account that, at each step, we need to solve a system with the matrix $A + \lambda I$. At the same time, focusing the attention on the accuracy (so neglecting the rate of convergence) one could expect that "large" values of λ should allow an improvement of it, since the linear systems with $A + \lambda I$ would be solved more accurately. The numerical experiments show that this is not true, as shown in Fig. 1, where we consider the problem BAART, taken out from the Hansen's Matlab toolbox `Regtools` (see [17] and [19]).

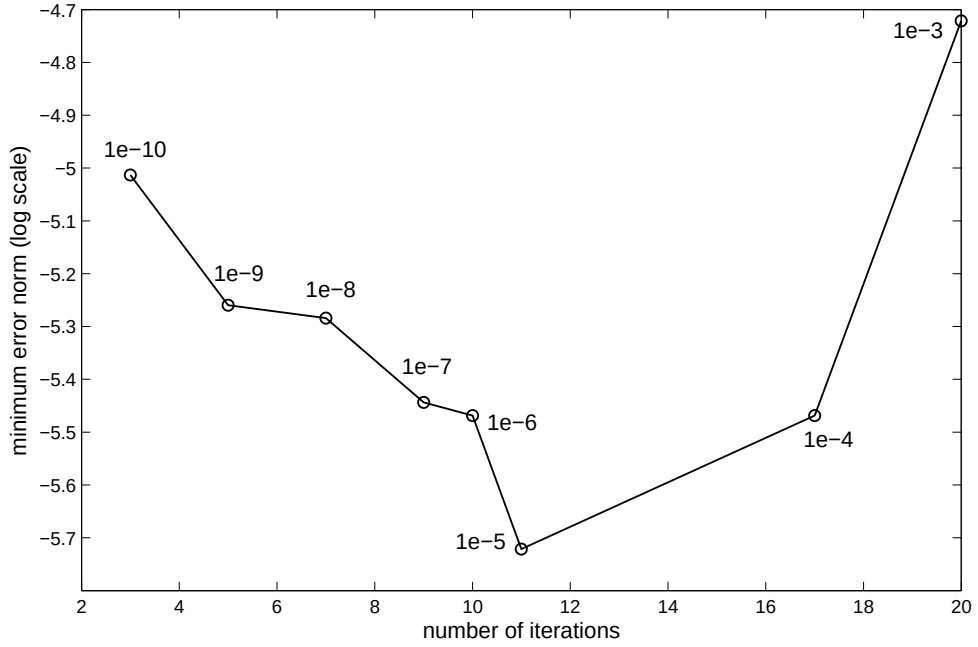

Figure 1: BAART(40) - Minimum attained error with respect to the number of iterations for different values of λ.

Indeed the diagram of Fig. 1 represents the standard situation, that is, increasing λ, we have a loss of accuracy. The behavior on the leftmost part of the diagram is clear since it is due to the conditioning of Z_λ for λ small. On the rightmost part we have again a loss of accuracy but now it depends on the numerical instability in the computation of $f_\lambda(Z_\lambda)$ for λ large (the problem can be easily observed even working scalarly). This observation leads us to consider the conditioning in the computation of $f_\lambda(Z_\lambda)b$ for having a good strategy to define λ.

The absolute and the relative condition number for the computation of $g(X)$ where g is a given smooth function and X a square matrix are given by (cf. [21] Chapter 3)

$$\kappa_a(g, X) = \lim_{\varepsilon \to 0} \sup_{\|E\| \leq \varepsilon} \frac{\|g(X + E) - g(X)\|}{\varepsilon}, \tag{35}$$

$$\kappa_r(g, X) = \kappa_a(g, X) \frac{\|X\|}{\|g(X)\|}, \tag{36}$$

and these definitions imply that

$$\|g(X+E)-g(X)\| \leq \kappa_a(g,X)\,\|E\| + O(\|E\|^2).$$

Proposition 7 *For the function $f_\lambda(z) = (1-\lambda z)^{-1}z$ we have the bound*

$$\kappa_r(f_\lambda, Z_\lambda) \leq \frac{\|(I-\lambda Z_\lambda)^{-2}\|\,\|Z_\lambda\|}{\left\|(Z_\lambda^{-1}-\lambda I)^{-1}\right\|}. \tag{37}$$

Proof. In order to derive first the absolute condition number we have

$$\begin{aligned} f_\lambda(Z_\lambda+E)-f_\lambda(Z_\lambda) &= \left[(Z_\lambda+E)^{-1}-\lambda I\right]^{-1} - (Z_\lambda^{-1}-\lambda I)^{-1}, \\ &= \left[(I+Z_\lambda^{-1}E)^{-1}Z_\lambda^{-1}-\lambda I\right]^{-1} - (Z_\lambda^{-1}-\lambda I)^{-1}, \\ &= \left[Z_\lambda^{-1}-\lambda I+\Lambda(Z_\lambda,E)\right]^{-1} - (Z_\lambda^{-1}-\lambda I)^{-1}, \end{aligned}$$

where

$$\Lambda(Z_\lambda,E) := \sum_{k=1}^{\infty}(-1)^k (Z_\lambda^{-1}E)^k Z_\lambda^{-1}.$$

