21

REGULARIZACIÓN

CONSTRAINED INVERSION

HEMOS ESTUDIADO ALGUNOS MÉTODOS PARA INVERTIR LA TRANSFORMADA INTEGRAL

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

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

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

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

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

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

PUEDE OCURRIR/SUELE OCURRIR, QUE

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

PROBLEMA MAL CONDICIONADO

SERIOS PROBLEMAS DE ESTABILIDAD

EN LA PRACTICA NUMERICA TENEMOS QUE RESOLVER

$$\underline{\underline{K}}\,\bar{f} = \bar{g}$$

O SEA

$$\underline{\underline{K}}\,\bar{f} - \bar{g} = \bar{0}.$$

DEBIDO A LOS ERRORES EN $\underline{\underline{K}}$ Y A LAS DIFICULTADES INHERENTES AL PROCESO DE INVERSION, TAL IGUALDAD ES DIFICIL DE SATISFACER CON SOLUCIONES $\bar{f}$ FISICAMENTE CORRECTAS

SE TRATA MINIMIZAR LA NORMA DEL RESIDUO $\bar{R}$ DE LA ECUACION APROXIMADA

$$\underline{\underline{K}}\,\bar{f} - \bar{g} = \bar{R}$$

ES DECIR — EN MEDIA CUADRATICA — SE TRATA DE MINIMIZAR $\underline{R}.\bar{R}$.

PERO, EVIDENTEMENTE NO CONOCEMOS $\bar{R}$ Y PERO SABEMOS QUE EN MUCHOS CASOS, LA PROPAGACION DE ERRORES

$$\frac{\delta f(y)}{f(y)} \rightarrow \frac{\delta f(x)}{f(x)}$$

PUEDE SER MONSTRUOSA.

QUIZÁS LA INFLUENCIA DE CIERTAS IMPRECISIONES EN LAS MEDIDAS SE PUEDA EVITAR RESOLVIENDO EL P.I.

$$K(y,x) \otimes f(x) = g(y)$$

VÍA MÍNIMOS CUADRADOS:

