Solving discrete ill–posed problems with Tikhonov regularization and generalized cross validation

Gérard MEURANT

November 2010

Introduction to ill–posed problems

We speak of a discrete ill–posed problem (DIP) when the solution is sensitive to perturbations of the data

Example:

$$A = \begin{pmatrix} 0.15 & 0.1 \\ 0.16 & 0.1 \\ 2.02 & 1.3 \end{pmatrix}, \quad c + \Delta c = A \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \begin{pmatrix} 0.01 \\ -0.032 \\ 0.01 \end{pmatrix}$$

The solution of the perturbed least squares problem (rounded to 4 decimals) using the QR factorization of A is

$$x_{QR} = \begin{pmatrix} -2.9977 \\ 7.2179 \end{pmatrix}$$

Why is it so?

The singular value decomposition (SVD) of A is

$$U = \begin{pmatrix} -0.0746 & 0.7588 & -0.6470 \\ -0.0781 & -0.6513 & -0.7548 \\ -0.9942 & -0.0058 & 0.1078 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 2.4163 & 0 \\ 0 & 0.0038 \\ 0 & 0 \end{pmatrix}$$

$$V = \begin{pmatrix} -0.8409 & -0.5412 \\ -0.5412 & 0.8409 \end{pmatrix}$$

The component $(u^2)^T \Delta c / \sigma_2$ (u^2 being the second column of U) corresponding to the smallest nonzero singular value is large being 6.2161

This gives the large change in the solution

More generally

$$Ax \approx c = \bar{c} - e$$

where A is a matrix of dimension $m \times n$, $m \geq n$ and the right hand side $\bar{c}$ is contaminated by a (generally) unknown noise vector e

- The standard solution of the least squares problem $\min \| c - Ax \|$ (even using backward stable methods like QR) may give a vector x severely contaminated by noise
- This may seems hopeless, but
- The solution is to modify the problem by regularization
- We have to find a balance between obtaining a problem that we can solve reliably and obtaining a solution which is not too far from the solution without noise

Examples of ill-posed problems

These examples were obtained with the Regutools Matlab toolbox from Per-Christian Hansen

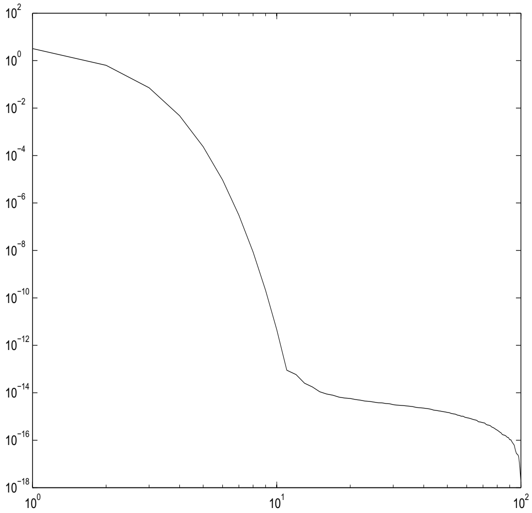
The Baart problem arises from the discretization of a first-kind Fredholm integral equation

$$\int_0^1 K(s,t) f(t)\, dt = g(s) + e(s)$$

e is the unknown noise. The kernel K and right-hand side g are given by

$$K(s,t) = \exp(s\cos(t)), \quad g(s) = 2\sinh(s)/s$$

and with integration intervals $s \in [0, \pi/2], t \in [0, \pi]$
The solution without noise is given by $f(t) = \sin(t)$
The square dense matrix A of order 100 is dense and its smallest and largest singular values are $1.7170\ 10^{-18}$ and 3.2286



Singular values for the `Baart` problem, $m = n = 100$

This distribution of singular values is typical of discrete ill-posed problems

The Phillips problem arises from the discretization of a first-kind Fredholm integral equation devised by D. L. Phillips
Let

$$\phi(x) = 1 + \cos(x\pi/3) \text{ for } |x| < 3, \;\; 0 \text{ for } |x| >= 3$$

The kernel K, the solution f and the right-hand side g are given by

$$K(s,t) = \phi(s-t), \quad f(t) = \phi(t)$$

$$g(s) = (6 - |s|)(1 + 0.5\cos(s\pi/3)) + 9/(2\pi)\sin(|s|\pi/3)$$

The integration interval is $[-6, 6]$
The square matrix A of order 200 is banded and its smallest and largest singular values are 1.3725 10^{-7} and 5.8029.