Hence

$$\begin{aligned} f_\lambda(Z_\lambda+E)-f_\lambda(Z_\lambda) &= \left[I+(Z_\lambda^{-1}-\lambda I)^{-1}\Lambda(Z_\lambda,E)\right]^{-1}(Z_\lambda^{-1}-\lambda I)^{-1} - (Z_\lambda^{-1}-\lambda I)^{-1}, \\ &= \sum_{j=0}^{\infty}(-1)^j (Z_\lambda^{-1}-\lambda I)^{-j}\Lambda(Z_\lambda,E)^j (Z_\lambda^{-1}-\lambda I)^{-1} - (Z_\lambda^{-1}-\lambda I)^{-1}, \end{aligned} \tag{38}$$

and finally

$$\|f_\lambda(Z_\lambda+E)-f_\lambda(Z_\lambda)\| \leq \left\|(Z_\lambda^{-1}-\lambda I)^{-1}Z_\lambda^{-1}EZ_\lambda^{-1}(Z_\lambda^{-1}-\lambda I)^{-1}\right\| + O(\|E\|^2),$$

so that

$$\kappa_a(f_\lambda, Z_\lambda) \leq \left\|(I-\lambda Z_\lambda)^{-2}\right\|,$$

that proves (37) using (36) and the definition of $f_\lambda(z)$. Note that by (38)

$$L(Z_\lambda,E) := (I-\lambda Z_\lambda)^{-1}E(I-\lambda Z_\lambda)^{-1}$$

is the Fréchet derivative of f_λ at Z_λ applied to E. ■

This Proposition simply shows that the problem is well conditioned for $\lambda \to 0$ and ill conditioned for $\lambda \gg 0$, that matches with the error analysis of Section 3. Of course the situation is opposite to what happens for the solution of the linear systems with $A+\lambda I$ during the Arnoldi process. Therefore the idea, confirmed by many numerical experiments, is to define λ such that $\kappa_r(f_\lambda, Z_\lambda) \approx \kappa(A+\lambda I)$, that is, to consider the bound (37) and solve the equation

$$\frac{\|(I-\lambda Z_\lambda)^{-2}\|\,\|Z_\lambda\|}{\left\|(Z_\lambda^{-1}-\lambda I)^{-1}\right\|} = \|(A+\lambda I)\|\left\|(A+\lambda I)^{-1}\right\|.$$

In the SPD case everything becomes clear since we have

$$\begin{aligned} \frac{\|(I-\lambda Z_\lambda)^{-2}\|\,\|Z_\lambda\|}{\left\|(Z_\lambda^{-1}-\lambda I)^{-1}\right\|} &= \frac{\lambda+\lambda_1}{\lambda_1} \\ \|(A+\lambda I)\|\left\|(A+\lambda I)^{-1}\right\| &= \frac{\lambda_N+\lambda}{\lambda_1+\lambda} \end{aligned}$$

that for $\lambda_1 \to 0$ leads to

$$\lambda = \sqrt{\lambda_1\lambda_N} + O(\lambda_1).$$

Remark 8 *If the underlying operator is bounded then one may consider the approximation*

$$\sqrt{\lambda_1\lambda_N} \approx \frac{1}{\sqrt{\kappa(A)}} \quad \textit{for } \lambda_1 \to 0.$$

Remark 9 *In the SPD case, taking* $\lambda^* = \sqrt{\lambda_1\lambda_N}$ *and putting it into (30)-(31), we find that the asymptotic convergence factor of the method is given by*

$$\|E_m\|^{1/m} \to \frac{1}{\widehat{r}} = \frac{\lambda_N^{1/4} - \lambda_1^{1/4}}{\lambda_N^{1/4} + \lambda_1^{1/4}} = \frac{\kappa(A)^{1/4} - 1}{\kappa(A)^{1/4} + 1}.$$

Remark 10 *The choice of* λ^* *has another interesting meaning. Indeed, let us consider the problem of the computation of* $g(A)b$ *with* g *singular only at 0 and* A *SPD. Using the transformation* $z = (a+\lambda)^{-1}$ *(cf. (3)), if the corresponding* $g^*(z) = g(z^{-1} - \lambda)$ *has a non-removable singularity at 0, then the optimal choice of* λ *is given by solving the equation*

$$c_0 = \frac{1}{2\lambda} \tag{39}$$

(cf. (27) and (28)), that is, the midpoint of $[0, 1/\lambda]$ *must be equal to the midpoint of* I_λ*, because in this way we have simultaneously* $\psi(-\widehat{r}) = 0$ *and* $\psi(\widehat{r}) = 1/\lambda$*. A straightforward computation shows that solving (39) leads exactly to* λ^**. For instance, in [24] the author uses the RD Arnoldi method to compute* $\sqrt{A}b$ *and obtains the same result even if following a different approach.*

Remark 11 *The condition number of* $A + \lambda^* I$ *is given by*

$$\kappa(A + \lambda^* I) = \frac{\lambda_N + \sqrt{\lambda_1\lambda_N}}{\lambda_1 + \sqrt{\lambda_1\lambda_N}} = \sqrt{\frac{\lambda_N}{\lambda_1}} = \sqrt{\kappa(A)}.$$

In the nonsymmetric case, the analysis is a bit more difficult but many numerical experiments have shown that just having information on the conditioning of A, the choice $\lambda \approx \kappa(A)^{-1/2}$ is generally satisfactory, that is, we are rather close to the minimum of a curve similar to the one of Fig. 1. For very ill-conditioned problems we suggest to define λ a bit larger, say in the range $10\kappa(A)^{-1/2} \div 100\kappa(A)^{-1/2}$, since the errors generated by the solution of the linear systems might be much larger than the machine precision.

