MINIMUM NASTINESS CURVE FITTING

BY
JOHN M. BLATT(1) AND HANOCH GUTFREUND

ABSTRACT

The "nastiness" of a function $\phi(x)$ is defined. We then discuss minimum nastiness interpolation to a set of given points (x_k, ϕ_k), as well as minimum nastiness curve fitting, where the given values ϕ_k have errors δ_k.

1. **Introduction.** Curve fitting is the art of drawing a smooth curve through given, experimental points so as to strike a compromise between the desire to come close to the points and the desire to have a smooth curve which is in accordance with whatever additional information we may possess about the true curve. For example, we may know from other considerations that the true curve $y(x)$ is such that y is positive, or we may have much more detailed information, e.g., $y(x)$ is a straight line.

The usual approach dates back to Gauss [1]. Let $f(a,x)$ be a family of curves depending upon a parameter a. Let $x_0, x_1, \cdots, x_K$ be the given abscissae, $y_0, y_1, \cdots, y_K$ be the measured ordinates, and $\delta_0, \delta_1, \cdots, \delta_K$ be the standard errors. The chi-square value of the fit is

$$X = \sum_{k=0}^{K} \frac{[y_k - f(a,x_k)]^2}{2(\delta_k)^2} \tag{1.1}$$

and the "likelihood" of the fit is

$$P(a) = \exp(-X). \tag{1.2}$$

We then vary the parameter a so as to maximize the likelihood, i.e., so as to minimize the sum of squares (1.1).

If the family of curves depends upon several parameters $a_1, a_2 \cdots, a_M$, say, there is no change in principle. The likelihood $P = \exp(-X)$ is now a function of these M parameters, and we maximize P, or minimize X, with respect to variations of all these M parameters.

This scheme works well if we possess large amounts of a priori information. e.g., if we know that the data should fall on a straight line. Statistical theory then enables us to find the best straight line, and to decide whether our expectation

Received August 29, 1966, and, in revised form, September 14, 1967.

(1) On study leave from the University of New South Wales, Kensington, NSW, Australia.

of a good straight line fit is confirmed, or contradicted, by the experimental values, within some preassigned confidence level.

But the scheme works very much less well if we possess no a priori information whatever about the fitting function, beyond a vague feeling that it ought to be "smooth". One may fit with polynomials of increasing degree, stopping when an increase of the degree of the polynomial no longer leads to a satisfactory improvement in the likelihood of the fit. However, many functions of simple functional form cannot be approximated well by polynomials of low degree. For example, $f(x) = \exp(-5x)$ in the interval $(0, 1)$ would require a polynomial of quite high degree to give a satisfactory fit. Low-order polynomial fitting prejudices us against such a perfectly reasonable function. Functional transformations are possible, e.g., we may try a fit to $\log f(x)$ rather than to $f(x)$, but there is an infinity of functional transformations and no simple way of selecting between them.

In this paper we present an alternative approach, which we believe to be new and worthy of further study. We start by defining the q-nastiness $N_q(\phi)$ of a given function $\phi(x)$, in such a way that a horizontal straight line has nastiness zero, and such that the nastiness increases as the curve becomes less "smooth". We then solve the minimum problem: find the curve of minimum nastiness passing exactly through a set of given points. We then proceed to the problem of minimum nastiness interpolation, so as to get a criterion for a reasonable choice of the parameter q. Thereafter we allow our points to have errors δ_k, and we discuss the variational problem of maximizing the "extended likelihood"

$$Q = F(N)\exp(-X) \tag{1.3}$$

where $F(N)$ is some monotonically decreasing, differentiable function of the nastiness N. In spite of the apparent arbitrariness involved in the choice of the function $F(N)$, it turns out that we are led to a mere *one-parameter family* of minimum nastiness curve fits. Finally, in the last section of the paper, we suggest what appears to us to be a reasonable choice of the function $F(N)$.

Before giving our definition of nastiness, in Section 2, let us conclude this introduction by showing that no simple definition of nastiness can be satisfactory. We restrict ourselves, without loss of generality, to functions $\phi(x)$ defined on the closed interval $(0, 1)$ of the x-axis. The simplest choice of a "nastiness" consistent with zero nastiness for horizontal straight lines and positive definite nastiness for all other curves is undoubtedly

$$M(\phi) = \int_0^1 \left(\frac{d\phi}{dx}\right)^2 dx. \tag{1.4}$$

Now consider the problem: Minimize $M(\phi)$ subject to $\phi(x_k) = y_k$ for a given set of points (x_k, y_k), all x_k in the interval $(0, 1)$. This is a straightforward problem

in the calculus of variations, with an exceedingly simple answer: we merely connect the given points by straight line segments![2]

The fact that this solution fails to have a continuous first derivative may be surprising, but should not be so. The existence of (1.4) requires that the first derivative be a square integrable function, no more than that. Piece-wise continuous first derivatives are consistent with this, and the variational solution makes use of this freedom.

This "solution" is however hardly what we mean intuitively by "smooth" function! The intuitive concept of smoothness requires that $\phi(x)$ have derivatives of all orders, i.e., that $\phi(x)$ is an *analytic* function of $z = x + iy$ over a region of the z-plane including the real line segment (0, 1).

One may be tempted to get around the difficulty by including some higher derivatives in an extended definition of nastiness. For example we may extend (1.4) to read

$$M_q(\phi) = \int_0^1 \left[\left(\frac{d\phi}{dx} \right)^2 + \left(\frac{q}{2} \right)^2 \left(\frac{d^2\phi}{dx^2} \right)^2 \right] dx \tag{1.5}$$

where q is a parameter. The resulting minimum problem is slightly more complicated to discuss but the solution is still inconsistent with our intuitive notions of smoothness, i.e., still not an analytic function. This time, the second derivatives have finite discontinuities at the given abscissae x_k.

