M2AN. MATHEMATICAL MODELLING AND NUMERICAL ANALYSIS
- MODÉLISATION MATHÉMATIQUE ET ANALYSE NUMÉRIQUE

DIDIER GIRARD

Practical optimal regularization of large linear systems

M2AN. Mathematical modelling and numerical analysis - Modélisation mathématique et analyse numérique, tome 20, n° 1 (1986), p. 75-87

<http://www.numdam.org/item?id=M2AN_1986__20_1_75_0>

MATHEMATICAL MODELLING AND NUMERICAL ANALYSIS
MODÉLISATION MATHÉMATIQUE ET ANALYSE NUMÉRIQUE
(Vol 20, n° 1, 1986, p 75 à 87)

PRACTICAL OPTIMAL REGULARIZATION OF LARGE LINEAR SYSTEMS (*)

by Didier GIRARD (1)

Communicated by P J LAURENT

Abstract — *We seek an estimate of $x^0 \in \mathbb{R}^n$, given a noisy value ($y \in \mathbb{R}^m$) of $y^0 = Kx^0$ where K is a known $m \times n$ matrix The regularization method produces a set of estimates x_ρ obtained by minimizing $J_\rho(x)$ in $\mathbb{R}^n$*

$$J_\rho(x) = \| Kx - y \|_m^2 + \rho \| Lx \|_{n-p}^2$$

where $\| \cdot \|_m$ is the Euclidean norm in $\mathbb{R}^m$, L is an a priori chosen $(n - p) \times n$ matrix and ρ is the regularization parameter

Three methods which give an automatic choice (ρ^*) of ρ are recalled We study here the numerical implementation of these practical approximations and present, for each of the three methods, an algorithm which estimates ρ^* and computes x_{ρ^*} (the minimizer of J_{ρ^*}) with a cost (number of operations and memory place) nearly identical to the minimum cost of the computation of one estimate $x_\rho = (K^t K + \rho L^t L)^{-1} K^t y$ (the preliminary computations, independent of y, are not taken in account)

This algorithm is applied to a tomographical picture reconstruction problem where K is a large sparse matrix ($m \gg n$) and $K^t K$ and L are block-circulant Simulations show that the cross-validation method is impressively efficient at choosing ρ^*

Résumé — *On recherche une approximation de $x^0 \in \mathbb{R}^n$ a partir de la connaissance d'une valeur approchée y de $y^0 = Kx^0$ où K est une matrice (m, n) connue La méthode de régularisation etudiee consiste à choisir une matrice $L(n - p, n)$ telle que la norme euclidienne $\| Lx \|_{n-p}$ mesure l'irrégularité de x, et à minimiser dans $\mathbb{R}^n$,*

$$J_\rho(x) = \| Kx - y \|_m^2 + \rho \| Lx \|_{n-p}^2$$

ou ρ est le paramètre de régularisation > 0 Trois critères de choix d'un paramètre ρ^* optimal sont rappelés On étudie ici la mise en œuvre numérique d'une telle approximation quand $m \gg n$ Après s'être ramené à un probleme equivalent a n équations, on propose une méthode de choix automatique de ρ^* pour chacun des 3 critères, et de calcul de x_{ρ^*} minimisant $J_{\rho^*}(x)$, pour un coût total (place en mémoire et temps de calcul) comparable au coût minimum de la résolution directe de l'équation normale $(K^t K + \rho L^t L) x_\rho = K^t y$, sans prendre en compte les pré-calculs indépendants de $K^t y$

Cette méthode est appliquée à un probleme de reconstruction d'image en tomographie où K est une grande matrice creuse, $K^t K$ et L sont bloc-circulantes On vérifie l'efficacité de la validation croisée comme critère de choix de ρ^*

(*) Received in June 1984

(1) Laboratoire d'Électronique et de Technologie de l'Informatique/MCTE, CENG, 38041 Grenoble Cedex et IMAG, Université Scientifique et Médicale de Grenoble