Tikhonov regularization

Replace the LS (unsolvable) problem by

$$\min_x \{\|c - Ax\|^2 + \mu\|x\|^2\}$$

where $\mu \geq 0$ is a regularization parameter to be chosen

For some problems (particularly in image restoration) it is better to consider

$$\min_x \{\|c - Ax\|^2 + \mu\|Lx\|^2\}$$

where L is typically the discretization of a derivative operator of first or second order

The solution x_μ of the regularized problem solves the linear system

$$(A^T A + \mu I)x = A^T c$$

The main problem in Tikhonov regularization is to choose the regularization parameter μ

- If μ is too small the solution is contaminated by the noise in the right-hand side
- if μ is too large the solution is a poor approximation of the original problem
- Many heuristic methods have been devised for choosing μ
- Most of these methods lead to the evaluation of bilinear forms with different symmetric matrices

Some heuristics for choosing μ

- Morozov's discrepancy principle
  Ask for the norm of the residual to be equal to the norm of the noise vector (if it is known)

$$\|c - A(A^T A + \mu I)^{-1} A^T c\| = \|e\|$$

  We have to consider $c^T A(A^T A + \mu I)^{-1} A^T c$ and $c^T A(A^T A + \mu I)^{-1} A^T A(A^T A + \mu I)^{-1} A^T c$

- The Gfrerer/Raus method

$$\mu^3 c^T (AA^T + \mu I)^{-3} c = \|e\|^2$$

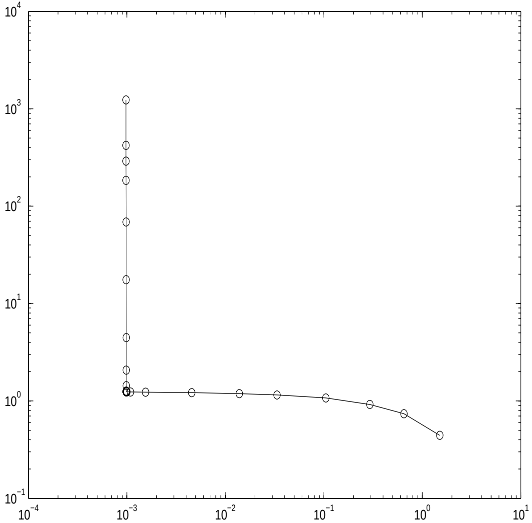
- The quasi–optimality criterion

$$\min[\mu^2 c^T A(A^T A + \mu I)^{-4} A^T c]$$

The L–curve criterion

- plot in log–log scale the curve $(\|x_\mu\|, \|b - Ax_\mu\|)$ obtained by varying the value of $\mu \in [0, \infty)$ where x_μ is the regularized solution
- In most cases this curve is shaped as an "L"
- Lawson and Hanson proposed to choose the value μ_L corresponding to the "corner" of the L–curve (the point of maximal curvature (see also Hansen; Hansen and O'Leary)

An example of L–curve



The L-curve for the `Baart` problem, $m = n = 100$, noise level$= 10^{-3}$

The L–curve

Here we have to consider

$$\|x_\mu\|^2 = c^T A(A^T A + \mu I)^{-2} A^T c$$

and

$$\begin{aligned}\|c - Ax_\mu\|^2 &= c^T c + c^T A(A^T A + \mu I)^{-1} A^T A(A^T A + \mu I)^{-1} A^T c \\ &\quad -2c^T A(A^T A + \mu I)^{-1} A^T c\end{aligned}$$

Generalized cross–validation

GCV comes from statistics (Golub, Heath and Wahba)

The regularized problem is written as

$$\min\{\|c - Ax\|^2 + m\mu\|x\|^2\}$$

where $\mu \geq 0$ is the regularization parameter and the matrix A is m by n

The GCV estimate of the parameter μ is the minimizer of

$$G(\mu) = \frac{\frac{1}{m}\|(I - A(A^T A + m\mu I)^{-1} A^T)c\|^2}{(\frac{1}{m}\mathrm{tr}(I - A(A^T A + m\mu I)^{-1} A^T))^2}$$

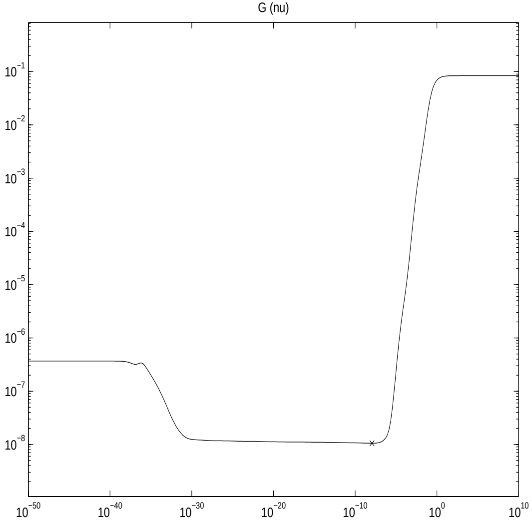
If we know the SVD of A and $m \geq n$ this can be computed as

$$G(\nu) = \frac{m\left\{\sum_{i=1}^{r} d_i^2 \left(\frac{\nu}{\sigma_i^2+\nu}\right)^2 + \sum_{i=r+1}^{m} d_i^2\right\}}{[m - n + \sum_{i=1}^{r} \frac{\nu}{\sigma_i^2+\nu}]^2}$$

