PHILOSOPHICAL TRANSACTIONS
OF
THE ROYAL SOCIETY
OF LONDON

A. MATHEMATICAL AND PHYSICAL SCIENCES

VOLUME **266** PAGES **123-192** NUMBER **1173**

5 March 1970 Price £1 17*s* (U.S. $4.80)

Uniqueness in the inversion of inaccurate gross Earth data

by G. Backus and F. Gilbert

PUBLISHED BY THE ROYAL SOCIETY
6 CARLTON HOUSE TERRACE LONDON S.W.1

NOTICE TO AUTHORS AND COMMUNICATORS

No scientific communication or paper is considered, read or published by the Royal Society unless it has been communicated by a Fellow or Foreign Member of the Society whose duty it shall be 'to satisfy himself that any letter, report or other paper which he may communicate is suitable to be read before the Society' (Statute 64).

Authors of papers submitted to the Royal Society are required to put their communications in as concise a form as possible. Papers should be typewritten, carefully revised and properly prepared as copy for the printer. Two copies of the typescript and of the illustrations should be submitted and it is desirable that a further copy be retained by the author.

The communicator shall be responsible for the suitable presentation of the paper and for its grammatical correctness. The Secretaries reserve the right to return a paper to the author for retyping or other adequate preparation if it is presented in a state which would make it difficult for the compositor to set.

Every paper must be accompanied by *three copies* of a brief summary not exceeding 5 % of the length of the paper.

Authors are required to send in all original drawings, diagrams or other illustrations, in a state suitable for direct photographic reproduction. They should be drawn on a large scale (to allow reduction to one-half or two-fifths linear for finished blocks) in Indian ink on Bristol board or on good quality tracing linen or tracing paper, with temporary lettering in blue pencil. Great care should be exercised in selecting only those that are essential. If unsatisfactory drawings are submitted authors may be required to have them redrawn by a professional artist. The second, photocopied, set of figures should have the necessary lettering inserted clearly in ink. When photographs are essential one set, for the use of the blockmaker, should be in the form of glossy, unmounted and unlettered prints, suitably protected against damage.

It is to be understood that papers are only accepted on the assumption that authors are prepared to conform with the instructions to authors prescribed by the Royal Society.

Copies of *Notes on the preparation of papers to be communicated to the Royal Society* will be sent to communicators and authors of papers on application to the Executive Secretary.

ASSOCIATE EDITORS FOR PHILOSOPHICAL TRANSACTIONS AND PROCEEDINGS OF THE ROYAL SOCIETY

(For Standing Orders see the current Year Book)

A. *Mathematical and physical sciences*

Professor D. R. Bates
Professor W. E. Burcham
Professor J. W. S. Cassels
Professor W. Cochran
Professor D. D. Eley
Professor A. W. Johnson
Professor M. J. Lighthill
Professor W. H. McCrea
Professor S. K. Runcorn
Professor J. Sutton
Dr D. Tabor
Professor D. H. Whiffen

UNIQUENESS IN THE INVERSION OF INACCURATE GROSS EARTH DATA

By G. BACKUS and F. GILBERT
Institute of Geophysics and Planetary Physics and Scripps Institution of Oceanography, University of California, San Diego, U.S.A.

(*Communicated by Sir Edward Bullard, F.R.S.—Received* 24 *January* 1969)

CONTENTS

A. FORMULATING THE PROBLEM

B. ABSOLUTE ERRORS

Vol. 266. A. 1173. (Price £1. 17*s*.; U.S. $4.80) 15 [Published 5 March 1970

C. RELATIVE ERRORS

D. NUMERICAL ILLUSTRATIONS

A gross Earth datum is a single measurable number describing some property of the whole Earth, such as mass, moment of inertia, or the frequency of oscillation of some identified elastic-gravitational normal mode. We suppose that a finite set $\mathscr{G}$ of gross Earth data has been measured, that the measurements are inaccurate, and that the variance matrix of the errors of measurement can be estimated. We show that some such sets $\mathscr{G}$ of measurements determine the structure of the Earth within certain limits of error except for fine-scale detail. That is, from some sets $\mathscr{G}$ it is possible to compute localized averages of the Earth structure at various depths. These localized averages will be slightly in error, and their errors will be larger as their resolving lengths are shortened. We show how to determine whether a given set $\mathscr{G}$ of measured gross Earth data permits such a construction of localized averages, and, if so, how to find the shortest length scale over which $\mathscr{G}$ gives a local average structure at a particular depth if the variance of the error in computing that local average from $\mathscr{G}$ is to be less than a specified amount. We apply the general theory to the linear problem of finding the depth variation of a frequency-independent local elastic dissipation (Q) from the observed damping rates of a finite number of normal modes. We also apply the theory to the nonlinear problem of finding density against depth from the total mass, moment and normal-mode frequencies, in case the compressional and shear velocities are known.

A. FORMULATING THE PROBLEM

1. Introduction

In two recent papers (Backus & Gilbert 1967, 1968; hereafter called Inverse I and Inverse II) we discussed techniques for extracting rigorous and geophysically useful conclusions about the internal structure of the Earth from an finite set of measured gross Earth data, such as frequencies of oscillation of normal modes, and seismic travel times between various sources and observers. Both Inverse I and Inverse II were idealized in that observational errors were neglected. The present paper is an examination of the effect such errors have on the conceptual question of what geophysical information is contained in a given set of gross Earth data.

In describing briefly the results already obtained in Inverse I and Inverse II we will use and supplement the terminology introduced in Inverse II: an n-dimensional Earth model is an ordered n-tuple of functions of position in the Earth, such as density, P-wave velocity and S-wave velocity ($n = 3$). A radial Earth model is one which depends only on radial distance from the centre of mass of the Earth, while a geographical Earth model also depends on latitude and longitude. A gross Earth functional is any rule for assigning a single real number to each member of a certain class of Earth models; examples are total mass, total moment of inertia, the frequency of oscillation of a particular normal mode, and the travel time of a particular seismic phase from a particular source to a particular observer. A gross Earth datum is the measured value of any gross Earth functional for the real Earth. If $\mathscr{G}$ is any finite set of gross Earth functionals $g_1, \ldots, g_N$, we say an Earth model m is '$\mathscr{G}$-acceptable' if the calculated values of $g_1(m), \ldots, g_N(m)$ agree with the values $\gamma_1, \ldots, \gamma_N$ measured for the real Earth.

Whenever g is a gross Earth functional and m_0 is an Earth model, we say that a second Earth model m is 'g-near to m_0' if $g(m) - g(m_0)$ is calculable with good accuracy from the first Fréchet derivative of g at m_0; i.e. if $g(m)$ is calculable from $g(m_0)$ via first-order perturbation theory, the contribution from higher order terms in $m - m_0$ being negligible. (The quantitative meaning of 'good accuracy' will depend, of course, on the accuracy of the observations $\gamma_1, \ldots, \gamma_N$, and the errors we are willing to accept in interpreting those data.) If second or higher order terms in $m - m_0$ make a significant contribution to $g(m) - g(m_0)$ we will say that m is 'g-far from m_0'. Evidently if g is linear then any two Earth models are g-near to one another. Given a finite set $\mathscr{G}$ of gross Earth functionals, we will say that Earth model m is '$\mathscr{G}$-near' to Earth model m_0 if m is g-near to m_0 for every g in $\mathscr{G}$. If m is g-far from m_0 for at least one g in $\mathscr{G}$, we will say that m is $\mathscr{G}$-far from m_0.

Clearly, at each epoch the set of all measured gross Earth data is finite. In Inverse I we showed that for any finite set $\mathscr{G}$ of well-behaved (i.e. independent, Fréchet differentiable) gross Earth functionals the set of $\mathscr{G}$-acceptable Earth models is either empty or an infinite dimensional manifold which, with any Earth model m_0, contains an infinite-parameter family of Earth models $\mathscr{G}$-near to m_0. We exploited this fact to automate the production of $\mathscr{G}$-acceptable Earth models, using Newton's method in Banach spaces (Bartle 1955), and we gave some sample results of machine calculations. Further results appear in Gilbert & Backus (1968). In short, Inverse I was directed to the problem of producing $\mathscr{G}$-acceptable Earth models.

Inverse II was directed to the question of uniqueness. An obvious, unavoidable non-uniqueness arises from the fact that given a finite set $\mathscr{G}$ of gross Earth functionals, there is a lower limit to the length scales of the structures resolvable by $\mathscr{G}$. If m_0 is any $\mathscr{G}$-acceptable Earth model and m is any Earth model which averages to zero over all but extremely small length scales, then it seems intuitively reasonable (and can be proved rigorously by the techniques of Inverse I) that $m_0 + m$

will be $\mathcal{G}$-acceptable, perhaps after a small distortion. Suppose that at some radius r this lack of resolution is the only source of the differences among the various $\mathcal{G}$-acceptable Earth models. Then there will be some smallest length l such that all $\mathcal{G}$-acceptable Earth models have approximately the same average over an interval of length l at radius r. We will say that $\mathcal{G}$ has resolving length l at radius r.

In Inverse II we showed how to determine whether a given set $\mathcal{G}$ of *linear* gross Earth functionals was capable of resolving structure near a given radius r, and, if so, how to estimate the resolving length of $\mathcal{G}$ at r. Thus, when the gross Earth functionals in $\mathcal{G}$ are linear and the corresponding gross Earth data are measured without observational error, Inverse II gives a complete solution to the problem of non-uniqueness in geophysical data inversion.

When some of the gross Earth functionals in $\mathcal{G}$ are nonlinear, and m_0 is any $\mathcal{G}$-acceptable Earth model, Inverse II shows how to learn whether there are $\mathcal{G}$-acceptable Earth models m which are $\mathcal{G}$-near to m_0 and differ significantly from m_0 at all length scales, or whether all $\mathcal{G}$-acceptable models $\mathcal{G}$-near to m_0 agree with m_0 except for fine-scale detail. Neither Inverse II nor the present paper contributes to the question of whether there are $\mathcal{G}$-acceptable Earth models $\mathcal{G}$-far from m_0. If such models exist, at present we can search for them only by various numerical search techniques described in Inverse I and Gilbert & Backus (1968), or by using the Monte Carlo methods introduced by Keilis-Borok & Yanovskaya (1967) and Levshin, Sabitova & Valus (1966). At present we know of no practical method for establishing that a given set $\mathcal{G}$ of nonlinear gross Earth functionals admits no $\mathcal{G}$-acceptable models $\mathcal{G}$-far from one another.*†

In both Inverse I and Inverse II the errors in the data were ignored. In the present paper we examine how such errors affect the resolving length of the data at various depths. We begin by summarizing and simplifying the procedures described in Inverse II for extracting localized information about the Earth from error-free data. Then we adapt the theory to data with error statistics which are known or can be estimated. Finally, we illustrate the theory by applying it to find dissipation (Q^{-1}) and density as functions of radius in the Earth. In all the numerical calculations reported here, v_P and v_S are assumed to be known. This assumption is unnecessary, but dropping it triples our computations and forces us to use a machine memory larger than that available at San Diego. We propose to report the results of the more extensive calculations in a later paper.

2. Extracting localized averages of Earth models from a given set of error-free gross Earth data

Let $\mathcal{G}$ be a finite set of gross Earth functionals, $g_1, \ldots, g_N$. Let $\gamma_1, \ldots, \gamma_N$ be the observed values of $g_1, \ldots, g_N$ for the real Earth; in this section we suppose that $\gamma_1, \ldots, \gamma_N$ have been measured without observational error. If m_E is the real Earth, we have $g_i(m_E) = \gamma_i$, $i = 1, \ldots, N$. The

* Press (1968), for a particular $\mathcal{G}$, found six $\mathcal{G}$-acceptable models among 5×10^6 randomly chosen models. He proposed, with some qualification, that this strengthened the suggestion that properties common to his six $\mathcal{G}$-acceptable models are common to all $\mathcal{G}$-acceptable models. We believe, on the contrary, that without some insight into the structure of the manifold of $\mathcal{G}$-acceptable models, Monte Carlo methods are not a practical way to preclude the existence of $\mathcal{G}$-acceptable models $\mathcal{G}$-far from one another. The set of Earth models under serious discussion in the current literature has at least 40 free parameters (for example the first 15 Fourier components of ρ and v_P and the first 10 of v_S, or the values of ρ and v_P at 15 radii and of v_S at 10 radii). If the selection problem consisted merely in picking the correct value of each parameter from among three possibilities, there would be 3^{40}, or 10^{19}, different Earth models to consider. In order to complete the calculations for 5×10^6 models on a large, modern digital computer within 20 h of machine time, Press was forced to assume without verification that all his models were $\mathcal{G}$-near to a single specified model.

† See also Asbel, Keilis-Borok & Yanovskaya (1966).

question is, what can we learn about m_E from the values of the N numbers $g_1(m_E), \ldots, g_N(m_E)$. In the present paper we will consider only one-dimensional spherical Earth models, so that m_E is a single real-valued function of r. The extension of the theory to n dimensional geographical Earth models is straightforward, but the computations become much more extensive.

(a) *Linear gross Earth functionals*

First let us examine the case that $\mathscr{G}$ consists entirely of linear gross Earth functionals $g_1, \ldots, g_N$. We may assume without loss of generality that $g_1, \ldots, g_N$ are linearly independent, since linearly dependent functionals can be winnowed from $\mathscr{G}$ without loss of information. The fact that $g_i(m)$ is linear in m implies the existence of a known function $G_i(r)$ such that for all m

$$g_i(m) = \int_0^1 m(r)\, G_i(r)\, \mathrm{d}r. \tag{2.1}$$

(As in Inverse I and Inverse II we choose units so that the radius of the Earth is 1.) The function $G_i(r)$ will be called the 'data kernel' for the functional g_i and the datum γ_i. In principle, linearity of g_i requires only that G_i be a distribution, or generalized function, but all the data kernels which have arisen in contemporary geophysical measurements to date are integrable functions (see Inverse I and Dahlen 1969), and for all functionals except travel times the data kernels are square integrable.

Now $m_E(r)$ describes the real Earth. We know about it only that it is one of infinitely many $\mathscr{G}$-acceptable Earth models m all satisfying

$$g_i(m) = \gamma_i \quad (i = 1, \ldots, N). \tag{2.2}$$

The known values of $g_i(m_E)$ can be regarded as generalized moments of m_E, moments with respect to the data kernels. Of course from these finitely many moments alone we cannot hope to calculate the separate values of $m_E(r)$, but we might hope to calculate weighted averages of the values of $m_E(r)$ at different r, with particularly heavy weighting close to some radius r_0 where we would like to estimate $m_E(r_0)$.