1. INTRODUCTION

Let K be a real $m \times n$ matrix. We consider the system of linear equations :

$$Kx = y^0 = y + \varepsilon \tag{1.1}$$

where we know only a noisy value ($y \in \mathbb{R}^m$) of the exact data vector y^0 and ε is a random noise vector. If we assume a « white noise » :

$$\begin{gathered} \varepsilon = (\varepsilon_1, \varepsilon_2, ..., \varepsilon_m)^t \quad \text{with} \quad E(\varepsilon_i) = 0\,, \\ E(\varepsilon_i\, \varepsilon_j) = \sigma^2\, \delta_{ij} \quad i, j = 1, ..., m \end{gathered} \tag{1.2}$$

then an intuitive idea is to say that given y and σ^2, there is only one constraint on x :

$$S(x) = \| Kx - y \|_m^2 \leqslant m.\sigma^2 \tag{1.3}$$

where $\| \cdot \|_m$ is the usual Euclidean norm in $\mathbb{R}^m$.

We are interested in the situations where either because σ is too large or because K is too ill-conditioned, a wide range of values of x may satisfy constraint (1.3), and therefore we need prior information concerning the solution. Defining the irregularity of x by $\| Lx \|_{n-p}$ where L is a chosen $(n - p) \times n$ matrix (when the origin of system (1.1) is the discretization of an integral equation, L is normally a discrete approximation of a derivative operator), we take as the solution the least irregular x satisfying (1.3). Except for some incoherent value of σ, such a solution minimizes $J_\rho(x)$ in $\mathbb{R}^n$:

$$J_\rho(x) = \| Kx - y \|_m^2 + \rho \| Lx \|_{n-p}^2 \tag{1.4}$$

where $\rho > 0$ is the regularization parameter that we must choose.

It is well known that when $N(K) \cap N(L) = \{ 0 \}$ (where $N(K)$ is the null space of K), for every $\rho > 0$, minimization of J_ρ has only one solution x_ρ given by :

$$(K^t K + \rho L^t L)\, x_\rho = K^t y\,. \tag{1.5}$$

The value of ρ is critical to the quality of x_ρ : if ρ is too small the data error induces a solution which is too irregular and if ρ is too large x_ρ is less sensitive to noise but Kx_ρ may be too far from the data.

In this paper we study methods of automatic choice of ρ and we show some important simplifications for the numerical implementation with $m \gg n$.

In § 2 we rapidly review three different methods. We shall see that the optimal ρ is the root of or minimizer of a real function of the variable ρ. As we have to approximate this optimal value by an iterative procedure, we need a fast algorithm for the evaluation of this function. For this evaluation, we have to compute $S(x_\rho) = \| Kx_\rho - y \|_m^2$ and except for the first method, the trace of the (m, m) matrix $A(\rho)$ satisfying $Kx_\rho = A(\rho)\, y$..

We will first transform our initial problem (1.4) into an equivalent problem (§ 3) with less than n equations :

$$\operatorname*{Min}_{x \in \mathbb{R}^n} (\tilde{S}(x) + \rho \| Lx \|_{n-p}^2)$$

where :

$$\tilde{S}(x) = \| \tilde{K}x - \tilde{y} \|_l^2, \quad \tilde{K} = F^t K, \quad \tilde{y} = F^t y, \quad l = \operatorname{rank}(K)$$

and F is any $m \times l$ matrix with columns f_i, $i = 1, ..., l$, forming an orthogonal basis of the range of K.

Then we will show in § 4 that one can choose F in order to obtain the following result : if L is invertible, we can compute a $n \times l$ matrix B, and l real values $d_i > 0$, $i = 1, ..., l$, using the singular value decomposition of $L^{-1t} K^t K L^{-1}$, such that :

$$\begin{aligned} \tilde{A}(\rho) &= D^2(D^2 + \rho I)^{-1} \\ \tilde{y} &= D^{-1} B^t K^t y \\ x_\rho &= B(D^2 + \rho I)^{-1} D\tilde{y} \end{aligned}$$

where $\begin{cases} D \text{ is the diagonal matrix } \begin{bmatrix} d_1 & & & \\ & d_2 & & \\ & & \ddots & \\ & & & d_l \end{bmatrix} \\ I \text{ is the identity in } \mathbb{R}^l \end{cases}$

If L is not invertible, we obtain (§ 5) similar expressions, with or without boundary values on x_ρ.