2. **Definition of nastiness.** No finite number of derivatives in a definition of the nastiness can ensure that the minimum nastiness problem leads to an analytic function of x. However, we can define the nastiness by an integral over an infinite series involving all derivatives of the function $\phi(x)$. Our definition is

$$N_q(\phi) = \int_0^1 dx \sum_{s=1}^{\infty} \left(\frac{q^s}{s!} \frac{d^s\phi}{dx^s} \right)^2. \tag{2.1}$$

The quantity q is a real parameter, whose value will be fixed later on, in specific cases. For the moment, we declare the integral (2.1), if it exists, to be the "q-nastiness" of $\phi(x)$.

The choice of N as a homogeneous quadratic form in ϕ has the desirable effect of making the associated minimum problem linear. Omission of the term $s = 0$ from the infinite sum in (2.1) means that N is invariant under a vertical displacement of the function, and $N = 0$ for all horizontal straight lines. Retention of the term $s = 1$ means that sloping straight lines have non-zero nastiness $(qm)^2$, where m is the slope of the line. This choice is made in accordance with the general

---

(2) To see this, obseve that the integral (1.4) can be written as a sum of integrals, the kth ntegral extending from x_{k-1} to x_k. Each of the sub-integrals is minimized by the straight line solution. Q.E.D.

preference for fitting of points by a constant (horizontal straight line) over fitting by a sloping straight line.

The parameter q in (2.1) has the dimension of a length, and is a scaling factor for the variable x. The choice of q affects the nature of the minimum nastiness solution, in a way to be discussed in Section 3.

The factors $s!$ in (2.1) are required in order that some very simple functions should have finite nastiness. The sth derivative of the function $1/(1+x)$ equals $(-)^s s!(1+x)^{-s-1}$. Without the factors $s!$ in (2.1), the infinite series would diverge for all non-zero values of q.

An interesting way of looking at the definition (2.1) is provided by the Taylor expansion of the function $\phi(z) = \phi(x+iy)$ around some point $x = x_0$ on the interval $(0,1)$ of the x-axis. If $\phi(x)$ is to have finite nastiness, the series in (2.1) must converge for every x in $(0,1)$. If this series converges, the Taylor series for $\phi(x_0+z)$ converges, with a radius of convergence at least equal to q. Hence, *a function of finite q-nastiness is analytic over the interior of the oval-shaped region $R(q)$ defined by*: every point z in $R(q)$ has its least distance to the line segment $(0,1)$ less than or equal to q in value. This region is bounded by two semi-circles of radius q, one to the right of $x=1$, the other to the left of $x=0$ and by two horizontal straight lines, at $\mathrm{Im}(z) = iq$ and at $\mathrm{Im}(z) = -iq$, respectively.

This relationship between the Taylor series and (2.1) can be used to express (2.1) as a contour integral, as follows:

$$N_q(\phi) = \frac{1}{2}\pi \int_0^{2\pi} dt \int_0^1 dx \left|\phi(x+qe^{it})\right|^2. \tag{2.1a}$$

However, we have not found this transformation particularly helpful.

Of considerable help, however, is the following simple theorem: Let $\phi(x)$ be defined on $(0,1)$. We define a new function $u(x)$ by subtracting the straight line connecting the extreme points, i.e.,

$$u(x) = \phi(x) - \phi(0)(1-x) - \phi(1)x \tag{2.2}$$

so that $u(0) = u(1) = 0$. A simple calculation then shows that the nastiness of ϕ and of u are related by

$$N_q(\phi) = N_q(u) + q^2[\phi(1) - \phi(0)]^2. \tag{2.3}$$

This simple relationship allows us to specialize to functions $u(x)$ with $u(0) = u(1) = 0$ without loss of generality.

3. **The function of minimum nastiness passing through given points.** Suppose we are given L points $(x_k, u_k), k = 1, 2, \cdots, L$, all x_k in the open interval $(0,1)$, plus the two extreme points $(0,0)$ and $(1,0)$. We wish to find the function $u(x)$ of minimum nastiness passing through these points, i.e., satisfying the conditions

$$u(0) = u(1) = 0 \qquad u(x_k) = u_k \qquad k = 1, 2, \cdots, L. \tag{3.1}$$

It turns out to be helpful to express $u(x)$ as a Fourier series:

$$u(x) = \sum_{m=1}^{\infty} c_m \sin(m\pi x). \tag{3.2}$$

We shall assume rapid convergence of this series, allowing various interchanges of limiting processes. This assumption will be verified at the end, on hand of the explicit solution.

We differentiate (3.2) term by term to obtain, for even s,

$$d^s u/dx^s = (-1)^{1/2s} \sum_{m=1}^{\infty} (m\pi)^s c_m \sin(m\pi x) \qquad s \text{ even}. \tag{3.3}$$

The result for odd values of s is similar: the sign in front is altered, and we get cosines instead of sines.