6 Numerical experiments

In order to test the efficiency of our method, that from now on we denote by RA (Rational Arnoldi), we consider here some numerical experiments where we compare it with other classical iterative solvers. The RA method have have been implemented in Matlab following the line of Algorithm 1 described below.

It is worth noting that we make use of the LU (or Cholesky) factorization to solve the linear system at each step. The reason is to reduce the computational cost since the factorization is computed only once at the beginning, taking also into account that $A + \lambda I$ should be relatively well conditioned. Anyway, for large scale non-sparse problems an iterative approach producing an inner-outer iteration should be considered.

We consider four classical test problems taken out from Hansen's Matlab toolbox `Regtools`, GRAVITY, FOXGOOD, SHAW and BAART. These discrete linear problems arise from the discretization of Fredholm integral equations of the first kind. In all experiments, we consider a

Algorithm 1 - RA Algorithm for solving $Ax = b$.

1: **Require** $A \in \mathbb{R}^{N\times N}$, $b \in \mathbb{R}^N, \lambda \in \mathbb{R}$

2: **Define** $f_\lambda = (1-\lambda z)^{-1} z$

3: **if** $(A+\lambda I)$ is SPD, **then Compute** L s.t. $(A+\lambda I) = L\,L^T$
else Compute L, U s.t. $(A+\lambda I) = L\,U$, **end if**

4: $v_1 \leftarrow b/\|b\|, V_1 \leftarrow [v_1]$

5: **for** $m = 1, 2, \ldots$ **do**

5.1: **Update** $H_m \in \mathbb{R}^{m\times m}$ by Arnoldi's algorithm
Remark: In the Arnoldi's algorithm, we compute $w_m = Z_\lambda v_m$
solving $(A+\lambda I)w_m = v_m$, that is $w_m = U^{-1}L^{-1}v_m$ or $w_m = (L^T)^{-1}L^{-1}v_m$.

5.2: **Compute** $f_\lambda(H_m)$ by Schur-Parlett algorithm

5.3: $x_m \leftarrow \|b\| V_m f_\lambda(H_m)\, e_1$

5.4: **Output** x_m, approximation of $f_\lambda(Z_\lambda)b = A^{-1}b$

5.5: **Update** $V_{m+1} = [v_1, \ldots, v_{m+1}] \in \mathbb{R}^{N\times(m+1)}$ orthonormal basis for
$K_{m+1}(Z_\lambda, b)$, by Arnoldi's algorithm

end for

noise-free right hand side, that is, we define $b = Ax$. The numerical results have been obtained with Matlab 7.9, on a single processor computer Intel Core2 Duo T5800.

Tables 1 and 2 below summarize the results. For comparison, we consider the codes ART, CGLS, LSQR_B and MR2 taken out from Hansen's toolbox, CG, GMRES and MINRES that are resident Matlab functions, and Riley's method. The number between parentheses beside the name of the test is the dimension of the system. In all tests λ_{RA} and λ_{Riley} denote the chosen values of the parameters for the RA and Riley's method respectively. Since no general indication about the choice of the parameter for Riley's method is available in the literature, in all experiments we heuristically select a nearly best one. In the tables we consider the minimum attained error norm *err*, the corresponding residual *res* and the number of iterations *nit*. Each method was stopped when the number of iterations reaches the dimension of the system. The missing numbers are due to the structure of the coefficient matrix (symmetric, SPD, and so on). For each problem the the condition number is approximatively 10^{20}.

The results of Tables 1 and 2 are of course encouraging, especially considering the accuracy with respect to the number of iterations. Indeed, both RA and Riley's method require a linear system to solve at each step, and so it is fundamental to keep the number of iterations low. However, it is worth pointing out that, in the experiments, such linear systems are solved with the LU or Cholesky factorization, so that most part of the computational cost is due to the first iteration.

A classical drawback of many iterative solvers for ill-conditioned problems is the so-called semi-convergence (see e.g. [3]), that is the iterations initially approach the exact solution but quite rapidly diverges. This phenomenon is very common in particular for iterative refinement methods (thus for Riley's and RA) where there is a heavy propagation of errors. Of course, unless a sharp error estimator is available, this undesired behavior can be quite dangerous for applications. In order to understand what we can do to face this problem, in Fig. 2 we consider the error behavior of the RA method for BAART changing the value of the parameter.

Looking at Fig. 2, we can observe that increasing λ the procedure becomes absolutely stable,

	GRAVITY(100)			FOXGOOD(80)		
λ_{RA}, λ_{Riley}	**1e-9**, 1e-11			**1e-8**, 1e-10		
	err	*res*	*nit*	*err*	*res*	*nit*
RA	**1.6e-5**	**8.1e-9**	**2**	**6.8e-7**	**2.9e-10**	**5**
CG	1.7e-4	7.5e-11	96			
ART	8.4e-2	5.8e-3	100	2.3e-3	8.8e-6	80
CGLS				6.3e-6	9.6e-14	80
LSQR_B	1.7e-3	2.0e-8	100	2.9e-6	1.1e-14	80
MR2	1.9e-3	2.3e-8	66	2.3e-6	1.6e-15	57
MINRES	1.8e-4	4.6e-11	100	2.0e-5	1.6e-15	80
RILEY	1.3e-3	8.0e-11	2	6.3e-6	5.2e-10	2

Table 1: Results for GRAVITY and FOXGOOD.

