James Cook Mathematical Notes

Volume 6, Issue Number 62
September, 1993

Editor and publisher: B. C. Rennie
Address: 66, Hallett Road,
Burnside, SA 5066
Australia.

The James Cook Mathematical Notes is published in 3 issues per year, dated January, May and September. The history of JCMN is that the first issue appeared in September 1975, and others at irregular intervals, all the issues up to number 31 being produced and sent out by the Mathematics Department of the James Cook University of North Queensland, of which I was then the Professor. In October 1983 this arrangement was beginning to be unsatisfactory, and I changed to publishing the JCMN myself, having three issues per year printed in Singapore and posted from there. I then set a subscription price of 30 Singapore dollars per year. When in 1985 I changed to printing in Australia I kept the same price, for the Singapore dollar is a stable currency.

In October 1992 it had become clear that the paying of subscriptions by readers is an inefficient operation. Bank charges for changing currency and for international transfers, with postage, together absorb most of the initial input of money. Therefore we have abandoned subscriptions as from the beginning of 1993, issue number 60. To those who want to give something in return for the JCMN, I ask them to make a gift to an animal welfare society in their own country. The animals of the world will be grateful and so will I.

Contributors, please tell me if and how you would like your address printed.

JCMN 62, September 1993

CONTENTS

FERMAT'S LAST THEOREM

T. C. S. Tao (Princeton University)

Fermat's Last Theorem has finally been proved. The result was announced by Professor Andrew Wiles of Princeton, in the last of his seminars on algebraic number theory in Cambridge. Apparently he revealed at the very end of the lecture that the theorem followed as a consequence of his work on the Taniyama-Weil conjecture. The reaction was first a stunned silence, then applause.

Wiles actually proved this result:

Theorem If E is a semi-stable elliptic curve defined over Q (the field of rationals), then E is modular.

This implies Fermat's Last Theorem, for if $a^p + b^p = c^p$, then the elliptic curve $y^2 = x(x - a^p)(x + b^p)$ over Q is semi-stable, and hence by Wiles's result there must exist a modular form of weight 2 on $\Gamma_0(2)$. But the work of Frey and Ribet has shown that there are no such forms.

The proof of Wiles's theorem is long and technical, with the full length of the proof said to be more than 100 pages (significantly more than the margin of a book on arithmetic!), and beyond the scope of my understanding. Though the proof has not been rigorously checked yet, many eminent mathematicians in the field (many of whom were in the audience when Wiles announced his result) have found the argument sound and plausible. Further, the new methods used in Wiles's proof of his theorem

have promise of potential for application to many problems in algebraic number theory.

Rumor has it that Wiles first found a partial breakthrough seven years ago, which would have proved a significant portion of the Taniyama-Weil conjecture and perhaps also part of Fermat's Last Theorem, but worked in near-secret for the following years to complete and perfect the result.

For those not familiar with Fermat's Last Theorem, it can be stated as follows: for each integer $n > 2$, there exists no set of positive integers a, b, c such that $a^n + b^n = c^n$. It has been one of the best-known problems in mathematics. Pierre Fermat (1601-1665) who was notorious for writing mathematical statements without proofs (but which, with very few exceptions, have happened to be correct), scribbled a claim in the margin of his copy of Diophantus's *Arithmetika* in about 1637 "I have found a truly wonderful proof of this theorem, but the margin is too small to contain it". Of course he never came up with the proof, and many mathematicians now believe that the proof he thought of was flawed (perhaps assuming unique factorization over the cyclotomic integers), especially considering that he later devoted some effort to proving special cases of the result. Nevertheless, it is called Fermat's Last "Theorem" perhaps because it is the last claim of his to be rigorously proved or disproved. FLT has attracted the attention of many famous mathematicians from the seventeenth century up to the present, as well as countless other mathematicians and amateurs, and the number of flawed proofs of

FLT known is enormous, with many universities having to assign a mathematician to field all the FLT claims submitted.

The first significant progress on FLT was made by Kummer in the nineteenth century, using techniques which would later lead to a new field, that of algebraic number theory. Then the work of twentieth century mathematicians showed the link with elliptic curves, leading to another rich field of mathematics being discovered.

QUICK INEQUALITY

Terry Tao (Princeton)

If the imaginary part of z is between $\pm\ \pi/2$, prove:

$$\tanh|z| \le |\tanh z|.$$

If $|z| < \pi/2$, prove $|\tan z| \le \tan|z|$.

QUOTATION CORNER 43

I think I can safely say that nobody understands quantum mechanics.

— Richard Feynman

MARKOFF'S INEQUALITY (JCMN 60, p.6221 & 61, p.6254)

Terry Tao (Princeton)

Let $f(x)$ be a real polynomial of degree n such that:

$$-1 \le f(x) \le 1 \quad \text{when} \quad -1 \le x \le 1 \qquad \ldots\ldots \quad (C)$$

Then $|f'(x)| \le n^2$ for all x in the interval.

This has been known since 1889, but the proofs available are rather long. Here is a more direct approach.

Take any positive integer n.

Theorem 1 Considering all polynomials:

$$f(x) = a_0 + a_1x + a_2x^2 + \ldots + a_nx^n,$$

that satisfy (C), each coefficient is bounded.

Proof $f(\cos\theta)$ can be expressed as $\frac{1}{2}c_0 + \Sigma_{r=1}^{n} c_r\cos r\theta$, where $c_r = \frac{1}{\pi}\int_{-\pi}^{\pi} f(\cos\theta)\cos r\theta\, d\theta$, which is between ± 2. Therefore $f(x) = \frac{1}{2}c_0 + \Sigma c_rT_r(x)$, where T_r is the Chebychev polynomial defined by $T_r(\cos\theta) = \cos r\theta$. The typical coefficient in $f(x)$ is therefore bounded by

$$|a_s| \le 2\ \Sigma_{r=0}^{n}|\text{coefficient of } x^s \text{ in } T_r(x)|.$$

Theorem 2 There is $B = B(n)$ such that $|f'(x)| \le B$ for all x in the closed interval $[-1, 1]$, for all f satisfying (C). Also there is f for which the bound is attained.

Proof The existence and attainment of a bound for each f is trivial. This bound is a continuous function of the point $(a_0, a_1, a_2, \ldots a_n)$ in a bounded closed set in a space of $n+1$ dimensions.

