JOURNAL OF MATERIALS SCIENCE 27 (1992) 3911–3924

Statistical approach to the solution of first-kind integral equations arising in the study of materials and their properties

J. D. WILSON
School of Mathematics, University of Bristol, University Walk, Bristol BS8 1TW, UK

There are a number of problems arising when studying the properties of materials, which require for their solution the inversion of a Fredholm first-kind integral equation. Examples include the determination of the distribution of adsorption energies on the surface of a solid and the evaluation of the distribution of pore radii of a solid from diffusion data. Such equations are, in practice, notoriously difficult to solve. This paper describes a general methodology for solving equations of this type. The method combines the ideas of regularization with a quadratic programming algorithm for minimizing quadratic expressions subject to non-negativity constraints. The condition of non-negativity is essential if we are to recover distribution functions for physical attributes of a solid. The method proposed is tested on simulated data for which the true solution to the equation is already known and on real data arising in each of the two situations described above. The method is shown to perform well in recovering the true solution for the simulated data and to produce results in the real data situations that are consistent with the data observed and with observations of related physical quantities.

1. Introduction

A statistical approach is presented to the solution of Fredholm first kind integral equations of the form

$$g(x) = \int_a^b K(x, y) f(y)\,dy \qquad c \le x \le d \quad (1)$$

The function $K(x, y)$ is known, as are the constants a, b, c and d and it is our aim to determine the unknown function f on the interval $[a, b]$ from a number of observations on the function g. This type of problem occurs quite frequently in all branches of science and our aim is to present an approach that is entirely general and applicable in a large number of situations. Our particular interest, however, is the solution of two specific problems arising in the material sciences. In the first, the aim is to use adsorption data to determine the distribution of adsorption energies on the surface of a solid and in the second, diffusion data are used to evaluate the distribution of the pore radii of a solid. Both problems will be considered in some detail in Sections 3 and 4, but first it is necessary to make some general comments about the nature of such problems and establish a general methodology.

Returning to Equation 1, if the function g were known exactly it might be possible to invert the integral equation analytically and find an exact form for f. In practice, however, g is unknown. We have, instead, a number of observations on g at different values of the variable x, each observation having a random error component. We therefore aim to determine f from a finite number of measurements g_i, $i = 1, \ldots, n$, where

$$g_i = g(x_i) + \varepsilon_i = \int_a^b K(x_i, y) f(y)\,dy + \varepsilon_i \qquad c \le x_i \le d, \quad i = 1, \ldots, n \quad (2)$$

and the ε_i, $i = 1, \ldots, n$, are independent random errors whose means are all zero and whose variances we shall denote σ_i^2, $i = 1, \ldots, n$, respectively. In general, solving the n equations given by Equation 2 to determine the solution f, is not a well-posed problem. There may be no solution f to fit the data or alternatively a solution may exist, but it may only be one of an infinite number of possibilities all equally valid but possibly very different. Moreover the problem is unstable, in the sense that small perturbations in g (due, for example, to random errors) may be magnified into very significant changes in the recovered solution function f.

One way of overcoming these difficulties is to use a method called regularization. Essentially the method of regularization consists of replacing the original problem by a stable minimization problem incorporating some positive parameter, α. The aim is to restrict the field of acceptable functions f, by enforcing some additional property of the solution, that of smoothness. By doing so we are able to find both a unique solution to the problem and one that is stable under small perturbations in our observations. The assumption of smoothness is, for the most part, not an unreasonable one, particularly in situations where the

0022–2461

functions f we seek are density functions. Moreover, we are able to adjust the degree of smoothness we require of f by adjusting the smoothing parameter α.

Using the method of regularization, we define our solution function f to be that which minimizes

$$R = \sum_{i=1}^{n} \left[g_i - \int_a^b K(x_i, y) f(y) \, dy \right]^2 + \alpha \int_a^b (Hf)(y) \, dy \quad (3)$$

The first term is the sum of squared errors and measures the closeness of $\int_a^b K(x_i, y) f(y) \, dy$ to the data g_i, $i = 1, \ldots, n$. The second term, for an appropriate choice of H, usually a function of the second or third derivatives of f, is known as the smoothing term and measures the degree of smoothness of f. The positive parameter α provides a means of trade-off between the degree of smoothness and the degree of consistency with the data. If α is chosen to be small then more emphasis is placed on the function estimate f providing a good fit to the data, than it being smooth. On the other hand, if α is chosen to be large, then more importance will be placed on attaining a smooth solution than one which closely reproduces the data.

The introduction of such smoothness criteria can transform an unstable problem into one that is stable. In doing so, however, we must be aware that we introduce a number of further questions. For example, how should we define smoothness, what function is an appropriate choice of H? Is the assumption of smoothness justified and if so, what size smoothing parameter α should be used to produce a smooth solution that is still consistent with the data? Furthermore, in the two problems that we shall be considering, the functions f that we aim to recover are essentially probability density functions. We must therefore constrain our estimate of the function f to be non-negative. We therefore seek that function f which minimizes Equation 3 subject to the constraint that $f \geqslant 0$.

In Section 2 we present a general methodology for dealing with constrained minimization problems of this kind. In particular we aim to offer an insight into the choice of an appropriate smoothing functional H. Section 2 will also include a discussion of previous research in this area. Regularization has found widespread acceptance in the materials science literature for handling integral equations of this kind. Merz [1], Britten *et al.* [2] and McEnaney *et al.* [3] used regularization as a means of determining adsorption energy distributions from adsorption isotherms, while Brown and Travis [4] and Mays and McEnaney [5] endorsed regularization as a means of estimating pore size distributions from diffusion data. Where the methods of these authors and the method recommended in this paper differs is in the choice of smoothing functional H and smoothing parameter α and in the technique employed to enforce the positivity constraints.

In Section 3 we shall focus on the problem of estimating adsorption energy distributions from adsorption isotherms. Particular attention will be directed towards finding a method of choosing the smoothing parameter α. This section builds very much on the work of McEnaney *et al.* [3] and will be concluded with the results obtained by applying our methodology to the real data set considered in [3] corresponding to the adsorption of argon at 77 K on a polyvinylidene chloride-based carbon at different relative pressures.

In Section 4 we shall consider the second of our two problems, that of determining pore size distributions from diffusion data. Owing to the nature of the kernel function, this problem is particularly difficult to solve and as such we need to recognize that any solution we find may owe more to our subjective choice of smoothing criteria than to the data. The section is concluded by considering a second real data set, namely that given in [5] for the diffusion rate of helium through graphite at a number of different pressures.

2. A general methodology

A general methodology is presented here for finding the required solution function f given observations on the function g. The method is compared with those adopted by previous researchers in the field, and the differences and similarities between the techniques discussed.

We recall that our aim is to find that f which satisfies the n equations in Equation 2 given the observations g_i, $i = 1, \ldots, n$. (Throughout the paper the convention will be adopted that vectors and matrices will be printed in bold.) We begin by discretizing the integral operator so that the n equations in Equation 2 can be expressed

$$g_i = \sum_{j=1}^{m} K(x_i, y_j) f(y_j) \delta_j + \varepsilon_i \quad i = 1, \ldots, n \quad (4)$$

where the δ_j are the weights of the quadrature scheme used to approximate the integral and the y_j are equally spaced on the interval (a, b) with $y_0 = a$ and $y_{m+1} = b$. Writing the equations in matrix form

$$\mathbf{g} = \mathbf{K}\mathbf{f} + \boldsymbol{\varepsilon} \quad (5)$$

where the quantity $\mathbf{K}$ is an $n \times m$ matrix operator whose ijth element is given by $K(x_i, y_j)\delta_j$ and $\mathbf{f}$ is the $m \times 1$ vector whose jth element corresponds to $f(y_j)$. The $n \times 1$ vector $\mathbf{g}$ is the known vector of observations and the $n \times 1$ vector $\boldsymbol{\varepsilon}$ is the vector of errors, here assumed to be a multivariate normal random variable with zero mean and variance given by $\text{var}(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{W}^{-1}$, where σ^2 is unknown and $\mathbf{W}$ is a known diagonal matrix (often the identity matrix $\mathbf{I}$).

Using regularization we choose our solution to be that $\hat{\mathbf{f}} \geqslant 0$ which minimizes the functional

$$\mathbf{R} = \mathbf{E} + \alpha \mathbf{S} \quad (6)$$

where $\mathbf{E}$ is defined to be the weighted sum of squared errors

$$\mathbf{E} = (\mathbf{g} - \mathbf{K}\mathbf{f})^{\mathrm{T}} \mathbf{W} (\mathbf{g} - \mathbf{K}\mathbf{f}) \quad (7)$$

Weighting the sum of squared errors by the matrix $\mathbf{W}$ ensures that we do not allow the error term to be swamped by the presence of data points with unnaturally large variability.

The second term, $\alpha\mathbf{S}$, in Equation 6 is the smoothing term consisting of a discrete matrix approximation to our chosen smoothing function. The choice of smoothing function is a subjective one which will depend not on the data but on our subjective assessment of the shape of the function f we are aiming to recover. What then are the properties we would like our solution to have and what smoothing function should we employ to induce these properties? A reasonable assumption to make is that our recovered solution should not exhibit erratic fluctuations or "wiggliness". A sensible smoothing function to choose in this case is $\int (f'')^2$, the integral of the squared second derivatives. This function measures the change in gradient of the function f. Clearly the more "wiggly" a function, the more its gradient will be changing and the larger will be the integral of squared second derivatives. Thus taking $\int (f'')^2$ as the smoothing function has the effect of penalising those functions whose gradients are constantly changing – the wiggly functions.