	SHAW(64)			BAART(120)		
λ_{RA}, λ_{Riley}	**1e-9**, 1e-10			**1e-8**, 1e-10		
	err	*res*	*nit*	*err*	*res*	*nit*
RA	**3.3e-3**	**2.0e-7**	**7**	**8.3e-6**	**1.3e-8**	**6**
GMRES				9.6e-6	1.4e-15	15
ART	7.7e-1	6.8e-2	64	3.4e-1	2.7e-2	120
CGLS	2.8e-2	5.1e-10	64	2.4e-2	1.7e-14	120
LSQR_B	2.8e-2	1.5e-10	62	2.4e-2	2.4e-15	120
MR2	1.6e-1	3.7e-6	15			
MINRES	1.0e-2	1.2e-11	64			
RILEY	9.6e-3	8.0e-10	2	1.3e-5	1.3e-10	2

Table 2: Results for SHAW and BAART.

even if we have to pay a small price in terms of accuracy. Therefore, for applications in which it is not possible to monitor in some way the accuracy step by step, the semi-convergence can be prevented taking $\kappa(A)^{-1/2} \ll \lambda \leq \kappa(A)^{-1/4}$, thus looking for a compromise between accuracy and stability. On the other side, reducing λ, the method is really fast but also highly unstable. This last consideration is particularly true for Riley's method, where, at least for these kind of problems, one always observes a rapid divergence after a couple of iterations, also for relatively large values of λ.

In this Section, we also look at another classical example coming out from approximation theory. We consider in particular the reconstruction of the Franke's bivariate test function via interpolation by means of Gaussian Radial Basis Functions (RBF) with shape coefficients equal to 1 (see e.g. [14] for a background). For simplicity, instead of scattered points, we consider here the very special case of a grid of 15×15 equally spaced points on the square $[0, 1] \times [0, 1]$ that leads to a SPD linear systems of dimension 225 whose condition number is about 10^{21}. In Fig. 3, the surfaces obtained with the Cholesky factorization, the CG and the RA method (with $\lambda = 10^{-11}$) are plotted. Since the exact solution of the system is unknown, we used the residual as a stopping criterion, so that the CG result corresponds to the iteration 190 (residual $\approx 1.6\text{e}-1$), while the RA result corresponds to the iteration 10 (residual $\approx 1.4\text{e}-1$).

While the result with the Cholesky factorization was expected (a similar test have been presented in [13]), the difficulties with Krylov methods were not. Indeed, the CG method has shown

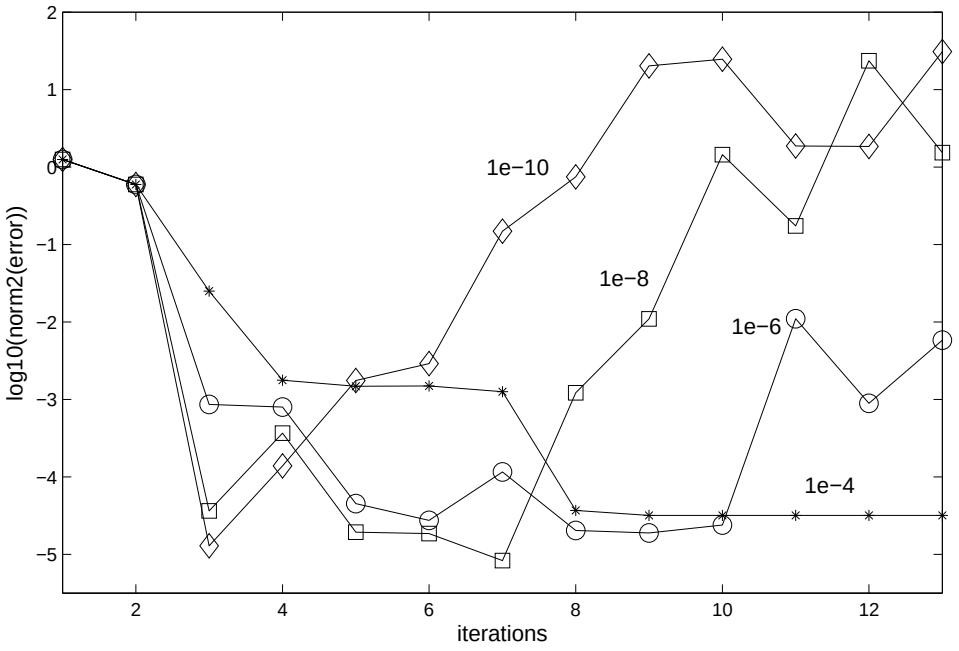

Figure 2: BAART(120) - Error behavior for $\lambda = 10^{-4}, 10^{-6}, 10^{-8}, 10^{-10}$.

to be the best Krylov method for this problem, but the results are poor if compared with those of the RA method. We have to point out that, for this case, the reconstruction given by the RA and the Riley's method are very similar.

