

21

REGULARIZACIÓN

HEMOS ESTUDIADO ALGUNOS MÉTODOS  
PARA INVERTIR LA TRANSFORMADA  
INTEGRAL

$$\int_A^B K(\eta, x) f(x) dx = g(\eta)$$

PERO PARA UN PROBLEMA DEFINIDO  
POR  $K(\eta, x)$ , LA FUNCIÓN OBSERVADA

$$g(\eta) = g^0(\eta) + \delta g(\eta)$$

SE COMPONE DE UNA PARTE  $g^0(\eta)$   
FÍSICAMENTE CORRECTA AFECTADA  
POR ERRORES ALEATORIOS  $\delta g(\eta)$

EN LA RESOLUCIÓN DEL PROBLEMA:  
CÁLCULO DE LA FUNCIÓN  $f(x)$   
SE TENDRÁ QUE LAS RESPECTIVAS  
PARTES INVERSA

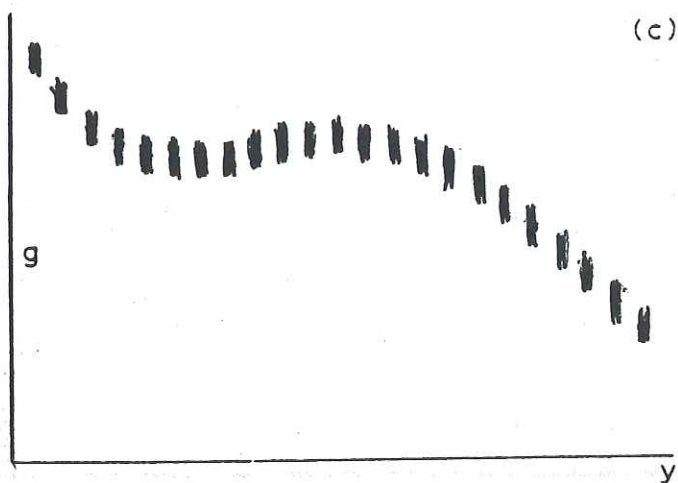
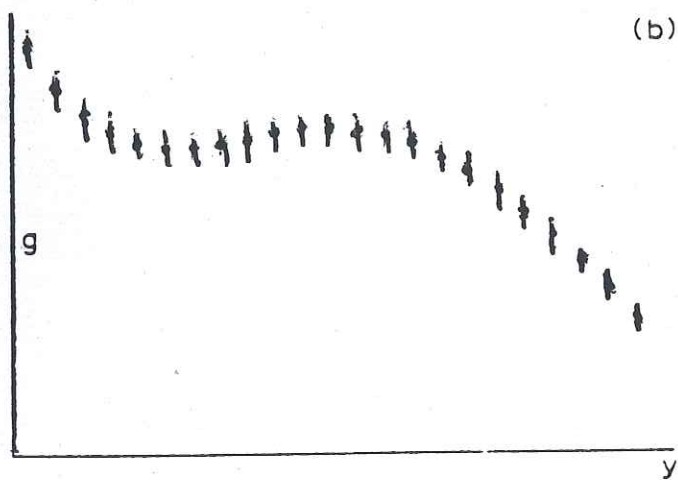
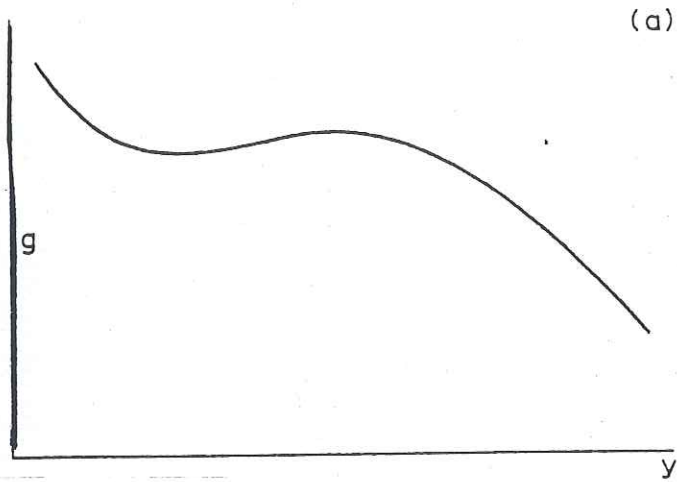
$$f(x) = f^0(x) + \delta f(x)$$

PUEDE OCURRIR / SUELE OCURRIR, QUE

$$\frac{\delta f(x)}{f(x)} \gg \frac{\delta g(\eta)}{g(\eta)}$$

PROBLEMA MAL CONDICIONADO  
SERIOS PROBLEMAS DE ESTABILIDAD

# RECONOCIMIENTO DE IMPRECISIONES EN LAS MEDIDAS



Schematic diagram illustrating the nature of measured "functions".