We can now concentrate our attention on the polynomial that attains the bound B for $|f'(x)|$. Call it $p(x)$.

<u>Theorem 3</u> The values $p(1)$ and $p(-1)$ are ± 1.

<u>Proof</u> Suppose that $p(1)$ is strictly between -1 and 1. By continuity we can find $\epsilon > 0$ such that the inequality $-1 \le p(x) \le 1$ holds in the enlarged interval $[-1, 1+\epsilon]$. Then we can shorten the interval linearly, in fact put $g(x) = p(x+\epsilon(x+1)/2)$. This $g(x)$ is a polynomial of degree n satisfying condition (C), but $|g'(x)|$ takes a value $(1+\epsilon/2)B$ in the interval $[-1, 1]$, a contradiction.

The reasoning for $p(-1)$ is similar.

Let A and B be the set of zeros in the interval $[-1, 1]$ of $p(x)-1$ and $p(x)+1$ respectively, each of these sets has $\le n$ points (in the algebraic sense, counting two for a double zero). Of all these zeros no more than two are simple zeros, in fact only $x=-1$ and $x=1$ can be simple zeros. Therefore A and B together cannot contain more than $n+1$ distinct points. Therefore we can find a set S (of at most n points) separating A and B, in the sense that between any point of A and any point of B there is a point of S.

<u>Theorem 4</u> The set S has n points.

<u>Proof</u> Suppose not. Let v be where $p'(x)$ attains one of the values $\pm B$. If v is not in A or in B, we can find a set S′ as follows:

Case 1 The points of A and B alternate (i.e. between two points of A is one of B and vice versa). The point v is between some a' in A and b' in B. Delete the point of S that separates a' from b' and replace it by v taken twice.

Call this set S′ and let $h(x)$ be a polynomial with these points as zeros, then for sufficiently small (positive or negative) β the polynomial $p(x)+\beta h(x)$ lies between smaller bounds than ± 1 and has the same derivative at v as $p(x)$, a contradiction.

Case 2 There are two points of A not separated by a point of B (or vice versa). Then S has $\le n-2$ points. Add v repeated to give S′, and proceed as before.

Finally we must deal with the case when v is in A (v in B is similar). Clearly $v = \pm 1$. Take the nearest point of S to v, delete it and replace it by v repeated to give S′, then proceed as before.

<u>Theorem 5</u> The polynomial $p(x)$ is a Chebychev polynomial (possibly with the sign changed).

<u>Proof</u> The two polynomials $n^2(1-p(x)^2)$ and $p'(x)^2(1-x^2)$ are both of degree $2n$, and they have the same zeros (the points of A and B), and the coefficient of x^{2n} is the same in both. Equating them and solving the differential equation for $p(x)$ gives the result.

<u>Markoff's inequality</u> If $f(x)$ of degree n satisfies (C) then $|f'(x)| \le n^2$ in the interval $-1 \le x \le 1$.

<u>Proof</u> $|f'(x)| \le \max|p'(x)| \le \max|(d/dx)T_n(x)|$

$T_n(\cos\theta) = \cos n\theta$, and $T'(\cos\theta)\sin\theta = n\sin n\theta$,

$|T_n'(\cos\theta)| = n\left|\dfrac{\sin n\theta}{\sin\theta}\right| \le n^2$, and so $|f'(x)| \le n^2$.

<u>Footnote</u> Two applications of the result above give $|f''(x)| \le n^2(n-1)^2$, but this bound is not the best possible. V. A. Markoff (brother of A. A.) showed $|f''| \le n^2(n^2-1)/3$.

BOOK REVIEW

JOURNEY INTO GEOMETRIES by Marta Sved

(Math. Assoc. of America, 1991, xvi + 182 pages, 6½ × 9 ins.)

The book consists of two parts, with a Foreword by Prof. H. S. M. Coxeter. Part 1 after an Introduction has 6 chapters, with these interesting titles:

1. Going Round in Circles (Circle geometry)
2. Reflections on Inversion (Inversion shown equivalent to reflection)
3. Dr. Whatif's Euclidean Geometry (Geometry of circles concurrent at a point O that vanishes)
4. A Hyperbolic T-Party (Introduction to the Non-Euclidean Hyperbolic Plane in contrast to Spherical Geometry)
5. Circle-Land Revisited (Poincaré Model of the hyperbolic plane)
6. Into the Shadows (Introduction to the Projective Plane and the Finite Projective Plane)

Each chapter contains delightful diagrams and illuminating illustrations, and is developed by lively conversations of an Alice (a student) with Master Lewis Carroll (her mathematical teacher at school) and with Dr Whatif (Master of 'What ... if ...' mathematics, as her present teacher), followed by problems and exercises. In Chapter 5 there appear two Hostesses, one is White who encourages Alice in lessons by providing her with delicate opera logical (log-ical) glasses for closer and more intimate observation of figures, and the other is Red and shows

her red eyes in trying to test her learning.

Part 2 contains Axiom Systems and solutions to the problems in Part 1, followed by a list of books with which the reader can follow up the journey.

Unfortunately there occur a good many minor misprints, which the author is expected to correct in her book's 2nd Edition. Such an entertaining, illustrative and illuminating book as this *Journey into Geometries* would have added value if it had an Index.

As a sample of the delightful dialogues, the reviewer is tempted to quote one: Dr. Whatif turning to Alice says (page 9):

"With Axioms, my dear, you need a gentle touch.
They should not say too little, they should not say too much,
And on one point above all, we have to be insistent,
Though Axioms need not be true, their set must be
consistent."

Sahib Ram Mandan.

SYMMETRIC SIMULTANEOUS EQUATIONS

(JCMN 59, p.6173 & 60, pp.6192-6194)

Harry Alexiev

The previous contribution discussed the equations:

$$x^2 - yz = a \qquad y^2 - zx = b \qquad z^2 - xy = c,$$

noting that the case $(a, b, c) = (1, \omega, \omega^2)$ had not been covered (ω being a complex cube root of unity). Now we shall deal with this case. Firstly we need:-

<u>Lemma</u> The solution of

$$\alpha + \beta + \gamma = 0$$
$$\alpha^2 + \beta^2 + \gamma^2 = 0$$

is (α, β, γ) = either $(k, k\omega, k\omega^2)$ or $(k, k\omega^2, k\omega)$, for any complex k.