7 Extension to Tikhonov regularization

In many applications it is often necessary to deal with ill-conditioned linear systems in which the right hand side is affected by noise. Defining e_b as a perturbation (of course unknown) of the right hand side b, one is forced to solve in some way

$$A\widetilde{x} = \widetilde{b}, \quad \widetilde{b} := b + e_b, \tag{40}$$

hoping that the computed solution of (40) is close to the solution of $Ax = b$. In this situation, the RA method does not seem to be so powerful and robust as in the noise-free case. Moreover, unless the noise level is very low, it is also difficult to design a strategy to define the parameter λ. Indeed, in order to adopt the theory of Section 5 based on the analysis of the conditioning, we should need, for instance, to construct an invertible linear filter F such that $Fe_b \approx 0$. In this way $F^{-1}Ax \approx \widetilde{b}$, and hence information on the choice of λ can be obtained considering $\kappa(F^{-1}A)$. Anyway this kind of approach is beyond the purpose of this paper, and we prefer to extend the idea of the RA method in order to make it able to work directly with Tikhonov regularization in its standard form.

As well known Tikhonov regularization is based on the solution of the minimization problem

$$\min_x \left(\left\| Ax - \widetilde{b} \right\|^2 + \lambda \left\| Hx \right\|^2 \right), \quad \lambda > 0, \tag{41}$$

where the matrix H is generally taken as an high-pass filter (e.g. the second derivative) so that the term $\|Hx\|^2$ plays the role of the penalization term in a constrained minimization. The main

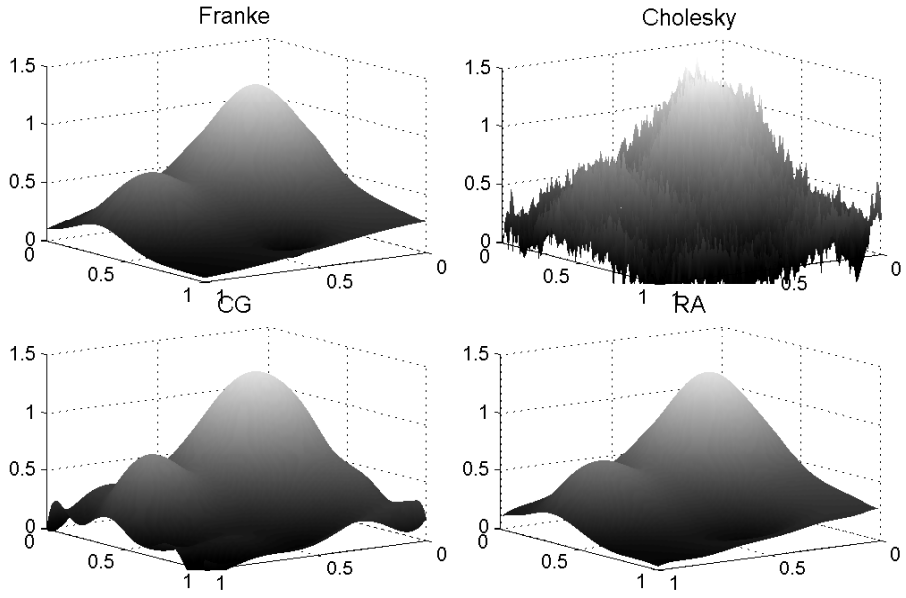

Figure 3: Interpolation of Franke's bivariate test function by means of Gaussian RBF.

problem is that the noise generally involves also frequencies of the exact solution so that it is not possible to solve (41) letting $\lambda \to \infty$ as in standard constrained minimization. Anyway, defining suitably λ (see [18] for a background), the corresponding solution x_λ is expected to be somehow similar to the desired noise-free solution. The problem (41) leads to the solution of the regularized system

$$(A^T A + \lambda H^T H) x_\lambda = A^T \widetilde{b}, \tag{42}$$

where the matrix $A^T A + \lambda H^T H$ is also expected to be better conditioned than A.

Following the idea of the RA method, and assuming to work with a non singular matrix H, we consider here the transformation

$$Z_\lambda = (A^T A + \lambda H^T H)^{-1}.$$

Since the exact solution can be written as $x = \left(A^T A\right)^{-1} A^T b$, we have

$$\begin{aligned} x &= \left(Z_\lambda^{-1} - \lambda H^T H\right)^{-1} A^T b, \\ &= f_\lambda(Q_\lambda) \left(H^T H\right)^{-1} A^T b, \end{aligned}$$

where

$$Q_\lambda = Z_\lambda \left(H^T H\right) = \left(\left(H^T H\right)^{-1} A^T A + \lambda I\right)^{-1}.$$

Note that we are assuming to work with the exact right hand side even if, in practice, the method is applied with $\widetilde{b}$.

Hence we can compute the solution working with the Arnoldi algorithm based on the construction of the Krylov subspaces $K_m(Q_\lambda, \left(H^T H\right)^{-1} A^T b)$. Thus, starting from $v_1 = v / \|v\|$, where v is the solution of

$$\left(H^T H\right) v = A^T b, \tag{43}$$

we need to compute, at each step of the algorithm, the vectors $w_j = Q_\lambda v_j$, $j \geq 1$, that is, we need to solve systems of the type

$$(A^T A + \lambda H^T H) w_j = \left(H^T H\right) v_j.$$

Note that by (43) and the arising definition of v_1, the first step of the Arnoldi algorithm yields the Tikhonov regularized solution x_λ (cf. (42)). Hence, also in this case, the procedure can be interpreted as an iterated Tikhonov regularization.