In § 6 we numerically apply the above results to a tomographical picture reconstruction problem. Here $m = 96 \times 128 \times 64$ and $n = 96 \times 128$ but K is very sparse and as $K^t K$ and L are block-circulant the preliminary computation of B and D is much simplified and their memorisation takes up only 97 64×64 matrices. Generalized cross-validation is implemented and simulations with generated noisy data show that it is very efficient.

2. METHODS FOR CHOOSING THE PARAMETER

These methods have been previously studied for the smoothing of data by Spline functions (see Reinsch [1], Wahba [4]) and they can be generalized [3] to any linear transformation of noisy data when this transformation is governed by a parameter which defines the degree of smoothing of these data. Here, noting $y_\rho = Kx_\rho$, the data vector predicted by x_ρ,

$A(\rho) = K(K^t K + \rho L^t L)^{-1} K^t$ is the $m \times m$ matrix satisfying $y_\rho = A(\rho) y$.

In the case of white noise with a known variance σ^2 (1.2) the generalisation of Reinsch's suggestion is to choose ρ_s so that :

$$S(x_{\rho_s}) = \| A(\rho_s) y - y \|_m^2 = m\sigma^2 . \tag{2.1}$$

If there is not such a ρ_s, the constraint $S(x) \leqslant m\sigma^2$ either is never satisfied or is always satisfied with $<$ instead of $\leqslant$. Generally one such ρ_s exists and it is then easy to show that x_{ρ_s} is the solution of the problem :

$$\underset{x \in \mathbb{R}^n / S(x) \leqslant m\sigma^2}{\text{Min}} \| Lx \|_{n-p}^2 .$$

Using the exact data vector y^0, Wahba has defined in [2] the optimum ρ_0 as the ρ which minimizes $R(\rho)$, the mean square error of prediction of this vector :

$$R(\rho) = \frac{1}{m} \| A(\rho) y - y^0 \|_m^2 \tag{2.2}$$

$R(\rho)$ cannot be known in practice, but we can easily demonstrate [4] that, with white noise, $\hat{R}(\rho)$, which is computable, is an unbiased estimate of $E(R(\rho))$:

$$\hat{R}(\rho) = \frac{1}{m} \| A(\rho) y - y \|_m^2 - \frac{\sigma^2}{m} \text{Trace}\,(A(\rho) - I)^2 + \frac{\sigma^2}{m} \text{Trace}\, A(\rho)^2 . \tag{2.3}$$

The search of ρ which minimizes $\hat{R}$ constitutes, then, a second method. The third one needs no statistical knowledge concerning the errors : it consists of choosing the ρ_v which minimizes $V(\rho)$:

$$V(\rho) = \frac{\frac{1}{m} \| A(\rho) y - y \|^2}{\left| \frac{1}{m} \text{Trace}\,(A(\rho) - I) \right|^2} . \tag{2.4}$$

This is Wahba's method of generalized cross-validation [3], [4]. It originates from a simple idea (Stone [5]) : one omits a datum and studies the validity of different ρ by measuring how the corresponding estimates x_ρ, computed without this datum, can predict it. We choose the ρ which gives the best prediction, on the average, for all the possible omissions. Wahba has also shown that when white noise is assumed, ρ_v is a good estimate of the minimizer of $E(R(\rho))$ (and is even better when m is large).