We now square (3.3) and integrate from $x = 0$ to $x = 1$. Using the orthogonality of the sine (or cosine, for odd s) functions we obtain the result

$$\int_0^1 (d^s u/dx^s)^2 \, dx = \frac{1}{2} \sum_{m=1}^{\infty} (m\pi)^{2s} (c_m)^2 \tag{3.4}$$

Substitution of (3.4) into (2.1) leads to

$$N_q(u) = \frac{1}{2} \sum_{s=1}^{\infty} \sum_{m=1}^{\infty} \frac{(qm\pi)^{2s}}{(s!)^2} (c_m)^2. \tag{3.5}$$

We interchange the order of the summations, and recognize that the sum over s is closely related to the power series for the Bessel function $I_0(z)$ of imaginary argument. The result is

$$N_q(u) = \frac{1}{2} \sum_{m=1}^{\infty} [I_0(2m\pi q) - 1](c_m)^2. \tag{3.6}$$

We must minimize (3.6) subject to the conditions arising from (3.1), i.e., subject to

$$\sum_{m=1}^{\infty} c_m \sin(m\pi x_k) = u_k \qquad k = 1, 2, \cdots, L. \tag{3.7}$$

These conditions are handled in the usual way by introduction of Langrange multipliers Z_k, with the result

$$c_m = [I_0(2m\pi q) - 1]^{-1} \sum_{k=1}^{L} Z_k \sin(m\pi x_k). \tag{3.8}$$

To determine the Z_k, we substitute (3.8) into (3.7). We define the L-by-L matrix A_{kl} by

$$A_{kl} = \sum_{m=1}^{\infty} \frac{\sin(m\pi x_k)\sin(m\pi x_l)}{I_0(2m\pi q) - 1} \tag{3.9}$$

to write the conditions on the Z_k in the form

$$\sum_{l=1}^{L} A_{kl} Z_l = u_k \qquad k = 1, 2, \cdots, L. \tag{3.10}$$

Let B_{kl} be the matrix inverse to A_{kl}, then (3.10) is solved by

$$Z_k = \sum_{l=1}^{L} B_{kl} u_l \qquad k = 1, 2, \cdots, L. \tag{3.11}$$

This completes the solution of the minimum problem. The matrix A is determined by the given abscissas x_k and the chosen value of q, through (3.9); its inverse, B, is then determined; this gives the Lagrange multipliers Z_k by (3.11), hence the Fourier coefficients c_m by (3.8), hence the actual function $u(x)$ by (3.2).

All that remains is to check that the Fourier series converges well enough to permit the various interchanges of limiting processes which have occurred. The coefficients c_m of (3.8) are bounded by

$$|c_m| \leqq \frac{Z}{I_0(2m\pi q)} \quad \text{with} \quad Z = \sum_{k=1}^{L} |Z_k|. \tag{3.12}$$

Since the Bessel function $I_0(x)$ behaves exponentially for large real x, the series of bounds (3.12) converges. The sine series (3.2) for $u(x)$ then converges absolutely and uniformly for all x in $(0, 1)$. Similar arguments suffice for justifying all our formal operations in the proof.

The minimum nastiness itself, (3.6), can be expressed rather simply in terms of the function values u_k. We insert (3.8) into (3.6) for *one* of the factors c_m to obtain

$$N_q(u) = \frac{1}{2} \sum_{m=1}^{\infty} \sum_{k=1}^{L} Z_k \sin(m\pi x_k)\, c_m = \frac{1}{2} \sum_{k=1}^{L} Z_k u_k. \tag{3.13}$$

Using (3.11) for Z_k gives the quadratic form

$$N_q(u) = \frac{1}{2} \sum_{k,l=1}^{L} u_k B_{kl} u_l. \tag{3.14}$$

We note that the ordinates u_k enter explicitly since the matrix B depends on the abscissae x_k only, not on the ordinates u_k.

We now drop some of the restrictive assumptions. First, suppose that the ordinates at $x = 0$ and $x = 1$ are not zero, i.e., we are given $L + 2$ points (x_k, ϕ_k)

$k = 0, 1, 2, \cdots, L+1$, with $x_0 = 0$ and $x_{L+1} = 1$, with none of the ϕ_k restricted to vanish. We define new given values u_k in accordance with the transformation (2.2) by

$$u_k = \phi_k - \phi_0(1 - x_k) - \phi_{L+1}x_k. \tag{3.15}$$

This gives $u_0 = u_{L+1} = 0$ and thus reduces to the problem we have just solved. The solution $\phi(x)$ of the minimum nastiness problem is related to $u(x)$ by (2.2), and the minimum nastiness itself is given by (2.3). This latter result can be re-written as a quadratic form in the given ordinates ϕ_k

$$N_q(\phi) = \frac{1}{2} \sum_{k,l=0}^{L+1} \phi_k C_{kl} \phi_l \tag{3.16}$$

where the matrix C is directly related to the matrix B of the simpler problem via:

$$C_{kl} = B_{kl} \qquad \text{for } 1 \leqq k, l \leqq L \tag{3.17a}$$

$$C_{k0} = C_{0k} = -\sum_{l=1}^{L} B_{kl}(1 - x_l) \qquad \text{for } 1 \leqq k \leqq L \tag{3.17b}$$

$$C_{k,L+1} = C_{L+1,k} = -\sum_{l=1}^{L} B_{kl}x_l \qquad \text{for } 1 \leqq k \leqq L \tag{3.17c}$$

$$C_{00} = 2q^2 + \sum_{k,l=1}^{L} (1 - x_k)B_{kl}(1 - x_l) \tag{3.17d}$$

$$C_{L+1,L+1} = 2q^2 + \sum_{k,l=1}^{L} x_k B_{kl} x_l \tag{3.17e}$$

$$C_{0,L+1} = C_{L+1,0} = -2q^2 + \sum_{k,l=1}^{L} x_k B_{kl}(1 - x_l). \tag{3.17f}$$

Next we remove the restriction that the function values at $x = 0$ and $x = 1$, the endpoints of the interval, should be given at all. It is just as easy to see how the results are changed if *any* one abscissa x_k and associated function value ϕ_k are removed from the set of given values. The function $\phi(x)$ of minimum nastiness must now fit function values ϕ_m at $x = x_m$ for all $m \neq k$, but $\phi(x)$ is unrestrained at x_k. Let ϕ_k be its actual value at $x = x_k$, a value that is so far unknown. Then the nastiness is given by (3.16) which we may now minimize with respect to the unknown ϕ_k. This procedure decides the best value of ϕ_k, and solves the problem.