TWOMEY PROPONE PARA LA  
FUNCIONAL

$$\tilde{P} \stackrel{H}{=} \bar{P}$$

LA SUMA DE LOS CUADRADOS DE LAS  
DIFERENCIAS ENTRE CADA PAR DE  
VALORES CONSECUTIVOS DE LA  
FUNCION  $f(x) \rightarrow \bar{P}$

O BIEN LA SUMA DE LOS CUADRADOS  
DE LAS DIFERENCIAS ENTRE MAS  
DE DOS PUNTOS.

For example, the first differences  $(f_1 - f_2), (f_2 - f_3), \dots, (f_{N-1} - f_N)$  are contained in  $Kf$  when:

$$R = \begin{bmatrix} 0 & 0 & & & & \\ 1 & -1 & 0 & & & \\ 0 & 1 & -1 & & & \\ & & & \ddots & & \\ & & & & 1 & -1 & 0 \\ & & & & 0 & 1 & -1 \end{bmatrix}; \quad R \cdot f = \begin{bmatrix} 0 \\ f_1 - f_2 \\ f_2 - f_3 \\ \vdots \\ f_{N-1} - f_N \end{bmatrix}$$

The quadratic measure  $\sum (f_{j-1} - f_j)^2$  is therefore given by the inner product of  $Rf$  with itself, i.e.:

$$q = (Rf)^T K f = f R^T R f$$

So the corresponding H is:

$$H = R^T \cdot R$$

Hence H is explicitly:

$$H = \begin{bmatrix} 1 & -1 & & & & \\ -1 & 2 & -1 & & & \\ & -1 & 2 & -1 & & \\ & & & \ddots & & \\ & & & & -1 & 2 & -1 \\ & & & & & -1 & 1 \end{bmatrix}$$

If sums of squares of second differences were used to obtain  $q$ , the result would be:

$$R = \begin{bmatrix} 0 & & & & & & \\ -1 & 2 & -1 & & & & \\ & -1 & 2 & -1 & & & \\ & & & & & & \\ & & & & & & \\ & & & & & & \\ & & & & & & \end{bmatrix}; \quad H = \begin{bmatrix} 1 & -2 & 1 & 0 & . & . & . \\ -2 & 5 & -4 & 1 & . & . & . \\ 1 & -4 & 6 & -4 & 1 & . & . \\ 0 & 1 & -4 & 6 & -4 & 1 & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \end{bmatrix}$$

We have, for *third differences*:

$$R = \begin{bmatrix} 0 & & & & & & \\ 0 & 0 & & & & & \\ 0 & 0 & 0 & & & & \\ 1 & -3 & 3 & -1 & & & \\ & 1 & -3 & 3 & -1 & & \\ & & 1 & -3 & 3 & 1 & \\ & & & . & . & . & . \\ & & & . & . & . & . \end{bmatrix}; \quad H = \begin{bmatrix} 1 & -3 & 3 & -1 & & & \\ -3 & 10 & -12 & 6 & -1 & & \\ 3 & -12 & 19 & -15 & 6 & & \\ -1 & 6 & -15 & 20 & -15 & 6 & \\ & -1 & 6 & -15 & 20 & -15 & \\ & & . & . & . & . & \\ & & . & . & . & . & \\ & & . & . & . & . & \end{bmatrix}$$

# ELECCION DEL PARAMETRO DE REGULARIZACION $\lambda$

In all cases the practical problem generally reduces to adopting a regularisation parameter that balances the size of the residual against the smoothness of the solution, as defined by  $\|Rf\|$ .

Of course oscillatory structure may also be present in the real source function for physical reasons but the filtering action of the kernel submerges the signature of such structure in the data noise. In order to be confident that only real structure is recovered therefore, suppression of oscillatory components by smoothing is essential.)

Speaking generally however, there is no universal strategy for selecting the optimum  $\lambda$ , and, within the limit  $0 < \lambda < \lambda_{\max}$ , it is probably best to regard  $\lambda$  as effectively undefined in any specific case

As far as the practical numerical solution is concerned, it should be noted that  $\lambda$  must be chosen large enough to maintain a well conditioned numerical procedure (Culham 1979).

A compromise must be reached but there appears to be no general strategy for choosing the optimum  $\lambda$ .

Yet in practical applications the tendency is to regularise the problem in a rather arbitrary fashion (see references in Culham 1979) without worrying overmuch about problems of convergence. In any specific inversion therefore it is worthwhile, at the very least, performing test-case numerical (or analytic) experiments with a wide variety of trial source functions to check whether, for a predetermined smoothing functional, the regularised solution converges adequately to the true solution as data errors are systematically reduced.

