A rational Arnoldi approach for ill conditioned linear systems

Paolo Novati
Department of Pure and Applied Mathematics,
University of Padova - Italy

Joint work with
Claude Brezinski (Lille - France)
Michela Redivo Zaglia (Padova - Italy)

June 2010

The problem

We consider ill-conditioned linear systems

$$Ax = b$$

We mainly focus the attention on **full-rank** problems in which the singular values of A decay gradually to zero.

- Discretization of compact operators, as in the case of Fredholm integral equations of the first kind.
- Vandermonde type systems arising from interpolation.

Standard approaches

- **Preconditioning**: typically used when A is structured (e.g. Toeplitz, block Toeplitz) or sparse.
- **Tikhonov regularization**: based on the solution of a regularized system

$$(A^T A + \lambda L^T L)x_\lambda = A^T b$$

where the parameter $\lambda > 0$ must be taken in order to filter out the unwanted (erroneous) SVD components due to perturbations of the input data A and b.

Outline

- main idea - extension of Riley's algorithm
- the rational Arnoldi method for function of matrices
- theoretical error analysis
- the choice of the regularization parameter
- numerical experiments

Riley's algorithm (Riley 1955)

iterative refinement process

Consider the regularized linear system

$$(A + \lambda I)\, x_\lambda = b$$

Defining

$$Z = (A + \lambda I)^{-1}$$

we have that

$$\begin{aligned} A^{-1} b &= (I - \lambda Z)^{-1} Z b \\ &= \frac{1}{\lambda} \sum_{k=1}^{\infty} (\lambda Z)^k b \end{aligned}$$

that leads to the recursion

$$x_{k+1} = y + \lambda Z x_k, \quad x_0 = 0, \quad y = Zb.$$

The method is generally referred to as an *iterated Tikhonov regularization* (Golub 1965, King and Chillingworth 1979, Hanke and Hansen 1993).

Main features of Riley's method

- Many regularized systems to be solved.
- λ can be taken in order to reduce heavily the conditioning.
- *Semiconvergence*: x_k initially approaches the solution but quite rapidly diverges (Bjorck 1996).
- We may have an effective refinement provided that we take λ "much larger" than the one arising from the classical analysis (L-curve, Generalized Cross Validation, etc.)
- For "large" λ, the Neumann series gives an extremely slow approximation of $(I - \lambda Z)^{-1} Zb$
- How to choose λ?

Extension of Riley's algorithm

Defining the function

$$f(z) = \left(\frac{1}{z} - \lambda\right)^{-1} = \frac{z}{1 - \lambda z}$$

we have that

$$x = A^{-1}b = f(Z)b, \quad Z = (A + \lambda I)^{-1}$$

Therefore, any iterative method for the computation of $f(Z)b$ is actually a refinement process. Using an iterative method that computes $f(Z)b$ in a relatively small number of steps (thus theoretically direct) may reduce the error before divergence.

The Arnoldi method

For the computation of $f(Z)b$ we use the standard Arnoldi method projecting the matrix Z onto the Krylov subspaces generated by Z and b, that is $K_m(Z,b) = span\{b, Zb, ..., Z^{m-1}b\}$.
For the construction of the subspaces $K_m(Z,b)$ the Arnoldi algorithm generates an orthonormal sequence $\{v_j\}_{j\geq 0}$, with $v_1 = b/\|b\|$, such that $K_m(Z,b) = span\{v_1, v_2, ..., v_m\}$ (here and below the norm used is always the Euclidean norm). For every m,

$$ZV_m = V_m H_m + h_{m+1,m} v_{m+1} e_m^T.$$

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

The Arnoldi method

The m-th Arnoldi approximation to $x = f(Z)b$ is defined as

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

At each step of the Arnoldi algorithm we have to compute the vector $w_j = Zv_j$, that is, to solve the system

$$(A + \lambda I) w_j = v_j.$$

Since $v_1 = b/\|b\|$, the corresponding w_1 is just the scaled solution of the regularized system. The method theoretically converges in a finite number of steps.
For the computation of $f(H_m)$ we employ the Schur-Parlett algorithm (Golub and Van.Loan 1983).