Extension of this scheme to the case where more than one function value is omitted from the given list is obvious, and leads to a simple problem in linear algebra.

In concluding this section, we discuss the role of the parameter q in the definition (2.1) of the nastiness. It is apparent from (3.9) that q enters explicitly into the solution of the minimum nastiness problem, as expected. The function of minimum

nastiness through a given set of points depends upon q. We now discuss the two limiting cases, large q and small q.

If q is large, the bound (3.12) on the Fourier coefficients is a rapidly decreasing function of the order m of the coefficient c_m. Thus the Fourier series converges rapidly, and can be approximated adequately by its first few terms. Excluding the outer values ϕ_0 and ϕ_{L+1} (since they do not enter the Fourier series), there are L function values ϕ_k to be fitted by a trigonometric polynomial. This requires, in general, L independent terms. Thus, in the limit of very large q, the function $\phi(x)$ consists of a straight line connecting the given values ϕ_0 and ϕ_{L+1}, plus a trigonometric polynomial with exactly L terms, chosen so that $\phi(x)$ passes through all the given points.

Quite apart from the rather special appearance of a low order trigonometric polynomial, the limiting case of high q gives rise to extremely large and "capricious" values of the nastiness. This can be seen from equation (3.6): in this limit, the coefficients c_m retain significant values until $m = L$, then become very small. Since the Bessel functions increase exponentially with large positive argument, the nastiness for high q is approximated well by

$$N_q(u) \cong \frac{1}{2}[I_0(2L\pi q) - 1](c_L)^2 \quad \text{(Limit of large } q\text{).} \tag{3.18}$$

The coefficient c_L depends only on the given function values, not on q, in this limit, since there are just enough coefficients retained to fit all the given function values. Hence $N_q(u)$ depends on q through the Bessel function in front, and this becomes exponentially large. Furthermore, small changes in the given function values can result in significant changes in c_L, so that the nastiness is not only disturbingly large, but a very sensitive function of the given information.

The opposite limiting case, q very small, is equally unsatisfactory. Over most of the range of integration in (2.1), the infinite series in the integrand can be replaced by its leading term, $s = 1$. In effect, then, we are reduced to the oversimplified definition (1.4) of the nastiness. This definition leads to a series of straight line segments connecting the given points. The sharp corners of this "solution" are inconsistent with convergence of the series in (2.1), so that the true solution must differ from the straight line segments at least in the immediate neighbourhood of the given points. The sharp corners must be smoothed over. It is easy to see, however, that this smoothing process takes place in distances of the order of q from the given abscissae x_k. Hence, if q is small compared to all distances between successive x_k, the minimum nastiness solution reduces to a series of straight line segments over most of the interval, the exceptions being a rounding off of the sharp corners in the q-neighbourhood of each given point (x_k, ϕ_k). Such a solution does not agree with our intuitive feeling of a "smooth curve through the given points".

4 **Minimum nastiness interpolation.** Although it is not often presented this way, interpolation is in essence a probabilistic question, with the choice of a "best" interpolation method depending upon the universe of functions to which we wish to interpolate. The usual polynomial interpolation is by no means "best for all "smooth functions".

Take as an example the function

$$f(x) = \frac{b^2}{(x-\frac{1}{2})^2 + b^2} \tag{4.1}$$

which would generally be considered a smooth function. It has a peak at $x = \frac{1}{2}$, of halfwidth b. We divide the interval $(0,1)$ into $K = L + 1$ equal intervals, each of length $1/K$, and make our "given" function values be

$$x_k = k/K \qquad \phi_k = f(x_k) \qquad k = 0, 1, 2, \cdots, K. \tag{4.2}$$

The question arises to what accuracy an interpolation method, based on these given values, reproduces the function values (4.1) for $x \neq x_k$.

In Figure 1, we show the result of polynomial interpolation, for $b = 0.1$ and $K = 10$. The interpolation polynomial fits the given values, as it must, and it approximates closely to the function in the immediate neighbourhood of the central peak. But away from this peak, the interpolation polynomial is nowhere near the function. Taking more given points, say $K = 20$, makes the maximum discrepancy worse, not better.

This does not contradict the theorem of Weierstrass [2] that every function, subject only to very mild conditions, can be approximated to any accuracy by a polynomial of sufficiently high order. In spite of frequent statements to the contrary, the Weierstrass theorem is not suitable as a basis for interpolation theory. The interpolation polynomials actually used are the polynomials of *lowest order* which fit the given function values. By assumption, we do not know the other function values to start with. Hence it is of no help at all to be informed that a polynomial of much higher degree could be drawn so as to fit $f(x)$ everywhere to high accuracy. So what? How do we find that superior polynomial from the given information?

One may assert that the interpolation polynomial of lowest order passing through all the given points is *by definition* the "smoothest curve through the points". This is possible logically, but we would then be forced to assert that the solid curve of Figure 1, rather than the dashed curve, is the "smoothest" fit to the given points.