<u>Proof</u> Substituting for γ, we find $\alpha^2 + \alpha\beta + \beta^2 = 0$, i.e. $(\alpha - \omega\beta)(\alpha - \omega^2\beta) = 0$, hence the result.

Now we return to our problem:

$$x^2 - yz = 1, \qquad y^2 - zx = \omega, \qquad z^2 - xy = \omega^2.$$

Then, $x^2 + y^2 + z^2 - xy - yz - zx = 1 + \omega + \omega^2 = 0.$

i.e. $(x-y)^2 + (y-z)^2 + (z-x)^2 = 0 = (x-y) + (y-z) + (z-x).$

By our lemma above: $x-y = k$, $y-z = k\omega$, $z-x = k\omega^2$ (or the second solution). Taking this first solution, note that $k = 0$ gives $x = y = z$, which is not allowable. Putting $x = y+k$ and $z = y-k\omega$, and using $x^2 - yz = 1$, we find $2yk + k^2 + yk\omega = 1$. This gives y, and so also x and z.

$$x = \frac{1 - k^2\omega^2}{k(2 + \omega)}, \qquad y = \frac{1 - k^2}{k(2 + \omega)}, \qquad z = \frac{1 - k^2\omega}{k(2 + \omega)}$$

Putting $k = (1-\omega)/p$, these take a simpler form:

$$x = p/3 + 1/p, \qquad y = p/3 + \omega/p, \qquad z = p/3 + \omega^2/p$$

where p is any non-zero complex number.

The second alternative given by the lemma (with $y-z = k\omega^2$ and $z-x = k\omega$) does not lead to a solution.

MORE SIMULTANEOUS EQUATIONS

Jordan Tabov
(Mathematical Institute, Bulgarian Academy of Sciences)

I met another interesting example (but not quite symmetric)

$$\frac{x - y\sqrt{x^2-y^2}}{\sqrt{1-x^2+y^2}} = a, \qquad \frac{y - x\sqrt{x^2-y^2}}{\sqrt{1-x^2+y^2}} = b.$$

Because $a^2 - b^2 = x^2 - y^2$, the answer is seen to be

$$x = \frac{a + b\sqrt{a^2-b^2}}{\sqrt{1-a^2+b^2}}, \qquad y = \frac{b + a\sqrt{a^2-b^2}}{\sqrt{1-a^2+b^2}}$$

If the original equation is $\varphi(x, y) = (a, -b)$ then the solution is $\varphi(a, -b) = (x, y)$.

These equations are related to the Lorentz transformation

$$x' = \frac{x - vt}{\sqrt{1 - v^2}}, \qquad t' = \frac{t - vx}{\sqrt{1 - v^2}}$$

in which x^2-t^2 is invariant $(= x'^2-t'^2)$. The inverse is

$$x = \frac{x' + vt'}{\sqrt{1 - v^2}}, \qquad t = \frac{t' + vx'}{\sqrt{1 - v^2}}$$

More generally, consider the simultaneous equations

$$\frac{x - yf(x^2-y^2)}{\sqrt{1-f^2(x^2-y^2)}} = a, \qquad \frac{y - xf(x^2-y^2)}{\sqrt{1-f^2(x^2-y^2)}} = b.$$

where f is any function with $|f| < 1$.

The solution (as before) can be obtained by the change $(x, y) \longleftrightarrow (a, -b)$. In fact it is:-

$$x = \frac{a + bf(a^2-b^2)}{\sqrt{1-f^2(a^2-b^2)}} \qquad y = \frac{b + af(a^2-b^2)}{\sqrt{1-f^2(a^2-b^2)}}.$$

SIMULTANEOUS SYMMETRIC EQUATIONS

In the Trinity College, Clare College and Trinity Hall Entrance Scholarship paper on the morning of Thursday, November 9th, 1899, in Cambridge, question 3 was as follows:-

3. If
$$y^3 - z^3 = ayz,$$
$$z^3 - x^3 = azx,$$
$$x^3 - y^3 = axy,$$
show that
$$x^3 + y^3 + z^3 = 3xyz.$$

In those days one would assume that the variables in such a problem were complex numbers, for fields and rings were certainly not taught to schoolboys, and in a sense had not yet been invented (van der Waerden's *Moderne Algebra* had not been written). On this assumption the proposition is untrue, as may be shown by the example where $x = y = 1$, $z = \exp(2\pi i/3)$ and $a = 0$.

Suppose we were to add the condition $a \neq 0$, would the proposition become true? Alternatively, suppose we took the variables all to be real, would the proposition then be true? More generally —— what was in the mind of the examiners?

If we took an arbitrary $a \neq 0$, then would the set of equations have a solution (x, y, z) in complex numbers, other than the trivial $x = y = z = 0$?

THE MASS OF MERCURY

Richard L. Branham, Jr.
(Centro Regional de Investigaciones Cientificas y Tecnologicas
Mendoza, Argentina)

If a planet has a satellite it is easy to find the planet's mass, Kepler's third law states that, if P is the period of revolution of the satellite round the planet, and d the distance of the satellite from the center of the planet (more precisely, half the sum of the maximum and minimum distances), then the sum of the masses m and M of the satellite and planet is given by

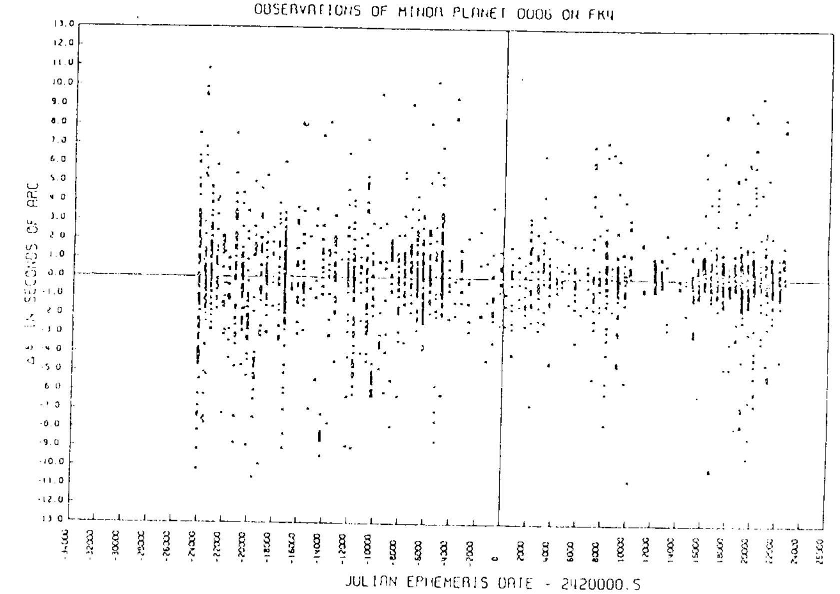
$$m + M = \frac{4\pi^2 d^3}{GP^2} \qquad \text{........... (1)}$$