We approximate this smoothing function in the natural way by taking finite second differences. Denoting the jth value of vector $\mathbf{f}$, namely $f(y_j)$, by $\mathbf{f}_j$, we approximate $\int (f'')^2 \mathrm{d}y$ by

$$\frac{1}{(\delta y)^4}\left\{\sum_{j=0}^{m+1}(\mathbf{f}_{j+1} - 2\mathbf{f}_j + \mathbf{f}_{j-1})^2 \delta y\right\} \qquad (8)$$

where δy is the width of the interval $[y_{j-1}, y_j]$, assumed equal for all j, and we assume $\mathbf{f}_{-1} = \mathbf{f}_0 = 0$ and $\mathbf{f}_{m+1} = \mathbf{f}_{m+2} = 0$. This last assumption has the effect of smoothing the function f to zero at the ends of the interval (a, b). This may not always be appropriate, for example, if we anticipate that our function f is likely to exhibit asymptotic behaviour at one of the endpoints. However, we have adopted this approximation of the smoothing function to be consistent with the method of Reference 3, which constrained the function f to be tied to zero at the ends of the interval on which it was defined. This will enable us more easily to make comparisons between the results obtained by our method and those obtained in Reference 3 on the same set of data.

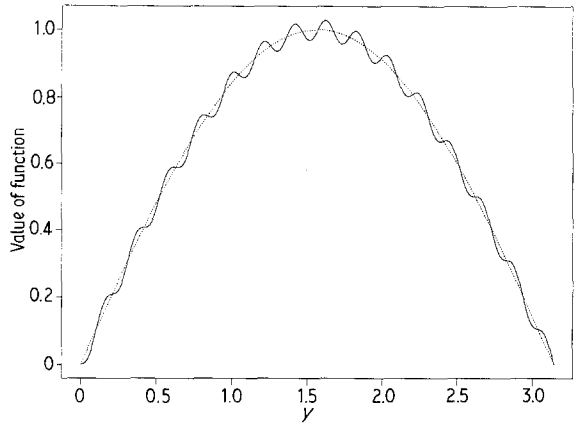

Figure 1 Graph of a smooth function $f(\ldots)$ plotted against a "wiggly" function h (——) with the same integrated square value.

Using matrix notation, Equation 8 can be written as $\mathbf{f}^T\mathbf{Q}\mathbf{f}$ where $\mathbf{Q}$ is the symmetric $m \times m$ matrix obtained by expanding Equation 8 and defining $\mathbf{Q}_{ij}$ to be the coefficient of $\mathbf{f}_i\mathbf{f}_j$, $i = 1, \ldots, m, j = 1, \ldots, m$, namely

$$\mathbf{Q} = \frac{1}{(\delta y)^3}\begin{bmatrix} 6 & -4 & 1 & 0 & \cdots & 0 & 0 & 0 \\ -4 & 6 & -4 & 1 & \cdots & 0 & 0 & 0 \\ 1 & -4 & 6 & -4 & \cdots & 0 & 0 & 0 \\ 0 & 1 & -4 & 6 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 6 & -4 & 1 \\ 0 & 0 & 0 & 0 & \cdots & -4 & 6 & -4 \\ 0 & 0 & 0 & 0 & \cdots & 1 & -4 & 6 \end{bmatrix} \qquad (9)$$

An alternative smoothing function and the one used by Britten *et al.* [2] and McEnaney *et al.* [3] is $\int f^2$. The matrix approximation to this is $\mathbf{Q} = \mathbf{I}$, the identity matrix. Careful consideration of this choice however, reveals that $\int f^2$ does not penalize "wiggliness" in the final solution. For example, consider the functions $f(y) = (1 + q^2)^{1/2} \sin y$ and $h(y) = \sin y + q \cos(2px + \pi/2)$, as depicted by the dotted and solid curves in Fig. 1, respectively (for constants $q = 0.03$ and $p = 15.0$). The function h is wiggly and f is smooth and yet it is easy to show that $\int_0^\pi f^2 = \int_0^\pi h^2$. The effect of choosing such a smoothing function is not to penalize wiggly functions but to penalize large values of the function f and particularly, as α increases, to flatten f down to the zero axis.

With this in mind, we choose $\mathbf{Q}$ to be the approximation to the integral of the squared second derivatives, as defined above. Using Equations 6 and 7

$$\mathbf{R} = (\mathbf{g} - \mathbf{K}\mathbf{f})^T\mathbf{W}(\mathbf{g} - \mathbf{K}\mathbf{f}) + \alpha\mathbf{f}^T\mathbf{Q}\mathbf{f} \qquad (10)$$

Expanding the brackets and neglecting terms constant in $\mathbf{f}$ the problem becomes one of minimizing

$$\mathbf{R}^* = -2\mathbf{g}^T\mathbf{W}\mathbf{K}\mathbf{f} + \mathbf{f}^T(\mathbf{K}^T\mathbf{W}\mathbf{K} + \alpha\mathbf{Q})\mathbf{f} \qquad (11)$$

subject to the non-negativity constraint

$$\mathbf{f}_j \geq 0 \quad j = 1, \ldots, m \qquad (12)$$

This is a constrained minimisation of a quadratic function in $\mathbf{f}$, with the added difficulty that we must choose a value for our smoothing parameter α. The choice of smoothing parameter will be considered in the next section. Here we observe that the problem defined in Equation 11 may be ideally solved using the quadratic programming algorithm, an optimization algorithm designed to minimize quadratic functions subject to linear and positivity constraints. The algorithm used is due to Wolfe [6] and makes use of the Simplex method widely used in linear programming problems. It has been implemented in Fortran 77 on a Sun Workstation. The quadratic programming algorithm will yield a unique solution for the global

minimum whenever $\mathbf{K}^T\mathbf{W}\mathbf{K} + \alpha\mathbf{Q}$ is positive definite. (A matrix $\mathbf{C}$ is said to be positive definite if $\mathbf{x}^T\mathbf{C}\mathbf{x} > 0$ for all non-zero vectors $\mathbf{x}$.) If $\mathbf{K}^T\mathbf{W}\mathbf{K} + \alpha\mathbf{Q}$ is only positive semidefinite then the solution need not be unique.

Before moving on to consider the choice of smoothing parameter, α, it is worthwhile to look at the way in which other authors have solved the minimization problem given in Equation 11. Merz [1] tackled the problem by ignoring the non-negativity constraints. The smoothing parameter α was selected using a procedure called generalized cross-validation as outlined by Wahba [7]. It is a well-documented procedure for choosing the smoothing parameter, when there are no constraints on $\mathbf{f}$. We shall consider its merits in the constrained case in Section 3. Having selected the parameter, Merz [1] chooses that $\mathbf{f}$ which minimises Equation 11 for this value of α. Ignoring the constraints, the required $\mathbf{f}$ is simply that which satisfies the linear simultaneous equations

$$(\mathbf{K}^T\mathbf{W}\mathbf{K} + \alpha\mathbf{Q})\mathbf{f} = \mathbf{K}^T\mathbf{W}\mathbf{g} \qquad (13)$$

A simulation study is carried out and gives sample solutions $\mathbf{f}$ for three different α values, one of which corresponds to the generalized cross-validation estimate of α. The solution obtained for this α, while being a closer approximation to the true function f than those obtained using other choices of α, nevertheless is problematical because it takes negative values at some points in its domain and it is not clear how to interpret such values. The author makes it clear that a method incorporating the constraints would be an improvement, were such a method to exist. The ideas of generalized cross-validation are useful ones and will be considered later.

Britten *et al.* [2] present an approach to the minimisation of Equation 10 which incorporates the non-negativity constraints and a method for choosing the smoothing parameter. The method chosen to determine $\mathbf{f}$ is an iterative procedure proposed by Butler *et al.* [8]. Using an optimization technique called gradient projection, it has the disadvantage that convergence to a solution may be slow and hampered by round-off errors. The choice of smoothing parameter is a technique due to Reference 8 which is based on the level of error in the data. In simulations it was found that the estimate of α thus obtained brought about systematic oversmoothing in all cases. The method also requires that the smoothing functional is of the form $\int f^2$, which as we saw earlier may not be the best function to choose for our purposes.

The final paper we shall consider is McEnaney *et al.* [3] which addressed the adsorption energy problem. Here the smoothing parameter is chosen to be the smallest value of α for which the non-negativity constraints are satisfied and for which the ends of the solution are tied to zero, that is

$$\mathbf{f}_j \geq 0 \qquad j = 2, \ldots, m-1 \qquad (14a)$$

$$\mathbf{f}_1 = \mathbf{f}_m = 0 \qquad (14b)$$

The smoothing procedure is defined as follows. The upper value of the domain of the solution function f, here the value of y_m at which f_m is to be evaluated, is fixed at an initial high value. The constraints on the solution vector $\mathbf{f}$, as given in Equations 14a and b, are relaxed and $\mathbf{f}$ found by solving the linear simultaneous equations in Equation 13 for $\mathbf{f}_2, \mathbf{f}_3, \ldots, \mathbf{f}_m$. With the value of $\mathbf{f}_1$ (the function f evaluated at $y = 0$) fixed at 0, it is found that for a given smoothing parameter α there is a maximum value of y_m for which the calculated solution $\mathbf{f}$ satisfies the constraints in Equations 14a and b. It is also found that there is a minimum α for which a sufficiently large y_m can be determined. The optimal smoothing parameter α is therefore said to be that which supplies the least smoothed solution with the widest range. The procedure is somewhat *ad hoc* and may involve a considerable amount of computation as many values of α may need to be tried before a suitable $\mathbf{f}$ is found. It is also questionable whether this method will be able to produce a function which is not unimodal, particularly a bimodal distribution with a small weight at the far end of the range, which may well be lost through this procedure. We shall return to this paper in the next section when we shall compare our results with those of McEnaney *et al.* [3] for their adsorption data.