The concept of minimum nastiness allows an alternative procedure: the "smoothest" function through a given set of points is now taken to be the

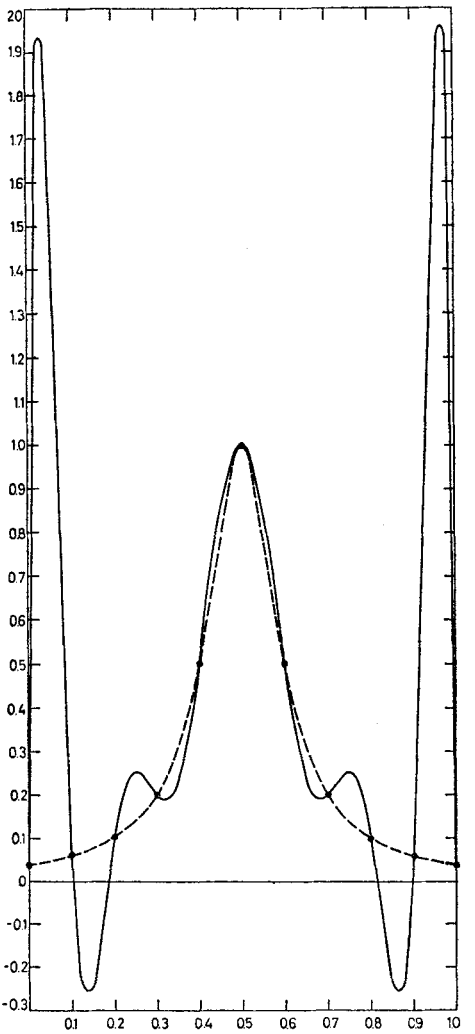

Figure 1. The dashed curve is function (4.1) with $b = 0.1$. The solid curve is the interpolation polynomial through the given points.

function of minimum nastiness through these points. In Figure 2, we show the minimum nastiness curve through the same points as in Figure 1, with the choice $q = 0.05$. It is apparent that minimum nastiness interpolation is preferable to polynomial interpolation for the function (4.1). In fact, it was impossible to separate the solid and dashed curves on this Figure.

Since the definition of nastiness contains a parameter q, we must make a choice regarding the value of q. The discussion at the end of Section 3 has shown that very large q and very small q are unsuitable. In the remainder of this paper

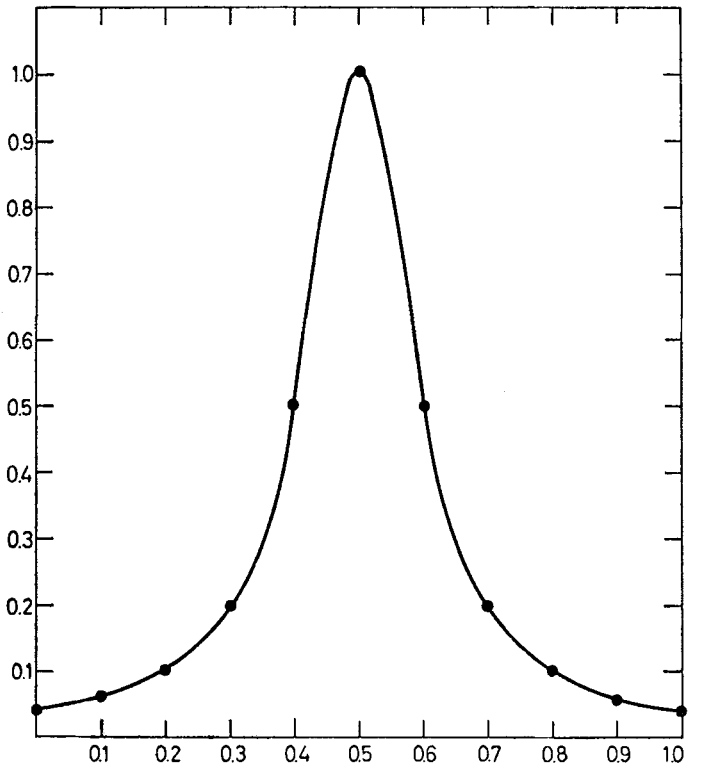

Figure 2. Minimum nastiness interpolation through the same points as in Figure 4.1. The highest deviation from function (4.1) is 0.008 at $x = 4.5$ and $x = 5.5$.

we shall restrict ourselves to equally spaced abscissae, i.e., the given x-values are always taken to be([3])

$$x_k = k/K \qquad k = 0, 1, 2, \cdots, K. \tag{4.3}$$

Large q then means $q \gg 1/K$, small q means $q \ll 1/K$, and both these choices lead to poor results. We have experimented with a number of functions $f(x)$, a number of interval sizes $h = 1/K$, and with various choices of q. The best minimum nastiness interpolation curves are obtained in the range

$$1/5K < q < 1/K \tag{4.4}$$

([3]) Our considerations retain their usefulness even if the points are not exactly equally spaced. But if there are really major differences in spacing (e.g., almost all the given points are crowded into a small part of the interval (0,1)), *no* constant value of the parameter q in (2.1) leads to satisfactory results. In such a case, the constant q in (2.1) should be replaced by a suitable function $q(x)$, such that $q(x)$ is proportional to the average spacing of the given points in the neighbohrhood of x. Unfortunately, the minimum nastiness problem with this generalized nastiness is very much harder to solve.

with a good compromise choice being

$$q = 1/2K. \tag{4.5}$$

It is of interest, perhaps, that the value (4.5) of q is the smallest value of q with the property: the Taylor series expansions around the various given points x_k, between them, define the interpolation function $\phi(x)$ on the whole interval $(0, 1)$, without the use of analytic continuation. If q falls below $1/2K$, the region in the middle between any two given points may not be reached by the Taylor expansion around either one of these given points.

Although examples exist, e.g., the function (4.1), for which minimum nastiness interpolation is distinctly superior to polynomial interpolation, we do not suggest the use of minimum nastiness interpolation as a practical interpolation method in the usual sense, of interpolation in a table of function values. For example, take the function $\exp(-5x)$ instead of the function (4.1). With $K = 20$ (21 given function values at equally spaced x_k), the maximum error from the interpolation polynomial is 7.6×10^{-17}. The maximum error of the minimum nastiness interpolation, even for the best q, is 2.5×10^{-3}; and this disappointing result is achieved after very much more labour.