3. EQUIVALENT PROBLEM

We remark that for every ρ the regularized solution x_ρ does not depend directly on y but on $K^t y \in \mathbb{R}^n$. In fact we can transform our problem into an equivalent problem made of l linear equations with $l = \text{rank}(K)$. Let P be the projection in $\mathbb{R}^m$ on the range of K, $R(K)$. Then,

$$\| Kx - y \|_m^2 = \| Kx - Py \|_m^2 + \| Py - y \|_m^2 \cdot$$

And, as $Kx - Py \in R(K)$, if F is any $m \times l$ matrix, with columns $f_1, f_2, \ldots, f_l \in \mathbb{R}^m$ forming an orthogonal basis of $R(K)$, then we have :

$$\| Kx - Py \|_m^2 = \| F^t(Kx - Py) \|_l^2 \quad \text{and} \quad F^t(Py - y) = 0 .$$

Then, letting $\tilde{K} = F^t K$, $\tilde{y} = F^t y$, the minimization (1.4) is equivalent to

$$\underset{x \in \mathbb{R}^n}{\text{Min}} \left(\| \tilde{K}x - \tilde{y} \|_l^2 + \rho \| Lx \|_{n-p}^2 \right) .$$

On the basis of this formulation we will now choose the parameter ρ.

When we assume white noise with variance σ^2 then a white noise is still satisfied by $\tilde{y}$ (because the covariance matrix of $F^t \varepsilon$ is $F^t(\sigma^2 I_m) F = \sigma^2 I_l$), and the methods 1 and 2 are applicable.

If we let $\tilde{y}^0 = F^t y_0$ then $\| Kx_\rho - y^0 \|_m^2$ is equal, if $y^0 \in R(K)$, to $\| \tilde{K}x_\rho - \tilde{y}^0 \|_l^2$ or else differs by a constant which is independant of ρ, so the optimum ρ (following Wahba (2.2)) is the same in the two problems. The generalized cross-validation applied to $\tilde{A}(\rho) = \tilde{K}(\tilde{K}^t \tilde{K} + \rho L^t L)^{-1} \tilde{K}^t$ and $\tilde{y}$ can give an estimate of this optimum.

In the numerical implementation of these methods, we have to evaluate $\tilde{S}(x_\rho) = \| \tilde{A}(\rho) \tilde{y} - \tilde{y} \|_l^2$ and Trace $(\tilde{A}(\rho))$ for several values of ρ. We see below how to choose F in order to simplify these evaluations.

4. AN ALGORITHM FOR OPTIMAL SOLUTIONS WHEN L IS INVERTIBLE

When L is invertible, we can write :

$$x_\rho = (\tilde{K}^t \tilde{K} + \rho L^t L)^{-1} \tilde{K}^t \tilde{y}$$
$$= L^{-1}(L^{-1t} \tilde{K}^t \tilde{K} L^{-1} + \rho I_n)^{-1} L^{-1t} \tilde{K}^t \tilde{y}. \quad (4.1)$$

Consider now the singular value decomposition of KL^{-1} (see Golub [6]) :

$$KL^{-1} = UDV^t$$

where U and V are $m \times l$ and $n \times l$ orthogonal matrices and D is the diagonal matrix $\text{diag}(d_1, d_2, ..., d_l)$, with $d_1 \geqslant d_2 \geqslant \cdots \geqslant d_l > 0$.

Then we can take $F = U$, so that :

$$\tilde{K} L^{-1} = DV^t = M \quad (4.2)$$

and

$$\tilde{y} = U^t y = D^{-1} V^t L^{-1t} K^t y. \quad (4.3)$$

Substituting (4.2) and (4.3) in (4.1) gives :

$$Lx_\rho = (M^t M + \rho I_n)^{-1} M^t \tilde{y} = M^t (MM^t + \rho I_l)^{-1} \tilde{y}$$
$$= VD(D^2 + \rho I_l)^{-1} D^{-1} V^t L^{-1t} K^t y.$$

Now let $B = L^{-1} V$. Then one obtains :

$$x_\rho = B \,\text{diag}\left(\frac{1}{d_i^2 + \rho}, i = 1, ..., l\right) B^t K^t y \quad (4.4)$$

$$\tilde{A}(\rho) = MM^t (MM^t + \rho I_l)^{-1} = \text{diag}\left(\frac{d_i^2}{d_i^2 + \rho}, i = 1, ..., l\right) \quad (4.5)$$

$$\tilde{A}(\rho)\,\tilde{y} - \tilde{y} = \text{diag}\left(\frac{d_i}{d_i^2 + \rho} - \frac{1}{d_i}, i = 1, ..., l\right) B^t K^t y. \quad (4.6)$$