3. The adsorption problem

3.1. Description of the problem

The aim is to determine the heterogeneity of a solid by a consideration of adsorption energies. Background information to this problem and a detailed review of recent progress in this area may be found in Jaroniec and Brauer [9]. The development that follows is based on that appearing in Reference 3. We begin by developing a mathematical model for the physical situation. It will be assumed, in common with much of the research into adsorption on heterogeneous solids, that the surface of the solid consists of areas of uniform adsorption energy y, each of which fills according to a local isotherm $K(x, y)$. The isotherm $K(x, y)$ is a function relating the fraction of sites covered at equilibrium, K, to pressure x and adsorption energy y. Assuming the size of the patches to be small relative to the total area of adsorption, it can be assumed that the density of adsorption energies $h(y)$, $a < y < b$, is continuous and the total adsorption isotherm $T(x)$ at pressure x is given by

$$T(x) = \int_a^b K(x, y)h(y)\,dy \qquad (15)$$

a Fredholm first-kind integral equation. This expression is referred to in the literature as the generalized adsorption isotherm. The function h is a density and therefore satisfies

$$h(y) \geq 0 \qquad a < y < b \qquad (16a)$$

and

$$\int_a^b h(y)\,dy = 1 \qquad (16b)$$

There are a number of possible functions currently in use for the local adsorption isotherm $K(x, y)$. A full

discussion of the advantages and limitations of the different functions may be found in Reference 9. The question of which function is most appropriate will not be addressed here. It suffices to say that the methodology used in this paper may be applied to any chosen kernel function $K(x, y)$. Here $K(x, y)$ will be taken to be the Langmuir isotherm, namely

$$K(x, y) = \frac{bx}{1 + bx} \tag{17}$$

where the factor b is given by

$$b = b_0 \exp(y/B) \tag{18}$$

The values of the constants b_0 and B are taken to be those given in Reference 3 for adsorption of argon on microporous carbons, namely $b_0 = 0.000\,006\,55$ and $B = 640.178$.

In practice, observations are not made on the total isotherm T but on the specific amount adsorbed at pressure x. Denoting this function by g, g is related to T by the expression

$$g = cT \tag{19}$$

where c is an unknown constant corresponding to the specific amount adsorbed in the total adsorption space. Writing $f(y) = ch(y)$, Equation 15 becomes

$$g(x) = \int_a^b K(x, y) f(y) \mathrm{d}y \tag{20}$$

Given observations on g subject to measurement or experimental error, the aim is to determine the non-negative function f and hence to evaluate the density h and the constant c using

$$\int_a^b f(y) \mathrm{d}y = c \int_a^b h(y) \mathrm{d}y = c \tag{21}$$

and

$$h(y) = \frac{f(y)}{\int f(y') \mathrm{d}y'} \tag{22}$$

The function f is determined using the general methodology of Section 2, that is by discretizing Equation 20, choosing an appropriate smoothing matrix $\mathbf{Q}$ (here we use that given in Section 2) and using quadratic programming to minimize the expression in Equation 11 subject to the non-negativity constraints. The only problem is that of choosing a suitable value for the smoothing parameter α in Equation 11. This problem is addressed in the next section.

3.2. Simulation study to investigate the choice of smoothing parameter

3.2.1. Method of simulation

The aim is to choose a level of smoothing which will produce a smooth solution while remaining true to the data. In order to investigate this problem a simulation study was conducted. The true distribution of adsorption energies f^* say, was chosen to be proportional to a gamma density with parameters 35 and 0.005. The value of the lower energy level $y_{\min}$ was taken to be three standard deviations below the mean and the upper energy level $y_{\max}$ was taken to be four standard deviations above the mean. This particular f^* was chosen for its similarity to the recovered site energy distributions given in Reference 3. The area under the curve f^*, corresponding to the specific amount adsorbed in the total adsorption space was approximately 293 units.

The values of the $x_i, i = 1, \ldots, n$ were taken to be the 25 relative pressures used in Reference 3, where relative pressure is defined to be the actual pressure divided by the saturated vapour pressure of supercooled liquid argon at 77 K. The number of points in the quadrature m was chosen to be 49, large enough for the vector $\mathbf{f}$ to be a reasonable approximation to the function f while not so large that the problem becomes unwieldy given the limited amount of data. We notice that $m > n$ and hence that the matrix $\mathbf{K}^T\mathbf{K}$ is only semi-definite with, in this case, at most 25 non-zero eigenvalues. The value of the local isotherm $K(x_i, y_i)$ at relative pressure $x_i, i = 1, \ldots, 25$ and site energy $y_j, j = 1, \ldots, m$ is given by

$$K(x_i, y_j) = \frac{x_i}{A \exp(-y_j/B) + x_i} \tag{23}$$

where $A = 5280.28$ and $B = 640.178$ using the values of the constants given in Reference 3. The data $\mathbf{g}_i$, $i = 1, \ldots, n$ was generated using the expression

$$\mathbf{g}_i = \sum_{j=1}^{m} K(x_i, y_j) f^*(y_j) \delta_j + \varepsilon_i \tag{24}$$

where $K(x_i, y_j)$ is as defined above and the y_j are equally spaced on the interval $(y_{\min}, y_{\max})$ with $y_{\min}$ taken to be y_0 and $y_{\max}$ taken to be y_{m+1}. Each δ_j corresponds to the length of the interval $[y_{j-1}, y_j]$ assumed equal for all j, namely $(y_{\max} - y_{\min})/(m + 1)$, and $f^*(y_j)$ is the value of our chosen energy distribution f^* evaluated at energy y_j. The errors ε_i are independent normal random variables with mean zero and common variance of 25.0 for all $i = 1, \ldots, n$ (i.e. $\sigma^2 = 25.0$ and $\mathbf{W}$ is the identity matrix). They are generated using marsaglia's polar method [10] and a congruential pseudo-random number generator [11].

Twenty different data sets were simulated using this model. One such data set is displayed in Fig. 2. For each of the twenty data sets generated we solved the constrained minimization problem in Equation 11 to find the vector $\hat{\mathbf{f}}$, our discrete approximation to the unknown energy distribution f, using the methodology given in Section 2 and a range of different values of the smoothing parameter α.

3.2.2. Results: some initial observations

For each of the twenty data sets a similar pattern of behaviour emerged as the smoothing parameter α varied. A single data set, that given by the plotted points in Fig. 2, will be used here as an example. The numerical results for a number of different values of the smoothing parameter α are given in Table I. The second column refers to the number of pivoting operations required in the quadratic programming procedure before a solution was reached. The third column gives an estimate of the specific amount adsorbed in

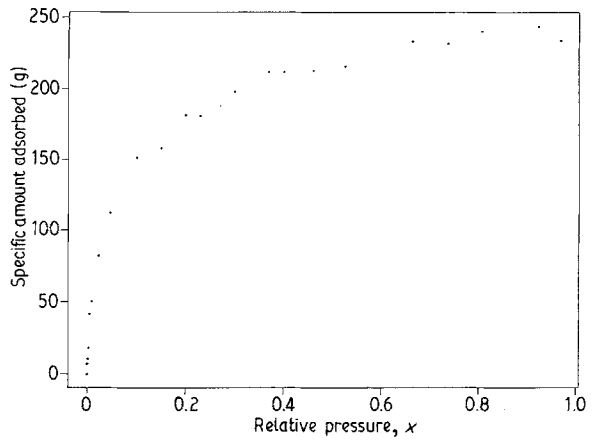

Figure 2 Plot of simulated data values corresponding to the specific amount adsorbed at 25 different relative pressures. (●) Generated data.

TABLE I Results for simulated data set 1

α	Number of pivots	Specific amount absorbed (4 sf)	Value of R^* (6 sf)
0.001	75	277.3	−25 562.6
0.1	75	277.3	−25 562.6
10^3	73	277.2	−25 562.6
10^5	76	277.3	−25 562.6
10^8	87	280.2	−25 562.4
10^9	87	284.5	−25 562.2
10^{10}	87	291.6	−25 562.0
10^{11}	77	302.4	−25 561.2
10^{12}	71	311.4	−25 556.8
10^{13}	81	310.5	−25 522.7
10^{14}	84	287.9	−25 292.8
10^{15}	87	252.3	−24 156.2

sf: significant figures

the total adsorption space, the true value of which we know in this case to be 293 units. The fourth column is the value of the function **R*** (as given in Equation 11) that we are minimizing. The recovered solution vectors $\hat{\mathbf{f}}$ for four different values of α are plotted in Fig. 3. The solid line in each case depicts the original f^* plotted on 49 points and the dotted line is a line drawn through the 49 points of the vector estimate $\hat{\mathbf{f}}$ recovered by quadratic programming.

If we consider the values of **R*** in Table I we notice that **R*** remains constant at a value of −25 562.6 for all values of α less than 10^8. Between $\alpha = 10^8$ and 10^{13} the value of **R*** changes relatively slowly. Thereafter **R*** begins to change quite dramatically as α increases. This can be seen most clearly by means of a graph, in which we plot **R*** against $\log_{10}\alpha$. See Fig. 4.