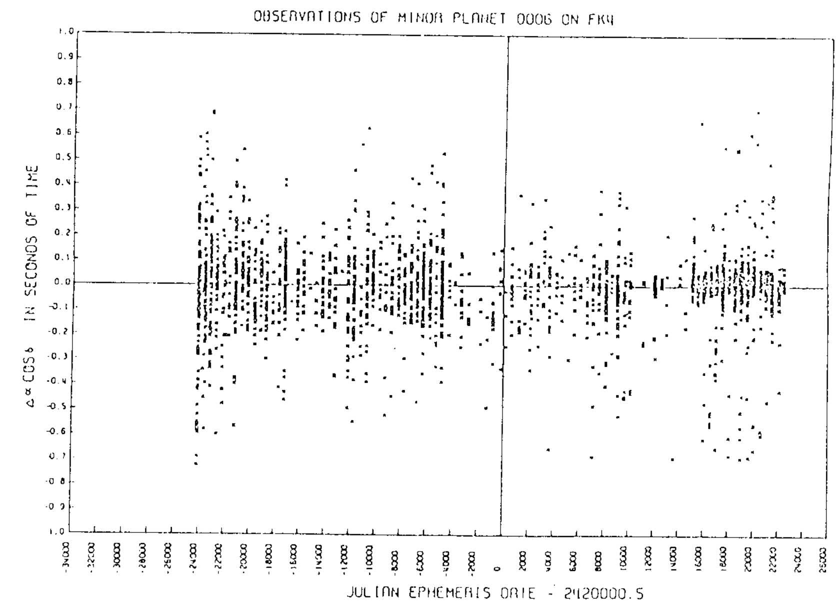
where G is Newton's gravitational constant. To find the sum of the masses we merely need to measure d and P.

But for a planet such as Mercury or Venus with no satellite, one must use indirect techniques. One studies the perturbations by the planet under study of another object, to infer the mass of the planet. In effect one uses a mathematical model for the solar system to calculate theoretical positions of the object perturbed, compares the actual observed positions with the theoretical positions, and uses the differences, called (O-C)'s, to correct for the mass of the perturbing body. One linearizes the non-linear equations of condition by a Taylor series expansion, truncated at the first order. This technique, called a differential correction by astronomers and the Gauss-Newton method for non-linear regression by statisticians, is discussed in most books on data reduction.

Consider Mercury's mass. The planet closest to Mercury, Venus, will be the most influenced by Mercury's mass. We calculate Venus's (O-C)'s and solve for Mercury's mass and perhaps other unknowns. Having the observations of Venus and a standard mathematical technique, what could go wrong?

Plenty. The method used to solve the equations in a differential correction is usually least squares (OLS), which given certain assumptions calculates an optimum solution. These assumptions are: 1) all of the experimental error resides in the observations themselves, whereas the equations of condition are error-free; 2) the observational errors are uncorrelated; 3) the observational error follows the normal distribution; and 4) the observational error exhibits no systematic trends with time (an assumption known as "homocedasticity"). Many of these assumptions, unfortunately, are often violated. Take the assumption of homocedasticity, the figure (opposite) shows (O-C)'s in right ascension and declination for observations made of the asteroid Hebe between 1847 and 1975. Modern observations are more precise than their nineteenth century counterparts; the (O-C)'s are heterocedastic. Observations of solar system objects, moreover, rarely lead to errors following the normal distribution, and are frequently correlated. Nor does one have to look far to find instances where the equations of condition as well as the observations incorporate error. When the latter situation prevails, to solve the mathematical system with OLS can lead to a biased solution; the correct procedure employs total least squares (TLS), which allows for error in the equations themselves. In short, on occasion we find in astronomy situations where none of





the assumptions allowing us to infer the optimality of an OLS solution is obeyed.

This is the probable reason for the fiascos such as Pluto's mass. In the year 1978 the official International Astronomical Union (IAU) mass for that planet was 1/3,000,000, taking the solar mass as unity, with a formal mean error of 20%. But that very year Pluto's satellite Charon was discovered. Equation (1) showed that the real mass is 1/130,000,000, meaning that the official mass was in error not by 20%, but by 4000%.

What about Mercury? The official IAU mass for this planet, based on numerous determinations including the Mariner flybys, is 1/6,023,600 ± 250. But if this mass is correct, then the density of the planet is anomalously high with respect to the other terrestrial planets and the Moon. Because the mass determinations have, with only one exception that I can find, been based on differential corrections, is there a chance that something has gone amiss and that the real mass, which if it were 1/9,000,000 would give concordant densities for all the terrestrial planets and the Moon, has been overestimated?

If we look at just optical observations, the possibility cannot be lightly dismissed. If the real mass is 1/9,000,000, but we use 1/6,000,000 in the mathematical model, a quick back-of-the-envelope calculation shows that the difference in the computed position of Venus could reach 0·3" of arc. This seems large, but observations of Venus have a mean error of about 1". A good routine for spectral analysis, however, should be able to

detect a 0·3" signature IF THE DATA DO NOT SUFFER FROM STRONG SYSTEMATIC ERRORS. Venus is observed in the daytime with a transit circle. Many corrections must be applied to the observations. One of these is the day-night correction. A transit circle defines a celestial reference system largely by nighttime observations of stars. Daytime observations must be reduced to the nighttime system. Venus's image is generally a crescent, and the observations must be reduced to the centre of the disc. And so forth. Many of these corrections are difficult to determine. As the upshot, systematic errors in the observations may easily swamp a 0·3" signature.

What saves the situation, and for me makes a mass much different from 1/6,000,000 for Mercury unsustainable, is radar. Venus is easily reached by radar, which has a formal precision of about 0·004" for a planet at Venus's distance. Although radar determines the distance much better than angular elements, a signature of 0·3" will be easily detected.