Theoretical error analysis

Theorem

Assume that the field of values of A, $F(A)$, is symmetric with respect to the real axis and contained in the right half complex plane. Let Ω be an ellipse (with associated conformal mapping $\psi(w) = \gamma w + c_0 + c_1/w$ and inverse ϕ) symmetric with respect to the real axis and such that $F(Z) \subseteq \Omega$ with $1/\lambda \notin \Omega$. Then for m large enough we have

$$\|E_m\| = \|x - x_m\| \leq 4eC \frac{R}{R-1} \frac{1}{\psi'(R)} K \frac{m+1}{R^m}$$

where $K = 1/\lambda^2$, $R = \phi(1/\lambda)$, and $C = 11.08$ ($C = 1$ if A is symmetric).

Sketch of the proof

Using the properties of the Arnoldi algorithm we know that for every $p_{m-1} \in \Pi_{m-1}$,

$$V_m p_{m-1}(H_m) e_1 = p_{m-1}(Z) b.$$

Hence for every $p_{m-1} \in \Pi_{m-1}$,

$$E_m = f(Z)b - p_{m-1}(Z)b - V_m(f(H_m) - p_{m-1}(H_m))e_1.$$

Since $\|V_m\| = 1$ we have that

$$\|E_m\| \leq 2C \|p_{m-1} - f\|_{F(Z)}$$

(Crouzeix 2007). Since $F(Z) \subseteq \Omega$, taking p_{m-1} as the $(m-1)$-th truncated Faber (Chebyshev) series, the thesis follows from classical results of complex approximation (Ellacott 1983), taking into account that f is singular only at $1/\lambda$.
Other approaches: Dunford-Taylor integral representation (Kniznherman 1991, Hochbruck et al 1998, Moret et al 2005, 2008), Faber transform (Beckermann and Reichel 2009).

The SPD case

Let $\lambda_1 \gtrsim 0$ and λ_N be respectively the smallest and the largest eigenvalue A. Then $F(A) = [\lambda_1, \lambda_N]$ and $F(Z) = \left[\frac{1}{\lambda_N+\lambda}, \frac{1}{\lambda_1+\lambda}\right] =: I_\lambda$. In this case

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

The conformal mapping ψ associated to I_λ is given by

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

where

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

The SPD case

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

$$\gamma R + c_0 + \frac{\gamma}{R} = \frac{1}{\lambda}$$

that is

$$R = u + \sqrt{u^2 - 1} \quad \text{where} \quad u = \frac{2\lambda_1\lambda_N}{\lambda(\lambda_N - \lambda_1)} + \frac{\lambda_N + \lambda_1}{\lambda_N - \lambda_1}$$