Note that we don't need to compute U, and D^2 and V are given by the singular value decomposition of $L^{-1t} K^t K L^{-1}$. We observe that this method is well adapted to the case where K is a large sparse matrix (so that we can easily compute $K^t y$) and when we need to solve (1.1) a great number of times with different data vectors, since we can then precompute (the most expensive part) and store the $n \times l$ matrix B and the l values d_i. Then the calculation of an optimal parameter and of the corresponding regularized solution is reduced to the four following steps :

1) compute $K^t y$

2) compute $z = B^t(K^t y)$ $(0(n^2)$ operations)

3) search by an iterative procedure for ρ^*, the zero or the minimizer of a function (e.g. $\tilde{V}(\rho)$), computable with $0(n)$ operations

4) compute $x_{\rho^*} = B \operatorname{diag}\left(\frac{1}{d_i^2 + \rho^*}, i = 1, ..., l\right) z$ $(0(n^2)$ operations).

So the whole algorithm requires the same memory space and about twice the number of operations as the calculation of x_ρ (with a prior chosen ρ and memorisation of $(K^t K + \rho L^t L)^{-1}$).

5.1 SOLUTIONS WITH AUXILIARY CONSTRAINTS WHEN L IS NOT INVERTIBLE

If L is not invertible, again we can easily obtain an expression of x_ρ similar to (4.1) as soon as we assume some auxiliary boundary value on the solution.

Since we choose L, we can assume its rank is $n - p$.

A frequently occurring example in the case of an integral equation of one variable is the discrete approximation of the derivative :

$$L = \begin{bmatrix} 1 & -1 & & & (0) \\ & 1 & -1 & & \\ & & \ddots & \ddots & \\ (0) & & & \ddots & \\ & & & & 1 \quad -1 \end{bmatrix}, p = 1 .$$

Let N be a $n \times p$ matrix with columns forming a basis of $N(L)$.

Every $x \in \mathbb{R}^n$ has an unique expression $x = L^\dagger w + Nz$ where $L^\dagger$ is the pseudo-inverse of L, $w = Lx \in \mathbb{R}^{n-p}$, $z \in \mathbb{R}^p$.

In order to be able to express x as a function of Lx, we must add p boundary values :

$$Z^t x = c \quad \text{where} \left| \begin{array}{l} Z \text{ is a } n \times p \text{ matrix} \\ Z^t N \text{ invertible} \\ c \in \mathbb{R}^p . \end{array} \right. \tag{5.1}$$

With (5.1), we get :

$$w = Lx \Leftrightarrow x = L^\dagger w + N(Z^t N)^{-1}(c - Z^t L^\dagger w)$$

$$x = (I - N(Z^t N)^{-1} Z^t) L^\dagger w + N(Z^t N)^{-1} c . \tag{5.2}$$

Let $G = (I - N(Z^t N)^{-1} Z^t) L^\dagger$. Thus the problem

$$\underset{x \in \mathbb{R}^n / Z^t x = c}{\text{Min}} (S(x) + \rho \| Lx \|^2_{n-p})$$

becomes :

$$\underset{w \in \mathbb{R}^{n-p}}{\text{Min}} (\| KGw - (y - KN(Z^t N)^{-1} c) \|^2_m + \rho \| w \|^2_{n-p}).$$

We then use our procedure of § 3 and § 4 with KG, I in place of K, L. Without repeating this procedure, we note only that the singular value decomposition introduced here is :

$$KG = UDV^t.$$

Then (4.4) gives :

$$w_\rho = V \operatorname{diag}\left(\frac{1}{d_i^2 + \rho}, i = 1, ..., l\right) V^t G^t K^t (y - KN(Z^t N)^{-1} c).$$