I have calculated solutions for Mercury's mass by techniques for nonlinear regression that do not linearize the equations of condition. When only optical data are used, it is possible to get a mass as small as 1/8,970,000. But when radar is incorporated the mass becomes close to the official IAU value. I therefore harbor little doubt as to the real mass for the planet Mercury: the mass is close to 1/6,000,000, and the planet is indeed anomalously dense.

FITTING

John Parker

(Oak Tree Cottage, Reading Road, Padworth Common, RG74QN)

In statistics we often want to fit a "best possible" line to a set of points in the plane; and in navigation there is the dual problem of fitting a "best possible" position to a set of Sumner lines (position lines) on the chart. Both these problems have generalizations in spaces of more than 2 dimensions.

The orthodox (and easiest) way to tackle these questions is by the "method of least squares", that is by choosing our point (or line or whatever) to minimize the sum of squares of the relevant perpendicular distances (or more generally the sum of squares of other quantities that measure the discrepancies between the proposed answer and the data). But there may be advantages in using the modulus instead of the square. It may be noted that Kendall and Stuart in their *Advanced Theory of Statistics* (Vol 2, p.286) discuss this question, saying that the choice of a method of fitting is essentially arbitrary, and that least squares is preferable because of the ease of calculation. However, in 1961 (when their book was published) computing was not as easy as it is now.

This will be a story about two statisticians, one, whom we shall call LS, uses the least squares method, and the other, MP, obtains his answers by minimizing the sum of perpendicular distances in the geometrical cases, or more generally by minimizing what in the jargon of functional analysis is called the L_1 norm.

<u>The case of one dimension</u> (which when examined carefully is not as simple as one might have hoped) Take the simplest case. Suppose that there is an unknown parameter, and for

it we have three estimates, x, y and z; these are samples from the random variables X, Y and Z. With (x, y, z) as data LS will say that the best conclusion to come to about the unknown parameter is to take it to be the mean, (x+y+z)/3, but MP will say that the best answer is the median, i.e. the middle one of the three values x, y and z when they are arranged in increasing order. Which of the two should we trust? LS or MP? An unanswerable question!!

In real life (though not when answering a question in an examination paper in mathematical statistics) we have very little idea about the random variables X, Y and Z, in other words we know very little about the accuracy (or, more precisely, the probability distribution of the errors) of our data.

Even when we know that the three random variables X, Y and Z all have the same distribution, it is still difficult to choose between LS and MP. If X, Y and Z all have the same Gaussian distribution then LS usually (in fact with probability 59%) gives a better answer than MP (i.e. an answer nearer to the true value). If, however, X, Y and Z all have the same Cauchy distribution then (as any good text-book explains) the mean put forward by LS will have the same Cauchy distribution, and MP's median will probably be a better answer (in fact it has probability 64% of being better).

If some of the three random variables are more accurate than others (but the analyst does not know which ones), then the median tends to perform better as a predictor of the true value, as we shall see now.

The table below (calculated by simulation, accuracy no more than 1%) gives the probability that the mean of three data values is nearer than the median to the true value of the unknown; it is calculated for various choices of the three random variables X, Y and Z. In each of the ten cases X, Y and Z are of the same type (given in the first column), and are all unbiassed, i.e. having mean equal to the true value of the unknown. Values in the second column are for X, Y and Z being identically distributed, and in the third column are for X as before but Y having twice the error of X, and Z three times, (formally stated, if t is the true value, then Y-t has the same distribution as 2X-2t, etc.). The five types of distribution considered are as follows; in each case we give a probability density f(x) with mean zero, the "probability distribution" F(x) being defined by

$$F(x) = \int_{-\infty}^{x} f(t)\,dt.$$

The variance is defined as $\int_{-\infty}^{\infty} x^2 f(x)\,dx$, and the mean error as $\int_{-\infty}^{\infty} |x| f(x)\,dx$.

In general, for three samples from any random variable, the probability density of the median is $6f(x)F(x)(1-F(x))$, and so the variance of the median is

$$6\int_{-\infty}^{\infty} x^2\, f(x)F(x)(1-F(x))dx$$

Now to define the 5 random variables that we shall consider.

<u>Uniform</u> $f(x) = \frac{1}{2}$ if $-1 < x < 1$, and zero otherwise. The variance is 1/3 and the mean error 1/2.

<u>Triangular</u> $f(x) = 1-|x|$ if $-1 < x < 1$ and zero otherwise. The variance is 1/6 and the mean error is 1/3.

<u>Gaussian</u> $f(x) = (2\pi)^{-1/2}\exp(-x^2/2)$. The variance is 1 and the mean error is $\sqrt{(2/\pi)} = 0{\cdot}7979$.

<u>Negative exponential</u> $f(x) = \frac{1}{2}\exp-|x|$. The variance is 2 and the mean error is 1. $F(x) = \frac{1}{2}\exp x$ if $x < 0$, and
$= 1 - \frac{1}{2}\exp-x$ if $x > 0$.

<u>Cauchy</u> $f(x) = 1/(\pi + \pi x^2)$. The variance and mean error are both infinite. $F(x) = \frac{1}{2} + (1/\pi)\tan^{-1}x$.

The five distributions are listed in this order because it is roughly the order of increasing fatness of tails; to measure this property it may be observed that the ratio (variance)/(square of mean error) takes the five values: 4/3, 3/2, $\pi/2$, 2, (indeterminate, but might be called ∞)

<u>Probability of the mean being nearer than the median</u>

Distributions of X, Y and Z	ratio of errors 1:1:1	ratio of errors 1:2:3
Uniform	70%	55%
Triangular	60%	48%
Gaussian (normal)	59%	48%
Negative exponential	47%	41%
Cauchy	36%	33%

The keen student may verify that the 70% in the top row of the table above is exact.

There are other ways of comparing two unbiassed estimators, for instance we may call one "better" than another if the estimates that it gives have a smaller variance than those given by the other; but this method has the awkward feature that three values (or indeed any other number) from a Cauchy distribution give a mean with infinite variance.