5. **Minimum nastiness curve fitting.** In curve fitting, unlike interpolation, the "given" points are not specified exactly but have errors attached. We shall denote the given values by y_k, their errors by e_k. The abscissae are by assumption equally spaced, see (4.3).

Let $\phi(x)$ be a "fitted curve", and let

$$\phi_k = \phi(x_k) \qquad k = 0, 1, 2, \cdots, K. \tag{5.1}$$

The chi-square value of the fit is then defined by

$$X = \sum_{k=0}^{K} \frac{(\phi_k - y_k)^2}{2e_k^2} \tag{5.2}$$

The usual maximum likelihood fit is obtained by minimizing this X. This procedure is of no interest to us, since for *any* set of values y_k we can produce a minimum nastiness curve which has $\phi_k = y_k$ exactly, and hence has $X = 0$.([4])

We therefore generalize the criterion for a "good fit" by including the nastiness $N(\phi)$ explicitly into the criterion. From now on, we shall use $q = 1/2K$ for the parameter q, without writing this down every time. We let $F(N)$ be some (not yet specified) monotonically decreasing function of N, and we take as our criterion of good fit the quantity

(4) A similar situation occurs in conventional curve fitting procedures. For example, we can make $X = 0$ in polynomial curve fitting by choosing a polynomial of sufficiently high order. Procedures are then required for deciding what is the maximum degree of the fitting polynomial warranted by the data.

$$Q = F(N)\exp(-X). \tag{5.3}$$

In this section, we shall see what conclusions can be drawn from (5.3) without specifying the function $F(N)$ any further. It turns out that these conclusions are surprisingly definite. The specification of a reasonable choice of $F(N)$ is left to section 6.

We maximize (5.3) in two steps: (1) Assume given ϕ_k at $x = x_k$, and vary $\phi(x)$ for $x \neq x_k$. (2) Vary the ϕ_k. Step (1) is identical with the minimum nastiness interpolation problem of Section 4, since given ϕ_k means given X, hence maximizing Q, (5.3), means minimizing N. Thus the solution of step 1 is the minimum nastiness interpolation curve through the points (x_k, ϕ_k). The nastiness of this curve is given by (3.16) as a quadratic form in the ϕ_k.

We now turn to step 2, variation of the ϕ_k. We take logarithm of (5.3) and differentiate with respect to ϕ_k. Let

$$G(N) = -\frac{d\ln F(N)}{dN} = -\frac{F'(N)}{F(N)} \tag{5.4}$$

where the minus sign has been introduced to make $G(N)$ a positive quantity. The minimization conditions can be written in the form

$$\phi_k + g(e_k)^2 \sum_{m=0}^{K} C_{km}\phi_m = y_k \qquad k = 0, 1, 2, \cdots, K \tag{5.5}$$

$$g = G\left(\frac{1}{2}\sum_{k,m=0}^{K} \phi_k C_{km}\phi_m\right). \tag{5.6}$$

This set of equations is non-linear in the unknowns ϕ_k and g, and is not even defined until we know what the function $G(N)$ is.

Nonetheless, we can obtain a surprisingly large amount of information from this system of equations as it stands, simply by ignoring Equation (5.6) and treating g as a positive parameter whose value is at our disposal. Once g is given, equation (5.5) is a set of $K+1$ *linear* equations in the $K+1$ unknown values ϕ_k, with a unique solution. Thus, *without specifying the function $F(N)$ in* (5.3), *we are led to a mere one-parameter family of minimum nastiness fitting curves $\phi(x)$ to the given points.*

One procedure, by no means unreasonable, is then: construct some typical members of this one-parameter family, and decide by inspection which of them you prefer as the "best fit" to the points. Though obviously a subjective criterion, it is surprisingly effective: most people arrive at rather similar conclusions.

A second procedure is to compute the chi-square value X for each of the fitting curves obtained, and to settle on that value of g, and hence that fitting curve $\phi(x)$, which leads to the highest still acceptable chi-square. This procedure can be

mechanized once the highest acceptable chi-square has been specified by the investigator.

A third procedure, discussed in Section 6, is to specify the function $F(N)$ in detail, and to solve the non-linear equations (5.5), (5.6) numerically.

Whichever procedure is used, however, the final fitting curve $\phi(x)$ is going to be one of the one-parameter family obtained from (5.5), and this family can be discussed in its own right. We now proceed to do so.

The parameter g can range from 0 to infinity. When $g = 0$, Equation (5.5) is solved by $\phi_k = y_k$, all k. Thus $g = 0$ yields the minimum nastiness interpolation curve to the given values y_k, without taking into account that these points have errors e_k. The nastiness of this extreme solution is given by

$$N_0 = \frac{1}{2} \sum_{k,m=0}^{K} y_k C_{km} y_m. \tag{5.7}$$

This is the maximum value of N for the entire family of fitting curves: as soon as we take into account that the points have errors, we are allowed to go to curves of lower nastiness than N_0.

Since the set of Equations (5.5) is continuous in the parameter g, small positive values of g lead to solutions in the neighbourhood of this extreme solution, i.e., to values of ϕ_k rather close to the y_k, though no longer precisely equal to the y_k.