A. N. Tikhonov<sup>[56-58]</sup> has introduced the concept of regularization of the solution of an incorrectly posed problem. This is taken to mean the construction of a family of correctly posed problems depending on a regularization parameter  $\alpha$ , which has the property that for  $\alpha \rightarrow 0$  and when the errors of the right member also simultaneously approach zero the solution of the correctly posed problem approaches the true solution of the incorrectly posed problem. Equation (11) was postulated and studied by Tikhonov, independently of Phillips, as a regularized equation with the regularization parameter  $\alpha$ . In its practical application the Tikhonov method is identical with the Phillips method, perhaps with the difference that the indefiniteness of the parameter  $\alpha$  follows naturally from the very idea of regularization, whereas in Phillips' approach it is a bit of an embarrassment, to a certain extent discrediting the method (on this see below). It must be remarked, by the way, that in a number of papers elaborating Tikhonov's idea, algorithms have been proposed for the determination of  $\alpha$ , but they can scarcely be regarded as having an real foundation, since the source or even the exact form of the a priori information remains an open question.



#### 6.2.4 Statistical regularisation

So far the recovery of the source function has been treated as an entirely deterministic problem, with no account taken of the stochastic nature of the measured data function. It is possible however, to construct regularisation methods using statistical rather than deterministic arguments. In this case the extra information required to stabilise the problem may be incorporated using Bayesian strategies by invoking a prior probability distribution for the unknown function. This is accomplished by introducing a regularising functional in much the same way as in the deterministic problem. An excellent review of the statistical approach, and ill posed problems in general, can be found in Turchin *et al* (1971). We give only a brief description here (see also Strand and Westwater 1968, Franklin 1970).

It is assumed, at the outset, that the errors associated with the measured function ( $\delta g_i$ ) are independent, normally distributed and have zero expectation. For the case in which the prior probability distribution is uniform, corresponding to trivial *a priori* information—that is essentially no correlation between adjacent source function values—the Bayesian strategy reduces to the classical method of least squares.

# OPTIMIZAR LA SEPARACION CUADRATICA MEDIA, ALREDEDOR DE UNA FUNCION $f(x) = S(x)$ CONOCIDA

## Forms of constraint when the statistics of the unknown distribution are known

In some inversion problems, nothing at all is known about the unknown function  $f(x)$ . One can obtain an inversion using some objective constraint such as minimization of variance, but the solution thus obtained has a considerable element of ambiguity about it. One can have some confidence about the general shape of the solution but finer structure is another question. In other situations a considerable amount of background exists — perhaps measurements by direct methods have been made, as is the case with atmospheric temperature profiles, which were measured by balloon-borne and aircraft-borne instruments long before satellite-borne indirect measurements were possible. These measurements showed, for example, that relatively steady gradients were to be expected through certain layers of the atmosphere while sharper excursions and reversals of gradient tended to occur at more or less predictable levels. Furthermore fluctuations from the norm at different levels tended to be correlated with each other, in the sense that some certain layer being warmer than average might imply a high probability of certain other layers being colder than average, and so on. It is possible to construct constraints which allow for such correlations — but it is important to realise that by so doing one is pushing the indirectly sensed solutions towards conformity with the body of past data obtained by more direct methods. Such a forcing is difficult to justify if it is applied, for example, in regions of the globe where no body of past data exist.

A straightforward method of taking into account the tendencies existing in a body of past data is to construct from the past data a suitable set of base functions for approximating the unknown  $f(x)$ . In a crude way this can be done by deriving the mean  $f(x)$  of all past data and finding a constrained solution which minimizes the mean-square departure from this mean.

To do this one merely uses for the quadratic form  $Q$  the square norm of  $(f - b)$ ,  $b$  being the expected (mean) value towards which we wish to bias  $f$ . In the usual way, the solution is obtained by finding the extremum of:

$$(Af - g)^*(Af - g) + \gamma(f - b)^*(f - b)$$

which is given by:

$$e_k^* A^* A f - e_k^* A^* g + \gamma e_k^* f - \gamma e_k^* b = 0 \quad (k = 1, 2, \dots, N)$$

or:

$$f = (A^* A + \gamma I)^{-1} (A^* g + \gamma b)$$

When there is a reasonable basis for choosing a particular  $b$ , this formula gives a useful improvement. But if  $b$  is unrealistic the attempt to force  $f$  to stay close to  $b$  results in spurious oscillations of the solution.

# MAXIMUM ENTROPY REGULARISATION

## 6.3.1 Maximum entropy solution

A method which has been used extensively in astronomical, two-dimensional, data image processing is the maximum entropy technique. Although not as well developed mathematically as the regularisation method, maximum entropy can be understood in much the same general terms. We shall consider only the main features of the method here: further details can be found in for example Gull and Daniel (1978).