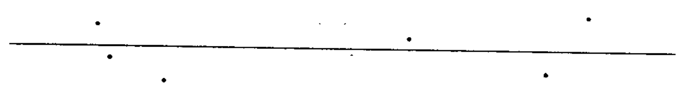
Ratio $\dfrac{\text{Variance of mean}}{\text{Variance of median}}$

Distributions	ratio of errors 1:1:1	ratio of errors 1:2:3
Uniform	$5/9 = 0{\cdot}556$	$80/93 = {\cdot}860$
Triangular	$140/207 = 0{\cdot}676$	$7840/7859 = 1{\cdot}00$
Gaussian	$\pi/(3\pi-3\sqrt{3}) = 0{\cdot}743$	$1{\cdot}07$
Negative exponential	$24/23 = 1{\cdot}04$	$1355200/936289 = 1{\cdot}447$
Cauchy	∞	∞

The bottom row is a little obscure. The mean and the median both have infinite variance, but one infinity is bigger than the other, in a sense. If instead of the three data values x, y and z, we had five or more, then the median would have finite variance.

<u>Two dimensions</u> Consider drawing a line to fit a set of points in the plane. The least squares solution from LS will be a line through the centroid, and in the language of mechanics it is the major principal axis of inertia of the set of points regarded as equal point masses. In general the line will not pass through any of the points.

The solution from MP, the line minimizing the sum of perpendicular distances, will in general be a line through two of the points. More precisely, there is always an optimal line, but it may not be unique, and every optimal line is through two points. This may be proved as follows.



Suppose that an optimal line, as in the picture above, does not pass through any of the points, then it may be moved parallel to itself without increasing the sum of the perpendicular distances, until it meets a point; therefore there is an optimal line through one point. Now take (if possible) an optimal line through only one point, and consider rotating it about that point. As long as the line does not meet any other point the sum of the perpendicular distances, as a function of the angle θ of rotation from the optimal position, is of the form $\Sigma\ c \sin|\theta-\alpha|$ which has second derivative negative and therefore cannot attain a minimum where $\theta = 0$. Therefore every optimal line is through two points.

For both LS and MP their algorithms lead in general to a unique answer, but there are exceptional cases. For example, if the problem is to draw the optimum line through the three vertices of an equilateral triangle, LS will give any line through the centroid, but MP's answer will be any one of the three sides.

<u>Fitting a linear function</u> A common problem (with some relation to the problem discussed above) is to fit the linear form $y = mx + c$ to a number of (inaccurate) observed y values, the x's being known. This was essentially the question discussed in MASS OF MERCURY (3) in JCMN 61, p.6240. Take a simple example of this problem.

Suppose that the data is (x, y, z), where x is an observation of f(0), y is an observation of f(1) and z is an observation of f(3). We want to estimate the slope m of the linear function $mt + c$ fitted to f(t). The least squares calculation leads to LS giving the estimate $(5z-y-4x)/14$ for m, and the estimate $(5x+3y-z)/7$ for c. However MP (who can get his answer just by drawing a line from the first to the third point on the graph) gives the estimates $(z-x)/3$ for m and x for c. As before, we ask what is the probability of the estimate for m from LS being nearer to the true value than that from MP.

If x, y and z are all normally distributed about their correct values, with equal variances, then both LS's estimate for m and MP's are normally distributed about the correct value, the former with a smaller variance (in the ratio 27/28). The probability of the least squares estimate being the nearer of the two to the right value is $1-(1/\pi)\text{arc tan } 6\sqrt{3} = 0{\cdot}53$.

As before we can calculate (by simulation) the probability of LS's answer being nearer than MP's to the true value, for different distributions of the random variables X, Y and Z, (now assumed the same).

Probability that LS is nearer than MP to the right answer for the slope m:-

Distribution	Prob	Distribution	Prob
Uniform	52%	Neg. Expon.	52%
Triangular	53%	Cauchy	50%
Gaussian	53%		

The lesson (or non-lesson) to be drawn from these figures is that our simple example tells us little about the general statistical questions of data analysis, because to find a slope from only three points there is not much choice available; no amount of statistical theory will take us far from the obvious choice of the slope from the first to the third point. What can be said about the problem with a large number of data points?

Higher dimensions In d dimensions, given n points, we want (if we follow the precepts of MP) to find the k-dimensional hyperspace (where $k < d < n$) that minimizes the sum of perpendicular distances from the given points. How many of the points must it contain? We saw above that the answer is 2 when $k = 1$, $d = 2$ and $n > 2$.

MORE FITTING (JCMN 62, p.6284 above)

The two statisticians (see FITTING above), LS who is devoted to least squares methods, and MP who in the geometrical problems minimises the sum of perpendicular distances, were of course in disagreement about what to do with measurements of a certain quantity given by five different experimental scientists. LS wanted to take the mean and MP wanted to take the median as the best possible conclusion from the data.

A third statistician, VM, believed in the middle way, the *via media*. He reasoned that with data arranged in order $x_1 < x_2 < x_3 < x_4 < x_5$, LS wanted $(x_1+x_2+x_3+x_4+x_5)/5$ but MP wanted x_3, so perhaps something like $(x_2 + 3x_3 + x_4)/5$ would avoid the excesses of LS who attached too much importance to the extreme values, as well as those of MP who gave too much weight to the middle value x_3 and discarded the information probably stored in its two neighbours.

If you have three horses and want to know the fastest of them, you should race them together a lot and either count which of them wins most often or calculate which has the best average speed. Let us do the same with the theories of these three statisticians. We set up random variables to give sets of five data values. For each trial we can pick out which of the three is the winner (i.e. gets nearer to the right answer than either of the other two), and we can find the error and squared error of each. The table below gives for each of the three statistician the probability of winning, and the expectation of error and the expectation of the square of the error (these last two entries normalised to sum = 1).

In our computer simulation the data for each trial come from similar distributions, with their errors in the ratio 1 : 2 : 3 : 4 : 5, because in the real world scientists are not equally accurate in their results.

Distribution	Test	LS	MP	VM
Uniform	Winning	30%	43%	27%
	Mean error	·39	·30	·31
	Mean square	·41	·30	·29
Triangular	Winning	36%	38%	26%
	Mean error	·36	·32	·32
	Mean square	·37	·33	·30
Gaussian	Winning	36%	38%	26%
	Mean error	·37	·32	·31
	Mean square	·39	·31	·30
Negative exponential	Winning	28%	43%	29%
	Mean error	·43	·28	·29
	Mean square	·50	·25	·25
Cauchy	Winning	19%	49%	32%
	Mean error	1	0	0
	Mean square	1	0	0

Consider the simpler problem —— given five data values all known to come from the same distribution, what is the best estimate for the mean of the distribution? The answer (or the non-answer) is well known; if the distribution is uniform you should take half the sum of the smallest and biggest, if Gaussian you should take the mean, and if the distribution is Cauchy you should take the median of the five values.