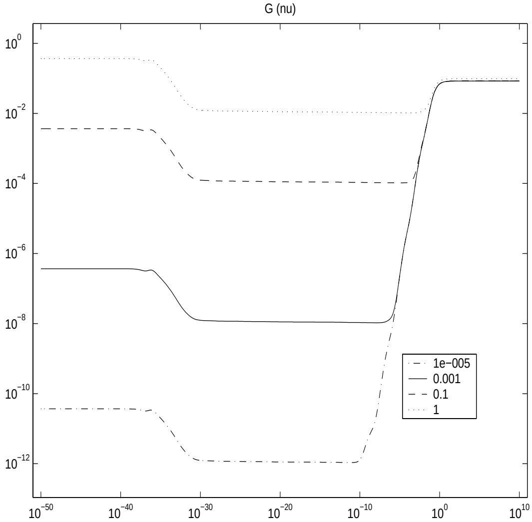
where $\nu = m\mu$

- G is almost constant when ν is very small or large, at least in log–log scale
- When $\nu \to \infty$, $G(\nu) \to \|c\|^2/m$
- When $\nu \to 0$ the situation is different wether $m = n$ or not

An example of GCV function



GCV function for the Baart problem, $m = n = 100$, noise level$=10^{-3}$



GCV functions for the Baart problem, $m = n = 100$ for different noise levels

The main problem is that the GCV function is usually quite flat near the minimum

For large problems we cannot use the SVD for computing the minimum

- First we approximate the trace in the denominator $\rightarrow \tilde{G}$
- Then we obtain bounds of all the quadratic forms in $\tilde{G}$
- Finally we have to locate the minimum of the lower and/or upper bounds

Approximation of the trace

Proposition (Hutchinson)

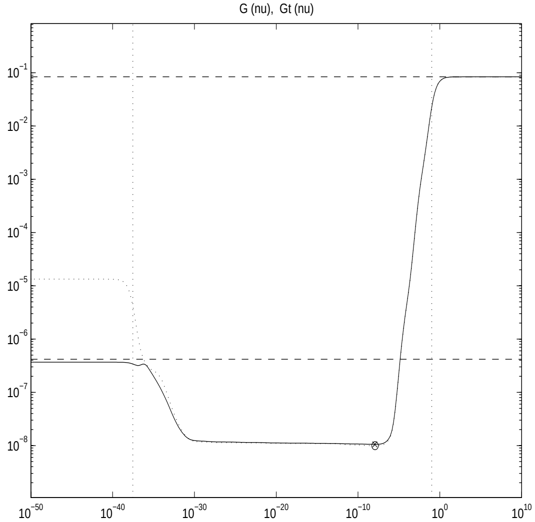
Let B be a symmetric matrix of order n with $\mathrm{tr}(B) \neq 0$
Let $\mathcal{Z}$ be a discrete random variable with values 1 and -1 with equal probability 0.5 and let z be a vector of n independent samples from $\mathcal{Z}$
Then $z^T Bz$ is an unbiased estimator of $\mathrm{tr}(B)$

$$E(z^T Bz) = \mathrm{tr}(B)$$

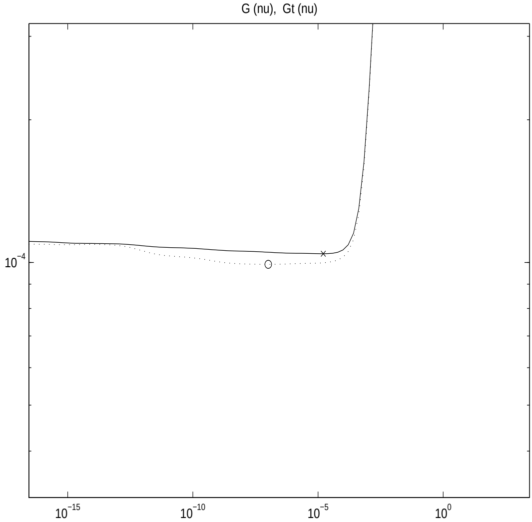
$$\mathrm{var}(z^T Bz) = 2 \sum_{i \neq j} b_{i,j}^2$$

where $E(\cdot)$ denotes the expected value and var denotes the variance

For GCV we just use one vector z and we have to bound a quadratic form



G (plain) and $\tilde{G}$ (dotted) functions for the `Baart` problem, $m = n = 100$, noise level$=10^{-3}$