The approach just presented (that we indicate by RAT, Rational-Arnoldi-Tikhonov) represents the natural extension of the RA method to the Tikhonov regularization but it does not exploit the symmetry of the the problem (42). In order to improve the method, we can observe that the matrix Q_λ is $H^T H$-symmetric and hence we could applied the Lanczos method for the matrix function with the $H^T H$-weighted inner product.

Alternatively we can consider the transformation

$$\widetilde{Q}_\lambda = (H^{-T} A^T A H^{-1} + \lambda I)^{-1},$$

that leads to the expression

$$x = H^{-1} f_\lambda(\widetilde{Q}_\lambda) H^{-T} A^T b.$$

The argument matrix $\widetilde{Q}_\lambda$ is now symmetric and hence it is possible to apply the Lanczos method. We denote this approach by RLT (Rational-Lanczos-Tikhonov). Starting from $v_1 = v/\left\|v\right\|$, where v is the solution of

$$H^T v = A^T b,$$

each iteration the Lanczos method will require the computation of the vectors $w_j = \widetilde{Q}_\lambda v_j$, $j \geq 1$, that is, to solve the systems

$$(A^T A + \lambda H^T H) s_j = H^T v_j,$$

and then to compute $w_j = H s_j$. We can observe that with respect to the RAT, this new approach does not introduce computational difficulties for what concerns the computation of the vectors w_j.

In order to appreciate the potential of this extensions we consider the test problems SHAW and BAART with a right hand side contaminated by an error e_b defined by

$$e_b = \frac{\delta \left\|b\right\|}{\sqrt{N}}\, u,$$

where δ is the relative noise level, and u is a vector containing random values drawn from a normal distribution with mean 0 and standard deviation 1. In the experiments, we define $\delta = 10^{-2}$ and $\delta = 10^{-3}$, and, as suggested in [9], we take as regularization matrix

$$H = \begin{pmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ & & & -1 & 2 \end{pmatrix} \in \mathbb{R}^{N \times N}.$$

Since the noise is randomly generated, for both examples we consider the mean results over 50 experiments, and we compare the RAT and the RLT method (with different values of the parameter λ) with GMRES, ART, LSQR_B and MR2. The results are collected in Table 3.

		SHAW(64)				BAART(120)			
		$\delta = 10^{-2}$		$\delta = 10^{-3}$		$\delta = 10^{-2}$		$\delta = 10^{-3}$	
	λ	*err*	*nit*	*err*	*nit*	*err*	*nit*	*err*	*nit*
RAT	1e-4	3.624	2.1	0.712	3.1	0.443	2.0	0.078	2.0
	1e-3	2.459	2.8	0.377	3.3	0.133	2.0	0.048	2.1
	1e-2	1.252	3.0	0.303	4.4	0.066	2.0	0.041	2.3
	1e-1	0.848	3.5	0.294	6.2	0.055	2.0	0.030	2.7
	1e-0	0.826	5.0	0.294	7.6	0.044	2.2	0.013	3.0
	1e+1	0.838	6.8	0.296	8.8	0.031	2.8	0.008	3.1
	1e+2	0.887	7.0	0.295	8.8	0.023	3.0	0.007	3.5
	1e+3	0.834	9.4	0.287	12.1	0.014	3.2	0.008	4.0
	1e+4	0.855	11.5	0.534	11.8	0.015	3.3	0.009	4.1
RLT	1e-4	3.713	2.1	1.290	3.0	0.057	2.0	0.056	2.0
	1e-3	3.234	2.4	0.441	3.3	0.057	2.0	0.056	2.1
	1e-2	1.606	3.0	0.318	4.2	0.057	2.0	0.052	2.2
	1e-1	1.241	4.1	0.297	6.1	0.056	2.1	0.037	2.8
	1e-0	0.895	4.7	0.298	7.5	0.052	2.2	0.020	2.9
	1e+1	0.888	6.5	0.298	8.2	0.039	2.7	0.009	3.1
	1e+2	0.899	6.9	0.296	8.5	0.027	3.0	0.008	3.8
	1e+3	0.900	7.2	0.295	11.2	0.014	3.2	0.008	4.0
	1e+4	0.895	8.2	0.298	12.2	0.015	3.3	0.009	4.0
GMRES		1.261	5.2	0.401	7.0	0.388	3.0	0.060	3.0
ART		1.151	8.0	0.810	40.7	0.332	120.0	0.338	120.0
LSQR_B		0.899	5.7	0.399	9.7	0.198	3.4	0.135	5.3
MR2		0.901	5.4	0.387	7.5				

Table 3: Minimum attained error and corresponding iteration number for SHAW and BAART with Gaussian noise of level $\delta = 10^{-2}$ and $\delta = 10^{-3}$

Similarly to the noise-free case, we also consider the stabilizing effect of a careful choice of λ. Indeed, in Figure 4 and 5 we plot the error behavior of some of the methods considered for the solution of SHAW(64) and BAART(120) respectively. Taking $\lambda = 100$ for the RAT and the RLT method, we can heavily reduce the problem of semi-convergence keeping at the same time a good level of accuracy contrary to other well performing methods such as GMRES and LSQR_B.