We find

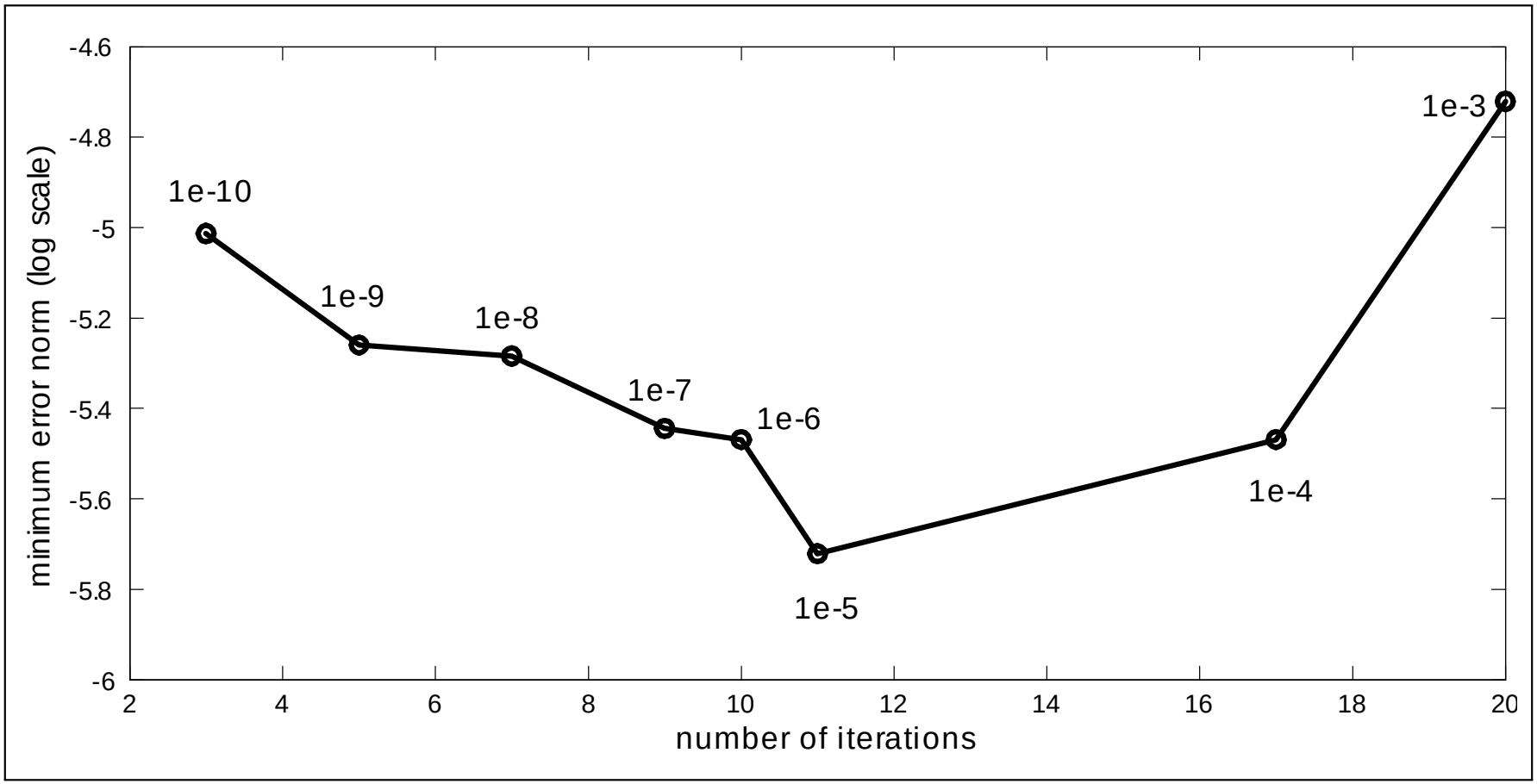
$$R \approx 1 + 2\sqrt{\frac{\lambda_1}{\lambda}} \quad \text{for } \lambda_1 \to 0 \text{ and } \lambda << \lambda_N$$

Thus R monotonically decreases with respect to λ and $R \to \infty$ for $\lambda \to 0$.

The choice of lambda

- The common strategy for iterated Tikhonov regularization is to find a compromise between the conditioning of $A + \lambda I$ and the rate of convergence (Fasshauer 2008).
- Focusing the attention on the accuracy one could expect that "large" values of λ should allow an improvement, since the linear systems with $A + \lambda I$ would be solved more accurately. The numerical experiments show that this is not true (semiconvergence).

An example



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

The conditioning of f(Z)

The absolute and the relative condition number for the computation of $g(X)$ where g is a given function and X a square matrix are given by (Higham 2008)

$$\begin{aligned} k_a(g,X) &= \lim_{\epsilon\to 0}\sup_{\|E\|\le\epsilon}\frac{\|g(X+E)-g(X)\|}{\epsilon}. \\ k_r(g,X) &= k_a(g,X)\frac{\|X\|}{\|g(X)\|} \end{aligned}$$

This definition implies that

$$\|g(X+E)-g(X)\| \le k_a(g,X)\,\|E\| + O(\|E\|^2)$$

Theorem

For the function $f(z) = (1-\lambda z)^{-1}z$ we have

$$k_r(f,Z) = \left\|(I-\lambda Z)^{-1}\right\|$$

The conditioning of f(Z)

Assuming that Z is independent of $A + \lambda I$ choose λ such that

$$\left\|(I - \lambda Z)^{-1}\right\| = \|(A + \lambda I)\| \left\|(A + \lambda I)^{-1}\right\|$$