G (plain) and $\tilde{G}$ (dotted) functions for the Baart problem, $m = n = 100$, noise level$=10^{-1}$

Computing bounds of quadratic forms

Most of the heuristics for finding a "good" value of μ involve quadratic forms

$$u^T f(A) u$$

where A is a symmetric matrix, u is a given vector and f is smooth function

How can we compute bounds for this quadratic form?

Quadratic forms

$$u^T f(A) u$$

Since A is symmetric

$$A = Q \Lambda Q^T$$

where Q is the orthonormal matrix whose columns are the normalized eigenvectors of A and Λ is a diagonal matrix whose diagonal elements are the eigenvalues λ_i. Then

$$f(A) = Q\, f(\Lambda)\, Q^T$$

In fact this is a definition of $f(A)$ when A is symmetric
Of course, usually we don't know Q and Λ. That's what makes the problem interesting!

$$
\begin{aligned}
u^T f(A) u &= u^T Q f(\Lambda) Q^T u \\
&= \gamma^T f(\Lambda) \gamma \\
&= \sum_{i=1}^{n} f(\lambda_i) \gamma_i^2
\end{aligned}
$$

This last sum can be considered as a Riemann–Stieltjes integral

$$
I[f] = u^T f(A) u = \int_a^b f(\lambda)\, d\alpha(\lambda)
$$

where the measure α is piecewise constant and defined by

$$
\alpha(\lambda) = \begin{cases} 0 & \text{if } \lambda < a = \lambda_1 \\ \sum_{j=1}^{i} \gamma_j^2 & \text{if } \lambda_i \leq \lambda < \lambda_{i+1} \\ \sum_{j=1}^{n} \gamma_j^2 & \text{if } b = \lambda_n \leq \lambda \end{cases}
$$

Gauss quadrature rules

The optimal quadrature rule of degree $2N - 1$ is called a Gauss quadrature

It was introduced by C.F. Gauss at the beginning of the nineteenth century

The general formula for a Riemann–Stieltjes integral is

$$I[f] = \int_a^b f(\lambda)\, d\alpha(\lambda) = \sum_{j=1}^{N} w_j f(t_j) + \sum_{k=1}^{M} v_k f(z_k) + R[f],$$

where the weights $[w_j]_{j=1}^N, [v_k]_{k=1}^M$ and the nodes $[t_j]_{j=1}^N$ are unknowns and the nodes $[z_k]_{k=1}^M$ are prescribed

see Davis and Rabinowitz; Gautschi; Golub and Welsch

- If $M = 0$, this is the Gauss rule with no prescribed nodes
- If $M = 1$ and $z_1 = a$ or $z_1 = b$ we have the Gauss–Radau rule
- If $M = 2$ and $z_1 = a, z_2 = b$, this is the Gauss–Lobatto rule

The term $R[f]$ is the remainder which generally cannot be explicitly computed
If the measure α is a positive non–decreasing function

$$R[f] = \frac{f^{(2N+M)}(\eta)}{(2N+M)!} \int_a^b \prod_{k=1}^{M} (\lambda - z_k) \left[\prod_{j=1}^{N} (\lambda - t_j) \right]^2 d\alpha(\lambda), \quad a < \eta < b$$

Note that for the Gauss rule, the remainder $R[f]$ has the sign of $f^{(2N)}(\eta)$
see Stoer and Bulirsch

If the sign of the derivative is known then we obtain bounds

The nodes and weights of the Gauss rule are computed using the three-term recurrence of the orthogonal polynomials associated with the measure α

The nodes are the eigenvalues of the tridiagonal (Jacobi) matrix given by the coefficients of the three-term recurrence

The weights are the squares of the first elements of the normalized eigenvectors

Other bounds can be obtained by using the Gauss-Radau quadrature rule (modifying one element of the Jacobi matrix)