$$\| K(y,x) \otimes f(x) - g(y) \|^2 \quad \text{mínimo}$$

EN TÉRMINOS VECTORIALES

$$\| \underline{\underline{K}}\, \bar{f} - \bar{g} \|^2 \quad \text{mínimo}$$

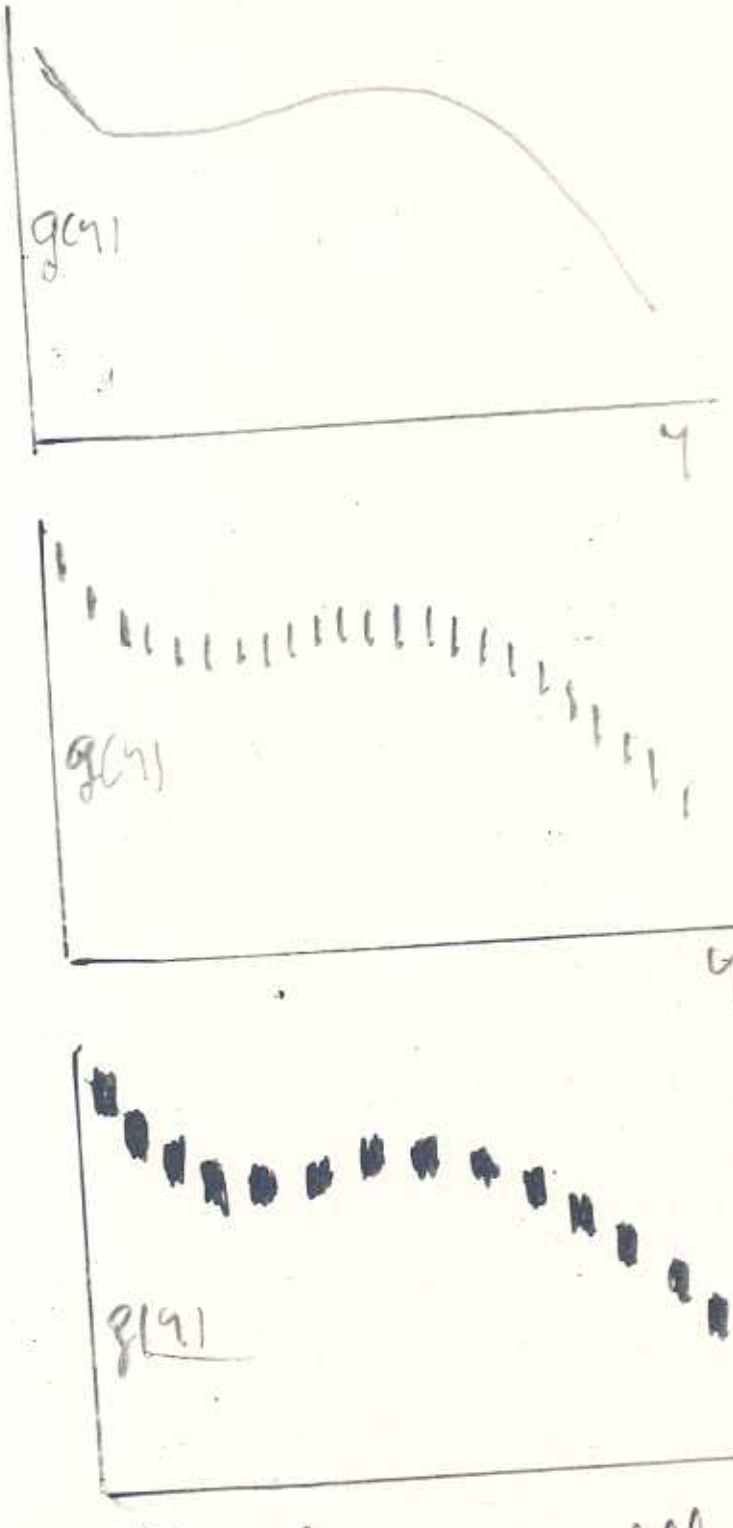
$$\left[\underline{\underline{K}}\, \bar{f} - \bar{g} \right]^T \left[\underline{\underline{K}}\, \bar{f} - \bar{g} \right] \quad \text{mínimo}$$

$$\left[\underline{f}\, \underline{\underline{K}}^T - \underline{g} \right] \left[\underline{\underline{K}}\, \bar{f} - \bar{g} \right] \quad \text{mínimo}$$

CALCULAREMOS LAS COMPONENTES f_J DEL VECTOR $\bar{f}$ DE ACUERDO CON ESA CONDICIÓN EXPLICITANDO

$$\left[(f_1, f_2, \dots f_J, \dots f_N)\, \underline{\underline{K}}^T - \underline{g} \right] \left[\underline{\underline{K}} \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_J \\ \vdots \\ f_N \end{pmatrix} - \bar{g} \right] \quad \text{mínimo}$$

CONSTRAINED LINEAR INVERSION



Schematic diagram illustrating the nature of measured functions.

DERIVANDO CON RESPECTO A CADA UNA DE ELLAS f_J $(J = 1, 2 \cdots N)$, E IGUALANDO A CERO:

$$\left(k_{J1}^T \; k_{J2}^T \cdots k_{JN}^T\right)\left[\underline{\underline{K}}\begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_J \\ \vdots \\ f_N \end{pmatrix} - \bar{q}\right] +$$

$$+\left[(f_1 f_2 \cdots f_J \cdots f_N)\,\underline{\underline{K}}^T - \underline{q}\right]\begin{pmatrix} k_{1J} \\ k_{2J} \\ \vdots \\ k_{nJ} \end{pmatrix} = 0$$

Los dos términos son iguales. Uno es el transpuesto del otro. Luego ambos son nulos

Teniendo en cuenta las derivadas para cada uno de los f_J $(J = 1, 2 \cdots N)$ tendremos

$$\underline{\underline{K}}^T \underline{\underline{K}}\,\bar{f} - \underline{\underline{K}}^T \bar{q} = 0$$

$$\underline{\underline{K}}^T \cdot \underline{\underline{K}}\,\bar{f} = \underline{\underline{K}}^T \bar{q}$$

QUE ES LA REPRESENTACIÓN MATRICIAL DE LA ECUACIÓN INTEGRAL, TRATANDO EL PROBLEMA CON OPTIMIZACIÓN: VÍA MÍNIMOS CUADRADOS DE LOS POSIBLES ERRORES EN LOS DATOS $\bar{q}(q_J)$

PERO AUN ASÍ NO EVITAMOS LA INFLUENCIA DE LAS INESTABILIDADES INTRÍNSECAS DE LA P.I.
EXISTENCIA DE VALORES PROPIOS MUY PEQUEÑOS.