Substituting this in (5.2) :

$$x_\rho = GV \operatorname{diag}\left(\frac{1}{d_i^2 + \rho}, i = 1, ..., l\right) V^t G^t K^t (y - KN(Z^t N)^{-1} c) + \\ + N(Z^t N)^{-1} c$$

and (4.5), (4.6) are now :

$$\tilde{A}(\rho) = \operatorname{diag}\left(\frac{d_i^2}{d_i^2 + \rho}, i = 1, ..., l\right)$$

$$\tilde{A}(\rho)\, \tilde{y} - \tilde{y} = \operatorname{diag}\left(\frac{d_i}{d_i^2 + \rho} - \frac{1}{d_i}, i = 1, ..., l\right) \times \\ \times V^t G^t K^t (y - KN(Z^t N)^{-1} c).$$

In comparison with § 4, we have to compute G in place of L^{-1} and subtract from the data vector a linear combination of the p columns of $KN(Z^t N)^{-1}$.

5.2 SOLUTIONS WITHOUT AUXILIARY CONSTRAINTS

Furthermore, if we don't want to add some auxiliary constraint (5.1), we can use the following relation which is always satisfied by x_ρ :

$$N^t K^t K x_\rho = N^t K^t y.$$

This relation is obtained by multiplying $(K^t K + \rho L^t L) x_\rho = K^t y$ by N^t.

We verify that $Z^t = N^t K^t K$ is such that $Z^t N$ is invertible. Indeed, $Z^t N = N^t K^t KN$ is symmetric and $z^t Z^t Nz = \| KNz \|^2$ is > 0, except for $Nz = 0$ hence $z = 0$, as long as $N(K) \cap N(L) = \{ 0 \}$.

In the previous results, the constant term in (5.2) is :

$$N(Z^t N)^{-1} c = N(N^t K^t KN)^{-1} N^t K^t y .$$

Let x_∞ be this term, since it's the limit of x_ρ when $\rho \to + \infty$.

We note that x_∞ is the element of $N(L)$ which minimizes $\| Kx - y \|$ for $x \in N(L)$ and G is now :

$$G = [I - N(N^t K^t KN)^{-1} N^t K^t K] L^\dagger .$$

6. A NUMERICAL APPLICATION : POSITRON EMISSION TOMOGRAPHY WITH TIME OF FLIGHT MEASUREMENTS

This tomograph has detectors placed in a circle around a positron-emitting object and we want to estimate the activity of the emitter at every point in the plane of the circle. The emission of one positron produces, at the same point, two gamma rays in opposite directions. Hence, if two detectors detect gamma rays at about the same time then we know that a positron emission has occurred on the line between these detectors. The time-of-flight measurement gives an approximate localisation of this emission along this line. In the tomograph of Leti [7] the acquisition system groups all detections in 96 directions uniformly distributed throughout 180°. For each direction there are 128 parallel lines and, on each line, 64 time-of-flight groupings. Using the rotational invariance of this geometry and reorganizing with 192 directions covering 360°, these $m = 96 \times 128 \times 64$ counts can be taken as approximations of :

$$y^0_{k,i,j} = \int g_j(v)\, f_k(u_i, v)\, dv \qquad \begin{matrix} k = 1, ..., 192 \\ i = 1, ..., 64 \\ j = 1, ..., 64 \end{matrix} \tag{6.1}$$

where

$f_k(u, v)$ is the expression of the search activity in the Cartesian frame where the 64 lines in the direction k satisfy :

$$u = u_i \,(= \text{distance from the center}) \quad i = 1, ..., 64$$

g_j is a known function given by : $g_j(v) = \int_{C_j} g(v - s)\, ds$ where C_j is the interval on 0 v of the positions corresponding to the time-of-flight class j and g is the probability density of localisation error (taken as a given normal density).

Even if we refine this model, the measurement $y_{k,l,j}$ is only a Poisson random variable of unknown rate $y^0_{k,l,j}$ and has a small signal to noise ratio since in practice, $y^0_{k,l,j}$ is approximately between 0 and 100. Hence it is not worth searching for an estimate of f with m degrees of freedom.

A discretization able to represent possible solutions well, and considerably simplifying the computation is the discretization on the polar grid which has the same geometry as that of the measurements.