If A is SPD then

$$\begin{aligned} \left\|(I - \lambda Z)^{-1}\right\|_2 &= \frac{\lambda + \lambda_1}{\lambda_1} \\ \|(A + \lambda I)\| \left\|(A + \lambda I)^{-1}\right\| &= \frac{\lambda_N + \lambda}{\lambda_1 + \lambda} \end{aligned}$$

and we obtain for $\lambda_1 \longrightarrow 0$

$$\begin{aligned} \lambda &= \sqrt{\lambda_1 \lambda_N} + O(\lambda_1) \\ &\approx \frac{1}{\sqrt{k(A)}} \text{ for } \lambda_N \approx 1 \end{aligned}$$

Remarks for the SPD case

- Let $\lambda^* = \sqrt{\lambda_1 \lambda_N}$. We have

$$k(A + \lambda^* I) = \sqrt{k(A)}$$

$$\|E_m\|^{1/m} \to \frac{1}{R} = \frac{k(A)^{1/4} - 1}{k(A)^{1/4} + 1}$$

- Let us consider the problem of the computation of $\phi(A)b$ with ϕ singular only at 0. If the corresponding $g(z) = \phi(z^{-1} - \lambda)$ has a non-removable singularity at 0, following our error analysis the optimal choice of λ is obtained by solving the equation

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

that is, the midpoint of $[0, 1/\lambda]$ must be equal to the midpoint of I_λ. The solution is given by λ^*. (Moret 2007 with $\phi(a) = \sqrt{a}$)

- λ^* is also the solution of

$$k((A + \lambda I)^{-1} A) = k(A + \lambda I)$$

Numerical experiments 1/3

From Hansen's Matlab toolbox regu

	GRAVITY(100) SPD			FOXGOOD(80) sym.		
λ_{RA}, λ_{Riley}	**1e-9**, 1e-11			**1e-8**, 1e-10		
	error	residual	iter.	error	residual	iter.
RA	**1.6e-5**	**8.1e-9**	**2**	**6.8e-7**	**2.9e-10**	**5**
CG	1.7e-4	7.5e-11	96			
GMRES						
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