In common with regularisation, the maximum entropy approach imposes a structure constraint on the unknown function. Specifically, a solution is sought that maximises a non-linear entropy functional ( $\sum_j f_j \ln f_j$ ) subject to the constraint that it 'fits the data' within the known observational uncertainties. In matrix form the problem to be solved reduces to

$$\alpha \| Kf_\alpha - \hat{g} \|_2^2 + \sum_{j=1}^n f_j \ln f_j = \max \quad (6.30)$$

where  $\alpha$  is an undetermined Lagrange multiplier (cf equation (6.8)). The entropy term clearly has the effect of smoothing the source function since, in the limit of zero data weighting ( $\alpha = 0$ ) the recovered  $f$  is everywhere constant, i.e. it is 'flat' and of undetermined amplitude. This corresponds, in the context of two-dimensional image processing, to an *a priori* image of uniform intensity containing minimum information, that is, 'zero structure'.

The computational method begins by taking a uniform solution, which is normalised to preserve the brightness of the image—as determined by the measured data function—and the user introduces more empirical information into the problem by increasing  $\alpha$  until the recovered source function is sufficiently structured to fit the data: in practice,  $\alpha$  is generally the smallest value consistent with say a  $\chi^2$  fitting test. The computational algorithm is developed by differentiating equation (6.30) with respect to  $f_k$  and expressing the resulting implicit equation as an iterative scheme for the unknown component. Note that the non-linear entropy constraint precludes a closed form matrix solution, in contrast to the regularisation method. After each iteration the maximum entropy solution is renormalised to preserve the measured brightness of the image.

It is clear from the above outline that the maximum entropy technique, when stripped of its image processing guise, has much in common with the method of regularisation. One essential difference is the non-linear smoothness constraint represented by the entropy functional. The view is sometimes advanced that maximum entropy has the advantage of implicitly requiring the positive definiteness of the source function. But this view can be misleading since zeroth-order regularisation also implies a positive definite function in the limit of the large  $\lambda$  (equation (6.15)), at least given a positive semi-definite kernel function.

To investigate the matter further let us determine an approximate matrix solution to equation (6.30), by linearising the maximum entropy constraint. Specifically, if we differentiate equation (6.30) with respect to  $f_k$  and use the approximation that  $\log f_k \approx (f_k - 1)$  valid for near flat pictures ( $f_k \approx 1$ ) we find the matrix solution is given by

$$(\alpha \mathbf{K}^T \mathbf{K} + \mathbf{I}) f_2 = \alpha \mathbf{K}^T \hat{g}. \quad (6.31)$$

This is just the solution for zeroth-order regularisation with  $\alpha^{-1}$  playing the role of the regularisation parameter. Letting  $\alpha \rightarrow 0$  shows that  $f_\alpha \rightarrow 0$  (i.e. flatness when normalised) through a sequence of parallel vectors,  $\alpha \mathbf{K}^T \hat{g}$ , of ever decreasing magnitude. The matrix solution is therefore consistent with the notion that the data-function determines the *form* of the unknown function for  $\alpha > 0$ . Observe also that a positive definite solution leads to a positive definite inversion.

Of course, some choice must be made for the unknown parameter  $\alpha$ , a practical detail of much importance. It is clear that the maximum entropy strategy outlined above determines the least structured solution consistent with the data. In this sense it is analogous to the Phillips criterion (equation (6.28)) for the deterministic regularisation problem. In practice, the recovered function will probably be oversmoothed (§6.2.4); nonetheless the inversion is consistent with the strategy of rejecting those high frequency components that are not unambiguously imposed by the data.

Finally, it is of interest to note the recent claim that, of all the regularisation methods, maximum entropy represents the only 'consistent approach' to the data inversion problem (e.g. Skilling 1984). Certainly, from the perspective of utilising prior information to construct an optimum filter for the problem at hand, the general validity of this claim is far from apparent. This is not to say that certain data sets and certain experimental procedures are not especially susceptible to the maximum entropy technique. However, it is easy to see that maximum entropy is quite incapable of providing the most effective stabilisation for many inversion problems. Consider, for example, the problem of numerical differentiation discussed in §5.3 (example B): this was most effectively stabilised by bounding the second derivative of the unknown function.

# PARA ACABAR REGLA DE ORO

If we do not go into the detailed analysis of the noise problem by finding the eigenvalues and eigenvectors of the matrix  $\tilde{A}A$ , it is still imperative that we should convince ourselves that the physical noise will not drown out our alleged solution. For this purpose we modify the given right side by random quantities of the order of magnitude of the errors of the measurements and observe the influence of this modification on our solution. If the solution changes by too large amounts as the result of this perturbation, we must come to the conclusion that our solution, although mathematically correct, cannot be considered an adequate solution of the given physical problem.

SINTE TIZAR  $g(\eta)$

CON LA  $f(x)$  OBTENIDA

CON ESTOS MÉTODOS