The Lanczos algorithm

For our piecewise constant measure α the Jacobi matrix is obtained by using iterations of the Lanczos algorithm with A starting from u

starting from a vector $\tilde{v}^1 = u/\|u\|$

$$\alpha_1 = (Av^1, v^1), \tilde{v}^2 = Av^1 - \alpha_1 v^1$$

and then, for $k = 2, 3, \dots$

$$\eta_{k-1} = \|\tilde{v}^k\|$$

$$v^k = \frac{\tilde{v}^k}{\eta_{k-1}}$$

$$\alpha_k = (v^k, Av^k) = (v^k)^T Av^k$$

$$\tilde{v}^{k+1} = Av^k - \alpha_k v^k - \eta_{k-1} v^{k-1}$$

The coefficients α_k are the diagonal entries of the Jacobi matrix and the η_k are the subdiagonal entries

In fact for our problem the matrix is of the form $A^T A$ and we use the Golub–Kahan bidiagonalization algorithm

The Golub–Kahan bidiagonalization algorithm

It is a special case of the Lanczos algorithm for $A^T A$

The first algorithm (LB1) reduces A to upper bidiagonal form

Let $q^0 = c/\|c\|$, $r^0 = Aq^0$, $\delta_1 = \|r^0\|$, $p^0 = r^0/\delta_1$
then for $k = 1, 2, \ldots$

$$u^k = A^T p^{k-1} - \delta_k q^{k-1}$$

$$\gamma_k = \|u^k\|$$

$$q^k = u^k/\gamma_k$$

$$r^k = Aq^k - \gamma_k p^{k-1}$$

$$\delta_{k+1} = \|r^k\|$$

$$p^k = r^k/\delta_{k+1}$$

If

$$P_k = \begin{pmatrix} p^0 & \cdots & p^{k-1} \end{pmatrix}, \quad Q_k = \begin{pmatrix} q^0 & \cdots & q^{k-1} \end{pmatrix}$$

and

$$B_k = \begin{pmatrix} \delta_1 & \gamma_1 & & \\ & \ddots & \ddots & \\ & & \delta_{k-1} & \gamma_{k-1} \\ & & & \delta_k \end{pmatrix}$$

then P_k and Q_k, which is an orthogonal matrix, satisfy the equations

$$\begin{aligned} AQ_k &= P_k B_k \\ A^T P_k &= Q_k B_k^T + \gamma_k q^k (e^k)^T \end{aligned}$$

and therefore

$$A^T A Q_k = Q_k B_k^T B_k + \gamma_k \delta_k q^k (e^k)^T$$

The second algorithm (LB2) reduces A to lower bidiagonal form

Let $p^0 = c/\|c\|$, $u^0 = A^T p^0$, $\gamma_1 = \|u^0\|$, $q^0 = u^0/\gamma_1$,
$r^1 = Aq^0 - \gamma_1 p^0$, $\delta_1 = \|r^1\|$, $p^1 = r^1/\delta_1$
then for $k = 2, 3, \ldots$

$$u^{k-1} = A^T p^{k-1} - \delta_{k-1} q^{k-2}$$

$$\gamma_k = \|u^{k-1}\|$$

$$q^{k-1} = u^{k-1}/\gamma_k$$

$$r^k = Aq^{k-1} - \gamma_k p^{k-1}$$

$$\delta_k = \|r^k\|$$

$$p^k = r^k/\delta_k$$

If

$$P_{k+1} = \begin{pmatrix} p^0 & \cdots & p^k \end{pmatrix}, \quad Q_k = \begin{pmatrix} q^0 & \cdots & q^{k-1} \end{pmatrix}$$

and

$$C_k = \begin{pmatrix} \gamma_1 & & & \\ \delta_1 & \ddots & & \\ & \ddots & \ddots & \\ & & \ddots & \gamma_k \\ & & & \delta_k \end{pmatrix}$$

a $k+1$ by k matrix, then P_k and Q_k, which is an orthogonal matrix, satisfy the equations

$$\begin{aligned} AQ_k &= P_{k+1} C_k \\ A^T P_{k+1} &= Q_k C_k^T + \gamma_{k+1} q^k (e^{k+1})^T \end{aligned}$$

Of course, by eliminating P_{k+1} in these equations we obtain

$$A^T A Q_k = Q_k C_k^T C_k + \gamma_{k+1} \delta_k q^k (e^k)^T$$

and

$$C_k^T C_k = B_k^T B_k = J_k$$

B_k is the Cholesky factor of J_k and $C_k^T C_k$

J_k is the tridiagonal Jacobi matrix for $A^T A$

The Golub and Von Matt algorithm

Let $s_z(\nu) = z^T(A^TA + \nu I)^{-1}z$, where z is a random vector
Using Gauss and Gauss–Radau we can obtain

$$g_z(\nu) \leq s_z(\nu) \leq r_z(\nu)$$

We can also bound
$s_c^{(p)}(\nu) = c^TA(A^TA + \nu I)^pA^Tc, \quad p = -1, -2$ satisfying

$$g_c^{(p)}(\nu) \leq s_c^{(p)}(\nu) \leq r_c^{(p)}(\nu)$$

We want to compute approximations of the minimum of

$$\tilde{G}(\mu) = m\frac{c^Tc - s_c^{(-1)}(\nu) - \nu s_c^{(-2)}(\nu)}{(m - n + \nu s_z(\nu))^2}$$