The opposite extreme, very large g, clearly weights the nastiness very heavily at the expense of the chi-square criterion. A formal mathematical discussion tends to become awkward since the matrix C_{km} cannot be inverted.(5)

However, in any special case the actual solution for large g can be found easily enough numerically, and we have found such solutions for many cases. They all fall into one pattern, and we therefore *conjecture* that this pattern is universal: to wit, all ϕ_k are approximately equal to a common value ϕ, and the fitting curve $\phi(x)$ is the horizontal straight line $\phi(x) = \phi$. The constant ϕ adjusts itself to minimize chi-square, so that ϕ is just the conventional weighted mean

$$\phi = \sum_{k=0}^{K} w_k y_k \tag{5.8}$$

with the weights

---

(5) Direct calculation, based on the definition (3.17), shows that the following identity holds:

$$\sum_{m=0}^{L} C_{km} = 0, \quad \text{for all } k.$$

Thus the matrix C_{km} has an eigenvector with eigenvalue 0.

$$w_k = \frac{(e_k)^{-2}}{\sum_{m=0}^{K} (e_m)^{-2}}. \tag{5.9}$$

The resulting nastiness is very close to zero, but the chi-square value is large and is given by

$$X = \frac{1}{2}\left(\overline{y^2} - (\bar{y})^2\right) \sum_{m=0}^{K} (e_m)^{-2}, \tag{5.10}$$

where

$$\overline{y^m} = \sum_{k=0}^{K} w_k (y_k)^m. \tag{5.11}$$

If this value of chi-square is considered acceptable, then the given points are consistent with a horizontal straight line, and this horizontal straight line also has minimum nastiness, namely $N = 0$. If, on the other hand, this value of chi-square is considered excessive, the best fitting curve must be obtained from lower values of the parameter g, for which the ϕ_k tend to approximate more closely to the given values y_k.

To illustrate the sort of curves one gets, we present a series of curves in Figure 3. There are 7 given points, 6 of which have values $y_k = 0$ with errors $e_k = 0.1$. The middle point, however, is "out of line", having $y_3 = 1$ and $e_3 = 0.3$. The 6 curves shown are all members of the one-parameter family defined by (5.5), with different values of the parameter g (g is listed next to each curve). We see

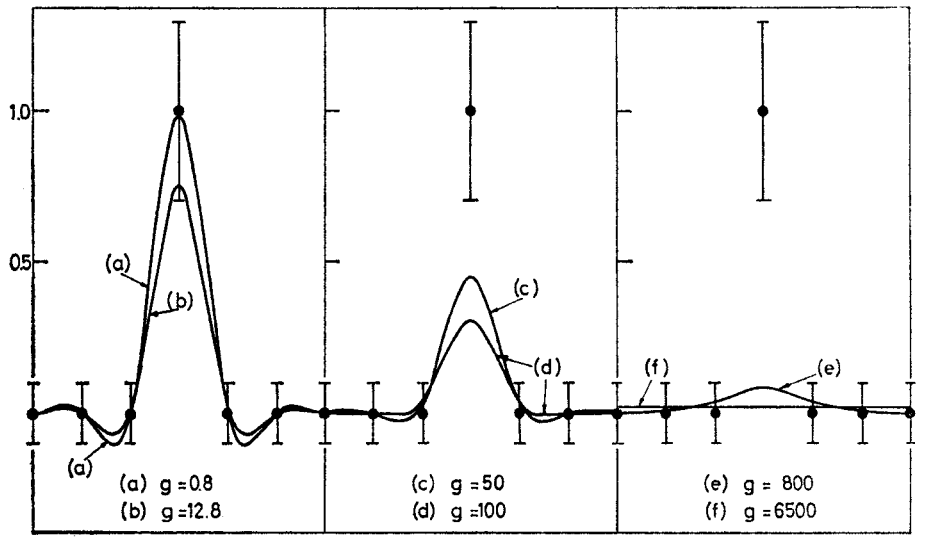

Figure 3. Minimum nastiness curve fitting for various values of g.

how we progress from a tight fit to the points, for small g, to the horizontal straight line (5.8) for very large g.

6. **A curve fitting algorithm.** The actual choice of the parameter g in the one-parameter family of curves obtained from (5.5) can be made by inspection of the results, by an a priori choice of a largest acceptable chi-square deviation criterion, or finally, by a specific choice of the function $F(N)$ in (5.3). In this section, we discuss this last alternative.

Looking at the curves of Figure 3, we note that it is most unlikely that we would wish to opt for an "intermediate" fit such as curves (c) or (d). Among the given points, one point is out of line. Either we believe that this deviation is real, in which case we opt for a curve such as (a) or (b), or we believe that the deviation is spurious, in which case we opt for a curve such as (e) or (f).

It is possible to find a family of functions $F(N)$ such that the criterion of goodness of fit, the quantity Q of Equation (5.3), has *two* local maxima, one corresponding to "belief", the other to "scepticism", regarding the one point that is out of line. By adjusting one parameter which, in essence, measures the extent of our scepticism, one or the other of these maxima becomes the absolute maximum, and hence the chosen solution.

The family which we have chosen to investigate in some detail is

$$F(N) = \left(1 + r\frac{N}{N_0}\right)^{-t} \tag{6.1}$$

where r and t are positive parameters. N_0 is the maximum nastiness (5.7) of the family of fitting curves.

With this choice of $F(N)$, the logarithmic derivative $G(N)$, (5.4), becomes

$$G(N) = \frac{rt}{N_0 + rN} \tag{6.2}$$

and we are now interested in simultaneous solution of the set of Equations (5.5), (5.6) for g and all the ϕ_k, with this choice for the function G in (5.6).

These equations can be solved on a computer by simple iteration: we choose some trial value N for the nastiness. This yields a trial value g through (5.6), $g = G(N)$. Insert this trial g into (5.5) and solve (5.5) for the ϕ_k. Insert these ϕ_k into the argument of the function G in (5.6). This yields a new value of g. Keep going until the process converges.