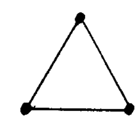
PLANE TOPOLOGY (JCMN 61 p.6256)

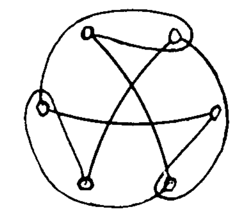
Terry Tao (Princeton)

The proposition suggested in the previous issue is untrue.

Rewording the question: In the plane are n nodes, numbered 0, 1, 2, ... n-1, by the residue classes mod n. There are n curves, numbered the same way, the curve C(r) joining tne nodes numbered r and r+1, and containing no other node. Each curve meets each other curve just once. The curve C(r) meets C(r-1) and C(r+1) at its ends, and where it meets each of the others they cross in the sense of going from one side to the other. For what n is such a figure possible? (The suggestion was that n could not be even)

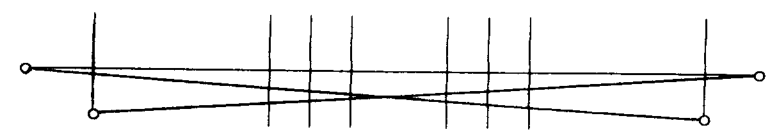
The answer is that n can be 3 or any number ≥ 5. To prove this, the following sketches indicate how n = 3 and n = 6 are possible.





Finally, if n is possible then so is n+2.

<u>Proof</u> Take the figure with n nodes and curves, and distort it so that curve C(1) is straight, as shown.



Now replace C(1) by three curves as shown above, adding two nodes. This gives the required figure of n+2 nodes and curves.

NON-DIFFERENTIABILITY

T. C. S. Tao

(Princeton University)

The function $g(\theta) = \frac{1}{3} + \Sigma_{n=1}^{\infty} (-2)^{-n}\cos(2^{n+1}\theta)$, which emerged in ORTHIC LIMIT POINT OF AN ISOSCELES TRIANGLE (JCMN 61, p.6250), is non-differentiable almost everywhere. To prove this it will be sufficient to consider only the function

$$F(x) = \Sigma_{n=1}^{\infty} (-2)^{-n}\cos(2^n\pi x).$$

<u>Lemma 1</u> If $a < x < b$ then $|\Sigma_{n=1}^{M} (-1)^n\sin(2^n\pi x)| > 7$. for some a, b and M.

<u>Proof</u> The integral mean of the square of the function can be made greater than 49 by taking $M = 99$, then the function will be continuous and have modulus > 7 on an open set of positive measure, therefore on an interval.

<u>Notation</u> $\{x\}$ denotes the non-integer part of x, i.e. $\{x\} = t$ where $0 \le t < 1$ and $x - t$ is an integer.

<u>Theorem</u> If c is such that $\{2^n c\} \in (a, b)$ for arbitrarily large n, then the function $F(x) = \Sigma_{n=1}^{\infty}(-2)^{-n}\cos(2^n\pi x)$ is non--differentiable at $x = c$.

<u>Proof</u> Put $F_N(x) = \Sigma_{n=1}^{N}(-2)^{-n}\cos(2^n\pi x)$,

then $F(x) - F_N(x) = (-2)^{-N-1}\cos(2^{N+1}\pi x) + \ldots$ has period 2^{-N},

and so if we put $h = h(N) = 2^{-N}$ we find

$F(x+h) - F(x) = F_N(x+h) - F_N(x)$, now use Taylor's theorem.

$= hF_N'(x) + (h^2/2)F_N''(x+\theta h)$ where $0 < \theta < 1$.

Put $\varphi(N) = \dfrac{F(c+h) - F(c)}{h} = F_N'(c) + hF_N''(c+\theta h)/2$.

We shall show that $\varphi(N)$ does not tend to a limit as $N \to \infty$.

Note that $F_N''(x) = -\pi^2\Sigma_{n=1}^{N}(-2)^n\cos(2^n\pi x)$ is between $\pm\, 2^{N+1}\pi^2$,

so that $hF_N''(c+\theta h)/2$ is between $\pm\,\pi^2$.

Therefore $\varphi(N+M) - \varphi(N)$ is between $F_{N+M}'(c) - F_N'(c) \pm 2\pi^2$.

But $F_{N+M}'(c) - F_N'(c) = -\pi\, \Sigma_{n=N+1}^{N+M}(-1)^n\sin(2^n\pi c)$

(now putting $g = 2^N c$)

$= (-1)^N\pi(\sin(2\pi g) - \sin(4\pi g) + \sin(8\pi g) + \ldots - (-1)^M\sin(2^M\pi g))$

and therefore $|F_{N+M}'(c) - F_N'(c)| > 7\pi$

by Lemma 1 because $\{g\} = \{2^N c\} \in (a, b)$.

Therefore $|\varphi(N+M) - \varphi(N)| > 7\pi - 2\pi^2 > 1$ for arbitrarily large N, proving our theorem.

<u>Corollary</u> The function F is non-differentiable almost everywhere, because almost all x are normal, which implies that the points $\{2^n x\}$ for $n = 0, 1, 2, \ldots$ are dense in the unit interval, and so infinitely many of them must be in the interval (a, b).

Is there any point where F is differentiable?

USEFUL INFORMATION

If your computer offers you a random variable uniform on the unit interval (0, 1), here is how you can get a normal (Gaussian) random variable.

Let U and V be random variables, both uniform on (0, 1). Then $\sqrt{-2\log U}\,\cos 2\pi V$ and $\sqrt{-2\log U}\,\sin 2\pi V$ are both normally distributed random variables, with mean = 0 and variance = 1. And they are independent.