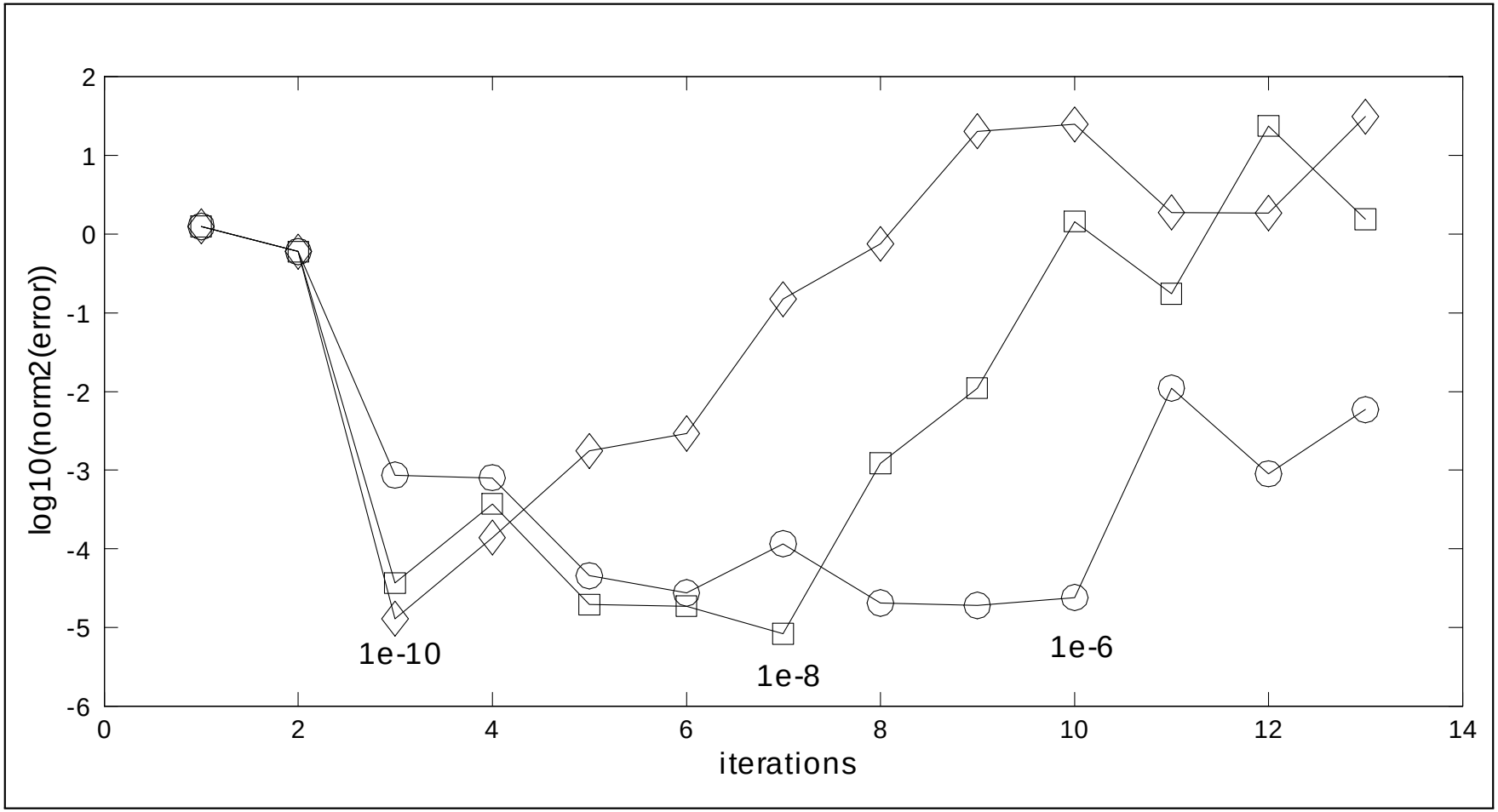
Numerical experiments 2/3

From Hansen's Matlab toolbox regu

	SHAW(64) sym.			BAART(120) nonsym.		
λ_{RA}, λ_{Riley}	**1e-9**, 1e-10			**1e-8**, 1e-10		
	error	residual	iter.	error	residual	iter.
RA	**3.3e-3**	**2.0e-7**	**7**	**8.3e-6**	**1.3e-8**	**6**
CG						
GMRES				8.4e-5	1.8e-15	112
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	1.4e-5	1.3e-10	2

Numerical experiments 3/3

From Hansen's Matlab toolbox regu



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

Final remarks

- RA is of course more expansive than classical iterative solvers.
- The choice of $\lambda \approx 1/\sqrt{k(A)}$ suggested by the analysis of the conditioning of $f(Z)$ seems to be confirmed by the results of the numerical experiments. Anyway in all examples we considered, it seemed that the method behaves quite similarly (w.r.t. the accuracy) for a rather large window of values, say $\left[\frac{\lambda}{100}, 100\lambda\right]$.
- The functions of matrices $f(H_m)$ are computed using the Schur-Parlett algorithm so the concept of "residual" is completely forgotten. As consequence, contrary to classical methods, the norm of the residual is typically closer to the norm of the error, at least before divergence.
- We are not still able to provide a reliable a-posteriori error bound. Our a-priori error analysis is too pessimistic at the beginning and (of course) does not consider the divergence.

Extension to the general Tikhonov regularization

Tikhonov regularization is based on the solution of the regularized system

$$(A^T A + \lambda H^T H) x_\lambda = A^T b$$

Defining

$$Z = (A^T A + \lambda H^T H)^{-1},$$

since $x = \left(A^T A\right)^{-1} A^T b$ we have

$$\begin{aligned} x &= \left(Z^{-1} - \lambda H^T H\right)^{-1} A^T b \\ &= f(Q) \left(H^T H\right)^{-1} A^T b \end{aligned}$$

where

$$\begin{aligned} Q &= Z \left(H^T H\right) \\ &= \left(\left(\left(H^T H\right)^{-1}\right) A^T A + \lambda I\right)^{-1} \end{aligned}$$

Extension to the general Tikhonov regularization

We can define an iterated Tikhonov process working with the Arnoldi algorithm based on the construction of the Krylov subspaces $K_m(Q, \left(H^T H\right)^{-1} A^T b)$. Starting from $v_1 = c / \|c\|$, where c is the solution of

$$\left(H^T H\right) c = A^T b$$

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

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

As before the first step of the Arnoldi algorithm yields the Tikhonov regularized solution x_λ.