If we consider now the solution functions $\hat{\mathbf{f}}$ obtained, we see that the $\hat{\mathbf{f}}$s corresponding to the αs in the region of constant **R*** are wiggly and undersmoothed (see in particular Fig. 3a corresponding to an α of 10^5). Those $\hat{\mathbf{f}}$s corresponding to the region of steeply changing **R***, for example the $\hat{\mathbf{f}}$ plotted in Fig. 3d corresponding to an α of 10^{15}, are considerably oversmoothed. It is in the region of slowly changing **R*** that it appears the most sensible values of α are obtained. Fig. 5 depicts the recovered solutions $\hat{\mathbf{f}}$ for

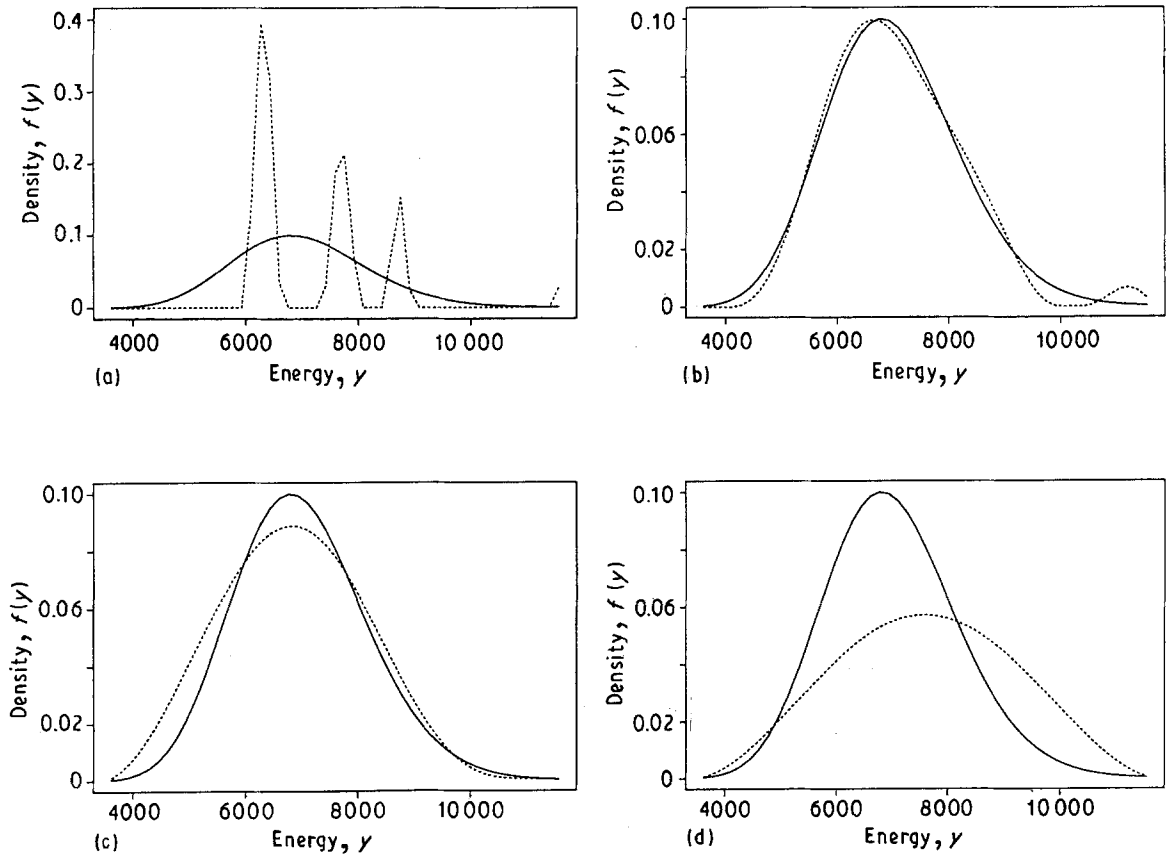

Figure 3 Graphs showing the recovered solutions for four different values of the smoothing parameter α (. . .) when plotted against the original function (——). α: (a) 10^5, (b) 10^{10}, (c) 10^{12}, (d) 10^{15}.

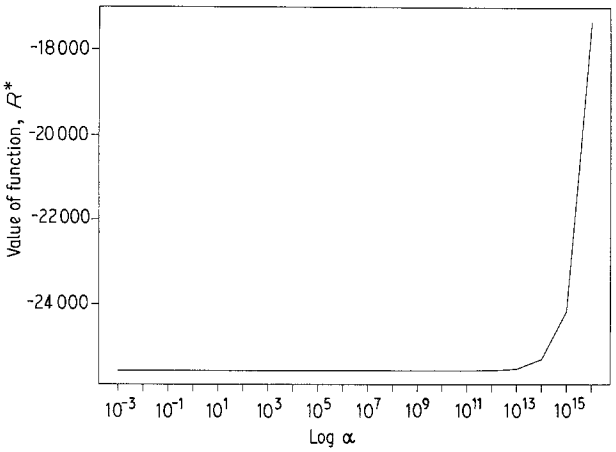

Figure 4 Graph of the value of the function R^* plotted against $\log_e\alpha$.

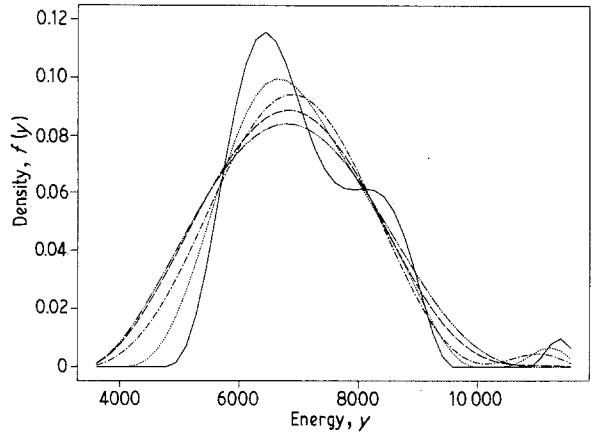

Figure 5 Graph of the recovered solutions $\hat{\mathbf{f}}$ for five different values of the smoothing parameter, α: (——) 10^9, (. . .) 10^{10}, (— - —) 10^{11}, (— —) 10^{12}, (— - - - —) 10^{13}.

successive values of α in this region. Returning to Fig. 3b and c, depicting the solutions $\hat{\mathbf{f}}$ for $\alpha = 10^{10}$ and 10^{12}, respectively, it is clear that both the solutions corresponding to α in the middle region yield good reconstructions of the original gamma function. An α of 10^9, on the other hand, would give an $\hat{\mathbf{f}}$ that is slightly undersmoothed and an α of 10^{14} would give a solution that is tending to be oversmoothed.

In practice, of course, the true function f^* is unknown. It might therefore be more useful to consider the values of $\hat{\mathbf{g}}$ that we might predict using our recovered solution $\hat{\mathbf{f}}$ and see if these are consistent with the actual data $\mathbf{g}$ observed. We construct our predicted data using the expression

$$\hat{\mathbf{g}}_i = \sum_{j=1}^{m} K(x_i, y_j)\hat{\mathbf{f}}_j\delta_j \qquad (25)$$

The graphs in Fig. 6 each depict a dotted line drawn through the predicted data $\hat{\mathbf{g}}$ plotted on 25 points and the original 25 observations. The predicted data $\hat{\mathbf{g}}$ depicted in Fig. 6a corresponds to the solution $\hat{\mathbf{f}}$ depicted in Fig. 3a, the predicted data in Fig. 6b to the solution in Fig. 3b and so on. We can quantify the difference between the predicted $\hat{\mathbf{g}}$ and the original data $\mathbf{g}$ using the following measure, which we shall call summed predicted error, namely

$$\text{SPE} = \sum_{i=1}^{n} (\mathbf{g}_i - \hat{\mathbf{g}}_i)^2 \qquad (26)$$

For this example, the summed predicted error values for a number of different αs are given in Table II. As we would expect, the summed predicted error increases with α, because as α increases we are giving more importance to smoothness and less importance to the data. Moreover the increase in the SPE is relatively slow until we reach smoothing parameters of about 10^{13} and 10^{14}. At this point the errors increase much more markedly. This corresponds to the level of α at which R^* begins to increase significantly and is indicative of oversmoothing. This effect is perhaps most visible when we consider the graphs of the predicted data in Fig. 6. From Fig. 6a, b and c it is clear that the three functions $\hat{\mathbf{f}}$ in Fig. 3a, b and c, obtained by quadratic programming from three very different values of α, all reproduce the original data very closely indeed. This is a very graphic illustration of the instability in the problem – that very different site energy distributions $\hat{\mathbf{f}}$ can yield almost identical sets of adsorption data. In the final figure, Fig. 6d, we see the effect on the predicted data of oversmoothing the site energy distribution. Both the value of the summed predicted error as given in Table II and the curve itself indicate that by increasing α too much we are producing a solution that is not consistent with the data and hence not to be recommended.

It is clear from our investigations that the use of such simple methods as the consideration of the shape of the solution curves, and quantities such as the value R^* and the summed predicted error, are able to give us an indication of the level of smoothing parameter that might be appropriate in a given situation, to yield solutions that are both smooth and consistent with the data. In the section that follows we shall consider how we might adapt the ideas of cross-validation to provide a further indication of the level of smoothing. However, before proceeding it might be useful to see how we might apply the criteria outlined above in choosing a solution for a second simulated data set.

One of the remaining nineteen data sets generated in Section 3.2.1 was selected arbitrarily and solution vectors $\hat{\mathbf{f}}$ obtained by quadratic programming for a number of different α values. The numerical results are

TABLE II Summed predicted error values for simulated data set 1

α	Summed predicted error (6 sf)
10^5	374.984
10^8	378.089
10^9	380.936
10^{10}	386.103
10^{11}	392.311
10^{12}	413.702
10^{13}	484.409
10^{14}	2143.45
10^{15}	8582.48