Any linearly weighted average of $m(r)$ has the form

$$\langle m, A \rangle = \int_0^1 m(r)\, A(r)\, \mathrm{d}r, \tag{2.3}$$

where the weighting function A is unimodular; that is

$$\int_0^1 A(r)\, \mathrm{d}r = 1. \tag{2.4}$$

Any unimodular function will be called an averaging kernel; the terms are interchangeable. In particular, we admit the possibility that some values of r will receive negative weights; i.e. we do not demand that an averaging kernel be non-negative.

For which averaging kernels $A(r)$ can we evaluate the averages $\langle m_E, A \rangle$, if we are given only the observed values $\gamma_1 = g_1(m_E), \ldots, \gamma_N = g_N(m_E)$? It is clear from equations (2.1) and (2.2) that if we let $a_1, \ldots, a_N$ be any constants and define

$$A(r) = \sum_{i=1}^{N} a_i G_i(r), \tag{2.5}$$

then for any $\mathscr{G}$-acceptable Earth model m, including m_E, we will have

$$\langle m, A \rangle = \sum_{i=1}^{N} a_i \gamma_i. \tag{2.6}$$

Thus if A is any unimodular linear combination of the data kernels, we can compute $\langle m_E, A\rangle$ directly from the gross Earth data. If A is not a linear combination of the data kernels, then it was shown in appendix A of Inverse II that $\langle m_E, A\rangle$ cannot be computed from the gross Earth data when $g_1, \ldots, g_N$ are linear functionals.

Now can we choose the coefficients $a_1, \ldots, a_N$ in (2.5) so that $A(r)$ not only is unimodular but also has most of its weight concentrated near a particular radius r_0 where we would like to have an estimate of the local average $\langle m_E\rangle_{r_0}$ of $m_E(r)$, averaged over some 'resolving length' $l(r_0)$? Can we arrange that $A(r)$ in (2.5) have a tall, narrow peak centred near r_0 and very small values elsewhere? In short, can we make $A(r)$ a good approximation to the Dirac delta distribution, $\delta(r-r_0)$, so that $\langle m, A\rangle$ is a good estimate of $m(r_0)$?

In Inverse II we considered a number of different quantitative measures of the deviation of an arbitrary unimodular function $A(r)$ from the ideal $\delta(r-r_0)$. In the present paper we shall consider only one such measure, which we call the 'spread of A from r_0' and denote by $s(r_0, A)$:

$$s(r_0, A) = 12\int_0^1 (r-r_0)^2 A(r)^2\,\mathrm{d}r. \tag{2.7}$$

Any unimodular function has the dimensions of an inverse length, so $s(r_0, A)$ is a length. The factor 12 is chosen to make $s(r_0, A)$ a measure of the width of the peak in $A(r)$ when $A(r)$ resembles $\delta(r-r_0)$. Specifically, if l is any small positive number, let A_{l,r_0} denote the unimodular step function which is 0 outside the interval $|r-r_0| \leqslant \frac{1}{2}l$ and is l^{-1} inside that interval. Then, because of the factor 12 in (2.7), $s(r_0, A_{l,r_0}) = l$. It is easy to verify that for other unimodular functions A with tall, narrow peaks centred at r_0 and with small weight elsewhere, such as normal curves, Cauchy distributions, linear and parabolic tent functions, etc., $s(r_0, A)$ as defined in (2.7) is close to other, more usual definitions of peak width.

For any fixed A, $s(r_0, A)$ is a quadratic polynomial in r_0 with a minimum value at $r_0 = c(A)$ where

$$c(A) = \int_0^1 rA(r)^2\,\mathrm{d}r \Big/ \int_0^1 A(r)^2\,\mathrm{d}r. \tag{2.8}$$

We will call $c(A)$ the 'centre' of A. By definition, the centre of A is that point from which the spread of A is least. The spread of A from $c(A)$ we will call the peak-width of A, or simply the 'width' of A, written $w(A)$:

$$w(A) = s(c(A), A). \tag{2.9}$$

From (2.7) it is clear that for any A and r_0 we have

$$s(r_0, A) = w(A) + 12[r_0 - c(A)]^2 \int_0^1 A(r)^2\,\mathrm{d}r. \tag{2.10}$$

Thus the spread of A from r_0 can be large either because the width of A is large or because the centre of A is far from r_0.

To show that $s(r_0, A)$ is a quantitative measure of the deviation of $A(r)$ from $\delta(r-r_0)$ we will show that $s(r_0, A)$ is small if and only if $\langle m, A\rangle$ is nearly $m(r_0)$ for all m which are well behaved in a sense about to be described.

For any $m(r)$ we define a norm $\|m\|_{r_0}$ as follows:

$$\|m\|_{r_0}^2 = \int_0^1 \left|\frac{m(r)-m(r_0)}{r-r_0}\right|^2 \mathrm{d}r. \tag{2.11}$$

If A is unimodular then

$$\langle m, A\rangle - m(r_0) = \int_0^1 [m(r)-m(r_0)]\,A(r)\,\mathrm{d}r.$$

If also $\|m\|_{r_0}$ is finite then Schwarz's inequality implies

$$|\langle m, A\rangle - m(r_0)| \leqslant \|m\|_{r_0}[\tfrac{1}{12}s(r_0, A)]^{\frac{1}{2}}. \tag{2.12}$$

Thus if $\{A_1, A_2, \ldots\}$ is a sequence of unimodular functions such that

$$\lim_{n\to\infty} s(r_0, A_n) = 0 \quad \text{then} \quad \lim_{n\to\infty} \langle m, A_n\rangle = m(r_0).$$

To prove the converse, we consider the space $\mathscr{L}_{r_0}$ of all functions $m(r)$ for which $\|m\|_{r_0} < \infty$. On this space each unimodular function A defines a linear functional $\mathscr{D}_{A,r_0}$ as follows: for any m in $\mathscr{L}_{r_0}$,

$$\mathscr{D}_{A,r_0}(m) = \langle m, A\rangle - m(r_0).$$

According to (2.12), $\mathscr{D}_{A,r_0}$ is bounded and its bound, $\|\mathscr{D}_{A,r_0}\|$, is no greater than $[\frac{1}{12}s(r_0, A)]^{\frac{1}{2}}$. In fact, if we take $m(r) = (r-r_0)^2A(r)$, then we have equality in (2.12), so

$$\|\mathscr{D}_{A,r_0}\| = [\tfrac{1}{12}s(r_0, A)]^{\frac{1}{2}}. \tag{2.13}$$

Now if A does a uniformly good job of estimating $m(r_0)$ for all m in $\mathscr{L}_{r_0}$ then $\|\mathscr{D}_{A,r_0}\|$ must be small, so $s(r_0, A)$ must be small. If $\{A_1, A_2, \ldots\}$ is a sequence of unimodular functions with the property that

$$\lim_{n\to\infty} \frac{\langle m, A_n\rangle - m(r_0)}{\|m\|_{r_0}} = 0$$

uniformly for all m in $\mathscr{L}_{r_0}$, then $\lim_{n\to\infty} s(r_0, A_n) = 0$. We conclude that an averaging kernel A makes $s(r_0, A)$ small if and only if, for that A and all m in $\mathscr{L}_{r_0}$, $\langle m, A\rangle$ is uniformly close to $m(r_0)$ when measured in units of $\|m\|_{r_0}$.

In Inverse II we fixed r_0 and sought that unimodular linear combination of the N data kernels which has the smallest spread from r_0. We denoted this optimal averaging kernel by $A_{r_0}(r)$, and its coefficients in (2.5) by $a_1(r_0), \ldots, a_N(r_0)$. If inspection of the graph of $A_{r_0}(r)$ shows it to be a good approximation to $\delta(r-r_0)$, or, what is the same thing, if $s(r_0, A) \ll 1$, then the $\langle m, A_{r_0}\rangle$ defined by (2.3) and calculated from (2.6) will be, roughly speaking, an average of $m(r)$ over an interval of length $w(A_{r_0})$ centred on $c(A_{r_0})$ and therefore near r_0. It is no distortion of language to call $\langle m, A_{r_0}\rangle$ a local average of m near r_0 with resolving length $w(A_{r_0})$. On the other hand, if A_{r_0} is not highly peaked near r_0, or, what is the same thing, if $s(r_0, A_{r_0})$ is not much less than 1, then we can conclude that the given gross Earth data do not enable us to compute an average of $m(r)$ which is any sense localized near r_0. At this particular depth the data admit other ambiguities of interpretation besides the expected lack of fine-scale resolution.

Now we introduce appropriate terminology. Let $\mathscr{G}$ be a set of linear gross Earth functionals with data kernels $G_1, \ldots, G_N$. Let A_{r_0} minimize $s(r_0, A)$ among all unimodular linear combinations A of the data kernels. If $s(r_0, A_{r_0}) \ll 1$ we say that $\mathscr{G}$ mean-determines m_{E}, the true Earth, at r_0, and that $\mathscr{G}$ is mean-decisive at r_0. If $s(r_0, A_{r_0})$ is not much less than 1, we say $\mathscr{G}$ is mean-indecisive at r_0. If $\mathscr{G}$ is mean-decisive at all r_0, we call $\mathscr{G}$ simply mean-decisive. In Inverse II, we gave examples of mean-decisive and mean-indecisive sets of gross Earth data.

(*b*) *Nonlinear gross Earth functionals*

If the gross Earth functional g_i is nonlinear, then (2.1) is no longer correct. We replace that equation by the assumption that g_i is Fréchet differentiable. This means that for any Earth model $m(r)$ there exist functions $G_1(r), \ldots, G_N(r)$ with the property that if $m'(r)$ is any other Earth model then for $i = 1, \ldots, N$

$$g_i(m') = g_i(m) + \int_0^1 [m'(r) - m(r)]\, G_i(r)\, \mathrm{d}r + O(m'-m)^2. \tag{2.14}$$

We will call G_i the 'data kernel for the functional g_i at the Earth model m'. In the nonlinear case these data kernels are different for different models m. Inverse I shows that all the usual gross Earth functionals are Fréchet differentiable, and gives the data kernels at radial Earth models. Dahlen (1969) gives data kernels at some non-radial, geographical Earth models.

Even in the nonlinear case, for any Earth model m we can still construct averaging kernels A, unimodular linear combinations of $G_1, \ldots, G_N$, the data kernels at m. Equation (2.6), however, will no longer be true. If A is given by equation (2.5), then equation (2.6) must be replaced by

$$\langle m, A\rangle = \sum_{i=1}^{N} a_i q_i, \tag{2.15}$$

where we define

$$q_i = \int_0^1 m(r)\, G_i(r)\, \mathrm{d}r. \tag{2.16}$$

Only when g_i is a linear functional do we have $q_i = g_i(m)$.

If A is a unimodular linear combination of the data kernels at m, and if m and m' are both $\mathscr{G}$-acceptable Earth models, $\langle m, A\rangle$ can differ from $\langle m', A\rangle$ if some of the functionals in $\mathscr{G}$ are nonlinear. However, the difference is of second order in $(m'-m)$. To see this we note that if m and m' are both $\mathscr{G}$-acceptable then $g_i(m) = g_i(m')$, so, from (2.14),

$$\int_0^1 m'(r)\, G_i(r)\, \mathrm{d}r - \int_0^1 m(r)\, G_i(r)\, \mathrm{d}r = O(m'-m)^2.$$

If we multiply this equation by a_i and sum from $i = 1$ to N, we obtain

$$\langle m', A\rangle - \langle m, A\rangle = O(m'-m)^2, \tag{2.17}$$

where A is given by (2.5). Equations (2.15) and (2.17) imply that

$$\langle m', A\rangle = \sum_{i=1}^{N} a_i q_i + O(m'-m)^2. \tag{2.18}$$

According to equation (2.17), if m' is $\mathscr{G}$-near to m then to a good approximation m' and m have the same averages with respect to any averaging kernel A which is a linear combination of the data kernels at m. Thus if m is a $\mathscr{G}$-acceptable Earth model and m_{E}, the real Earth, is $\mathscr{G}$-near to m, then to a good approximation we have $\langle m_{\mathrm{E}}, A\rangle = \langle m, A\rangle$ for any such averaging kernel A. Therefore, we can obtain a good estimate for any average $\langle m_{\mathrm{E}}, A\rangle$ whose averaging kernel A is a unimodular linear combination of the data kernels at an Earth model m which is $\mathscr{G}$-near to m_{E}.

Let m be any particular Earth model, and let $G_1, \ldots, G_N$ be the data kernels at m. Suppose that whenever an Earth model m' has $g_i(m') = g_i(m)$ for every g_i in $\mathscr{G}$, then m' is $\mathscr{G}$-near to m. In that case, we will say that '$\mathscr{G}$ isolates m'. If m is $\mathscr{G}$-acceptable and $\mathscr{G}$ isolates m, then m_{E}, the real Earth, is $\mathscr{G}$-near to m, so $\langle m, A\rangle$ is a good estimate for the average $\langle m_{\mathrm{E}}, A\rangle$ as long as A is a linear combination of $G_1, \ldots, G_N$.

Let A_{r_0} be that averaging kernel which minimizes $s(r_0, A)$ among all unimodular linear combinations of $G_1, \ldots, G_N$, the data kernels at m. If $s(r_0, A_{r_0}) \ll 1$ we say that $\mathscr{G}$ is 'mean-decisive at m near r_0'. If $s(r_0, A_{r_0}) \ll 1$ for all r_0, we say that $\mathscr{G}$ is 'mean-decisive at m'.

If m is a $\mathscr{G}$-acceptable Earth model, and $\mathscr{G}$ isolates m, and $\mathscr{G}$ is mean-decisive at m, then among the averages $\langle m_{\mathrm{E}}, A\rangle$ for which the data give good estimates are averages localized near every radius r_0. In this local average sense, the gross Earth data described by $\mathscr{G}$ determine the real Earth.