Neither direct nor least squares methods work well in ill-conditioned systems — as we have just seen. Since in either case the root cause is the existence of very small eigenvalues in A or A^*A, the only procedures which can be expected to improve the situation are those which in some way increase the magnitude of the eigenvalues. Since A is computed from the kernels associated with the various measurements we can change the eigenvalues only if we change the question which we are asking of the measured data and since error in *g* has been seen to be the other problem element, it is appropriate that the rephrased question should involve the error.

RECONOCIMIENTO DE LA IMPRECISIÓN EN LAS MEDIDAS

If we make M measurements at $y_1, y_2, \ldots, y_m$, $g(y)$ or g is thereby defined only at those points and to within the measurement error. It is not defined in the sense of Fig. 6.1a. One could in fact define a set of error bars (Fig. 6.1b) or zones (Fig. 6.1c) and say that $g(y)$ is arbitrary except that it passes through each of the bars or zones; the relationship:

$$g(y) = \int_a^b K(y, x)\, f(x)\, dx$$

implies that there exists in the f, x domain a set (probably infinite) of $f(x)$ which are associated (through the integral equation) with $g(y)$'s which pass through the required bars (zones). To within the limits set by our measurements these are *all* "solutions"; the ambiguity can only be removed by imposing an additional condition or criterion (not deriving from the measurements) which enables one of the set of possible $f(x)$ to be selected. We may for example ask for the smoothest $f(x)$ or the $f(x)$ with the smallest maximum deviation from the mean — but it is most important to realise that the additional condition is arbitrary. The measurements in themselves give no basis for suggesting that $f(x)$ is likely to be smooth or anything else. We select *arbitrarily* the smoothest $f(x)$ to represent all of the set of possible $f(x)$.

BUSCAREMOS UNA SOLUCIÓN QUE NO SERÁ EXTRICTAMENTE SOLUCIÓN DE LA ECUACIÓN ANTERIOR, PERO SÍ QUE PUEDE CONSIDERSE COMO TAL: SATISFACE A LA ECUACIÓN ANTERIOR, DENTRO DE UNOS INTERVALOS DE ERROR ACEPTABLES

We begin by discussing the well known method of regularisation. This non-classical technique was first developed by Phillips (1962) and Twomey (1963) and, independently, along more mathematical lines by Tikonov (1963).

In the non-classical approach to the inversion problem this difficulty is recognised at the outset: the information which is fundamentally 'lacking in the data' and which is required to stabilise the problem is introduced in the form of structure constraints on the source function. In this way, through the introduction of *a priori* information, the non-classical approach effectively transforms the original ill posed problem $f = \mathcal{K}^{-1} g$, into an inversion problem that is stable.

ES DECIR, NOS OLVIDAMOS "EN CIERTO MODO" DE LOS DATOS $g(q)$, Y OBLIGAMOS A LA SOLUCIÓN $f(x)$ QUE SE PORTE BIEN

This solution process can be looked at in the following way: a fixed value e^2 *is assigned to* $|\epsilon|^2 = |\mathbf{A}f - g|^2$ *— this is the quantity which measures the "goodness" of* f *as a possible solution of the equation* $\mathbf{A}f = g$. *Of all possible vectors a small subset will have* $|\mathbf{A}f - g|^2 \leqslant e^2$; *provided* e^2 *is of the same order as the uncertainties involved all vectors in the subset are acceptable as possible solutions. We now select from this subset a unique* f *which is the smoothest as judged by the measure* $q(f)$.