We define

$$L_0(\nu) = m\frac{c^Tc - r_c^{(-1)}(\nu) - \nu r_c^{(-2)}(\nu)}{(m - n + \nu r_z(\nu))^2}$$

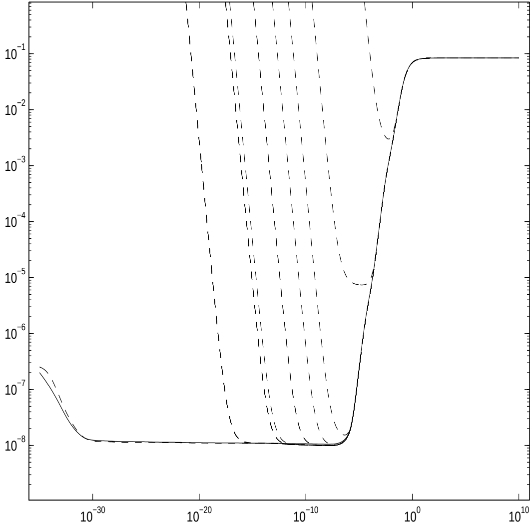
$$U_0(\nu) = m\frac{c^Tc - g_c^{(-1)}(\nu) - \nu g_c^{(-2)}(\nu)}{(m - n + \nu g_z(\nu))^2}$$

These quantities L_0 and U_0 are lower and upper bounds for the estimate of $G(\mu)$

We can also compute estimates of the derivatives of L_0 and U_0

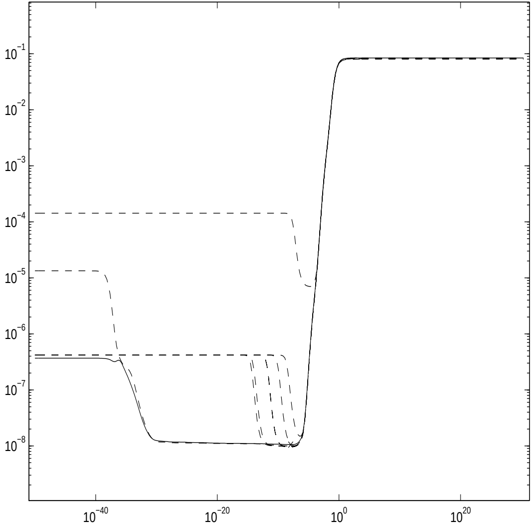
These bounds improve with the number of Lanczos iterations

The upper bound does not have the right asymptotic behavior when $m = n$ and $\nu \to 0$



G (plain) and $\tilde{G}$ (dashed) functions and upper bounds for the Baart problem, $m = n = 100$, noise level$=10^{-3}$

To obtain a better behavior we add a term $\|c\|^2$ to the denominator



G (plain) and $\tilde{G}$ (dashed) functions and upper bounds for the Baart problem, $m = n = 100$, noise level$=10^{-3}$

Optimization of the algorithm

- We choose a (small) value of ν (denoted as ν_0)
- When

$$\left| \frac{U_k^0(\nu_0) - U_{k-1}^0(\nu_0)}{U_{k-1}^0(\nu_0)} \right| \leq \epsilon_0$$

  we start computing the minimum of the upper bound

The algorithm for finding the minimum is as follows

- We work in log–log scale and compute only a minimizer of the upper bound
- We evaluate the numerator of the approximation $\tilde{G}$ by computing the SVD of B_k once per iteration
- We compute 50 samples of the function on a regular mesh
- We locate the minimum, say the point k, we then compute again 50 samples in the interval $[k-1\ \ k+1]$

- We use bisection for computing a local minimum in this interval
- After locating a minimum ν_k with a value of the upper bound U_k^0 at iteration k, the stopping criteria is

$$\left|\frac{\nu_k - \nu_{k-1}}{\nu_{k-1}}\right| + \left|\frac{U_k^0 - U_{k-1}^0}{U_{k-1}^0}\right| \leq \epsilon$$