Since the $m \times n$ matrix K (with $n = 96 \times 128$) of the system obtained is very sparse (only about 10^7 non-zero entries) we can compute $K^t y$ with a reasonably low cost. The computation of the $n \times n$ matrix $K^t K$ is possible with the rotational invariance of the measurements and the discretisation : indeed with an adequate arrangement of the coefficients x_i, $i = 1, ..., n$, representing f, $K^t K$ is a block-circulant matrix (see Girard [8] for more details) :

$$K^t K = \begin{bmatrix} M_1 & M_2 & \dots & M_{96} & M_{97} & M_{96} & \dots & M_2 \\ M_2 & M_1 & M_2 & & & & & \\ & \ddots & \ddots & \ddots & & & & M_{96} \\ & & \ddots & \ddots & \ddots & & & M_{97} \\ & & & \ddots & \ddots & \ddots & & \\ & & & & \ddots & \ddots & \ddots & M_2 \\ M_2 & M_3 & & & & & M_2 & M_1 \end{bmatrix}$$

where M_k are symmetric 64×64 matrices.

We take as the irregularity of the solution the quadratic form :

$$x^t Fx = x^t Sx + x^t G_1^t SG_1 x + x^t G_2^t SG_2 x$$

where
- *a)* S is the matrix diag $(s_i, i = 1, ..., n) \Big/ \sum_{i=1}^{n} s_i x_i^2$ approximates $\int f^2$
- *b)* G_1 (resp. G_2) is the matrix of discretisation of $\frac{\partial f}{\partial r}$ $\left(\text{resp. } \frac{1}{r}\frac{\partial f}{\partial \theta}\right)$ on the same polar grid, so that :

  $x^t G_1^t SG_1 x + x^t G_2^t SG_2 x$ approximates $\int \left(\frac{\partial f}{\partial r}\right)^2 + \frac{1}{r^2}\left(\frac{\partial f}{\partial \theta}\right)^2$.

Thus F has also the same structure as $K^t K$.

These two matrices can be block-diagonalized by the discrete Fourier transform and we can compute the matrix B of (4.4) with L taken as the symmetric

square root of F. We obtain B in the following form :

$$B = W \begin{bmatrix} B_1 & & & & & & & \\ & B_2 & & & & & (0) & \\ & & \ddots & & & & & \\ & & & B_{96} & & & & \\ & & & & B_{97} & & & \\ & & & & & B_{96} & & \\ & & & & & & \ddots & \\ (0) & & & & & & & B_2 \end{bmatrix}$$

where $\begin{cases} a)\ B_k \text{ are } 64 \times l_k \text{ matrices } (l_k \leqslant 64 \text{ and } l_1 + 2(l_2 + \cdots + l_{96}) + l_{97} = l) \\ b)\ \text{multiplying } z \in \mathbb{R}^n \text{ by } W \text{ involves 64 parallel discrete Fourier transforms of order 192.} \end{cases}$

This reconstruction algorithm has been implemented on a VAX 11/780. The regularization parameter is choosen by the method of generalized cross-validation. The minimization of $\tilde{V}(\rho)$ (3.1, 4.5, 4.6) uses the golden section search method. In practice, about twenty iterations are sufficient.

In the numerical simulation presented, the measurements are generated with an analytic computation of exact data $y^0_{k,i,j}$ (6.1) where f is a given constant on four non intersecting discs (so that the average value of $y^0_{k,i,j}$ is about 4), and are perturbed with Poisson random noise.

We have represented in figure 1 the Euclidean distance $\tilde{S}(\rho)$ between the predicted data and the noisy data, the cross-validation function $\tilde{V}(\rho)$ that we minimize, and the distance $\tilde{R}(\rho)$ between the predicted data and the exact data. Letting $z = B^t K^t y$, $z^0 = B^t K^t y^0$, these functions are computed by :

$$\tilde{R}(\rho) = \frac{1}{l} \| \tilde{A}(\rho) \tilde{y} - \tilde{y}^0 \|_l^2 = \frac{1}{l} \sum_{i=1}^{l} \frac{1}{d_i^2} \left| \frac{d_i^2}{d_i^2 + \rho} z_i - z_i^0 \right|^2$$

$$\tilde{S}(\rho) = \| \tilde{A}(\rho) \tilde{y} - \tilde{y} \|_l^2 = \rho^2 \sum_{i=1}^{l} \frac{1}{d_i^2} \left| \frac{z_i}{d_i^2 + \rho} \right|^2$$

$$\tilde{V}(\rho) = \frac{1}{l} \frac{\tilde{S}(\rho)}{\left(\frac{\rho}{l} \sum_{i=1}^{l} \frac{1}{d_i^2 + \rho} \right)^2}.$$