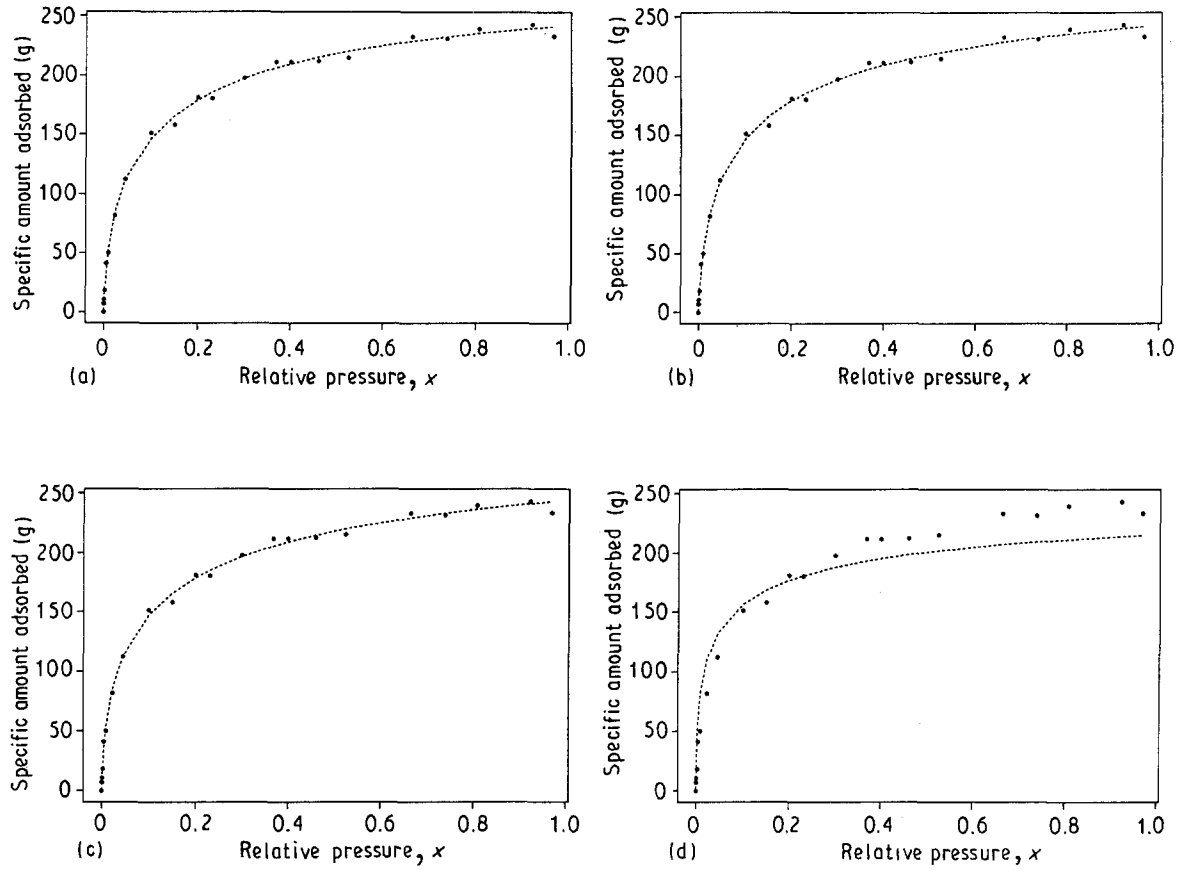

Figure 6 Graphs of the predicted data for four different values of the smoothing parameter α (. . .) plotted against the original observations, α: a) 10^5, b) 10^{10}, c) 10^{12}, d) 10^{15}.

TABLE III Results for simulated data set 2

α	Number of pivots	Specific amount adsorbed (4 sf)	Value of R^* (6 sf)	Summed predicted error (6 sf)
10^8	107	455.2	−25 125.9	472.212
10^9	87	384.8	−25 125.3	481.943
10^{10}	77	327.8	−25 124.8	499.803
10^{11}	75	316.8	−25 124.2	506.092
10^{12}	73	317.3	−25 120.4	510.624
10^{13}	79	311.2	−25 085.6	577.252
10^{14}	84	286.4	−24 853.2	1715.08

contained in Table III. A similar range of α values appear to be appropriate here as in the previous example. An α of 10^{13} gives unacceptably high predicted errors suggesting oversmoothing, while αs of 10^9 and 10^{10} yield curves which are still quite wiggly. An α of 10^{11} or 10^{12} seems to yield a smooth solution consistent with the data. If we compare the solution obtained for an α of 10^{12} with the true gamma function f^* (Fig. 7) we can see that our criteria has produced a very good approximation to the original function.

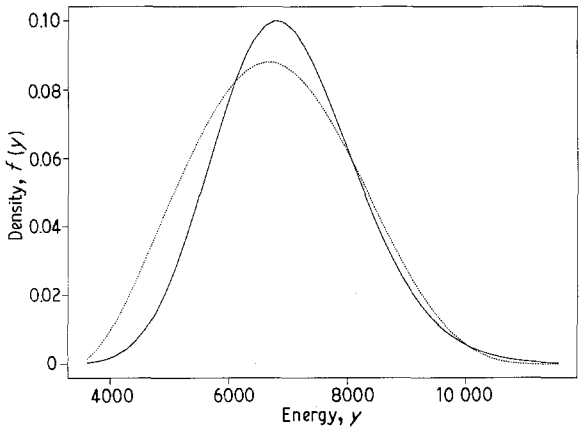

Figure 7 Graph of the recovered solution for $\alpha = 10^{12}$ (. . .) plotted against the original function (——) for the second simulated data set.

3.2.3. Using the ideas of generalized cross-validation

One widely recognized technique for finding an appropriate value for the smoothing parameter is that of cross-validation [7]. The principle underlying cross-validation is to leave out the data points one at a time and then to choose that value of α which yields solutions which best predict the missing data points given the remainder. Defining $\mathbf{f}_\alpha^{(i)}$ to be that solution to the constrained minimization problem obtained by

using smoothing parameter α and omitting data point i we choose as our parameter α, that which minimizes the function

$$\mathrm{XVSC}(\alpha) = \frac{1}{n}\sum_{i=1}^{n}((\mathbf{K}\mathbf{f}_{\alpha}^{(i)})_i - \mathbf{g}_i)^2 \qquad (27)$$

Essentially the procedure is seeking that α yielding solutions $\mathbf{f}_{\alpha}^{(1)}, \ldots, \mathbf{f}_{\alpha}^{(n)}$, which when premultiplied by $\mathbf{K}$ best predict the data $\mathbf{g}_1, \ldots, \mathbf{g}_n$.

In special cases the expression given in Equation 27 has a very much simpler computational form. In particular, if we were minimizing the regularization expression in Equation 11 without requiring non-negativity constraints on the $\mathbf{f}$, the solution $\hat{\mathbf{f}}_{\alpha}$ would be linear in the observations and found by solving the linear simultaneous equations

$$(\mathbf{K}^{\mathrm{T}}\mathbf{W}\mathbf{K} + \alpha\mathbf{Q})\hat{\mathbf{f}}_{\alpha} = \mathbf{K}^{\mathrm{T}}\mathbf{W}\mathbf{g} \qquad (28)$$

Defining the predicted vector of observations $\hat{\mathbf{g}}_{\alpha}$, given solution $\hat{\mathbf{f}}_{\alpha}$, by $\hat{\mathbf{g}}_{\alpha} = \mathbf{K}\hat{\mathbf{f}}_{\alpha}$, we can then find a matrix $\mathbf{A}(\alpha)$ with the property that

$$\hat{\mathbf{g}}_{\alpha} = \mathbf{A}(\alpha)\mathbf{g} \qquad (29)$$

namely

$$\mathbf{A}(\alpha) = \mathbf{K}(\mathbf{K}^{\mathrm{T}}\mathbf{W}\mathbf{K} + \alpha\mathbf{Q})^{-1}\mathbf{K}^{\mathrm{T}}\mathbf{W} \qquad (30)$$

With this definition a simple computational form for the cross-validation score and for a rotation invariant modification, the generalized cross-validation score, may be obtained. The generalized cross-validation score for the simple unconstrained case is given by

$$\mathrm{GXVSC}(\alpha) = \frac{1/n\mathrm{RSS}(\alpha)}{\{1 - 1/n\mathrm{Tr}[\mathbf{A}(\alpha)]\}^2} \qquad (31)$$

where $\mathrm{RSS}(\alpha)$ is the residual sums of squares given by

$$\mathrm{RSS}(\alpha) = (\mathbf{g} - \mathbf{K}\hat{\mathbf{f}}_{\alpha})^{\mathrm{T}}\mathbf{W}(\mathbf{g} - \mathbf{K}\hat{\mathbf{f}}_{\alpha}) \qquad (32)$$

and $\mathrm{Tr}\,[\mathbf{A}(\alpha)]$ is the trace of the matrix $\mathbf{A}(\alpha)$ as defined in Equation 30. The α for which the function $\mathrm{GXVSC}(\alpha)$ is minimized is known as the generalized cross-validation estimate of α.

Unfortunately constraining the vector $\hat{\mathbf{f}}_{\alpha}$ to be non-negative when minimising Equation 11 means that the solution $\hat{\mathbf{f}}_{\alpha}$ is no longer linear in the observations $\mathbf{g}$ and the score function no longer has the simple computational form given in Equation 31. Wahba [12] derives a form for the generalized cross-validation score when there is the problem of constraints to consider. However, its evaluation demands considerable computation. An approximation to that score function is therefore defined in Reference 12, which we shall denote $V(\alpha)$ given by

$$V(\alpha) = \frac{1/n\mathrm{RSS}(\alpha)}{\left(1 - 1/n\sum_{i=1}^{n}\partial\hat{\mathbf{g}}_{\alpha,i}/\partial\mathbf{g}_i\right)^2} \qquad (33)$$

Here $\mathrm{RSS}(\alpha)$ is the residual sums of squares given by

$$\mathrm{RSS}(\alpha) = (\mathbf{g} - \mathbf{K}\hat{\mathbf{f}}_{\alpha})^{\mathrm{T}}\mathbf{W}(\mathbf{g} - \mathbf{K}\hat{\mathbf{f}}_{\alpha}) \qquad (34)$$

where $\hat{\mathbf{f}}_{\alpha}$ is the solution to the constrained minimization problem in Equation 11. The partial derivative term in the denominator seeks to quantify the change in the predicted value of the ith data point $\hat{\mathbf{g}}_{\alpha,i}$ given a small change in the value of the ith observation $\mathbf{g}_i$. In the unconstrained case when $\hat{\mathbf{g}}_{\alpha}$ is linear in the observations $\mathbf{g}$, the partial derivative $\partial\hat{\mathbf{g}}_{\alpha,i}/\partial\mathbf{g}_i$ is simply the ith diagonal element of the matrix $\mathbf{A}(\alpha)$ and $V(\alpha)$ reduces to the score function in Equation 31. In this paper we shall employ a further approximation to evaluate the partial derivative. If we define $\hat{\mathbf{g}}_{\alpha}[\mathbf{g}]$ to be the predicted vector $\hat{\mathbf{g}}_{\alpha}$ given a smoothing parameter α and an initial set of observations $\mathbf{g}$, then we approximate the ith partial derivative by

$$\partial\hat{\mathbf{g}}_{\alpha,i}/\partial\mathbf{g}_i \approx \frac{\hat{\mathbf{g}}_{\alpha,i}[\mathbf{g} + \boldsymbol{\delta}_i] - \hat{\mathbf{g}}_{\alpha,i}[\mathbf{g}]}{\zeta_i} \qquad (35)$$

Here the vector $\boldsymbol{\delta}_i$ is an $n \times 1$ vector with one non-zero element ζ_i in the ith position. The ζ_i are small positive real numbers of the order of 10^{-5}. With this approximation the score function $V(\alpha)$ requires a constrained quadratic minimisation for the evaluation of each of the n partial derivatives. Hence determining the score function is a lengthy procedure.