GCV algorithms, Baart problem

	noise	μ	$\|c - Ax\|$	$\|x - x_0\|$	t (s)
vm	10^{-7}	$9.6482\ 10^{-15}$	$9.8049\ 10^{-8}$	$5.9424\ 10^{-2}$	0.38
	10^{-5}	$9.7587\ 10^{-12}$	$9.8566\ 10^{-6}$	$6.5951\ 10^{-2}$	0.18
	10^{-3}	$1.2018\ 10^{-8}$	$9.8573\ 10^{-4}$	$1.5239\ 10^{-1}$	0.16
	10^{-1}	$1.0336\ 10^{-7}$	$9.8730\ 10^{-2}$	1.6614	–
gm-opt	10^{-7}	$1.0706\ 10^{-14}$	$9.8058\ 10^{-8}$	$5.9519\ 10^{-2}$	0.18
	10^{-5}	$1.0581\ 10^{-11}$	$9.8588\ 10^{-6}$	$6.5957\ 10^{-2}$	0.27
	10^{-3}	$1.3077\ 10^{-8}$	$9.8582\ 10^{-4}$	$1.5205\ 10^{-1}$	0.14
	10^{-1}	$1.1104\ 10^{-7}$	$9.8736\ 10^{-2}$	1.6227	–

GCV algorithms, Phillips problem

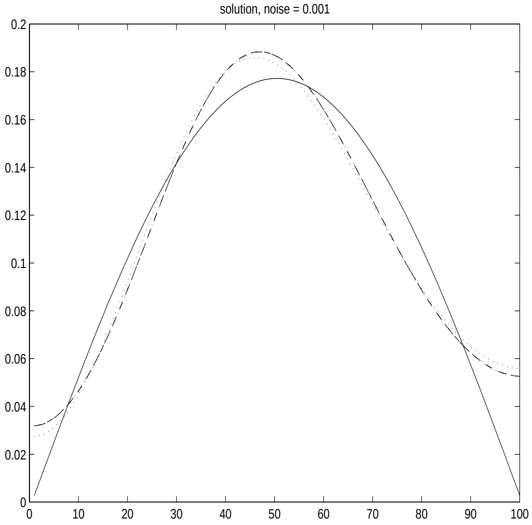
	noise	μ	$\|c - Ax\|$	$\|x - x_0\|$	t (s)
vm	10^{-7}	$8.7929\ 10^{-11}$	$9.0162\ 10^{-8}$	$2.2391\ 10^{-4}$	29.50
	10^{-5}	$4.5432\ 10^{-9}$	$9.0825\ 10^{-6}$	$2.2620\ 10^{-3}$	6.09
	10^{-3}	$4.3674\ 10^{-7}$	$9.7826\ 10^{-4}$	$1.0057\ 10^{-2}$	1.14
	10^{-1}	$3.8320\ 10^{-5}$	$9.8962\ 10^{-2}$	$9.3139\ 10^{-2}$	0.16
gm-opt	10^{-7}	$1.6343\ 10^{-10}$	$1.1260\ 10^{-7}$	$2.2163\ 10^{-4}$	15.30
	10^{-5}	$5.3835\ 10^{-9}$	$9.1722\ 10^{-6}$	$2.1174\ 10^{-3}$	6.09
	10^{-3}	$4.1814\ 10^{-7}$	$9.7737\ 10^{-4}$	$1.0375\ 10^{-2}$	0.66
	10^{-1}	$4.1875\ 10^{-5}$	$9.9016\ 10^{-2}$	$9.0659\ 10^{-2}$	0.22

Comparisons of methods

Baart problem, $n = 100$

noise	meth	μ	$\|c - Ax\|$	$\|x - x_0\|$
10^{-3}	μ opt	$2.7826\ 10^{-8}$	$2.3501\ 10^{-3}$	$1.5084\ 10^{-1}$
	vm	$1.2018\ 10^{-8}$	$9.8573\ 10^{-4}$	$1.5239\ 10^{-1}$
	gm-opt	$1.3077\ 10^{-8}$	$9.8582\ 10^{-4}$	$1.5205\ 10^{-1}$
	gcv	$9.4870\ 10^{-9}$	$9.8554\ 10^{-4}$	$1.5362\ 10^{-1}$
	disc	$8.4260\ 10^{-8}$	$1.0000\ 10^{-3}$	$1.5556\ 10^{-1}$
	gr	$1.7047\ 10^{-7}$	$1.0235\ 10^{-3}$	$1.6373\ 10^{-1}$
	lc	$4.5414\ 10^{-9}$	$9.8524\ 10^{-4}$	$1.6028\ 10^{-1}$
	qo	$1.2586\ 10^{-8}$	$9.8450\ 10^{-4}$	$6.6072\ 10^{-1}$