Neither Inverse II nor the present paper contributes to determining whether a set $\mathscr{G}$ of nonlinear gross Earth functionals isolates a particular Earth model. Both papers are directed to determining whether $\mathscr{G}$ is mean-decisive at particular Earth models. That is, in both papers we study the manifold of all $\mathscr{G}$-acceptable Earth models near a particular model; we do not attempt to delineate the overall global structure of that manifold in Hilbert space. If we find that $\mathscr{G}$ is mean-decisive at a $\mathscr{G}$-acceptable model m, then we can regard the computed local averages $\langle m, A_{r_0}\rangle$ as estimates of the true local averages $\langle m_{\mathrm{E}}, A_{r_0}\rangle$ only if we assume that m_{E} is $\mathscr{G}$-near to m. For convenience, we shall make this assumption throughout the present paper, but it is essential to keep in mind that if there are two $\mathscr{G}$-acceptable Earth models which are $\mathscr{G}$-far from one another, then without more data we cannot say which model resembles the real Earth. Of course if all the functionals in $\mathscr{G}$ are linear, then every Earth model is $\mathscr{G}$-near to m; $\mathscr{G}$ isolates every Earth model.

(c) *Algebraic statement of the problem of extracting local averages from error-free gross Earth data*

We are given a finite set $\mathscr{G}$ of gross Earth functionals and perfectly accurate measurements of the corresponding gross Earth data. In the linear case, $\mathscr{G}$ alone determines a set of data kernels, $G_1, \ldots, G_N$. In the nonlinear case, we fix a $\mathscr{G}$-acceptable Earth model m and obtain the data kernels for $g_1, \ldots, g_N$ at m. Then we define

$$u_i = \int_0^1 G_i(r)\,\mathrm{d}r, \tag{2.19}$$

$$S_{ij}^{(p)} = 12\int_0^1 r^p G_i(r)\,G_j(r)\,\mathrm{d}r,$$

and

$$S_{ij}(r_0) = 12\int_0^1 (r-r_0)^2\,G_i(r)\,G_j(r)\,\mathrm{d}r. \tag{2.20}$$

Clearly

$$S_{ij}(r_0) = r_0^2 S_{ij}^{(0)} - 2r_0 S_{ij}^{(1)} + S_{ij}^{(2)}. \tag{2.21}$$

The $N\times N$ matrices $S_{ij}^{(p)}$ and $S_{ij}(r_0)$ are symmetric, and the linear independence of $G_1, \ldots, G_N$ (already noted in the linear case and assumed henceforth for nonlinear gross Earth functionals) assures us that $S_{ij}^{(p)}$ and $S_{ij}(r_0)$ are also positive definite.

The condition on $a_1, \ldots, a_N$ which makes an averaging kernel (2.5) unimodular is

$$\sum_{i=1}^{N} a_i u_i = 1. \tag{2.22}$$

The spread of the averaging kernel (2.5) from r_0 is

$$s(r_0;\, a_1, \ldots, a_N) = \sum_{i,j=1}^{N} a_i a_j S_{ij}(r_0). \tag{2.23}$$

The centre of that averaging kernel is

$$c(a_1, \ldots, a_N) = \sum_{i,j=1}^{N} a_i a_j S_{ij}^{(1)} \Big/ \sum_{i,j=1}^{N} a_i a_j S_{ij}^{(0)}, \tag{2.24}$$

and its width is

$$w(a_1, \ldots, a_N) = \sum_{i,j=1}^{N} a_i a_j S_{ij}^{(2)} - \left[\sum_{i,j=1}^{N} a_i a_j S_{ij}^{(1)}\right]^2 \Big/ \sum_{i,j=1}^{N} a_i a_j S_{ij}^{(0)}. \tag{2.25}$$

For a fixed r_0, the optimal averaging kernel A_{r_0} is that unimodular linear combination of the data kernels which minimizes $s(r_0, A)$. Thus A_{r_0} is obtained by inserting in (2.5) that N-tuple of

coefficients $(a_1, \ldots, a_N)$ which minimizes the positive-definite quadratic form (2.23) subject to the constraint (2.22). The solution was given in Inverse II. The value of $\langle m, A_{r_0} \rangle$ is computed from (2.15); when all the functionals in $\mathscr{G}$ are linear this is equivalent to (2.6).

3. Extracting localized averages from erroneous gross Earth data

So far we have assumed that the gross Earth data $\gamma_1, \ldots, \gamma_N$ were measured with perfect accuracy. Once an averaging kernel A of the form (2.5) was constructed, the average (2.3) of m with respect to that kernel could be computed without error, via (2.6) when all the functionals in $\mathscr{G}$ were linear, and via (2.15) when some were nonlinear. Of course the gross Earth data will never be free of errors, but if we know the statistics of the errors then the theory of the propagation of errors gives us the statistics of the resulting error in $\langle m, A \rangle$.

The presence of errors in the data, however, may change our idea of what an optimal averaging kernel should be. The choice of $a_1, \ldots, a_N$ in (2.5) which produces A_{r_0}, the available averaging kernel most nearly like $\delta(r - r_0)$, may also produce very large amounts of cancellation in (2.6) or (2.15). Then any fractional errors in $\gamma_1, \ldots, \gamma_N$ will produce much larger fractional errors in $\langle m, A_{r_0} \rangle$. We must investigate whether there are coefficients $a_1, \ldots, a_N$ which produce via (2.5) an averaging kernel A only slightly less than optimally like $\delta(r - r_0)$, and which greatly reduce the cancellation in (2.6), so as to give us greatly improved accuracy in $\langle m, A \rangle$. Very inaccurate knowledge of a very highly and accurately localized average of m_{E} is not particularly useful. We are willing to sacrifice some resolution in the local average if we can greatly improve the accuracy with which we know its value.

(a) *Expressions for the errors and their statistics*

If we make errors $\Delta\gamma_1, \ldots, \Delta\gamma_N$ in measuring the gross Earth data $\gamma_1, \ldots, \gamma_N$, what error $\Delta\langle m_{\mathrm{E}}, A \rangle$ will result in our value for $\langle m_{\mathrm{E}}, A \rangle$ when A is given by equation (2.5)? In case all the gross Earth functionals in $\mathscr{G}$ are linear, we will compute $\langle m_{\mathrm{E}}, A \rangle$ from (2.6), so we have immediately

$$\Delta\langle m_{\mathrm{E}}, A \rangle = \sum_{i=1}^{N} a_i \Delta\gamma_i. \tag{3.1}$$

In case some functionals in $\mathscr{G}$ are nonlinear, we must suppose that our '$\mathscr{G}$-acceptable' Earth model m has values of $g_i(m)$ which agree with the erroneous gross Earth data, since only those data are available to us. The functions G_i are data kernels at m, and our averaging kernels A in (2.5) are linear combinations of those data kernels. We would like to know $\langle m_{\mathrm{E}}, A \rangle$, but we know only $\langle m, A \rangle$. The error, $\Delta\langle m_{\mathrm{E}}, A \rangle = \langle m, A \rangle - \langle m_{\mathrm{E}}, A \rangle$, is

$$\Delta\langle m_{\mathrm{E}}, A \rangle = \sum_{i=1}^{N} a_i \int_0^1 [m(r) - m_{\mathrm{E}}(r)]\, G_i(r)\, \mathrm{d}r. \tag{3.2}$$

If γ_i^E is the true value of γ_i for the Earth, our error of measurement is $\Delta\gamma_i = \gamma_i - \gamma_i^E$. If m_{E} is $\mathscr{G}$-near to m, then from (2.14)

$$\Delta\gamma_i = \int_0^1 [m(r) - m_{\mathrm{E}}(r)]\, G_i(r)\, \mathrm{d}r; \tag{3.3}$$

$\mathscr{G}$-nearness means simply that we can neglect the second-order term in (2.14). Then, combining (3.2) and (3.3) we obtain (3.1) even in the nonlinear case. Thus as long as our $\mathscr{G}$-acceptable Earth model m is $\mathscr{G}$-near to the true Earth model m_{E}, equation (3.1) is valid to first order in $m - m_{\mathrm{E}}$ whether the functionals in $\mathscr{G}$ are linear or not.

Of course we do not know the errors $\Delta\gamma_i$ (if we did, they would not be errors; we would remove them), but as in any error analysis we may assume that by repeated measurements of $\gamma_1, \ldots, \gamma_N$ we have learned something about the statistics of the errors. We will assume that the N-tuple of errors of measurement $(\Delta\gamma_1, \ldots, \Delta\gamma_N)$, has a probability distribution whose mean is $(0, 0, \ldots, 0)$ and whose second moments exist. That is, we assume that

$$\overline{\Delta\gamma_1} = \ldots = \overline{\Delta\gamma_N} = 0$$

and that the averages (expected values)

$$E_{ij} = \overline{(\Delta\gamma_i)\,(\Delta\gamma_j)} \tag{3.4}$$

exist. The $N \times N$ variance matrix E_{ij} is clearly symmetric and positive semidefinite (Cramér 1946). We assume that no linear relation among $\Delta\gamma_1, \ldots, \Delta\gamma_N$ is automatically satisfied in all measurements of $\gamma_1, \ldots, \gamma_N$. That is we assume that the N-tuples $(\Delta\gamma_1, \ldots, \Delta\gamma_N)$ are not restricted to lie with probability 1 in a subspace of dimension $N-1$. Then the matrix E_{ij} is positive definite. We also assume that we have an estimate of E_{ij}, and henceforth we will not distinguish between our estimate and the true variance matrix, so we will assume that the whole $N \times N$ matrix E_{ij} is known.

For any choice of coefficients $a_1, \ldots, a_N$ in (2.5) we can use (3.1) and (3.4) to write down the variance $\overline{(\Delta\langle m_E, A\rangle)^2}$ of the error in our estimate of $\langle m_E, A\rangle$:

$$\overline{(\Delta\langle m_E, A\rangle)^2} = \sum_{i,j=1}^{N} a_i a_j E_{ij}.$$

Then we may use the square root of this variance as an estimate of the error ϵ which we commit when we compute $\langle m_E, A\rangle$ from the gross Earth data by means of (2.6) or (2.15). Since we know E_{ij}, this error is completely determined by the coefficients $a_1, \ldots, a_N$, so we may write it as

$$\epsilon(a_1, \ldots, a_N)^2 = \sum_{i,j=1}^{N} a_i a_j E_{ij}. \tag{3.5}$$

(*b*) *Choosing optimal averaging kernels for erroneous data*

We can now give a quantitative proposal for choosing averaging kernels (2.5) when the gross Earth data contain errors. We fix r_0 and find that averaging kernel A_{r_0} which, among all unimodular linear combinations of the data kernels, minimizes $s(r_0, A)$. The coefficients of A_{r_0} in (2.5) we denote by $a_i(r_0)$. We suppose that $s(r_0, A_{r_0}) \ll 1$ so that $\langle m, A_{r_0}\rangle$ is indeed a local average of m, localized near r_0. The error in $\langle m_E, A_{r_0}\rangle$ produced by estimating it as $\langle m, A_{r_0}\rangle$, has a variance ϵ_0^2 given by substituting $a_i(r_0)$ in (3.5).

Suppose we feel that this variance is unacceptably large. To reduce it, we agree to accept averaging kernels (2.5) with spreads from r_0 slightly greater than $s(r_0, A_{r_0})$ but still much less than 1. This means that we choose a number s larger than $s(r_0, A_{r_0})$ but much less than 1 and we agree to accept any averaging kernel (2.5) whose coefficients $a_1, \ldots, a_N$ satisfy (2.22) and

$$\sum_{i,j=1}^{N} a_i a_j S_{ij} \leqslant s. \tag{3.6}$$

(Here and subsequently we write S_{ij} for $S_{ij}(r_0)$ when no ambiguity is possible.) Among these acceptable averaging kernels will be some with values of $\epsilon(a_1, \ldots, a_N)^2$ smaller than ϵ_0^2. Which acceptable averaging kernel produces the smallest such error? That is, which N-tuple of

coefficients $a_1, \ldots, a_N$ satisfying (2.22) and (3.6) minimizes $\epsilon(a_1, \ldots, a_N)$ as defined by (3.5)? This is the algebraic statement of the problem of decreasing the error variance in $\langle m_E, A \rangle$ by permitting the spread of A from r_0 to increase above its least possible value, $s(r_0, A_{r_0})$.

We hold r_0 fixed, and in the case of nonlinear gross Earth functionals we restrict attention to Earth models $\mathscr{G}$-near to a particular $\mathscr{G}$-acceptable model m, at which we compute the data kernels $G_1, \ldots, G_N$. Then u_i, S_{ij} and E_{ij} are all fixed. If we specify s, then the minimum of $\epsilon(a_1, \ldots, a_N)$ under the constraints (2.22) and (3.6) is determined as a function of s alone, say $\epsilon(s)$. If we increase s, then the set of N-tuples which satisfy (2.22) and (3.6) will increase. The minimum value of (3.5) over the larger, inclusive set is certainly no greater than its minimum over the smaller, included set, so if $s_1 < s_2$ we infer that $\epsilon(s_1) \geqslant \epsilon(s_2)$.

There remains the question of how we agree on some number s as an acceptable spread from r_0 for our averaging kernels. Clearly our choice of s will depend on what the given gross Earth data permit. To make the choice, we must know how much accuracy we gain for a given increase in spread from r_0. That is, we want to know the whole function $\epsilon(s)$. Then we might use different values of s for different purposes. The situation is very like that of spectral analysis of stationary time series, where very narrow spectral windows produce large uncertainty in the spectral estimates (Parzen 1962), and the choice of window is different for different applications. We will call the graph of $\epsilon(s)$ the absolute error tradeoff curve at r_0, since it tells us how much we can decrease the absolute error ϵ in our estimate of an average of m_E at r_0 by permitting the resolving length s of that average to increase.

Finally we note a second way to formulate our problem. We could choose a number $\epsilon > 0$ and refuse to consider any averaging kernels (2.5) unless they satisfied

$$\sum_{i,j=1}^{N} a_i a_j E_{ij} \leqslant \epsilon^2. \tag{3.7}$$

That is, we could restrict attention to averaging kernels A such that $\langle m_E, A \rangle$ could be computed from the data with an error variance no greater than ϵ^2. Then the problem would be to find among such kernels the one with least spread from r_0. We would seek the coefficients $a_1, \ldots, a_N$ which minimize $s(r_0; a_1, \ldots, a_N)$ in (2.23), subject to the constraints (2.22) and (3.7). This minimum spread would be a function of ϵ, say $s(\epsilon)$. Evidently $s(\epsilon)$ is a monotone non-increasing function of ϵ.

In § 5 we will see that the two formulations of our problem lead to the same averaging kernels and that $s(\epsilon)$ is the function inverse to $\epsilon(s)$. In this sense, the two formulations are equivalent. We prefer and will use the first formulation, with s, the spread from r_0, as the independent variable.