In what follows we shall consider again the solutions obtained using the example data set $\mathbf{g}$ plotted in Fig. 2 and evaluate two different score functions for each of a number of different αs. The first we shall denote $V_{\mathrm{L}}(\alpha)$. It is essentially that measure given in Equation 31 which assumes linearity and is defined

$$V_{\mathrm{L}}(\alpha) = \frac{(\mathbf{g} - \mathbf{K}\hat{\mathbf{f}}_{\alpha})^{\mathrm{T}}\mathbf{W}(\mathbf{g} - \mathbf{K}\hat{\mathbf{f}}_{\alpha})}{(1 - 1/n\mathrm{Tr}[\mathbf{K}(\mathbf{K}^{\mathrm{T}}\mathbf{W}\mathbf{K} + \alpha\mathbf{Q})^{-1}\mathbf{K}^{\mathrm{T}}\mathbf{W})]} \qquad (36)$$

The only difference between this and Equation 31 is that the solution vector $\hat{\mathbf{f}}_{\alpha}$ in the numerator is that obtained by a constrained minimization of Equation 11 not by solving the simultaneous linear equations in Equation 28. The denominator is exactly as in Equation 31 and therefore independent of the observations $\mathbf{g}$. The second score function is the function $V(\alpha)$ defined in Equation 33 with the approximation for the partial derivative term given by Equation 35. Notice that both measures therefore have the same numerator. It is only the denominator which varies. We anticipate that $V(\alpha)$ will yield a better approximation to the score function derived in Reference 12 for constrained problems than the score function $V_{\mathrm{L}}(\alpha)$ and therefore be a more instructive means of finding an appropriate level of smoothing, despite its disadvantages in terms of computational effort.

The values of the two scores are given in Table IV for varying levels of α. Observe first how the numerator corresponding to the sum of squared re-

TABLE IV Values of the score functions for simulated data set 1

α	$1/n$ RSS(α) (5 sf)	$V_{\mathrm{L}}(\alpha)$ (5 sf)	$V(\alpha)$ (5 sf)
10^8	0.604 95	1.190 9	0.978 98
10^9	0.609 49	1.066 8	0.944 77
10^{10}	0.617 76	0.971 33	0.928 72
10^{11}	0.627 69	0.897 78	0.897 78
10^{12}	0.661 91	0.869 23	0.869 23
10^{13}	0.774 60	0.943 01	0.923 79

TABLE V Results for real adsorption data

α	Number of pivots	Specific amount adsorbed (4 sf)	Value of R^* (6 sf)	Summed predicted error (6 sf)
10^7	115	3917	−1052250	25.6419
10^9	88	1554	−1052250	26.2726
10^{11}	69	657.6	−1052250	29.4593
10^{12}	74	497.1	−1052240	32.7914
10^{13}	72	418.9	−1052230	40.2755
10^{14}	76	363.9	−1052190	62.4481
10^{15}	72	324.9	−1052100	124.852
10^{16}	79	303.4	−1051860	215.755
10^{17}	80	294.4	−1050180	281.244
10^{18}	78	281.7	−1034900	968.329

siduals increases as α increases. This term is proportional to the summed predicted errors we calculated in the previous section and reflects the fact that as α increases we are seeking smoother solutions at the expense of consistency with the data. The most important thing to notice from the table, however, is the very close similarity in the values of the two score functions, particularly for larger α. This suggests that the constraints are ceasing to be active for larger α and the solution obtained by a constrained minimization approximates closely the solution that would be obtained simply by solving the linear simultaneous equations in Equation 28. The similarity in the score values, is not just a feature of this data set but of all the data sets we have considered. The minimum value of the score function for the subset of α values that we have considered occurs at the same value of α, namely $\alpha = 10^{12}$, for both score functions $V_L(\alpha)$ and $V(\alpha)$. This value of the smoothing parameter falls in the interval of feasible α values as given by the observations in the previous section. Perhaps this should not surprise us. If we consider $V_L(\alpha)$ more closely, it is clear that the denominator increases slowly and steadily with α. The rate of increase of the numerator, however, goes up as α increases yielding a unique minimum for the score function. The point at which the numerator increases substantially with α corresponds to the point at which R^* increases substantially and we obtain oversmoothed solutions.

If we consider now the solution $\hat{\mathbf{f}}_\alpha$ corresponding to an α of 10^{12}, that depicted by the dotted curve in Fig. 3c, it is clear that this solution gives a very good approximation to the original gamma function from which the data was generated and moreover produces a predicted data vector $\hat{\mathbf{g}}_\alpha$ which fits the original data very closely indeed (Fig. 6c). In the light of these results and the observed behaviour for the other simulated data sets we draw two conclusions.

Firstly our investigations suggest that generalized cross-validation scores can, and do, provide useful information about an appropriate level of smoothing. Secondly, in all the examples considered the two scores have been very similar and have yielded the same values of the smoothing parameter. In view of the computational complexity of the second score function $V(\alpha)$, it may be adequate in most cases to evaluate only the simpler score function $V_L(\alpha)$ as a means of determining a sensible degree of smoothing.

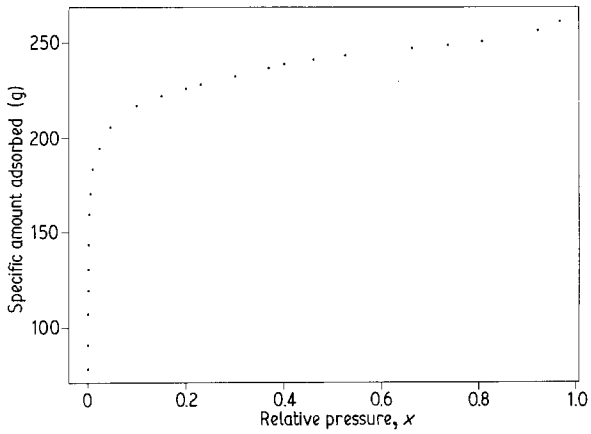

Figure 8 Plot of real data values corresponding to the specific amount adsorbed at 25 relative pressures.

TABLE VI Values of the score functions for the real adsorption data

α	$1/n$ RSS(α) (5 sf)	$V_L(\alpha)$ (5 sf)	$V(\alpha)$ (5 sf)
10^9	1.0509	2.8716	2.5416
10^{10}	1.0939	2.5735	2.4582
10^{11}	1.1783	2.4125	2.3781
10^{12}	1.3116	2.3711	2.3535
10^{13}	1.6111	2.6091	2.6024
10^{14}	2.4976	3.6732	3.6732

3.3. Applying the quadratic programming algorithm to real data

The aim in this section is to use the quadratic programming algorithm on the real adsorption data, originally used in Reference 3 to recover site energy distributions. The real adsorption data are plotted in Fig. 8 against relative pressure. The variance matrix $\mathbf{W}^{-1}$ for the errors in the data is assumed to be the identity matrix $\mathbf{I}$.

In Table V we see the numerical results of quadratic programming for varying values of the smoothing parameter α. Table VI gives the values of the two score functions for different values of α and Fig. 9 depicts the solution curves for different values of α. Both score functions suggest a smoothing parameter value of 10^{12}. If we consider the values of R^* as given in Table V we can see that 10^{12} falls in the region of

slowly varying R^* and has a relatively small value of the summed predicted error, namely 32.7914. From Fig. 9 the solution proposed can be seen to be approximately bimodal. Bimodality is a feature of many of the solutions obtained using αs in the region, although as the level of smoothing increases (see the curve corresponding to $\alpha = 10^{16}$) the bimodal behaviour is progressively smoothed away. Fig. 10a depicts the solution $\hat{\mathbf{f}}$ obtained for an α of 10^{12} and Fig. 10b compares the real data (plotted as points) with the predicted data $\hat{\mathbf{g}}$ (the dotted line) obtained using this chosen solution. It is clear that the solution $\hat{\mathbf{f}}$ obtained produces a good fit to the original data.

Comparison of the solution in Fig. 10a with those given in Reference 3 for their three different methods, one being the regularization technique outlined in Section 2, reveals quite significant differences. Most noticeably the solutions obtained in Reference 3 are unimodal. The first two methods assume unimodality in order to find the solution, so this should come as no surprise, but the regularization technique does not. No quantitative measure is given by the authors to assess the goodness of fit produced by their recovered solutions. However, the authors observe that for all three methods used the results produce residuals that are biased, the model tending to overpredict the amount adsorbed at low pressures and underpredict that observed at high pressures. The authors suggest that this might indicate a limitation in the Langmuir equation as the local isotherm. However, there may be an alternative explanation. If we return to the results obtained from our simulated data in Section 3.2, and in particular look at the predicted data corresponding to an α of 10^{15}, as displayed in Fig. 6d, we can observe the same phenomena. In the simulation study it occurred when the site distribution was oversmoothed. It is possible that the regularization technique adopted by McEnaney *et al.* [3] has given rise to an α which is too large, resulting in a degree of oversmoothing, which might account for the biased residuals.