Our experience indicates that this simple iteration converges rapidly in most cases. The resulting solution gives values of g and the ϕ_k which satisfy (5.5) and (5.6), i.e., which correspond to at least a local maximum of the criterion (5.3). If all initial values of N, in the whole range from $N = 0$ to $N = N_0$, lead to one and the same final solution, then this solution is the true overall maximum of Q.

It sometimes happens, however, that we are led to one local maximum by starting from small N, and to quite a different local maximum by starting from N near N_0. The solution with small nastiness N is the sceptic's solution, such as (e) or (f) in Figure 3. The solution with nastiness N near N_0 is the believer's solution, such as (a) or (b) in the Figure. In such a case, the computer is programmed to determine the actual value of the goodness of fit criterion Q for both solutions, and to select the solution with the larger Q.

It is instructive to estimate in a rough manner the actual values of Q for these two types of solution. The sceptic's solution has N near 0, and thus $F(N)$, equation (6.1), near unity. The chi-square value X is large however, given by equation (5.10) to a first approximation. Let us denote this chi-square value by X_0. It is of course the largest value of X among all the fitting curves. We obtain the estimate

$$Q \cong \exp(-X_0) \quad \text{for the sceptic's solution.} \tag{6.3}$$

The other extreme, the believer's solution, has points ϕ_k close to the given points y_k, so that the chi-square deviation is small. On the other hand, the nastiness of the fitting curve is now close to its maximum possible value, N_0, so that we obtain the rough estimate

$$Q \cong (1+r)^{-t} \quad \text{for the believer's solution.} \tag{6.4}$$

We must now make a choice, just when we wish to switch from belief to scepticism, or vice versa. Let us suppose that we are prepared to accept a chi-square value X_1, but no higher value of X. We would then opt for the sceptic's solution in case $X_0 < X_1$, and we would opt for the believer's solution if $X_0 > X_1$.

We can achieve this result automatically by choosing the parameter r in such a way that the estimates (6.3) and (6.4) become equal when $X_0 = X_1$. That is, *if the degree of our scepticism is measured by a maximum acceptable chi-square value equal to* X_1, *then we should choose the parameter* r *in* (6.1) *to equal*

$$r = \exp(X_1/t) - 1. \tag{6.5}$$

In this way, we have used the parameter r to measure the extent of our scepticism, and we remain with only one free parameter, t.

We have made numerical experiments with t-values of $\frac{1}{2}$, 1 2, and 4, with the following results: (1) If t is large, say $t = 4$ and to a considerable extent already for $t = 2$, the resulting choice of best fit is "indecisive". Instead of opting in a clearcut way for either the sceptic's solution or the believer's solution, we find the iteration converging onto some intermediate solution which is an uneasy,

timid compromise satisfying nobody. (2) If t is small, say $t = \frac{1}{2}$, the process is decisive enough, with a sharp and clear crossover between nearly complete belief and nearly complete scepticism. But for too small values of t, the believer solution becomes "ultra-orthodox": the differences $\phi_k - y_k$ between fitted values and given function values turn out to be much less than the errors e_k.

On the whole, choices of t between $\frac{1}{2}$ and 1 lead to the most reasonable results. Since $t = 1$ leads to such a very simple functional form, we adopt $t = 1$ as our compromise best choice. With this choice, *the criterion of goodness of fit is the quantity*

$$Q = \frac{\exp(-X)}{1 + rN/N_0} \tag{6.6}$$

where X is the usual chi-square deviation, N_0 is given by (5.7), *and r is related to the maximum acceptable chi-square deviation, X_1, through*

$$r = \exp(X_1) - 1. \tag{6.7}$$

We emphasize that once X_1 has been chosen (corresponding to choice of a "confidence level" in conventional curve fitting procedures) there exist no further parameters for minimum nastiness curve fitting. The criterion (6.6) is quite definite and selects one unique best fit for any given set of points with errors.

A great deal of additional work is required to bring minimum nastiness curve fitting to a level of technical elaboration comparable to the state of the art in polynomial curve fitting. We hope, however, that the present paper has demonstrated that minimum nastiness curve fitting is worth further investigation.

As one example of possible further work, we mention the application to the smoothing of histograms: a histogram defines successive areas under the true function. We may associate the midpoints of the intervals in the histogram with values Y_k corresponding to these areas, and obtain a minimum nastiness curve fit to these values. The resulting fitting curve is the integral of the desired function. Since the fitting functions in minimum nastiness curve fitting are analytic in a region surrounding the given interval of the x-axis, we can differentiate explicitly to obtain the desired function.

Another application of minimum nastiness curve fitting may lie in the area of multi-dimensional scaling in psychology. [3]. There we need a smooth function relating "dissimilarities" to "distances" in an abstract space. The concept of nastiness may allow us to put the concept of "smoothness" of a function into more precise mathematical form, suitable for such applications.

One of us (J.M.B.) is grateful to the John F. Kennedy Memorial Foundation for the Weizmann Institute of Science for the grant of a John F. Kennedy Senior Fellowship, during the tenure of which this work was done.

REFERENCES

1. K. F. GAUSS, *Theoria combinationis observationum erroribus minimus obnoxiae*, Werke, Band IV, p.1, Göttingen 1823.

2. K. T. W. Weierstrass, *Über die analytische Darstellbarkeit sogenannter willkurlicher Funktionen reeller Argumente.* Sitzungsber. K. Akad. Wiss. Berlin, 1885. pp. 633–639, 789–805.

3. R. N. Shepard, *The analysis of proximities: Multidimensional scaling with an unknown distance function*, Psychometrica, **27**, 125–139, 219–246. 1962.

4. J. B. Kruskal, *Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis*, Psychometrica, **29**, 1, 1964.

WEIZMANN INSTITUTE, REHOVOTH,
AND
THE HEBREW UNIVERSITY OF JERUSALEM