(c) *Relative errors*

Instead of minimizing the absolute error $\epsilon(a_1, \ldots, a_N)$ in $\langle m_E, A \rangle$ subject to the constraints (2.22) and (3.6), we could just as well minimize the relative error ρ defined by

$$\rho^2 = \overline{(\Delta \langle m_E, A \rangle)^2} / \langle m_E, A \rangle^2.$$

If we estimate $\langle m_E, A \rangle$ as $\langle m, A \rangle$ then from (2.15) and (3.5) we have

$$\rho(a_1, \ldots, a_N)^2 = \sum_{i,j=1}^{N} a_i a_j E_{ij} \Big/ \left[\sum_{i=1}^{N} a_i q_i \right]^2. \tag{3.8}$$

Usually we would expect minimizing ρ to have nearly the same effect as minimizing ϵ, but if the constraint (3.6) is sufficiently weak we will see that differences are possible. At any rate the

question must be investigated. We denote by $\rho(s)$ the minimum value of $\rho(a_1, \ldots, a_N)$ when $(a_1, \ldots, a_N)$ is subject to the constraints (2.22) and (3.6). We will call the graph of $\rho(s)$ the relative error tradeoff curve at r_0.

(*d*) *The effect of variations in the observational errors*

Suppose two different observers measure the same gross Earth data with different instruments or different skills, so that they produce different error variance matrices $E_{ij}^{(1)}$ and $E_{ij}^{(2)}$. How will their error tradeoff curves be related? The two observers use the same gross Earth functionals, so they have the same u_i and S_{ij}. Since $\epsilon(s)$ and $\rho(s)$ are defined as the minima of (3.5) and (3.8) subject to the constraints (2.22) and (3.6), it is easy to compare the two observers; we are minimizing their absolute and relative errors under the same constraints.

For example, if $E_{ij}^{(2)} - E_{ij}^{(1)}$ is positive definite, then

$$\epsilon^{(2)}(a_1, \ldots, a_N) > \epsilon^{(1)}(a_1, \ldots, a_N) \quad \text{and} \quad \rho^{(2)}(a_1, \ldots, a_N) > \rho^{(1)}(a_1, \ldots, a_N)$$

for any N-tuple $(a_1, \ldots, a_N)$. Since $\epsilon(a_1, \ldots, a_N)$ is continuous and the set of N-tuples satisfying the constraints is compact, $\epsilon^{(2)}(s) > \epsilon^{(1)}(s)$. Similarly, $\rho^{(2)}(s) \geqslant \rho^{(1)}(s)$ with strict inequality unless $\rho^{(1)}(s) = +\infty$.

In case there is a constant k such that $E_{ij}^{(2)} = k^2 E_{ij}^{(1)}$ then clearly

$$\epsilon^{(2)}(s) = k\epsilon^{(1)}(s) \tag{3.9}$$

and

$$\rho^{(2)}(s) = k\rho^{(1)}(s). \tag{3.10}$$

These results are particularly noteworthy. They show that the shape of the absolute or relative error tradeoff curve depends only on the ratios of the matrix elements E_{ij} and not on their absolute sizes. Evidently the same is true of the optimal averaging kernels at r_0 for any particular s.

B. ABSOLUTE ERRORS

4. The geometry of absolute errors

This section is devoted to developing a more compact notation for and some geometrical insight into the problem of minimizing the absolute error in $\langle m_E, A \rangle$ subject to the constraints (2.22) and (3.6).

(*a*) *Notation*

We regard the ordered N-tuple $(a_1, \ldots, a_N)$ as a vector $\boldsymbol{a}$ in the N-dimensional real vector space $\mathscr{R}^N$ consisting of such N-tuples. We define an inner (dot) product in the usual way: if $\boldsymbol{f} = (f_1, \ldots, f_N)$ and $\boldsymbol{g} = (g_1, \ldots, g_N)$ then

$$\boldsymbol{f} \cdot \boldsymbol{g} = \sum_{i=1}^{N} f_i g_i. \tag{4.1}$$

If we write $\boldsymbol{u}$ for the N-tuple $(u_1, \ldots, u_N)$ defined by (2.19), then the constraint (2.4) or (2.22) becomes

$$\boldsymbol{u} \cdot \boldsymbol{a} = 1. \tag{4.2}$$

Any $N \times N$ matrix K_{ij} defines a linear operator, $\boldsymbol{K}$: $\mathscr{R}^N \to \mathscr{R}^N$ as follows: $\boldsymbol{K}$ sends the vector $\boldsymbol{a}$ into the vector $\boldsymbol{K} \cdot \boldsymbol{a}$ whose ith component is

$$(\boldsymbol{K} \cdot \boldsymbol{a})_i = \sum_{j=1}^{N} K_{ij} a_j. \tag{4.3}$$

If K_{ij} is symmetric and positive definite, so is $\boldsymbol{K}$. Thus the matrices E_{ij} of (3.4) and S_{ij} of (2.20) define symmetric, positive definite linear operators $\boldsymbol{E}$: $\mathscr{R}^N \to \mathscr{R}^N$ and $\boldsymbol{S}$: $\mathscr{R}^N \to \mathscr{R}^N$.

Following Gibbs (1901) we regard linear operators as second-order tensors. Then $\boldsymbol{K}.\boldsymbol{a}$ can be interpreted as the dot product of a second order tensor on the left with a first-order tensor on the right. If $\boldsymbol{K}^T$ is the transpose of $\boldsymbol{K}$ then

$$\boldsymbol{a}.\boldsymbol{K} = \boldsymbol{K}^T.\boldsymbol{a},$$

so if $\boldsymbol{K}$ is symmetric, i.e. if $\boldsymbol{K}^T = \boldsymbol{K}$, then $\boldsymbol{a}.\boldsymbol{K} = \boldsymbol{K}.\boldsymbol{a}$. The symbol $\boldsymbol{K}.\boldsymbol{L}$ will denote the tensor dot product,

$$(\boldsymbol{K}.\boldsymbol{L})_{ij} = \sum_{k=1}^{N} K_{ik}L_{kj};$$

then $\boldsymbol{K}.\boldsymbol{L}$ is also the linear operator which arises from applying $\boldsymbol{L}$ to a vector and $\boldsymbol{K}$ to the result.

In this notation the constraint (3.6) becomes

$$\boldsymbol{a}.\boldsymbol{S}.\boldsymbol{a} \leqslant s, \tag{4.4}$$

and (3.5) becomes

$$\epsilon(\boldsymbol{a})^2 = \boldsymbol{a}.\boldsymbol{E}.\boldsymbol{a}. \tag{4.5}$$

We seek that $\boldsymbol{a}$ which minimizes the $\epsilon(\boldsymbol{a})^2$ in (4.5) subject to the constraints (4.2) and (4.4). We denote the minimum by $\epsilon(s)^2$.

Names will be required for several different sets of points $\boldsymbol{a}$ in $\mathscr{R}^N$. If $\boldsymbol{K}$ is a positive definite, symmetric operator and k is a positive number, we denote by $\mathscr{E}(\boldsymbol{K}, k)$ the set of all points $\boldsymbol{a}$ in $\mathscr{R}^N$ which satisfy

$$\boldsymbol{a}.\boldsymbol{K}.\boldsymbol{a} \leqslant k. \tag{4.6}$$

This set is an N dimensional solid ellipsoid centred on the origin. Its boundary is the $(N-1)$ dimensional ellipsoidal hypersurface consisting of all points $\boldsymbol{a}$ which satisfy

$$\boldsymbol{a}.\boldsymbol{K}.\boldsymbol{a} = k. \tag{4.7}$$

We denote this boundary by $\partial\mathscr{E}(\boldsymbol{K}, k)$. The interior of $\mathscr{E}(\boldsymbol{K}, k)$, denoted by $\mathscr{E}^0(\boldsymbol{K}, k)$, is the set of all points $\boldsymbol{a}$ which satisfy

$$\boldsymbol{a}.\boldsymbol{K}.\boldsymbol{a} < k. \tag{4.8}$$

For any $\boldsymbol{a}$ on $\partial\mathscr{E}(\boldsymbol{K}, k)$ the vector

$$\boldsymbol{n}_{\boldsymbol{K}}(\boldsymbol{a}) = \boldsymbol{K}.\boldsymbol{a} \tag{4.9}$$

is half the gradient of $\boldsymbol{a}.\boldsymbol{K}.\boldsymbol{a}$ with respect to $\boldsymbol{a}$ and hence is normal to the boundary surface $\partial\mathscr{E}(\boldsymbol{K}, k)$ at $\boldsymbol{a}$, and points out of $\mathscr{E}(\boldsymbol{K}, k)$.

(b) *Strict convexity of ellipsoids*

For our purposes the most important property of $\mathscr{E}(\boldsymbol{K}, k)$ is given by

LEMMA 1. *If $k > 0$ and $\boldsymbol{K}$ is symmetric and positive definite, then $\mathscr{E}(\boldsymbol{K}, k)$ is strictly convex.*

By this we mean that if $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ are two *different* points in $\mathscr{E}(\boldsymbol{K}, k)$ and $\boldsymbol{a}$ is any point between $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ on the straight line segment joining $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$, then $\boldsymbol{a}$ is in $\mathscr{E}^0(\boldsymbol{K}, k)$. The proof is simple. There are positive numbers α_1 and α_2 such that

$$\alpha_1 + \alpha_2 = 1 \tag{4.10}$$

and

$$\boldsymbol{a} = \alpha_1\boldsymbol{a}_1 + \alpha_2\boldsymbol{a}_2.$$

Since $\boldsymbol{K}$ is symmetric,

$$\boldsymbol{a}.\boldsymbol{K}.\boldsymbol{a} = \alpha_1^2\boldsymbol{a}_1.\boldsymbol{K}.\boldsymbol{a}_1 + 2\alpha_1\alpha_2\boldsymbol{a}_1.\boldsymbol{K}.\boldsymbol{a}_2 + \alpha_2^2\boldsymbol{a}_2.\boldsymbol{K}.\boldsymbol{a}_2.$$

Since $\boldsymbol{K}$ is positive-definite, Schwarz's inequality implies

$$|\boldsymbol{a}_1.\boldsymbol{K}.\boldsymbol{a}_2| \leqslant (\boldsymbol{a}_1.\boldsymbol{K}.\boldsymbol{a}_1)^{\frac{1}{2}}(\boldsymbol{a}_2.\boldsymbol{K}.\boldsymbol{a}_2)^{\frac{1}{2}},$$

with equality only when $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ are linearly dependent. Thus

$$\boldsymbol{a}.\boldsymbol{K}.\boldsymbol{a} \leqslant [\alpha_1(\boldsymbol{a}_1.\boldsymbol{K}.\boldsymbol{a}_1)^{\frac{1}{2}}+\alpha_2(\boldsymbol{a}_2.\boldsymbol{K}.\boldsymbol{a}_2)^{\frac{1}{2}}]^2, \tag{4.11}$$

with equality only when $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ are linearly dependent. Now we must consider three cases:

(i) $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ are in $\mathscr{E}(\boldsymbol{K},k)$ and at least one of them is in $\mathscr{E}^0(\boldsymbol{K},k)$. Then by (4.10) the right side of (4.11) is strictly less than k, so $\boldsymbol{a}$ is in $\mathscr{E}^0(\boldsymbol{K},k)$.

(ii) $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ are both in $\partial\mathscr{E}(\boldsymbol{K},k)$ but are linearly independent. Then the right side of (4.11) is equal to k, but the inequality in (4.11) is strict, so again $\boldsymbol{a}$ is in $\mathscr{E}^0(\boldsymbol{K},k)$.

(iii) $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ are both in $\partial\mathscr{E}(\boldsymbol{K},k)$ and are linearly dependent. Then

$$\boldsymbol{a}_1 = \kappa\boldsymbol{a}_2 \quad \text{and} \quad \boldsymbol{a}_1.\boldsymbol{K}.\boldsymbol{a}_1 = \boldsymbol{a}_2.\boldsymbol{K}.\boldsymbol{a}_2$$

so $\kappa^2 = 1$. Since $\boldsymbol{a}_1 \neq \boldsymbol{a}_2$, we must have $\boldsymbol{a}_1 = -\boldsymbol{a}_2$ and $\boldsymbol{a} = (\alpha_1-\alpha_2)\boldsymbol{a}_1$. Then

$$\boldsymbol{a}.\boldsymbol{K}.\boldsymbol{a} = (\alpha_1-\alpha_2)^2 k < k,$$

so again $\boldsymbol{a}$ is in $\mathscr{E}^0(\boldsymbol{K},k)$.