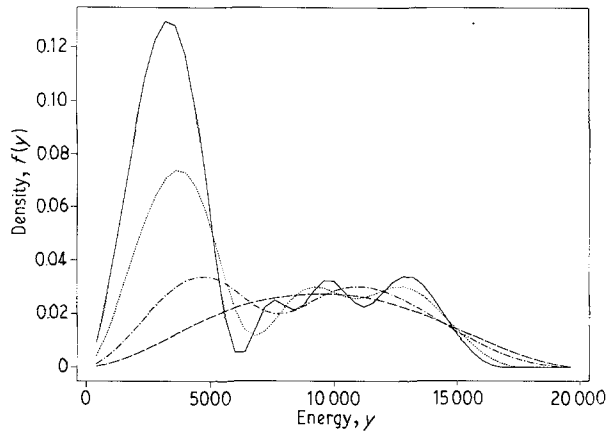

Figure 9 Graph of the recovered solutions $\hat{\mathbf{f}}$ for the data given in Fig. 8 and for four different values of the smoothing parameter, α: (——) 10^{11}, (– – –) 10^{12}, (— – —) 10^{14}, (— —) 10^{16}.

4. The pore size problem

The aim is to establish the pore structure of a particular type of graphite, in order to predict structure related properties such as strength and corrosion resistance and to assess its suitability for different applications. To do this, the gas transport properties of the solid are measured by exposing the graphite to two gases, pure helium and pure argon at varying pressures, x. Diffusion of both gases occurs through the slab of graphite and by measuring the proportions of helium and argon coming off at each side, it is possible to measure the diffusion rate of each gas through the solid for any given pressure x. From these observations an attempt was made to determine the pore structure of the solid.

The graphite is modelled as a "bundle" of cylindrical pores of varying radii, all parallel, non-intersecting and oriented in the direction of the diffusive flow. This model will enable an expression for the diffusion rate of the gas is be found in terms of the distribution of the pore radii of the solid. The departure of the model from reality will need to be compensated for at a later stage in the calculations. The density of the radii of the graphite pores is denoted $h(y), 0 < y < ymax$ where $ymax$ is some large value chosen to exceed all possible pore radii.

With this cylindrical pore structure model the dependence of the gas diffusion rate on the porosity of the solid, the gases involved, pressure and pore size y, can be described by

$$g(x) = \int_0^{ymax} K(x, y) f(y) \mathrm{d}y \qquad (37)$$

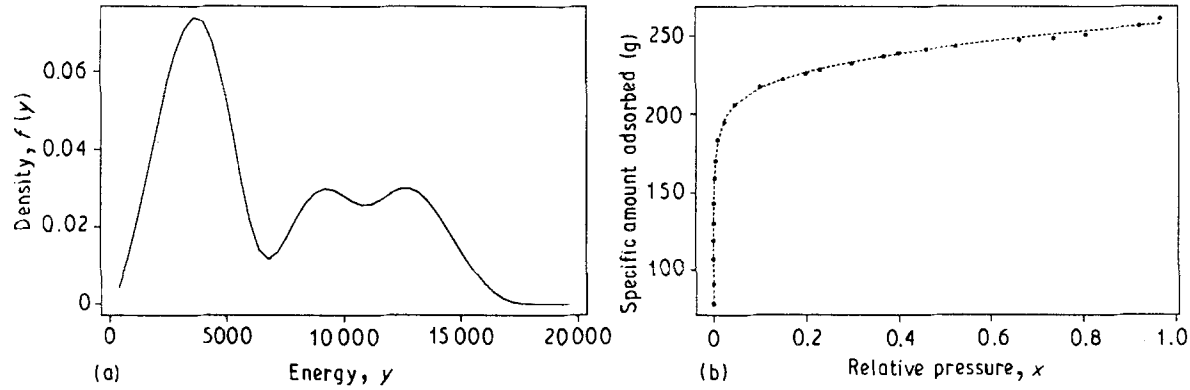

Figure 10(a) Recommended recovered solution $\hat{\mathbf{f}}$ for the real data set given in Fig. 8, corresponding to a smoothing parameter of $\alpha = 10^{12}$. (b) Predicted data obtained from the solution in Fig. 10a (. . .) plotted against the original observations.

where the quantity $g(x)$ represents the diffusivity rate at pressure x for some specified gas G [5]. The kernel function $K(x, y)$ is a function of both pressure and pore radius, y

$$K(x, y) = ay^2 \ln\left(\frac{bxy + 1}{cxy + 1}\right) \quad (38)$$

where a, b and c are known constants which depend on the gas and the type of graphite. For the gas helium and for y measured in micrometres and x measured in kilopascals, the constants take the values $a = 1.5514$, $b = 0.11253$ and $c = 0.036082$.

The quantity $f(y)$ is defined to be $N_p h(y)/q$ where N_p is the number of pores per cubic metre, and q is called the tortuosity factor. In a real solid, the pores are neither all cylindrical nor oriented in the direction of the gas diffusion. For example, pores may twist and turn, connect together or vary in width along the whole length. The tortuosity factor is designed to compensate for the departure of the true pore structure from the model assumptions. It is defined to be the diffusion rate that would occur through a cylindrical pore oriented in the direction of diffusive flow, divided by the actual diffusion rate that does occur through the real pore of the same effective radius. By virtue of the definition, it is impossible to determine the constants N_p and q. However

$$\int_0^{ymax} f(y)\,dy = \frac{N_p}{q}\int_0^{ymax} h(y)\,dy = \frac{N_p}{q} \quad (39)$$

since h is a probability density function. Hence, as in the previous problem, it is a simple procedure to evaluate N_p/q and the pore size density $h(y)$, $0 < y < ymax$, once the function f has been determined.

Before proceeding further it is worth considering what happens to the kernel function as pressure x varies. For x very small it can be shown that

$$K(x, y) \approx ay^3(b - c)x \quad (40)$$

and hence using the integral Equation 37 that

$$g(x) \approx a(b - c)x\int_0^{ymax} y^3 f(y)\,dy \quad (41)$$

so that $g(x)$ is approximately proportional to x. When pressure x is large it can be shown that

$$k(x, y) \approx \left(a \ln\frac{b}{c}\right)y^2 \quad (42)$$

and hence that

$$g(x) \approx \left(a \ln\frac{b}{c}\right)\int_0^{ymax} y^2 f(y)\,dy \quad (43)$$

Thus $g(x)$ is a constant for x sufficiently large. Indeed it is proportional to the second moment of the function f. It follows that, for sufficiently low and sufficiently high pressures x, we are unable to determine much about the solution f apart from its third and second moments, respectively.

The nature of the kernel function makes this problem a particularly difficult one to solve, as information about the unknown function f is only attainable for a very limited range of pressures. A further insight into the very difficult nature of this particular problem can be gained by considering the eigenvalues of the matrix $\mathbf{K}^T\mathbf{K}$ where $\mathbf{K}$ is the matrix approximation to the integral operator in Equation 37. In this problem, observations have been made at 23 pressures and the function is evaluated on 49 points as before. Hence $\mathbf{K}$ is a 23×49 matrix. As a result we know that the 49×49 matrix $\mathbf{K}^T\mathbf{K}$ will have at least 26 zero eigenvalues. It is the remaining 23 however, that are of most interest.

To calculate the eigenvalues a routine was used involving Householder reduction and the QL algorithm [13]. If $e_1, e_2, e_3, \ldots, e_{49}$ denote the 49 eigenvalues in order of decreasing magnitude, the eigenvalues are given to four significant figures by

$$e_1 = 0.2115$$
$$e_2 = 0.000\,034\,53$$
$$e_3 = 0.000\,000\,024\,48$$
$$e_j < 10^{-9} \text{ for all other } j.$$

These results indicate only too clearly the problem we are facing. Not only are the 26 eigenvalues we would expect, zero, but 20 of the remaining eigenvalues are less than10^{-9} and even e_2 and e_3 are fairly inconsequential compared to e_1. Because the 23 non-zero eigenvalues of $\mathbf{K}^T\mathbf{K}$ correspond to the squares of the moduli of the 23 singular values of $\mathbf{K}$, this implies that at most three of the columns of $\mathbf{K}$ are effectively linearly independent. The quantity $\mathbf{K}$ is a linear operator whose kernel or null space is approximately a 46-dimensional subspace of a 49-dimensional domain. Without introducing the added criteria of smoothness, an enormous number of completely different solutions will be able to reproduce the data. The effect of smoothing is to restrict our function f to some subspace of the domain consisting of "smooth" functions and in so doing restrict the number of feasible solutions. However, it is evident that in such a situation, the solution will owe as much to the information in the smoothing matrix as to the information in the $\mathbf{K}$ matrix, and hence recovery of solutions may well be poor.

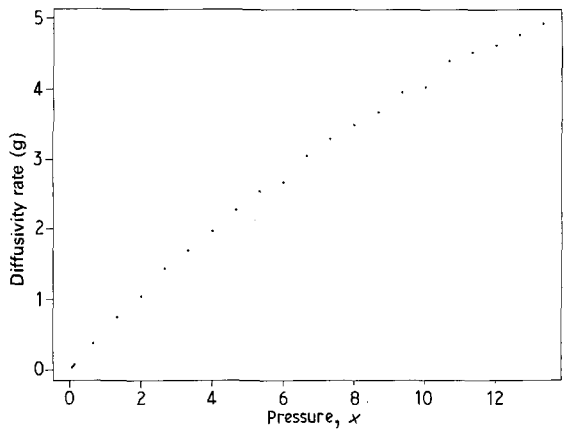

Figure 11 Plot of real data values corresponding to the diffusivity rate of helium through graphite at 23 different pressures.

It should be stressed that in the light of these observations any solution obtained can at best be only approximate. However, it will be interesting to investigate the problem nevertheless and to make comparisons with the solution obtained in Reference 5. The real diffusion data as used in this latter paper takes the following form. Each data value $\mathbf{g}_i$, $i = 1, \ldots, 23$ corresponds to the mean of nine different observations at pressure x_i, $i = 1, \ldots, 23$ of the diffusivity rate of helium. The data are plotted in Fig. 11. In recovering the pore radii distribution, the lower limit of the pore radii, $y_{\min}$, was taken to be zero, while the upper limit $y_{\max}$ was taken to be 1.40 (based on physical considerations in the problem). The number of points m, as stated above, was taken to be 49.