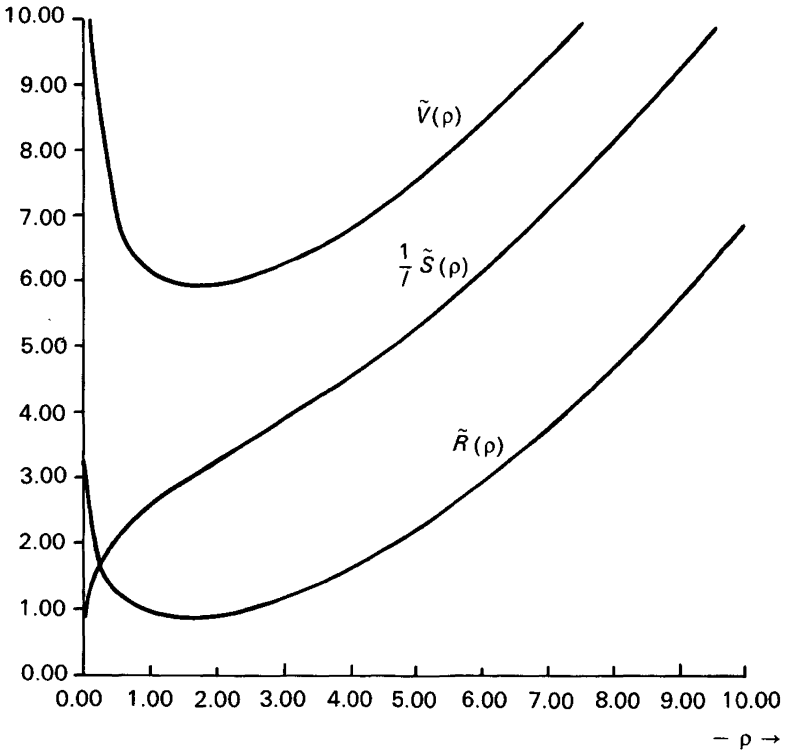

Figure 1.

Figure 1 shows that the cross-validation method produces a very good estimate of the optimal regularization parameter.

We have performed this experiment using several different f of the same type that used in the simulation described above. Each simulation performed demonstrated a similar, high efficiency of the cross-validation method and we verified that the optimal parameter was dependent on the signal-to-noise ratio of the measurements and on the shape of f.

REFERENCES

[1] C. H. REINSCH, *Smoothing by spline functions II.* Numer. Math., 16 (1971), pp. 451-454.

[2] G. WAHBA, *Smoothing noisy data with spline functions.* Numer. Math., 24 (1975), pp. 383-393.

[3] G. WAHBA, *Practical approximate solutions to linear operator equations when the data are noisy.* SIAM, J. Num. Anal. 14 (1977), 651-667.

[4] P. CRAVEN, G. WAHBA, *Smoothing noisy data with spline functions.* Numer. Math. 31 (1979) 377-403.

[5] M. STONE, *Cross-validatory choice and assessment of statistical prediction (with discussion).* J. Roy. Statist. Soc., Ser. B, 36 (1974) pp. 111-147.

[6] G. GOLUB, C. REINSCH, *Singular value decomposition and least square solution.* Numer. Math. 14, 403-420 (1970).

[7] R. ALLEMAND *et al., A new time-of-flight method for positron computed tomography* in D. A. B. Lindberg and P. L. Reichertz, Eds., Biomedical Images and Computers. Berlin : Springer, 1980.

[8] D. GIRARD, *Les méthodes de régularisation optimale et leurs applications en tomographie.* Thesis, University of Grenoble, 1984.