(c) *Hyperplane sections of ellipsoids*

Now for any non-zero vector $\boldsymbol{f}$ in $\mathscr{R}^N$ and any real number z we denote by $\mathscr{H}_z(\boldsymbol{f})$ the $(N-1)$ dimensional affine hyperplane consisting of all points $\boldsymbol{a}$ in $\mathscr{R}^N$ which satisfy $\boldsymbol{f}.\boldsymbol{a} = z$. Evidently $\mathscr{H}_z(\boldsymbol{f})$ is convex. We abbreviate $\mathscr{H}_1(\boldsymbol{f})$ by $\mathscr{H}(\boldsymbol{f})$. For any two sets $\mathscr{A}$ and $\mathscr{B}$ we denote by $\mathscr{A}\cap\mathscr{B}$ their set-theoretic intersection, consisting of all points common to both $\mathscr{A}$ and $\mathscr{B}$. Then we define

$$\left.\begin{aligned}\mathscr{E}(\boldsymbol{f},\boldsymbol{K},k) &= \mathscr{H}(\boldsymbol{f})\cap\mathscr{E}(\boldsymbol{K},k),\\ \partial\mathscr{E}(\boldsymbol{f},\boldsymbol{K},k) &= \mathscr{H}(\boldsymbol{f})\cap\partial\mathscr{E}(\boldsymbol{K},k),\\ \mathscr{E}^0(\boldsymbol{f},\boldsymbol{K},k) &= \mathscr{H}(\boldsymbol{f})\cap\mathscr{E}^0(\boldsymbol{K},k).\end{aligned}\right\} \tag{4.12}$$

The set $\mathscr{E}(\boldsymbol{f},\boldsymbol{K},k)$ consists of all points $\boldsymbol{a}$ satisfying (4.6) and $\boldsymbol{f}.\boldsymbol{a} = 1$, while $\partial\mathscr{E}(\boldsymbol{f},\boldsymbol{K},k)$ consists of all points $\boldsymbol{a}$ satisfying (4.7) and $\boldsymbol{f}.\boldsymbol{a} = 1$, and $\mathscr{E}^0(\boldsymbol{f},\boldsymbol{K},k)$ consists of all points $\boldsymbol{a}$ satisfying (4.8) and $\boldsymbol{f}.\boldsymbol{a} = 1$. Since the intersection of two convex sets is convex, both $\mathscr{E}(\boldsymbol{f},\boldsymbol{K},k)$ and $\mathscr{E}^0(\boldsymbol{f},\boldsymbol{K},k)$ are convex. Moreover, from Lemma 1 we have immediately

LEMMA 2. *If $k > 0$ and $\boldsymbol{K}$ is symmetric and positive definite and $\boldsymbol{f}$ is not the zero vector then $\mathscr{E}(\boldsymbol{f},\boldsymbol{K},k)$ is strictly convex.*

Intuitively, $\mathscr{E}(\boldsymbol{f},\boldsymbol{K},k)$ is an $(N-1)$-dimensional solid ellipsoid in $\mathscr{H}(\boldsymbol{f})$ and $\partial\mathscr{E}(\boldsymbol{f},\boldsymbol{K},k)$ is its $(N-2)$-dimensional boundary surface, while $\mathscr{E}^0(\boldsymbol{f},\boldsymbol{K},k)$ is its interior. For a fixed non-zero $\boldsymbol{f}$ and fixed positive definite, symmetric $\boldsymbol{K}$ we must study the family of ellipsoids $\mathscr{E}(\boldsymbol{f},\boldsymbol{K},k)$ parametrized by k. Geometrical intuition suggests that if k is sufficiently small then $\mathscr{E}(\boldsymbol{K},k)$ is confined so close to the origin that it has no points in common with $\mathscr{H}(\boldsymbol{f})$, but that as k increases from 0 it reaches a value $k_{\min}$ for which $\mathscr{E}(\boldsymbol{K},k_{\min})$ touches $\mathscr{H}(\boldsymbol{f})$ at a single point of tangency, $\boldsymbol{a}_{f,K}$. If $k \leqslant k_{\min}$, $\mathscr{E}^0(\boldsymbol{f},\boldsymbol{K},k)$ is empty, while if $k > k_{\min}$, $\mathscr{E}^0(\boldsymbol{f},\boldsymbol{K},k)$ is non-empty, has $\boldsymbol{a}_{f,K}$ as its centre, and grows with k.

It is easy to verify these conjectures. If $\mathscr{E}(\boldsymbol{K},k_{\min})$ is tangent to $\mathscr{H}(\boldsymbol{f})$ at $\boldsymbol{a}_{f,K}$ then $\boldsymbol{f}$ must be parallel to the normal to $\mathscr{E}(\boldsymbol{K},k_{\min})$ at $\boldsymbol{a}_{f,K}$. From (4.9) it follows that there must be a constant κ such that

$$\boldsymbol{K}.\boldsymbol{a}_{f,K} = \kappa\boldsymbol{f}. \tag{4.13}$$

Since $\boldsymbol{K}$ is positive definite, it has an inverse. Then, using $\boldsymbol{f}.\boldsymbol{a}_{f,K} = 1$ we can write immediately

$$\boldsymbol{a}_{f,K} = \frac{\boldsymbol{K}^{-1}.\boldsymbol{f}}{\boldsymbol{f}.\boldsymbol{K}^{-1}.\boldsymbol{f}} \tag{4.14}$$

and

$$k_{\min} = \boldsymbol{a}_{f,K}.\boldsymbol{K}.\boldsymbol{a}_{f,K} = \frac{1}{\boldsymbol{f}.\boldsymbol{K}^{-1}.\boldsymbol{f}}. \tag{4.15}$$

Now any $\boldsymbol{a}$ in $\mathscr{H}(\boldsymbol{f})$ can be written in the form

$$\boldsymbol{a} = \boldsymbol{a}_{f,K} + \boldsymbol{b}, \tag{4.16}$$

where

$$\boldsymbol{f}.\boldsymbol{b} = 0 \tag{4.17}$$

and $\boldsymbol{b}$ is uniquely determined by $\boldsymbol{a}$ and equations (4.16) and (4.17). From (4.14), (4.16) and (4.17) we infer immediately that

$$\boldsymbol{a}.\boldsymbol{K}.\boldsymbol{a} = k_{\min} + \boldsymbol{b}.\boldsymbol{K}.\boldsymbol{b}, \tag{4.18}$$

where $k_{\min}$ is given by (4.15). Because $\boldsymbol{K}$ is positive definite, it is now clear from (4.18) that $\mathscr{E}(\boldsymbol{f}, \boldsymbol{K}, k)$ is empty if $k < k_{\min}$, while $\mathscr{E}(\boldsymbol{f}, \boldsymbol{K}, k_{\min})$ consists of the single point $\boldsymbol{a}_{f,K}$, and if $k > k_{\min}$ then $\mathscr{E}^0(\boldsymbol{f}, \boldsymbol{K}, k)$ is non-empty and is centred on the point described by $\boldsymbol{b} = \boldsymbol{0}$, or $\boldsymbol{a} = \boldsymbol{a}_{f,K}$. It is also clear that when $\boldsymbol{a}$ is confined to $\mathscr{H}(\boldsymbol{f})$ then the least value of $\boldsymbol{a}.\boldsymbol{K}.\boldsymbol{a}$ is $k_{\min}$ and occurs at $\boldsymbol{a}_{f,K}$. Figure 1 depicts the geometry for $N = 3$.

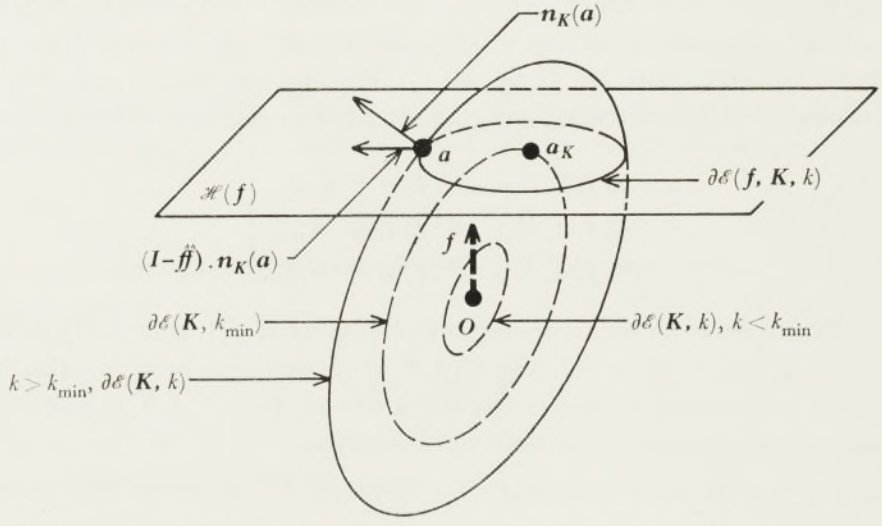

FIGURE 1. Illustration of the notation for hyperplane sections of ellipsoids in the case $N = 3$.

If $\boldsymbol{a}$ is any point on $\partial\mathscr{E}(\boldsymbol{K}, k)$, we know that $\boldsymbol{n}_K(\boldsymbol{a})$ in (4.9) is an outward normal to $\partial\mathscr{E}(\boldsymbol{K}, k)$ at $\boldsymbol{a}$. If $\boldsymbol{a}$ is on $\partial\mathscr{E}(\boldsymbol{f}, \boldsymbol{K}, k)$ then the orthogonal projection of $\boldsymbol{n}_K(\boldsymbol{a})$ onto $\mathscr{H}(\boldsymbol{f})$ is an outward normal to $\partial\mathscr{E}(\boldsymbol{f}, \boldsymbol{K}, k)$ lying in (i.e. tangent to) the hyperplane $\mathscr{H}(\boldsymbol{f})$. This orthogonal projection is

$$(\boldsymbol{I}-\hat{\boldsymbol{f}}\hat{\boldsymbol{f}}).\boldsymbol{n}_K(\boldsymbol{a}) = (\boldsymbol{I}-\hat{\boldsymbol{f}}\hat{\boldsymbol{f}}).\boldsymbol{K}.\boldsymbol{a}, \tag{4.19}$$

where $\boldsymbol{I}$ is the identity tensor in $\mathscr{R}^N$ and $\hat{\boldsymbol{f}}$ is $\boldsymbol{f}/(\boldsymbol{f}.\boldsymbol{f})^{\frac{1}{2}}$, the unit vector in the direction of $\boldsymbol{f}$. Figure 1 shows the geometrical situation for the special case $N = 3$.

All the foregoing remarks about $\mathscr{E}(\boldsymbol{f}, \boldsymbol{K}, k)$ apply verbatim to $\mathscr{E}(\boldsymbol{u}, \boldsymbol{S}, s)$ and $\mathscr{E}(\boldsymbol{u}, \boldsymbol{E}, \epsilon^2)$. Thus we have

$$\boldsymbol{a}_S = \frac{\boldsymbol{S}^{-1}.\boldsymbol{u}}{\boldsymbol{u}.\boldsymbol{S}^{-1}.\boldsymbol{u}}, \tag{4.20}$$

$$s_{\min} = \boldsymbol{a}_S . \boldsymbol{S} . \boldsymbol{a}_S = \frac{1}{\boldsymbol{u} . \boldsymbol{S}^{-1} . \boldsymbol{u}}, \tag{4.21}$$

$$(\boldsymbol{I} - \hat{\boldsymbol{u}}\hat{\boldsymbol{u}}) . \boldsymbol{n}_S(\boldsymbol{a}) = (\boldsymbol{I} - \hat{\boldsymbol{u}}\hat{\boldsymbol{u}}) . \boldsymbol{S} . \boldsymbol{a}, \tag{4.22}$$

and

$$\boldsymbol{a}_E = \frac{\boldsymbol{E}^{-1} . \boldsymbol{u}}{\boldsymbol{u} . \boldsymbol{E}^{-1} . \boldsymbol{u}}, \tag{4.23}$$

$$\epsilon^2_{\min} = \boldsymbol{a}_E . \boldsymbol{E} . \boldsymbol{a}_E = \frac{1}{\boldsymbol{u} . \boldsymbol{E}^{-1} . \boldsymbol{u}}, \tag{4.24}$$

$$(\boldsymbol{I} - \hat{\boldsymbol{u}}\hat{\boldsymbol{u}}) . \boldsymbol{n}_E(\boldsymbol{a}) = (\boldsymbol{I} - \hat{\boldsymbol{u}}\hat{\boldsymbol{u}}) . \boldsymbol{E} . \boldsymbol{a}. \tag{4.25}$$

In the foregoing equations and subsequently in this paper we abbreviate $\boldsymbol{a}_{u,S}$ and $\boldsymbol{a}_{u,E}$ as $\boldsymbol{a}_S$ and $\boldsymbol{a}_E$. It is understood that $\boldsymbol{S}$ depends on r_0, but we regard r_0 as fixed and so do not show this dependence explicitly.

(d) Elementary remarks about $\epsilon(s)$, the tradeoff curve between absolute error and spread

We can now reformulate in geometrical terms the problem of minimizing $\epsilon(a_1, \ldots, a_N)$, as defined by (3.5), subject to the constraints (2.22) and (3.6). We simply seek the greatest lower bound, $\epsilon(s)$, of the values of the function $\epsilon(\boldsymbol{a})$ defined by (4.5), when $\boldsymbol{a}$ is confined to $\mathscr{E}(\boldsymbol{u}, \boldsymbol{S}, s)$. If $s < s_{\min}$, we have seen that $\mathscr{E}(\boldsymbol{u}, \boldsymbol{S}, s)$ is empty, and we will agree that the greatest lower bound of the empty set is $+\infty$. There simply are no unimodular linear combinations of the data kernels whose spread from r_0 is less than $s_{\min}$.

If $s = s_{\min}$ then $\mathscr{E}(\boldsymbol{u}, \boldsymbol{S}, s_{\min})$ consists of the single point $\boldsymbol{a}_S$ which produces, via (2.5), the optimal averaging kernel A_{r_0} considered in Inverse II. The error variance for $\langle m_E, A_{r_0} \rangle$ is

$$\epsilon(s_{\min})^2 = \boldsymbol{a}_S . \boldsymbol{E} . \boldsymbol{a}_S,$$

which we abbreviate as $\epsilon_{\max}{}^2$.

If $s \geqslant s_{\min}$ then $\epsilon(s)^2$ is finite and non-negative. If $s_1 < s_2$ then $\mathscr{E}(\boldsymbol{u}, \boldsymbol{S}, s_1) \subseteq \mathscr{E}(\boldsymbol{u}, \boldsymbol{S}, s_2)$ so $\epsilon(s)$ is a monotonically non-increasing function of s. In particular, $\epsilon(s) \leqslant \epsilon_{\max}$ whenever $s \geqslant s_{\min}$. As s increases, $\mathscr{E}(\boldsymbol{u}, \boldsymbol{S}, s)$ expands outward from $\boldsymbol{a}_S$ in $\mathscr{H}(\boldsymbol{u})$, and $\epsilon(s)$ steadily decreases (so far we have proved only that it never increases). Eventually, s becomes large enough that $\partial\mathscr{E}(\boldsymbol{u}, \boldsymbol{S}, s)$ reaches $\boldsymbol{a}_E$. The value of s at which this occurs we call $s_{\max}$; clearly

$$s_{\max} = \boldsymbol{a}_E . \boldsymbol{S} . \boldsymbol{a}_E.$$

But $\boldsymbol{a}_E . \boldsymbol{E} . \boldsymbol{a}_E$ is $\epsilon^2_{\min}$, the least value which $\boldsymbol{a} . \boldsymbol{E} . \boldsymbol{a}$ takes anywhere in $\mathscr{H}(\boldsymbol{u})$. If $s \geqslant s_{\max}$ then $\boldsymbol{a}_E$ is in $\mathscr{E}(\boldsymbol{u}, \boldsymbol{S}, s)$, so the minimum value of $\boldsymbol{a} . \boldsymbol{E} . \boldsymbol{a}$ in $\mathscr{E}(\boldsymbol{u}, \boldsymbol{S}, s)$ is $\epsilon^2_{\min}$. If $s \geqslant s_{\max}$, then $\epsilon(s) = \epsilon_{\min}$. Willingness to accept averaging kernels A whose spread from r_0 is larger than $s_{\max}$ does not enable us to reduce the error variance in $\langle m_E, A \rangle$.