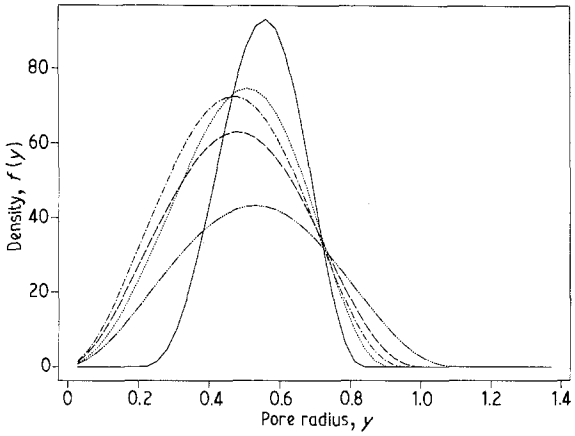

Figure 12 Graph of the recovered solutions $\hat{\mathbf{f}}$ for the data given in Fig. 11 and for five different values of the smoothing parameter, α: (——) 10^{-11}, (– – –) 10^{-10}, (— – —) 10^{-9}, (— —) 10^{-8}, (— – – – —) 10^{-7}.

We use exactly the same technique to find a solution as we used for the adsorption problem, namely discretizing the integral equation in Equation 37 and minimizing the regularization expression R^*, as given in Equation 11, for a number of different α values. The numerical results of the quadratic programming algorithm for different αs are given in Table VII, the values of the score functions in Table VIII and the solution vectors plotted in Fig. 12. The score functions suggest that an α of 10^{-9} might be an appropriate one to choose. Once again we observe the similarity in the values of the score functions, but this time we note that $V(\alpha)$ does not have a single minimum but a number of local minima. The possibility of this occurring gives yet another reason why it is both simpler and perhaps more instructive to consider only the score function $V_L(\alpha)$.

If we consider the value of R^* and the solution $\hat{\mathbf{f}}$ (Fig. 13a) in the usual way, this value of α produces a smooth curve in the region of slowly varying R^*. Evaluating the vector $\hat{\mathbf{g}}$ of predicted data values given this solution $\hat{\mathbf{f}}$, we obtain the result depicted by the dotted line in Fig. 13b (plotted against the real data). A good fit to the original data is produced.

It is interesting to compare this solution with the one obtained in Reference 5 using the method described in Section 2. In Reference 5 the recovered pore radii density (after normalization) was normal with mean 0.45 μm and standard deviation 0.15 μm. This shows a gratifying similarity to our chosen solution. Consideration of the value of the area under the curve corresponding to the constant N_p/q is even more encouraging. Our chosen solution produced a value for N_p/q of 35.18 mm^{-3} (to two decimal places). In Reference 5, values for N_p and q are given which are

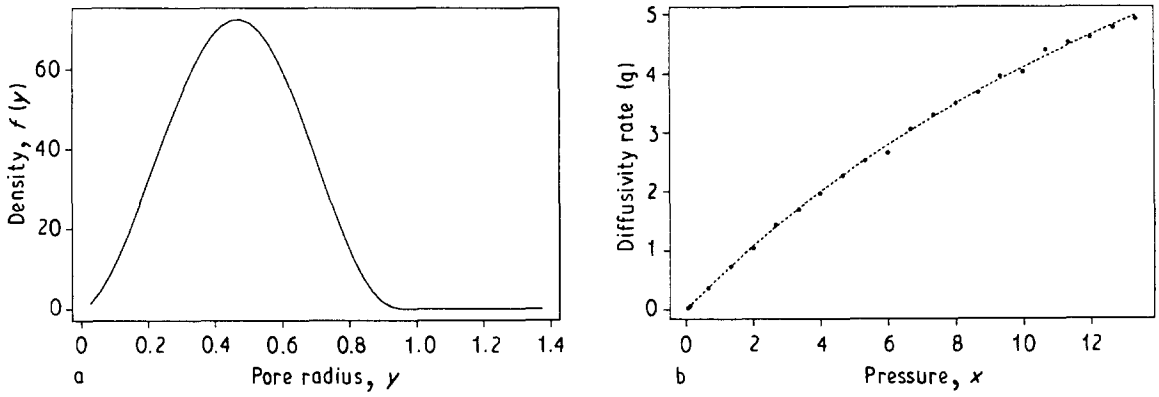

Figure 13(a) Recommended recovered solution $\hat{\mathbf{f}}$ for the real data set given in Fig. 11, corresponding to a smoothing parameter of $\alpha = 10^{-9}$. (b) Predicted data obtained from the solution in (. . .) plotted against the original observations.

TABLE VII Results for real diffusion data

α	Number of pivots	Specific amount adsorbed (3 sf)	Value of R^* (6 sf)	Summed predicted error (6 sf)
10^{-12}	59	25.1	−216.014	0.050 0965
10^{-11}	77	27.2	−216.013	0.050 1935
10^{-10}	110	33.4	−216.014	0.050 3625
10^{-9}	109	35.2	−216.012	0.050 4680
10^{-8}	108	31.9	−216.004	0.052 2389
10^{-7}	128	24.3	−215.967	0.070 0862

TABLE VIII Values of the score functions for the real diffusion data

α	$1/n$ RSS(α) (5 sf)	$V_L(\alpha)$ (5 sf)	$V(\alpha)$ (5 sf)
10^{-12}	0.002 1780	0.002 9064	0.002 6121
10^{-11}	0.002 1824	0.002 8753	0.002 6188
10^{-10}	0.002 1897	0.002 8014	0.002 6262
10^{-9}	0.002 1943	0.002 6738	0.002 6107
10^{-8}	0.002 2711	0.002 7260	0.002 6396
10^{-7}	0.003 0458	0.003 6151	0.003 4517

calculated from molecular diffusivity and permeability data. The values given yield a value for N_p/q of 33 mm^{-3}, to two significant figures. Not only is our estimate of N_p/q very close to the actual value, it is also an improvement on that obtained in Reference 5, namely a value of 27 mm^{-3}.

5. Conclusions

A method is presented for the solution of Fredholm first kind integral equations, when the solution we are seeking is a density function and therefore non-negative. Such equations occur frequently in scientific problems. Here we have addressed ourselves to two specific problems found in the material science literature. Other problems (for example, the extraction of site-energy distributions from temperature programmed desorption data [2]) may also be tackled in this way. The method uses the ideas of regularization combined with a quadratic programming algorithm for minimizing quadratic expressions subject to non-negativity constraints. A smoothing parameter α in the regularization expression provides a means of "trade-off" between the degree of smoothness of our solution and the degree of consistency it yields with our observed data.

An approximate value for the smoothing parameter is chosen by combining the ideas of generalized cross-validation with a consideration of the data predicted using our solution and its closeness to the original observations. Two approximations to the generalized cross-validation function are considered. The first, $V_L(\alpha)$, is a modification of the cross-validation function that would be used if the regularization expression was to be minimized without constraints. The second, $V(\alpha)$, is an approximation to that generalized cross-validation function developed specifically for the constrained problem. In all cases investigated, the two functions yielded the same value of the parameter α for the subsets of α values considered. In view of the considerable computational cost involved in evaluating $V(\alpha)$ for apparently no extra information, it is recommended that only the simpler $V_L(\alpha)$ need be calculated.

The method proposed is tested on both simulated and real data. The results on the simulated data show that the procedure produces solutions that are a very good approximation to the true solution function. Results obtained from the real data show that the technique gives rise to solutions, that while being smooth are nevertheless consistent with the observed data. Moreover in the diffusion problem, the solution obtained yields a value for the unknown constant N_p/q which is consistent with that obtained by experimentation.

Acknowledgements

The initial stages of this work were carried out under the supervision of Professor B. W. Silverman and with the financial support of the Science and Engineering Research Council. The author is particularly grateful to Dr Tim Mays for many useful conversations regarding the physical aspects of the problem and for supplying the real data which made this work possible. She would also like to thank Professor Peter Green for his many useful comments on earlier drafts of this paper.

References

1. P. H. MERZ, *J. Comp. Phys.* **38** (1980) 64.
2. J. A. BRITTEN, B. J. TRAVIS and L. F. BROWN, *AIChE Symp. Ser.* **79** (1983) 7.
3. B. McENANEY, T. J. MAYS and P. D. CAUSTON, *Langmuir* **3** (1987) 695.
4. L. F. BROWN and B. J. TRAVIS, *Chem. Engng Sci.* **38** (1983) 843.
5. T. J. MAYS and B. McENANEY, in "Extended abstracts of the 17th biennial conference on carbon", Lexington, USA 1985, pp. 104–105.
6. P. WOLFE, *Econometrica* **27** (1959) 382.
7. G. WAHBA, *SIAM J. Numer. Anal.* **14** (1977) 651.
8. J. P. BUTLER, J. A. REEDS and S. V. DAWSON, *ibid.* **18** (1981) 381.
9. M. JARONIEC and P. BRAUER, *Surf. Sci. Rep.* **6** (1986) 65.
10. G. MARSAGLIA, in "Information Theory Statistical Decision Functions Random Processes: Transactions of the Third Prague Conference", edited by J. Kozesnik (Czechoslovak Academy of Sciences, Prague, 1962) pp. 499–510.
11. B. A. WICHMANN and I. D. HILL, *J. Roy. Statist. Soc. Ser. C* **31** (1982) 188.
12. G. WAHBA, *Statistical Decision Theory and Related Topics III* **2** (1982) 383.
13. J. H. WILKINSON and C. REINSCH, in "Handbook for Automatic Computation, Volume II, Linear Algebra" (Springer Berlin, 1971) p. 227.

Received 25 July
and accepted 10 September 1991