The Method of Regularisation

In the method of regularisation the extra information required to stabilise the inversion, or to complete the definition of the problem, is introduced by way of a 'smoothness condition' on the source function. A solution is then obtained by bounding an appropriate *linear* functional of the source function, say $\mathcal{R}f$, subject to the classical constraint $\|\mathcal{K}f - \hat{g}\|$ being minimised. In other words the problem is reduced to solving

$$\|\mathcal{K}f - \hat{g}\|^2 + \lambda \|\mathcal{R}f\|^2 = \min \tag{6.8}$$

where λ, the regularisation parameter, emerges as a Lagrange multiplier in the minimisation procedure.

R ES UN OPERADOR TAL QUE

$R \cdot f(x) \equiv R \otimes f(x)$

DESCRIBE EL COMPORTAMIENTO DE LA FUNCIÓN $f(x)$

LUEGO VEREMOS ALGUNOS

PRIMERA PERVERSIÓN SEMÁNTICA

SE TRATA DE OPTIMIZAR LA CONDICIÓN DE CURVA SUAVE : SMOOTH

CON UNA LIGADURA $\underline{\underline{K}}\, f = \bar{q}$

EL PROBLEMA FUNDAMENTAL HA PASADO A CONDICIÓN DE LIGADURA.

LA FUNCIONAL (OPERADOR) $R \cdot \bar{f}$

SE DESCRIBE, EN TERMINOS PRACTICOS COMO

$$\underline{\underline{R}} \cdot \bar{f}$$

UNA MATRIZ DE REGULARIZACION, QUE ACTUA SOBRE LA SOLUCION BUSCADA $\bar{f}$

(LUEGO VEREMOS ALGUNOS EJEMPLOS)

SE TRATA PUES DE MINIMIZAR

$$\| \underline{\underline{K}} \bar{f} - \bar{g} \|^2 + \lambda \| \underline{\underline{R}} \bar{f} \|^2 \quad \text{MINIMO}$$

λ PARAMETRO DE REGULARIZACION (DE MOMENTO PARAMETRO LIBRE)

PARA $\lambda = 0$ SOLUCION A MINIMOS CUADRADOS

LA FORMA DE $\| \underline{\underline{K}} \bar{f} - \bar{g} \|^2$

LA VIMOS ANTERIORMENTE

$$\| \underline{\underline{R}} \bar{f} \|^2 = \underline{f} \, \underline{\underline{R}}^T \underline{\underline{R}} \, \bar{f} = \underline{f} \, \underline{\underline{H}} \, \bar{f}$$

SIENDO $\underline{\underline{H}} = \underline{\underline{R}}^T \cdot \underline{\underline{R}}$

SE TRATA PUES DE MINIMIZAR

$$[\underline{f}\,\underline{\underline{K}}^T - \underline{q}]\cdot|\underline{\underline{K}}\,\bar{f} - \bar{q}] + \lambda\ \underline{f}\,\underline{\underline{H}}\,\bar{f} \quad \text{MINIMO}$$

DERIVANDO CON RESPECTO A TODAS Y CADA UNA DE LAS COMPONENTES f_j DEL VECTOR $\bar{f}$ E IGUALANDO A CERO

SE OBTIENE EL SISTEMA

$$[\underline{\underline{A}}^T\underline{\underline{A}} + \lambda\,\underline{\underline{H}}]\,\bar{f} = \underline{\underline{A}}^T\bar{g}$$

QUE ES EL SISTEMA QUE DEBEMOS RESOLVER UNA VEZ QUE DISPONEMOS DE LA MATRIZ DE REGULARIZACIÓN $\underline{\underline{R}}$:

$$\underline{\underline{H}} = \underline{\underline{R}}^T\cdot\underline{\underline{R}}$$

This is the equation for constrained linear inversion. The usual procedure for applying this equation is to choose several values for λ and then post-facto decide the most appropriate value for λ by computing the residual $|A\,f - g|$; if this is appreciably larger than the overall error in g due to all causes (experimental error, quadrature error, etc.) then λ is too large — the solution has been constrained too much; if $|A\,f - g|$ is smaller than the estimated error in g one has an underconstrained solution — the error component or part of it has been inverted and spurious oscillations put into the solution.