We can summarize these remarks as follows: we have

$$\left.\begin{aligned} s_{\min} &= \boldsymbol{a}_S . \boldsymbol{S} . \boldsymbol{a}_S = (\boldsymbol{u} . \boldsymbol{S}^{-1} . \boldsymbol{u})^{-1}, \\ s_{\max} &= \boldsymbol{a}_E . \boldsymbol{S} . \boldsymbol{a}_E = \frac{\boldsymbol{u} . \boldsymbol{E}^{-1} . \boldsymbol{S} . \boldsymbol{E}^{-1} . \boldsymbol{u}}{(\boldsymbol{u} . \boldsymbol{E}^{-1} . \boldsymbol{u})^2}, \\ \epsilon^2_{\min} &= \boldsymbol{a}_E . \boldsymbol{E} . \boldsymbol{a}_E = (\boldsymbol{u} . \boldsymbol{E}^{-1} . \boldsymbol{u})^{-1}, \\ \epsilon^2_{\max} &= \boldsymbol{a}_S . \boldsymbol{E} . \boldsymbol{a}_S = \frac{\boldsymbol{u} . \boldsymbol{S}^{-1} . \boldsymbol{E} . \boldsymbol{S}^{-1} . \boldsymbol{u}}{(\boldsymbol{u} . \boldsymbol{S}^{-1} . \boldsymbol{u})^2}. \end{aligned}\right\} \tag{4.26}$$

If $s < s_{\min}$, $\epsilon(s) = +\infty$. If $s = s_{\min}$, $\epsilon(s) = \epsilon_{\max}$. If $s_{\min} < s < s_{\max}$ then $\epsilon(s)$ is monotonically non-increasing and lies between $\epsilon_{\max}$ and $\epsilon_{\min}$. If $s \geqslant s_{\max}$, $\epsilon(s) = \epsilon_{\min}$.

Evidently the only interesting range of s is $s_{\min} \leqslant s \leqslant s_{\max}$. What if $s_{\min} = s_{\max}$? Then from (4.20), (4.23) and (4.26) we have

$$(\boldsymbol{a}_S . \boldsymbol{S} . \boldsymbol{a}_E)^2 = (\boldsymbol{a}_S . \boldsymbol{S} . \boldsymbol{a}_S)\,(\boldsymbol{a}_E . \boldsymbol{S} . \boldsymbol{a}_E).$$

Since $\boldsymbol{S}$ is positive definite, it has a Schwarz inequality (Halmos 1958), so this last equation implies that $\boldsymbol{a}_S$ and $\boldsymbol{a}_E$ are linearly dependent. Since they are both in $\mathscr{H}(\boldsymbol{u})$ they are equal. Similarly, $\epsilon_{\min} = \epsilon_{\max}$ if and only if $\boldsymbol{a}_E = \boldsymbol{a}_S$. In the very unlikely event that $\boldsymbol{a}_E = \boldsymbol{a}_S$, the presence of errors in the data will have no influence on our choice of an optimal averaging kernel (2.5). We will have $\epsilon(s) = \epsilon_{\min}$ for all $s \geqslant s_{\min}$, and for any $s \geqslant s_{\min}$ the optimal A is A_{r_0}, its coefficients in (2.5) being the components of $\boldsymbol{a}_S$.

In §§ 4 and 5 we will assume that $\boldsymbol{a}_S \neq \boldsymbol{a}_E$, so that $s_{\min} < s_{\max}$ and $\epsilon_{\min} < \epsilon_{\max}$.

(e) *The complete geometrical theory of* $\epsilon(s)$

We have already discussed $\epsilon(s)$ for all s except those in the open interval $s_{\min} < s < s_{\max}$. Now we consider this interval. We will prove from the geometry of hyperplane sections of ellipsoids that $\epsilon(s)$ is a continuous, strictly monotonic decreasing function on the closed interval $s_{\min} \leqslant s \leqslant s_{\max}$. We will also give a geometrical description of $\mathscr{E}(\boldsymbol{u}, \boldsymbol{S}, s)$ and $\mathscr{E}(\boldsymbol{u}, \boldsymbol{E}, \epsilon(s)^2)$ which enables us to calculate $\epsilon(s)$. All these results stem from

LEMMA 3. *If* $s_{\min} \leqslant s$ *then* $\mathscr{E}(\boldsymbol{u}, \boldsymbol{S}, s) \cap \mathscr{E}(\boldsymbol{u}, \boldsymbol{E}, \epsilon(s)^2)$ *contains exactly one point, which we denote by* $\boldsymbol{a}(s)$. *This point lies on* $\partial\mathscr{E}(\boldsymbol{u}, \boldsymbol{E}, \epsilon(s)^2)$. *When* $s_{\min} \leqslant s \leqslant s_{\max}$, $\boldsymbol{a}(s)$ *also lies on* $\partial\mathscr{E}(\boldsymbol{u}, \boldsymbol{S}, s)$, *and consequently is a point of external tangency of those two ellipsoids.*

Proof. Suppose that $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ are different points in $\mathscr{E}(\boldsymbol{u}, \boldsymbol{S}, s) \cap \mathscr{E}(\boldsymbol{u}, \boldsymbol{E}, \epsilon(s)^2)$. Let $\boldsymbol{a}' = \frac{1}{2}(\boldsymbol{a}_1 + \boldsymbol{a}_2)$. According to lemma 2, $\boldsymbol{a}'$ is in $\mathscr{E}^0(\boldsymbol{u}, \boldsymbol{S}, s) \cap \mathscr{E}^0(\boldsymbol{u}, \boldsymbol{E}, \epsilon(s)^2)$. But then the facts that $\boldsymbol{a}'$ is in $\mathscr{E}(\boldsymbol{u}, \boldsymbol{S}, s)$ and $\boldsymbol{a}' . \boldsymbol{E} . \boldsymbol{a}' < \epsilon(s)^2$ contradict the definition of $\epsilon(s)$. Thus

$$\mathscr{E}(\boldsymbol{u}, \boldsymbol{S}, s) \cap \mathscr{E}(\boldsymbol{u}, \boldsymbol{E}, \epsilon(s)^2)$$

contains at most one point. Now $\mathscr{E}(\boldsymbol{u}, \boldsymbol{S}, s)$ is compact and $\boldsymbol{a} . \boldsymbol{E} . \boldsymbol{a}$ depends continuously on $\boldsymbol{a}$, so there is at least one point in $\mathscr{E}(\boldsymbol{u}, \boldsymbol{S}, s)$ where $\boldsymbol{a} . \boldsymbol{E} . \boldsymbol{a}$ assumes its greatest lower bound, $\epsilon(s)^2$. Evidently this point lies in $\mathscr{E}(\boldsymbol{u}, \boldsymbol{S}, s) \cap \mathscr{E}(\boldsymbol{u}, \boldsymbol{E}, \epsilon(s)^2)$, so it is the unique point in that intersection and we may call it $\boldsymbol{a}(s)$ without ambiguity. Evidently $\boldsymbol{a}(s)$ is in $\partial\mathscr{E}(\boldsymbol{u}, \boldsymbol{E}, \epsilon(s)^2)$. When $s \geqslant s_{\max}$, clearly $\boldsymbol{a}(s) = \boldsymbol{a}_E$, but when $s < s_{\max}$ this is not so. If $\boldsymbol{a}(s) \neq \boldsymbol{a}_E$ then in every neighbourhood (relative to $\mathscr{H}(\boldsymbol{u})$) of $\boldsymbol{a}(s)$ there are points $\boldsymbol{a}'$ where $\boldsymbol{a}' . \boldsymbol{E} . \boldsymbol{a}' < \epsilon(s)^2$. If $\boldsymbol{a}(s)$ were in $\mathscr{E}^0(\boldsymbol{u}, \boldsymbol{S}, s)$, this open set would be a neighbourhood of $\boldsymbol{a}(s)$, and again we would have a contradiction of the definition of $\epsilon(s)^2$. Thus if $s_{\min} \leqslant s < s_{\max}$, $\boldsymbol{a}(s)$ is in $\partial\mathscr{E}(\boldsymbol{u}, \boldsymbol{S}, s)$. By the definition of $s_{\max}$, $\boldsymbol{a}(s)$ is in $\partial\mathscr{E}(\boldsymbol{u}, \boldsymbol{S}, s)$ when $s = s_{\max}$. This completes the proof of lemma 3.

Now we can prove

LEMMA 4. *Suppose that* $\mathscr{E}(\boldsymbol{u}, \boldsymbol{S}, s)$ *is externally tangent to* $\mathscr{E}(\boldsymbol{u}, \boldsymbol{E}, \epsilon^2)$. *Then* $s_{\min} \leqslant s \leqslant s_{\max}$ *and* $\epsilon_{\min} \leqslant \epsilon \leqslant \epsilon_{\max}$ *and* $\epsilon^2 = \epsilon(s)^2$, *so the point of external tangency is* $\boldsymbol{a}(s)$.

Proof. Since neither ellipsoid is empty, $s_{\min} \leqslant s$ and $\epsilon_{\min} \leqslant \epsilon$. If $s > s_{\max}$ then $\mathscr{E}(\boldsymbol{u}, \boldsymbol{S}, s)$ contains $\boldsymbol{a}_E$, the centre of $\mathscr{E}(\boldsymbol{u}, \boldsymbol{E}, \epsilon^2)$, so external tangency is impossible. A similar argument shows that $\epsilon \leqslant \epsilon_{\max}$. Next, since the two ellipsoids are strictly convex, an argument like the proof of uniqueness in lemma 3 shows that there is exactly one point of external tangency, say $\boldsymbol{a}'$. In the two degenerate cases $\boldsymbol{a}' = \boldsymbol{a}_S$ and $\boldsymbol{a}' = \boldsymbol{a}_E$ the theorem is obvious, so we assume that $\boldsymbol{a}'$ is neither $\boldsymbol{a}_S$ nor $\boldsymbol{a}_E$. Then $s_{\min} < s < s_{\max}$, and $\epsilon_{\min} < \epsilon < \epsilon_{\max}$. Because $\boldsymbol{a}' . \boldsymbol{E} . \boldsymbol{a}' = \epsilon^2$, evidently $\epsilon(s) \leqslant \epsilon$.

If $\epsilon(s) < \epsilon$, then $\mathscr{E}(\boldsymbol{u}, \boldsymbol{S}, s)$ and $\mathscr{E}(\boldsymbol{u}, \boldsymbol{S}, \epsilon(s)^2)$ have no common point, contrary to lemma 3. Hence $\epsilon = \epsilon(s)$ and, by lemma 3, $\boldsymbol{a}' = \boldsymbol{a}(s)$.

Lemmas 3 and 4 establish that our problem is symmetric with respect to $\boldsymbol{S}$ and $\boldsymbol{E}$. In particular then, if $\epsilon_{\min} \leqslant \epsilon \leqslant \epsilon_{\max}$ there is precisely one s in $s_{\min} \leqslant s \leqslant s_{\max}$ such that $\epsilon = \epsilon(s)$. Therefore, since $\epsilon(s)$ is non-increasing, it must be strictly decreasing and continuous on $s_{\min} \leqslant s \leqslant s_{\max}$. Figure 2 shows, for $N = 3$, the relation between the ellipsoids $\partial\mathscr{E}(\boldsymbol{u}, \boldsymbol{S}, s)$ and $\partial\mathscr{E}(\boldsymbol{u}, \boldsymbol{E}, \epsilon(s)^2)$ and the point $\boldsymbol{a}(s)$. The solid curve in that figure running from $\boldsymbol{a}_S$ to $\boldsymbol{a}_E$ is the path traced out in $\mathscr{H}(\boldsymbol{u})$ by $\boldsymbol{a}(s)$ as s increases from $s_{\min}$ to $s_{\max}$.

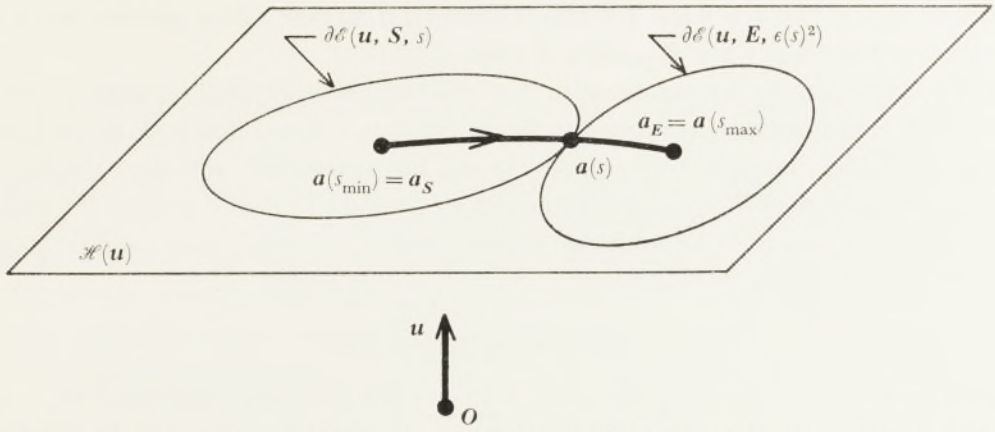

FIGURE 2. The path followed in $\mathscr{H}(\boldsymbol{u})$ by the coefficient vector $\boldsymbol{a}(s)$ of the optimal averaging kernel $A(r_0, r)$ as s, the spread of A from r_0, increases from $s_{\min}$ to $s_{\max}$ and ϵ, the absolute error in $\langle m_E, A\rangle$, decreases from $\epsilon_{\max}$ to $\epsilon_{\min}$.

5. THE ALGEBRA OF ABSOLUTE ERRORS

The external tangency of $\partial\mathscr{E}(\boldsymbol{u}, \boldsymbol{S}, s)$ and $\partial\mathscr{E}(\boldsymbol{u}, \boldsymbol{E}, \epsilon(s)^2)$ at $\boldsymbol{a}(s)$ gives us a means of calculating $\boldsymbol{a}(s)$ and hence $\epsilon(s)$.