8 Conclusions

Our experience with the RA, RAT and RLT methods leads us to consider these methods as reliable alternatives to the classical iterative solvers for ill-conditioned problems. Since they actually are iterative refinement processes, the attainable accuracy is almost never worse that the other solvers. While this property could be somehow expected, maybe the most important feature of these methods is their robustness. Indeed, contrary to other iterative refinement processes such as the Riley's algorithm or other Krylov solvers, the methods work pretty well for a large window of values of λ. Hence, having a good error estimator or working with applications in which it is possible to monitor the result step by step, one may reduce λ in order to save computational work; in the opposite case, one may increase λ slowing down the method but assuring a stable convergence. To this purpose, we intend to use, in a forthcoming work, the estimates of the norm

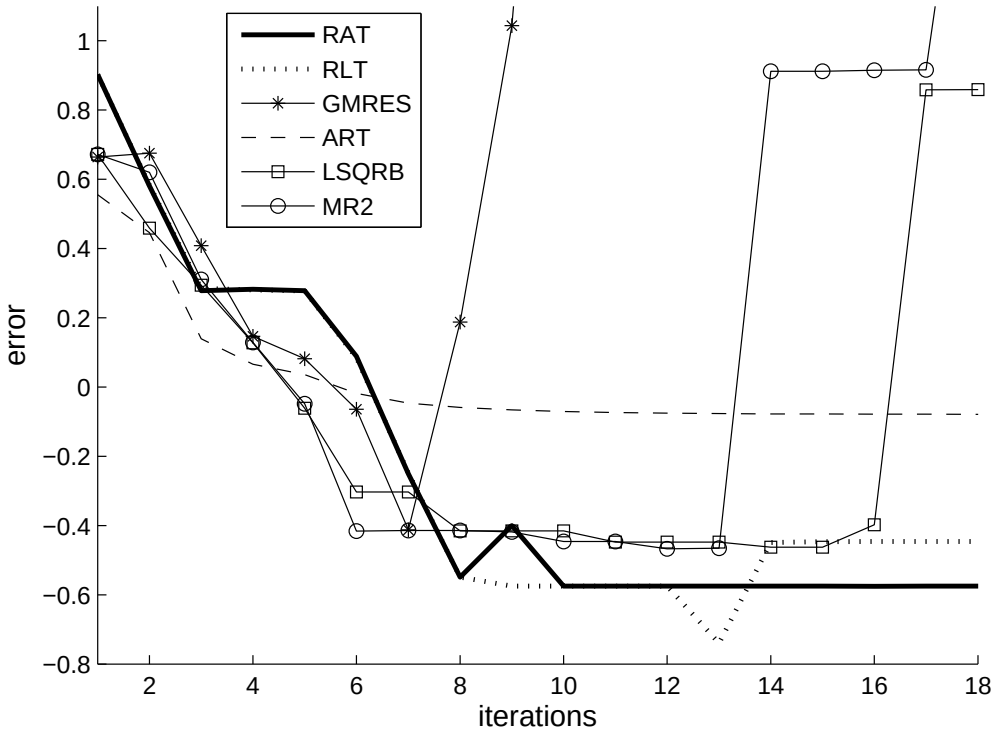

Figure 4: Error behavior for SHAW(64) with noise. RAT and RLT are implemented with $\lambda = 100$.

of the error described in [5] and [8] which are based on an extrapolation procedure of the moments of the matrix of the system with respect to the residuals of the iterative method.

Acknowledgement: The authors are grateful to Marco Donatelli, Igor Moret, Giuseppe Rodriguez, and Marco Vianello for many helpful discussions and comments.

References

[1] B. Beckermann, L. Reichel, Error estimation and evaluation of matrix functions via the Faber transform, SIAM J. Numer. Anal., 47 (2009) 3849–3883.

[2] B. Beckermann, M. Crouzeix, Operators with numerical range in a conic domain, Archiv der Mathematik, 88 (2007) 547-559.

[3] A. Björck, *Numerical Methods for Least Squares Problems*, SIAM, Philadelphia, 1996.

[4] C. Brezinski, Outlines of Padé approximation, in *Computational Aspects of Complex Analysis*, H. Werner et al. eds., Reidel, Dordrecht, 1983, pp. 1–50.

[5] C. Brezinski, Error estimates for the solution of linear systems, SIAM J. Sci. Comput., 21 (1999) 764–781.

[6] C. Brezinski, M. Redivo–Zaglia, unpublished notes (2002).