ENCONTRAMOS UN SISTEMA PRACTICAMENTE IGUAL AL QUE TENÍAMOS QUE RESOLVER CUANDO PLANTEABAMOS DIRECTAMENTE EL PROBLEMA OPTIMIZANDO, A MÍNIMOS CUADRADOS, LOS POSIBLES ERRORES SOBRE $\underline{g}(q)$.

SE ENCUENTRA ESTE CASO CUANDO EL PARAMETRO DE REGULARIZACIÓN λ TOMA EL VALOR $\lambda = 0$

PERO EN ESTE CASO: MÍNIMOS CUADRADOS EL SISTEMA ERA TAN INESTABLE COMO EL SISTEMA CORRESPONDIENTE AL PLANTEO DIRECTO

$$\underline{\underline{K}}\,\bar{f} = \bar{q}$$

LA INESTABILIDAD AQUI ES CONSECUENCIA DE QUE CIERTOS DE LOS AUTOVALORES DE LA MATRIZ $\underline{\underline{K}}$ SON (CASI SIEMPRE) MUY PEQUEÑOS CON RESPECTO A LOS OTROS

EN EL CASO

$$\underline{\underline{K}}^T \cdot \underline{\underline{K}}\,\bar{f} = \underline{\underline{K}}^T \bar{q}$$

ESTAMOS IGUAL O PEOR.

AHORA, CON LA REGULARIZACIÓN, AL AÑADIR EL TERMINO $\lambda \underline{\underline{H}} = \lambda \underline{\underline{R}}^T \underline{\underline{R}}$ AL TERMINO $\underline{\underline{K}}^T \cdot \underline{\underline{K}}$

TENDREMOS SIEMPRE AUTOVALORES MAYORES

A TODOS SE LES AÑADE UNA CANTIDAD PROPORCIONAL A λ

LA ESTABILIDAD SERA MUCHO MAYOR

PERO TENEMOS QUE CONTROLAR EL RESULTADO - DE ACUERDO CON EL VALOR DEL PARÁMETRO λ - PARA QUE CON LA FUNCIÓN $\varphi(x)$ OBTENIDA, NO SEA

$$\underline{\underline{K}} \varphi - \bar{g}$$

MUCHO MAYOR QUE LOS POSIBLES ERRORES QUE PODEMOS ADMITIR PARA $g(y)$

MATRICES DE REGULARIZACIÓN

SE TRATA AHORA DE PRECISAR EN TERMINOS NUMÉRICOS COMO ES ESA MEDIDA QUE DA CUENTA DEL SMOOTHNESS DE LA FUNCIÓN $f(x)$.

COMO DIJIMOS ANTERIORMENTE LO EXPRESAREMOS EN TERMINOS LINEALES COMO UN OPERADOR R MATRICIAL QUE ACTUA SOBRE EL VECTOR $\underline{f}$: FUNCIÓN $f(x)$ EN TERMINOS DE ORDENADAS DISCRETAS

DIFFERENT SMOOTHING MATRICES (MATRICES DE REGULARIZACION) ARE, OF COURSE, ASSOCIATED WITH DIFFERENT FORMS OF THE REGULARIZING FUNETIONAL

POR EJEMPLO PHILLIPS (1962)

PROPONE

$$Q(f) = \int_a^b [f''(x)]^2 dx$$

ADMITIENDO UNA DIVISION REGULAR DEL INTERVALO DE INTEGRACION $(a\ b)$ EN $N-1$ INTERVALOS DE ANCHURA Δ

$$\Delta = \frac{b-a}{N-1}$$

$x_1 = a \qquad x_2 = x_1 + \Delta \; \ldots$

$x_N = b = x_{N-1} + \Delta$

Y ADMITIENDO UNA INTEGRACION TRAPEZOIDAL:

$$Q(f) = \Delta \sum_{j=1}^{N} [f''(x_j)]^2$$

PERO, NUMERICAMENTE

$$f''(x_j) = \frac{1}{2\Delta^2} [f_{j-1} - 2f_j + f_{j+1}]$$

$$f''(x_1) = f''(x_2) \qquad f''(x_N) = f''(x_{N-1})$$

$$f_k \equiv f(x_k)$$

$$q(f) = \frac{1}{4\Delta^2} \left\{ \begin{array}{l} [f_1 - 2f_2 + f_3]^2 \\ + \sum_{j=2}^{N-1} [f_{j-1} - 2f_j + f_{j+1}]^2 \\ + [f_{N-2} - 2f_{N-1} + f_N]^2 \end{array} \right\}$$