(a) *Algebraic statement of the geometrical problem*

Suppose that $\partial\mathscr{E}(\boldsymbol{u}, \boldsymbol{S}, s)$ and $\partial\mathscr{E}(\boldsymbol{u}, \boldsymbol{E}, \epsilon^2)$ are externally tangent at $\boldsymbol{a}$. Then in the hyperplane $\mathscr{H}(\boldsymbol{u})$ the outward normals to those two ellipsoids at $\boldsymbol{a}$ must be antiparallel. Those outward normals are given by (4.22) and (4.25). Hence there is a *positive* constant α such that

$$\alpha(\boldsymbol{I} - \hat{\boldsymbol{u}}\hat{\boldsymbol{u}}) . \boldsymbol{E} . \boldsymbol{a} + (\boldsymbol{I} - \hat{\boldsymbol{u}}\hat{\boldsymbol{u}}) . \boldsymbol{S} . \boldsymbol{a} = 0. \tag{5.1}$$

If we define $\lambda = [\boldsymbol{u} . (\alpha\boldsymbol{E} + \boldsymbol{S}) . \boldsymbol{a}]/(\boldsymbol{u} . \boldsymbol{u})$ then (5.1) is

$$(\boldsymbol{S} + \alpha\boldsymbol{E}) . \boldsymbol{a} = \lambda\boldsymbol{u}. \tag{5.2}$$

We also have

$$\boldsymbol{a} . \boldsymbol{S} . \boldsymbol{a} = s \tag{5.3}$$

and

$$\boldsymbol{u} . \boldsymbol{a} = 1. \tag{5.4}$$

Equations (5.2), (5.3) and (5.4) are two scalar equations and a vector equation for the two scalar unknowns α and λ and the vector unknown $\boldsymbol{a}$. If $s_{\min} < s < s_{\max}$, then lemma 3 assures us that those equations have at least one solution $\alpha, \lambda, \boldsymbol{a}$, with $\alpha > 0$ and $\boldsymbol{a} = \boldsymbol{a}(s)$.

Conversely, suppose that for some s we have found a solution α, λ, $\boldsymbol{a}$ of (5.2), (5.3) and (5.4) which has $\alpha > 0$. Then we claim that $s_{\min} < s < s_{\max}$ and $\boldsymbol{a} = \boldsymbol{a}(s)$. To prove this converse, first we define $\epsilon^2 = \boldsymbol{a} . \boldsymbol{E} . \boldsymbol{a}$. Clearly, $\epsilon^2 \geqslant \epsilon_{\min}^2$. From (5.3), $\mathscr{E}(\boldsymbol{u}, \boldsymbol{S}, s)$ is non-empty, so $s \geqslant s_{\min}$. If $s = s_{\min}$ then (5.3) implies that $\boldsymbol{a} = \boldsymbol{a}_S = s_{\min}\boldsymbol{S}^{-1} . \boldsymbol{u}$. Then (5.2) becomes

$$\alpha\boldsymbol{E} . \boldsymbol{S}^{-1} . \boldsymbol{u} = (\lambda/s_{\min} - 1)\boldsymbol{u}.$$

Since $\alpha > 0$, the foregoing equation would imply the linear dependence of $\boldsymbol{a}_S$ and $\boldsymbol{a}_E$, and the equality of $s_{\min}$ and $s_{\max}$. This contradiction with our hypothesis shows that $s \neq s_{\min}$, so $s > s_{\min}$. Similarly, $\epsilon > \epsilon_{\min}$. Now we dot $\boldsymbol{I} - \hat{\boldsymbol{u}}\hat{\boldsymbol{u}}$ on the left of both sides of (5.2). The result is (5.1), and since $\alpha > 0$ this means that $\boldsymbol{a}$ is a point of *external* tangency of $\mathscr{E}(\boldsymbol{u}, \boldsymbol{S}, s)$ and $\mathscr{E}(\boldsymbol{u}, \boldsymbol{E}, \epsilon^2)$. It follows from lemma 4 that $\boldsymbol{a} = \boldsymbol{a}(s)$, $\epsilon = \epsilon(s)$, and $s_{\min} \leqslant s \leqslant s_{\max}$. Since we already know that $s_{\min} < s$ and $\epsilon_{\min} < \epsilon$ the strict monotonicity of $\epsilon(s)$ requires $s_{\min} < s < s_{\max}$. We summarize the foregoing remarks as

THEOREM 1. *Suppose* $s_{\min} < s_{\max}$. *Then for any* s *in* $s_{\min} < s < s_{\max}$, *equations* (5.2), (5.3) and (5.4) *have a solution* α, λ, $\boldsymbol{a}$ *with* $\alpha > 0$ *and* $\boldsymbol{a} = \boldsymbol{a}(s)$. *Conversely, if for some* s *those equations have a solution* α, λ, $\boldsymbol{a}$ *with* $\alpha > 0$ *then* $s_{\min} < s < s_{\max}$ *and* $\boldsymbol{a} = \boldsymbol{a}(s)$.

There is a simple alternative argument to show that $\boldsymbol{a}(s)$ satisfies (5.2) for some α and λ. We know that $\boldsymbol{a}(s)$ minimizes $\boldsymbol{a}.\boldsymbol{E}.\boldsymbol{a}$ subject to the constraints $\boldsymbol{u}.\boldsymbol{a} = 1$ and $\boldsymbol{a}.\boldsymbol{S}.\boldsymbol{a} \leqslant s$. Lemma 3 permits us to replace inequality by equality in the last constraint, and then the method of Lagrange multipliers gives (5.2). This argument does not show that among the $2N-2$ solutions α, λ, $\boldsymbol{a}$ of (5.2), (5.3) and (5.4) we want that solution with $\alpha > 0$, a fact which will lie at the root of our technique for solving those equations.

(*b*) *Solving the algebraic problem*

Since the solution of (5.2), (5.3) and (5.4) with $\alpha > 0$ is uniquely determined by s, we can write it as $\alpha(s)$, $\lambda(s)$, $\boldsymbol{a}(s)$. If we fix s and try to find α, λ and $\boldsymbol{a}$ the problem looks slightly more complicated than an ordinary eigenvalue problem. The structure of (5.2) suggests that the computation will be simpler if we regard α rather than s as the independent variable and try to find solutions $s(\alpha)$, $\lambda(\alpha)$, $\boldsymbol{a}(\alpha)$ of (5.2), (5.3) and (5.4). We know from theorem 1 that we will find at most one solution, and that if a solution exists it will satisfy $s_{\min} < s(\alpha) < s_{\max}$ and $\boldsymbol{a}(\alpha) = \boldsymbol{a}(s(\alpha))$. If we define

$$\epsilon(\alpha)^2 = \boldsymbol{a}(\alpha).\boldsymbol{E}.\boldsymbol{a}(\alpha) \tag{5.5}$$

then $\boldsymbol{a}(\alpha)$ will be the point of external tangency of $\partial\mathscr{E}(\boldsymbol{u}, \boldsymbol{E}, \epsilon(\alpha)^2)$ and $\partial\mathscr{E}(\boldsymbol{u}, \boldsymbol{S}, s(\alpha))$.

For which positive α can we find a solution $s(\alpha)$, $\lambda(\alpha)$, $\boldsymbol{a}(\alpha)$ to (5.2), (5.3) and (5.4)? It will shortly appear that the answer is any positive α whatever. First we note that in the sense of dimensional analysis (Bridgeman 1963) $\boldsymbol{S}$ and $\boldsymbol{E}$ have different physical dimensions. We choose a positive constant w such that, roughly speaking, $\boldsymbol{S}$ and $w\boldsymbol{E}$ will be of comparable numerical size. A particularly convenient choice of w will be described later, but w can be chosen at will. Having chosen it, we replace α by a new independent variable, θ, defined thus:

$$\alpha = w \tan\theta. \tag{5.6}$$

The domain of the independent variable θ is the finite interval $0 < \theta < \frac{1}{2}\pi$, corresponding to the infinite interval $0 < \alpha < \infty$.

If we write $\lambda(\theta)$ as $\beta(\theta)\sec\theta$ then (5.2) becomes

$$(\boldsymbol{S}\cos\theta + w\boldsymbol{E}\sin\theta).\boldsymbol{a}(\theta) = \beta(\theta)\boldsymbol{u}.$$

For any θ in the closed interval

$$0 \leqslant \theta \leqslant \tfrac{1}{2}\pi \tag{5.7}$$

we define the operator

$$\boldsymbol{W}(\theta) = \boldsymbol{S}\cos\theta + w\boldsymbol{E}\sin\theta. \tag{5.8}$$

Then equations (5.2), (5.3) and (5.4) take the very simple form

$$\boldsymbol{W}(\theta).\boldsymbol{a} = \beta\boldsymbol{u}; \tag{5.9}$$

$$\boldsymbol{u}.\boldsymbol{a} = 1, \tag{5.10}$$

$$s = \boldsymbol{a}.\boldsymbol{S}.\boldsymbol{a}. \tag{5.11}$$

The operator $\boldsymbol{W}(\theta)$ is symmetric; and so long as θ is in (5.7) the positiveness of w implies that $\boldsymbol{W}(\theta)$ is positive definite. Then $\boldsymbol{W}(\theta)$ has a positive definite inverse, $\boldsymbol{W}(\theta)^{-1}$. For any fixed θ in (5.7) we can write, from (5.9),

$$\boldsymbol{a}(\theta) = \beta(\theta)\, \boldsymbol{W}(\theta)^{-1}.\boldsymbol{u}.$$

Then from (5.10), $$\beta(\theta) = (\boldsymbol{u}.\boldsymbol{W}(\theta)^{-1}.\boldsymbol{u})^{-1}, \tag{5.12}$$

so $$\boldsymbol{a}(\theta) = \frac{\boldsymbol{W}(\theta)^{-1}.\boldsymbol{u}}{\boldsymbol{u}.\boldsymbol{W}(\theta)^{-1}.\boldsymbol{u}}. \tag{5.13}$$

Then $$s(\theta) = \boldsymbol{a}(\theta).\boldsymbol{S}.\boldsymbol{a}(\theta) \tag{5.14}$$

and $$\epsilon(\theta)^2 = \boldsymbol{a}(\theta).\boldsymbol{E}.\boldsymbol{a}(\theta). \tag{5.15}$$

Now the algebraic problem is solved. For any θ in (5.7), equations (5.9), (5.10) and (5.11) have a unique solution $s(\theta), \beta(\theta), \boldsymbol{a}(\theta)$. If $0 < \theta < \frac{1}{2}\pi$ then $s_{\min} < s(\theta) < s_{\max}$. Clearly, $\boldsymbol{a}(0) = \boldsymbol{a}_S$ and $\boldsymbol{a}(\frac{1}{2}\pi) = \boldsymbol{a}_E$, so $s(0) = s_{\min}$ and $s(\frac{1}{2}\pi) = s_{\max}$. Moreover, $\boldsymbol{W}(\theta)^{-1}$ is continuous and $\boldsymbol{u}.\boldsymbol{W}(\theta)^{-1}.\boldsymbol{u}$ is positive in $0 \leqslant \theta \leqslant \frac{1}{2}\pi$, so $\boldsymbol{a}(\theta)$, $s(\theta)$ and $\epsilon(\theta)^2$ are continuous in that interval. We conclude that as θ increases from 0 to $\frac{1}{2}\pi$, $\boldsymbol{a}(\theta)$ moves in $\mathscr{H}(\boldsymbol{u})$ continuously from $\boldsymbol{a}_S$ to $\boldsymbol{a}_E$ while $\epsilon(\theta)^2$ changes continuously from $\epsilon^2_{\max}$ to $\epsilon^2_{\min}$, $s(\theta)$ changes continuously from $s_{\min}$ to $s_{\max}$, and the pair $(s(\theta), \epsilon(\theta))$ trace out a parametric representation of the curve $\epsilon(s)$ for $s_{\min} \leqslant s \leqslant s_{\max}$.

We know from theorem 1 that for any s in $s_{\min} < s < s_{\max}$ there is exactly one θ in $0 < \theta < \frac{1}{2}\pi$ which permits a solution to (5.9), (5.10) and (5.11). Consequently the relation between θ and s is one-to-one, and $s(\theta)$ must be a monotonically increasing function of θ.

(c) *The shape of $\epsilon(s)$, the tradeoff curve of absolute error against spread*

Further information about the shape of $\epsilon(s)$ can be deduced by calculating $\partial_\theta s(\theta)$ and $\partial_\theta \epsilon(\theta)^2$, the derivatives of $s(\theta)$ and $\epsilon(\theta)^2$ with respect to θ. From (5.14) and (5.15),

$$\tfrac{1}{2}\partial_\theta s(\theta) = \boldsymbol{a}(\theta).\boldsymbol{S}.\partial_\theta \boldsymbol{a}(\theta), \tag{5.16}$$

$$\tfrac{1}{2}\partial_\theta [w\epsilon(\theta)^2] = \boldsymbol{a}(\theta).w\boldsymbol{E}.\partial_\theta \boldsymbol{a}(\theta). \tag{5.17}$$

To calculate $\partial_\theta \boldsymbol{a}(\theta)$ from (5.13) we must find $\partial_\theta[\boldsymbol{W}(\theta)^{-1}]$. If we apply ∂_θ to the equation $\boldsymbol{W}(\theta)^{-1}.\boldsymbol{W}(\theta) = \boldsymbol{I}$ we obtain

$$\partial_\theta[\boldsymbol{W}(\theta)^{-1}].\boldsymbol{W}(\theta) + \boldsymbol{W}(\theta)^{-1}.\partial_\theta \boldsymbol{W}(\theta) = \boldsymbol{0},$$

or $$\partial_\theta[\boldsymbol{W}(\theta)^{-1}] = -\boldsymbol{W}(\theta)^{-1}.\partial_\theta \boldsymbol{W}(\theta).\boldsymbol{W}(\theta)^{-1}. \tag{5.18}$$

Then from (5.13)

$$\partial_\theta \boldsymbol{a}(\theta) = \frac{\boldsymbol{W}^{-1}.\boldsymbol{u}(\boldsymbol{u}.\boldsymbol{W}^{-1}.\partial_\theta \boldsymbol{W}.\boldsymbol{W}^{-1}.\boldsymbol{u})}{(\boldsymbol{u}.\boldsymbol{W}^{-1}.\boldsymbol{u})^2} - \frac{\boldsymbol{W}^{-1}.\partial_\theta \boldsymbol{W}.\boldsymbol{W}^{-1}.\boldsymbol{u}}{(\boldsymbol{u}.\boldsymbol{W}^{-1}.\boldsymbol{u})}. \tag{5.19}$$