ORTHIC LIMIT OF A TRIANGLE

(JCMN 58, p.6138, JCMN 60, p.6209 and JCMN 61, p.6250)

T. C. S. Tao (Princeton University)

Represent points of the plane by complex numbers in the usual way. Take a triangle ABC of unit circumradius with origin at the circumcenter. Use the letters A, B and C to represent either the points or the corresponding complex numbers. By rotation we may make the product ABC = 1. The orthic triangle A′B′C′ has vertices the feet of the perpendiculars from A, B and C to the opposite sides, given by

$$A' = \frac{A+B+C}{2} - \frac{BC}{2A} = \frac{A+B+C-A^{-2}}{2}$$
$$B' = \frac{A+B+C}{2} - \frac{CA}{2B} = \frac{A+B+C-B^{-2}}{2}$$
$$C' = \frac{A+B+C}{2} - \frac{AB}{2C} = \frac{A+B+C-C^{-2}}{2}$$

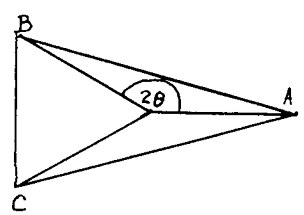
(See JCMN 30, p.3133, or, if your stock of old copies does not go back to December 1982, the calculation is not hard)
The altitudes meet at H = A+B+C, and the points A′, B′ and C′ are on the nine-point circle with center at N = (A+B+C)/2, and with circumradius half that of ABC.

The orthic limit of the triangle is

$$\frac{A+B+C}{2} - \frac{1}{4}(A^{-2}+B^{-2}+C^{-2}) + \frac{1}{8}(A^4+B^4+C^4) - \ldots$$

because the partial sums of this series are the nine-point centers of the triangles of the sequence, each nine-point center being the circumcenter of the next triangle of the sequence.

From this formula we can derive the result found in the previous issue for the orthic limit of an isosceles triangle. Take vertices A = 1, $B = \exp(2i\theta)$ and $C = \exp(-2i\theta)$.
$A + B + C = 1 + 2\cos 2\theta$,
$A^{-2}+B^{-2}+C^{-2} = 1 + 2\cos 4\theta$, etc.
This gives the series



$$\left(\frac{1}{2} - \frac{1}{4} + \frac{1}{8} - \ldots\right) + \cos 2\theta - \frac{1}{2}\cos 4\theta + \frac{1}{4}\cos 8\theta - \ldots$$
$$= \frac{1}{3} + \Sigma_{k=0}^{\infty}(-2)^{-k}\cos(2^{k+1}\theta)$$

for the position of the orthic limit. Subtracting $\cos 2\theta$ gives the height of the orthic limit above the base of the triangle, this height is $\frac{1}{3} + \Sigma_{k=1}^{\infty}(-2)^{-k}\cos(2^{k+1}\theta)$, which in the notation of the previous contribution is

$$\sin 2\theta\, f(\tan\theta) = g(\theta),$$

because the triangle has base = $2\sin 2\theta$ and base angles θ.

Are there extensions to higher dimensions? Does the circumsphere of the feet of the altitudes of a tetrahedron have interesting properties? Can you use quaternions for this question, as you use complex numbers in 2 dimensions?

QUOTATION CORNER 44

Biologists think they are biochemists,
Biochemists think they are chemists,
Chemists think they are chemical physicists,
Chemical physicists think they are physicists,
Physicists think they are God,
God thinks he is a mathematician.
—— Anonymous

USELESS INFORMATION

$2^{216193} \times 3^3 \times 14503 - 1$ is a prime number.

$2^{11235} \times 3 \times 5 \times 11 \times 10343 \pm 1$ form a prime pair.

MEDICAL RESEARCH IN AUSTRALIA (2)

About 30 years ago Dr McBride in Australia reported having found correlation between babies being born deformed and their mothers having taken what was then a new drug called Thalidomide, sometimes labelled Distaval. Other workers soon confirmed his results and there was widespread publicity over the affair. It was estimated that 12,000 babies had been affected. The drug was said to have been banned all over the world.

The drug companies needed to prevent another such disaster. As many as possible of the brochures sent out to the medical profession advertising thalidomide were collected and destroyed. Then strings were pulled to ensure that Dr McBride was never given any recognition or award or honour by universities or governments; that was easy, but they needed to do more. Dr McBride was given funds to set up a small research institute, which survived quietly for many years, then one of the employees announced to the newspapers that Dr McBride had been dishonest in his publications, that he had been "massaging" the experimental data to make them look better. There was a formal enquiry and the mass media tried to convince the general public (largely successfully) that Dr McBride should be regarded as disgraced. Few people now know that most experimental scientists massage their data before publishing, though Captain Cook had remarked on that usage in 1776 (see JCMN 61, p.6236). A certain amount of massaging is often justified, but it is hard to be dogmatic on where to draw the line.

Modern research scientists face a slightly different temptation, to massage the data to improve their chances of getting a research grant for the next year's work. This consideration is particularly strong in the medical sciences, for the donors of research grants are often drug companies with a concern about what is published. In Cambridge University the Glaxo Professorship of Molecular Parasitology has recently been created with an endowment of £850,000 from Glaxo Holdings plc. In Australia and the U.K. the governments are pressing the universities to finance their research laboratories by contracts and grants from industry. In these circumstances would a university research scientist, remembering the fate of Dr McBride, announce that a drug put out by a big company was harmful? How many have ever done so?

More recently (June and July 1993) it has been reported (in the London *Weekly Telegraph*) that thalidomide is being given to leprosy patients in Brazil, and that the same birth deformities are being found as those observed long ago by Dr McBride.

QUOTATION CORNER 45

The weakness of democracy is that the assumption that all men are equal and capable of equal contribution to the common good is flawed.

—— Lee Kuan Yew, speech in Tokyo, November 1992.

EDUCATION IN SOUTH AUSTRALIA

Channel Seven (television) News last month screened a controversial list of the "top South Australian schools" —— those attended by the top 250 candidates in the 1991 public matriculation examinations.

The government body concerned at first refused to release the information. Journalists obtained it only after an appeal to the State Ombudsman. Channel Seven reports that the State government has now decided to change the law so that no such information will ever be released again.

A little over half the schools in the list were private (i.e. not State Government) though private schools make up only about a quarter of the schools in South Australia.

—— Newspaper report in March.

Another feature emerged in June, that the University of Adelaide, in using public examination marks to decide which applicants to admit to the university, has for the last three years had a policy (quaintly called the *Fairway scheme*) of adding extra marks to the scores of candidates from certain schools. There was a precedent elsewhere in Australia; a few years ago in the A.C.T. all girls had two marks added to their scores for university entrance, but this scheme was abandoned after two years.