QUE PERMITE GENERAR FACILMENTE LA MATRIZ H PARA

$$q(f) \rightarrow \underline{\underline{H}}\, \bar{f}$$

Y RESOLVER

$$\left[\underline{\underline{K}}^T \underline{\underline{K}} + \lambda \underline{\underline{H}} \right] \bar{f} = \underline{\underline{K}}^T \bar{q}$$

TWOMEY PROPONE PARA LA FUNCIONAL (1963)

$$Q(f) \rightarrow \underline{f}^T \underline{\underline{H}}\, \underline{f}$$

LA SUMA DE LOS CUADRADOS DE LAS DIFERENCIAS ENTRE CADA PAR DE VALORES CONSECUTIVOS DE LA FUNCIÓN

$$f(x) \rightarrow \bar{f}$$

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

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

$$R = \begin{bmatrix} 0 & 0 & & & & & \\ 1 & -1 & 0 & & . & . & . \\ 0 & 1 & -1 & & . & . & . \\ & & . & & & & \\ & & & . & & & \\ & & & & 1 & -1 & 0 \\ & & & & 0 & 1 & -1 \end{bmatrix} ; \quad \underline{\underline{R}}\,\bar{f} = \begin{bmatrix} 0 \\ f_1 - f_2 \\ f_2 - f_3 \\ . \\ . \\ f_{N-1} - f_N \end{bmatrix}$$

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

$$q = (\underline{\underline{R}}\,\bar{f})^T \underline{\underline{R}}\,\bar{f} = \underline{f}\,\underline{\underline{R}}^T \underline{\underline{R}}\,\bar{f}$$

So the corresponding H is:

$$H = \underline{\underline{R}}^T \underline{\underline{R}}$$

Hence H is explicitly:

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

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*:

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

Another commonly used expression for q is given by the sums of the squares of the elements of f; use of this constraint selects that f which has the least value of $\Sigma f_i^2 = f^* f$; this quantity corresponds to the power in a signal.

$$q = \Sigma_i f_i^2;$$

H is evidently the identity matrix.

the case of zero-order regularisation,

$$(\mathbf{K}^T\mathbf{K} + \lambda\mathbf{I})f = \mathbf{K}^T\hat{g}$$

where $\mathbf{I}$ is the identity matrix of order n.

A closely related constraint is given by the variance of f, $\Sigma(f_i - \bar{f})^2$, $\bar{f}$ being the average $N^{-1}\ \Sigma_{i=1}^{N}\ f_i$. The appropriate H is readily obtained in each case:

$$q = \Sigma_i\ (f_i - \bar{f})^2;$$

Then the differences from the average are given by:

$$R = \begin{bmatrix} (1-N^{-1}) & -N^{-1} & -N^{-1} & -N^{-1} & -N^{-1} & \dots & \\ -N^{-1} & (1-N^{-1}) & -N^{-1} & -N^{-1} & -N^{-1} & \dots & \\ -N^{-1} & -N^{-1} & (1-N^{-1}) & -N^{-1} & -N^{-1} & \dots & \\ . & . & . & (1-N^{-1}) & & & \\ . & . & . & & & & \\ -N^{-1} & -N^{-1} & . & . & . & \dots & (1-N^{-1}) \end{bmatrix}$$

$$= N^{-1}\|N\delta_{ij} - 1\| \qquad \left(\delta_{ij} = \begin{cases} 0, i \neq j \\ 1, i = j \end{cases}\right)$$

Thus the diagonal elements of H are given by:

$$h_{ii} = (N-1)\,N^{-2} + (1-N^{-1})^2 = N^{-2}[N-1+(N-1)^2] = N^{-2}(N^2-N)$$
$$= (1-N^{-1})$$

and the off-diagonal elements are:

$$h_{ij} = (N-2)\,N^{-2} - 2(1-N^{-1})\,N^{-1} = N^{-2}[N-2-2(N-1)] = -N^{-1}$$

So H is equal to R.. (This matrix therefore has the interesting property that it equals itself squared; it is in fact the sum of the identity matrix and a singular matrix containing $-N^{-1}$ in every position.)