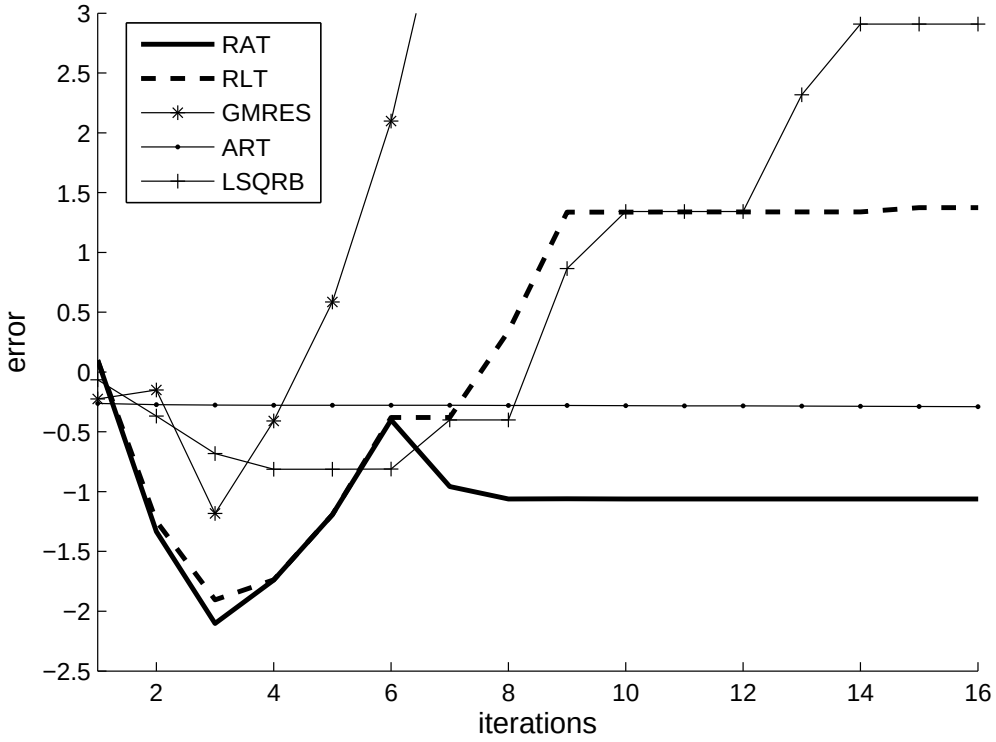

Figure 5: Error behavior for BAART(120) with noise. RAT and RLT are implemented with $\lambda = 100$.

[7] C. Brezinski, M. Redivo–Zaglia, G. Rodriguez, S. Seatzu, Extrapolation techniques for ill-conditioned linear systems, Numer. Math. 81 (1998) 1-29.

[8] C. Brezinski, G. Rodriguez, S. Seatzu, Error estimates for linear systems with applications to regularization, Numer. Algorithms, 49 (2008), 85–104.

[9] D. Calvetti, L. Reichel, A. Shuibi, Tikhonov regularization of large symmetric problems, Numer. Linear Algebra Appl. 12 (2005) 127–139.

[10] M. Crouzeix, Numerical range and numerical calculus in Hilbert space, J. Functional Analysis, 244 (2007) 668–690.

[11] S.W. Ellacott, Computation of Faber series with application to numerical polynomial approximation in the complex plane, Math.Comp., 40 (1983) 575–587.

[12] J.v.d. Eshof, M. Hochbruck, Preconditioning Lanczos approximations to the matrix exponential, SIAM J. Sci. Comp., 27 (2005) 1438–1457.

[13] G. Fasshauer, Tutorial on Meshfree Approximation Methods with Matlab, Slides for 6 Lectures, Dolomites Research Notes on Approximation, Vol. 1, 2008.

[14] G. Fasshauer, *Meshfree Approximation Methods with MATLAB*, World Scientific Publishers, Singapore, 2007.

[15] G.H. Golub, Numerical methods for solving linear least squares problems. Numer. Math., 7 (1965) 206–216.

[16] G.H. Golub, C.F. Van Loan, *Matrix Computations*, 3rd ed., Johns Hopkins University Press, Baltimore, 1996.

[17] P.C. Hansen, Regularization Tools: A Matlab package for analysis and solution of discrete ill-posed problems, Numer. Algorithms, 6 (1994) 1–35.

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

[19] P.C. Hansen, Regularization Tools, Version 4.0 for Matlab 7.3, Numer. Algorithms, 46 (2007) 189–194.

[20] M. Hanke, P.C. Hansen, Regularization methods for large-scale problems, Surveys Math. Indust., 3 (1993) 253–315.

[21] N.J. Higham, *Functions of Matrices: Theory and Computation*, SIAM, Philadelphia, 2008.

[22] D. Kincaid, W. Cheney, *Numerical Analysis: Mathematics of Scientific Computing*, 3rd ed., Brooks/Cole, Pacific Grove, 2002.

[23] J.T. King, D. Chillingworth, Approximation of generalized inverses by iterated regularization, Numer. Funct. Anal. Optim. 1 (1979) 499–513.

[24] I. Moret, Rational Lanczos approximations to the matrix square root and related functions, Numer. Linear Algebra Appl., 16 (2009) 431–445.

[25] I. Moret, P. Novati, The computation of functions of matrices by truncated Faber series, Numer. Func. Anal. and Optimiz., 22 (2001) 697–719.

[26] I. Moret, P. Novati, RD-rational approximations of the matrix exponential, BIT, 44 (2004) 595–615.

[27] A. Neumaier, Solving ill-conditioned and singular linear systems: a tutorial on regularization, SIAM Rev., 40 (1998) 636–666.

[28] S.P. Nørsett, Restricted Padé approximations to the exponential function, SIAM J. Numer. Anal., 15 (1978) 1008–1029.

[29] J.D. Riley, Solving systems of linear equations with a positive definite, symmetric, but possibly ill-conditioned matrix, Math. Tables Aids Comput., 9 (1955) 96–101.

[30] Y. Saad, Analysis of some Krylov subspace approximations to the matrix exponential operator, SIAM J. Numer. Anal., 29 (1992) 209–228.

[31] V.I. Smirnov, N.A. Lebedev, *Functions of a Complex Variable - Constructive Theory*, Iliffe Books, London, 1968.

[32] J.L. Walsh, *Interpolation and Approximation by Rational Functions in the Complex Domain*, AMS, Providence, 1965.