μ opt is the value giving (approximately) the minimum error

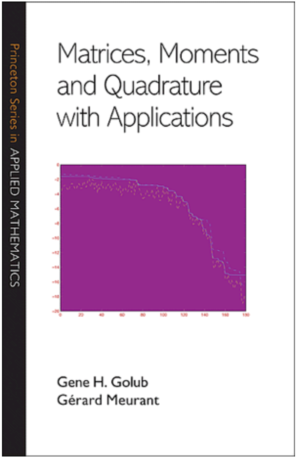


Solutions for the Baart problem, $m = n = 100$, noise level=10^{-3}, solid=unperturbed solution, dashed=vm, dot–dashed=gm-opt

Phillips problem, $n = 200$

noise	meth	μ	$\|c - Ax\|$	$\|x - x_0\|$
10^{-5}	μ opt	$1.3725\ 10^{-7}$	$2.9505\ 10^{-14}$	$1.6641\ 10^{-3}$
	vm	$4.5432\ 10^{-9}$	$9.0825\ 10^{-6}$	$2.2620\ 10^{-3}$
	gm-opt	$5.3835\ 10^{-9}$	$9.1722\ 10^{-6}$	$2.1174\ 10^{-3}$
	gcv	$3.1203\ 10^{-9}$	$8.9283\ 10^{-6}$	$2.6499\ 10^{-3}$
	disc	$1.2107\ 10^{-8}$	$1.0000\ 10^{-5}$	$1.6873\ 10^{-3}$
	gr	$4.1876\ 10^{-8}$	$1.5784\ 10^{-5}$	$1.9344\ 10^{-3}$
	lc	$3.6731\ 10^{-14}$	$2.4301\ 10^{-6}$	$7.9811\ 10^{-1}$
	qo	$1.5710\ 10^{-8}$	$1.0542\ 10^{-5}$	$1.6463\ 10^{-3}$

For more details, see the book published by Princeton University Press in 2010



D. Calvetti, G.H. Golub and L. Reichel, *Estimation of the L–curve via Lanczos bidiagonalization*, BIT, v 39 n 4, (1999), pp 603–619

D. Calvetti, P.C. Hansen and L. Reichel, *L–curve curvature bounds via Lanczos bidiagonalization*, Elec. Trans. Numer. Anal., v 14, (2002), pp 20–35

H. Gfrerer, *An a posteriori parameter choice for ordinary and iterated Tikhonov regularization of ill–posed problems leading to optimal convergence rates*, Math. Comp., v 49, (1987), pp 507–522

G.H. Golub, M. Heath and G. Wahba, *Generalized cross–validation as a method to choosing a good ridge parameter*, Technometrics, v 21 n 2, (1979), pp 215–223

G. H. Golub and W. Kahan, *Calculating the singular values and pseudo-inverse of a matrix*, SIAM J. Numer. Anal., v 2 (1965), pp 205–224

G.H. Golub and U. von Matt, *Tikhonov regularization for large scale problems*, in Scientific Computing, G.H. Golub, S.H. Lui, F. Luk and R. Plemmons Eds., Springer, (1997), pp 3–26

G.H. Golub and U. von Matt, *Generalized cross–validation for large scale problems*, in Recent advances in total least squares techniques and errors in variable modeling, S. van Huffel ed., SIAM, (1997), pp 139–148

M. Hanke and T. Raus, *A general heuristic for choosing the regularization parameter in ill–posed problems*, SIAM J. Sci. Comput., v 17, (1996), pp 956–972

P.C. Hansen, *Regularization tools: a Matlab package for analysis and solution of discrete ill–posed problems*, Numer. Algo., v 6, (1994), pp 1–35

P.C. Hansen and D.P. O'Leary, *The use of the L–curve in the regularization of discrete ill–posed problems*, SIAM J. Sci. Comput., v 14, (1993), pp 1487–1503

P.C. Hansen, T.K. Jensen and G. Rodriguez, *An adaptive pruning algorithm for the discrete L-curve criterion*, J. Comp. Appl. Math., v 198 n 2, (2007), pp 483–492

C.L. Lawson and R.J. Hanson, *Solving least squares problems*, SIAM, (1995)

A.S. Leonov, *On the choice of regularization parameters by means of the quasi–optimality and ratio criteria*, Soviet Math. Dokl., v 19, (1978), pp 537–540

V.A. Morozov, *Methods for solving incorrectly posed problems*, Springer, (1984)

A.N. Tikhonov, *Ill–posed problems in natural sciences*, Proceedings of the international conference Moscow August 1991, (1992), TVP Science publishers.

A.N. Tikhonov and V.Y. Arsenin, *Solutions of ill–posed problems*, (1977), Wyley