Many other formulae may be arrived at to give other measures of smoothness, but in practice they differ very little in the final result — indeed, if they gave very different final results this would represent an ambiguity which would invalidate the whole method.

More general constraints (e.g. Tikonov 1963) can be constructed by taking a linear superposition of derivatives.

$$q(f) = \alpha_0 f + \alpha_1 f' + \alpha_2 f'' + \ldots .$$

A. N. Tikhonov[56-58] 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 α, which has the property that for $\alpha \to 0$ 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 α. In its practical application the Tikhonov method is identical with the Phillips method, perhaps with the difference that the indefiniteness of the parameter α 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 α, 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 Baysian 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 (δg_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 Baysian strategy reduces to the classical method of least squares.

OPTIMIZAR LA SEPARACIÓN CUADRÁTICA MEDIA, ALREDEDOR DE UNA FUNCIÓN $f_0(x) = 5(x)$ 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 $\overline{f(x)}$ 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:

$$(\mathrm{A}f - g)^*(\mathrm{A}f - g) + \gamma(f - b)^*(f - b)$$

which is given by:

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

or:

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

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

$$\alpha \| \boldsymbol{K}\boldsymbol{f}_\alpha - \hat{\boldsymbol{g}} \|_2^2 + \sum_{j=1}^{n} f_j \ln f_j = \max \tag{6.30}$$

where α 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 α until the recovered source function is sufficiently structured to fit the data: in practice, α is generally the smallest value consistent with say a χ^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 λ (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 $\log f_k \simeq (f_k - 1)$ valid for near flat pictures ($f_k \simeq 1$) we find the matrix solution is given by

$$(\alpha \mathbf{K}^{\mathrm{T}}\mathbf{K} + \mathbf{I})\boldsymbol{f}_2 = \alpha \mathbf{K}^{\mathrm{T}}\hat{\boldsymbol{g}}. \tag{6.31}$$

This is just the solution for zeroth-order regularisation with α^{-1} playing the role of the regularisation parameter. Letting $\alpha \to 0$ shows that $f_\alpha \to \boldsymbol{0}$ (i.e. flatness when normalised) through a sequence of parallel vectors, $\alpha \mathbf{K}^{\mathrm{T}}\hat{\boldsymbol{g}}$, 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 α, 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.

Measures of smoothness

Suppose $q(f)$ is some non-negative scalar measure of the deviations from smoothness in f; if f is varied until $q(f)$ becomes a minimum, the resulting f may be completely smooth in the sense that $q(f)$ will be zero.

Now let q be incorporated with the least squares procedure so that one minimizes not $(Af - g)^2$ as in the least squares method, but $|Af - g|^2 + \gamma q(f)$, where γ is a parameter which can be varied from zero to infinity. Obviously with $\gamma \to \infty$ minimization leads to $q(f) = 0$, i.e. a perfectly smooth f (judged by the measure q)

The early work along these lines, by Phillips (1962), suggested choosing to minimise the norm of the second derivative of the solution. In other words, of all the solutions that can 'fit the data' the one is chosen that minimises

$$q(f) = \| f'' \|_2^2 = \int_a^b [f''(x)]^2 \, dx$$

subject to the classical condition that $\| Kf - \hat{g} \|_2^2 \simeq \| \delta g \|_2^2$. This form of structure condition is quite natural since by considering only 'sufficiently smooth' functions there is a tendency to eliminate oscillatory solutions that arise through data noise.

ELECCION DEL PARAMETRO DE REGULARIZACION λ

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 λ, and, within the limit $0 < \lambda < \lambda_{max}$, it is probably best to regard λ as effectively undefined in any specific case

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

TRIVIAL

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

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.

VER COMENTARIOS FINALES

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.