From (5.13) it is clear that $[\boldsymbol{a}(\theta).\boldsymbol{W}(\theta).\boldsymbol{a}(\theta)][\boldsymbol{u}.\boldsymbol{W}(\theta)^{-1}.\boldsymbol{u}] = 1$, so, for any θ,

$$\boldsymbol{u} = \frac{\boldsymbol{W}(\theta).\boldsymbol{a}(\theta)}{\boldsymbol{a}(\theta).\boldsymbol{W}(\theta).\boldsymbol{a}(\theta)}.$$

Thus we can write (5.19) as

$$\partial_\theta \boldsymbol{a} = \frac{(\boldsymbol{a}.\partial_\theta \boldsymbol{W}.\boldsymbol{a})\boldsymbol{a} - (\boldsymbol{a}.\boldsymbol{W}.\boldsymbol{a})\boldsymbol{W}^{-1}.\partial_\theta \boldsymbol{W}.\boldsymbol{a}}{\boldsymbol{a}.\boldsymbol{W}.\boldsymbol{a}}. \tag{5.20}$$

Now by differentiating the definition (5.8) with respect to θ and multiplying first by $\sin\theta$ and then by $\cos\theta$, we obtain

$$\left.\begin{aligned} \sin\theta\, \partial_\theta \boldsymbol{W}(\theta) &= \boldsymbol{W}(\theta)\cos\theta - \boldsymbol{S}, \\ \cos\theta\, \partial_\theta \boldsymbol{W}(\theta) &= w\boldsymbol{E} - \boldsymbol{W}(\theta)\sin\theta. \end{aligned}\right\} \tag{5.21}$$

Equations (5.20) and (5.21) imply

$$\sin\theta\,\partial_\theta \boldsymbol{a} = \frac{(\boldsymbol{a}.\boldsymbol{W}.\boldsymbol{a})\,\boldsymbol{W}^{-1}.\boldsymbol{S}.\boldsymbol{a} - (\boldsymbol{a}.\boldsymbol{S}.\boldsymbol{a})\boldsymbol{a}}{\boldsymbol{a}.\boldsymbol{W}.\boldsymbol{a}}, \tag{5.22}$$

$$\cos\theta\,\partial_\theta \boldsymbol{a} = \frac{(\boldsymbol{a}.w\boldsymbol{E}.\boldsymbol{a})\boldsymbol{a} - (\boldsymbol{a}.\boldsymbol{W}.\boldsymbol{a})\,\boldsymbol{W}^{-1}.w\boldsymbol{E}.\boldsymbol{a}}{\boldsymbol{a}.\boldsymbol{W}.\boldsymbol{a}}. \tag{5.23}$$

Then from (5.16) and (5.17) we have

$$\tfrac{1}{2}\sin\theta\,\partial_\theta s(\theta) = \frac{(\boldsymbol{a}.\boldsymbol{S}.\boldsymbol{W}^{-1}.\boldsymbol{S}.\boldsymbol{a})\,(\boldsymbol{a}.\boldsymbol{W}.\boldsymbol{a}) - (\boldsymbol{a}.\boldsymbol{S}.\boldsymbol{a})^2}{\boldsymbol{a}.\boldsymbol{W}.\boldsymbol{a}}, \tag{5.24}$$

$$\tfrac{1}{2}\cos\theta\,\partial_\theta s(\theta) = \frac{(\boldsymbol{a}.\boldsymbol{S}.\boldsymbol{a})\,(\boldsymbol{a}.w\boldsymbol{E}.\boldsymbol{a}) - (\boldsymbol{a}.\boldsymbol{W}.\boldsymbol{a})\,(\boldsymbol{a}.\boldsymbol{S}.\boldsymbol{W}^{-1}.w\boldsymbol{E}.\boldsymbol{a})}{\boldsymbol{a}.\boldsymbol{W}.\boldsymbol{a}}, \tag{5.25}$$

$$\tfrac{1}{2}\sin\theta\,\partial_\theta[w\epsilon(\theta)^2] = \frac{(\boldsymbol{a}.w\boldsymbol{E}.\boldsymbol{W}^{-1}.\boldsymbol{S}.\boldsymbol{a})\,(\boldsymbol{a}.\boldsymbol{W}.\boldsymbol{a}) - (\boldsymbol{a}.\boldsymbol{S}.\boldsymbol{a})\,(\boldsymbol{a}.w\boldsymbol{E}.\boldsymbol{a})}{\boldsymbol{a}.\boldsymbol{W}.\boldsymbol{a}}, \tag{5.26}$$

$$\tfrac{1}{2}\cos\theta\,\partial_\theta[w\epsilon(\theta)^2] = \frac{(\boldsymbol{a}.w\boldsymbol{E}.\boldsymbol{a})^2 - (\boldsymbol{a}.\boldsymbol{W}.\boldsymbol{a})\,(\boldsymbol{a}.w\boldsymbol{E}.\boldsymbol{W}^{-1}.w\boldsymbol{E}.\boldsymbol{a})}{\boldsymbol{a}.\boldsymbol{W}.\boldsymbol{a}}. \tag{5.27}$$

Geometrical arguments have already led us to the conclusion that $\partial_\theta s(\theta) > 0$ if $0 < \theta < \frac{1}{2}\pi$. Schwarz's inequality for the positive definite operator $\boldsymbol{W}^{-1}$, applied to the vectors $\boldsymbol{W}.\boldsymbol{a}$ and $\boldsymbol{S}.\boldsymbol{a}$, immediately enables us to conclude from (5.24) that $\partial_\theta s(\theta) > 0$ if $0 < \theta \leqslant \frac{1}{2}\pi$, while (5.25) with $\boldsymbol{a} = \boldsymbol{a}_S$ shows that $\partial_\theta s(0) = 0$. Geometry has also led us to conclude that $\epsilon(s)$ is monotone decreasing, so $\partial_\theta[w\epsilon(\theta)^2] < 0$ if $0 < \theta < \frac{1}{2}\pi$. From (5.27) we verify this inequality for $0 \leqslant \theta < \frac{1}{2}\pi$, and we see that $\partial_\theta[w\epsilon(\frac{1}{2}\pi)^2] = 0$. It follows that the graph of $\epsilon^2(s)$ has a vertical tangent at $(s_{\min}, \epsilon^2_{\max})$ and a horizontal tangent at $(s_{\max}, \epsilon^2_{\min})$.

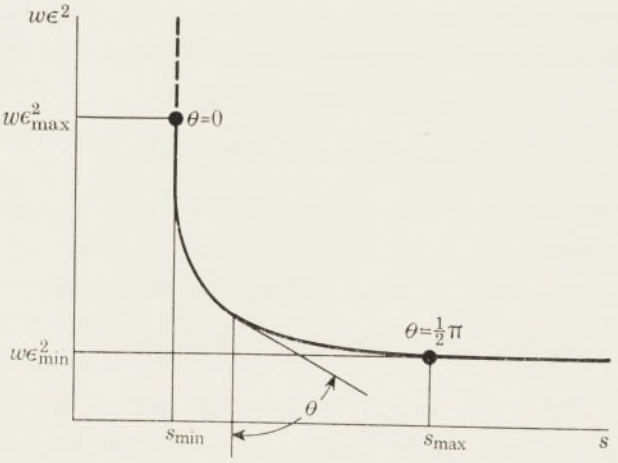

FIGURE 3. A schematic picture of a typical tradeoff curve of absolute error against spread. The factor w is chosen as described in the text, so that $w\epsilon^2$ has the same dimensions as s, that is, a length.

Comparison of (5.25) and (5.26) shows that

$$\cos\theta\,\partial_\theta s(\theta) + \sin\theta\,\partial_\theta[w\epsilon(\theta)^2] = 0, \tag{5.28}$$

or

$$\frac{\mathrm{d}(w\epsilon^2)}{\mathrm{d}s} = -\cot\theta. \tag{5.29}$$

Therefore if we plot the curve of $w\epsilon^2$ against s, then θ is the angle between the tangent to this curve and the negative $w\epsilon^2$ axis. Since θ is a monotonic increasing function of s, it follows that the

graph of $\epsilon(s)^2$ is convex toward the origin. That is, $\mathrm{d}^2[\epsilon(s)^2]/\mathrm{d}s^2 > 0$. As we have already seen, this curve has a vertical tangent at $s_{\min}$ and a horizontal tangent at $s_{\max}$; its appearance is schematically that sketched in figure 3. Evidently a very small loss of resolution near $s_{\min}$ leads to a very great improvement in the error, and a very small increase in the error near $\epsilon_{\min}$ leads to a very great improvement in resolution. In general, we would expect to find it advantageous to work well away from either end of the tradeoff curve.

(*d*) *The choice of w*

If w is not carefully chosen, most of the absolute error tradeoff curve, $\epsilon(s)$, will have θ very close to 0 or to $\frac{1}{2}\pi$. To avoid this possibility, we choose w so that equal changes in θ produce equal dimensionless displacements along the tradeoff curve at $\theta = 0$ and $\theta = \frac{1}{2}\pi$. That is, we choose w so that

$$\frac{\partial_\theta s(\frac{1}{2}\pi)}{s_{\max} - s_{\min}} = -\frac{\partial_\theta[\epsilon(0)^2]}{\epsilon_{\max}^2 - \epsilon_{\min}^2}. \tag{5.30}$$

From (5.24) and (5.27) we have

$$\partial_\theta s(\tfrac{1}{2}\pi) = \frac{2(M-1)\, s_{\max}^2}{w\epsilon_{\min}^2} \quad \text{and} \quad \partial_\theta[\epsilon(0)^2] = -\frac{2w(N-1)\,\epsilon_{\max}^4}{s_{\min}},$$

where $$M = \frac{\epsilon_{\min}^2}{s_{\max}^2}(\boldsymbol{a}_E.\boldsymbol{S}.\boldsymbol{E}^{-1}.\boldsymbol{S}.\boldsymbol{a}_E) \quad \text{and} \quad N = \frac{s_{\min}}{\epsilon_{\max}^4}(\boldsymbol{a}_S.\boldsymbol{E}.\boldsymbol{S}^{-1}.\boldsymbol{E}.\boldsymbol{a}_S).$$

Then $$w = \frac{s_{\max}}{\epsilon_{\max}^2}\left(\frac{M-1}{N-1}\right)^{\frac{1}{2}}\left[\frac{(\epsilon_{\max}^2/\epsilon_{\min}^2)-1}{(s_{\max}/s_{\min})-1}\right]^{\frac{1}{2}}. \tag{5.31}$$

With this choice of w, we have

$$\begin{aligned}\frac{\partial_\theta s(\frac{1}{2}\pi)}{s_{\max} - s_{\min}} &= -\frac{\partial_\theta[\epsilon(0)^2]}{\epsilon_{\max}^2 - \epsilon_{\min}^2} \\ &= 2\frac{s_{\max}\epsilon_{\max}^2}{s_{\min}\epsilon_{\min}^2}\left\{\frac{(M-1)\,(N-1)}{[(s_{\max}/s_{\min})-1]\,[(\epsilon_{\max}^2/\epsilon_{\min}^2)-1]}\right\}^{\frac{1}{2}}.\end{aligned} \tag{5.32}$$

The foregoing choice of w is convenient but not necessary. The error tradeoff curve $\epsilon(s)$ and the vector-valued function $\boldsymbol{a}(s)$ are independent of w. The choice of w simply governs the parametrization of s by θ.

(*e*) *Summary on optimizing averaging kernels to reduce absolute error*

Now we summarize our method for calculating the absolute error tradeoff curve $\epsilon(s)$, the vector $\boldsymbol{a}(s)$ and the corresponding optimal averaging kernel A_s in (2.5).

We fix r_0 and choose a constant w_{r_0} in any way we please, one convenient choice being (5.31). Then for each θ in the closed interval $0 \leqslant \theta \leqslant \frac{1}{2}\pi$ we calculate $\boldsymbol{W}_{r_0}(\theta)$ from (5.8), $\boldsymbol{a}_{r_0}(\theta)$ from (5.13), $s_{r_0}(\theta)$ from (5.14) and $\epsilon_{r_0}(\theta)$ from (5.15). The vector $\boldsymbol{a}_{r_0}(\theta)$ is the only point common to $\partial\mathscr{E}(\boldsymbol{u}, \boldsymbol{S}(r_0), s_{r_0}(\theta))$ and $\partial\mathscr{E}(\boldsymbol{u}, \boldsymbol{E}, \epsilon_{r_0}(\theta)^2)$, and is the only point of external tangency of those two $(N-2)$-dimensional ellipsoidal hypersurfaces in $\mathscr{H}(\boldsymbol{u})$. The minimum value of $\boldsymbol{a}.\boldsymbol{E}.\boldsymbol{a}$ in $\mathscr{E}(\boldsymbol{u}, \boldsymbol{S}(r_0), s_{r_0}(\theta))$ occurs at $\boldsymbol{a}_{r_0}(\theta)$ and is $\epsilon_{r_0}(\theta)^2$. The minimum value of $\boldsymbol{a}.\boldsymbol{S}(r_0).\boldsymbol{a}$ in $\mathscr{E}(\boldsymbol{u}, \boldsymbol{E}, \epsilon_{r_0}(\theta)^2)$ occurs at $\boldsymbol{a}_{r_0}(\theta)$ and is $s_{r_0}(\theta)$. As θ increases from 0 to $\frac{1}{2}\pi$, $\boldsymbol{a}_{r_0}(\theta)$ moves continuously (but in general not in a straight line) from $\boldsymbol{a}_S(r_0)$ to $\boldsymbol{a}_E$, while $s_{r_0}(\theta)$ increases monotonically from $s_{\min}(r_0)$ to $s_{\max}(r_0)$ and $\epsilon_{r_0}(\theta)$ decreases monotonically from $\epsilon_{\max}(r_0)$ to $\epsilon_{\min}$. The curve of $w_{r_0}\epsilon_{r_0}(s)^2$ against s

is convex toward the origin, vertical at $s_{\min}(r_0)$, and horizontal at $s_{\max}(r_0)$, and its tangent at any point makes the angle θ with the negative axis of $w\epsilon^2$.

The complete symmetry between the pair $\boldsymbol{E}, \epsilon^2$ and the pair $\boldsymbol{S}, s$ is now apparent. As suggested at the end of § 3*b*, we can obtain our optimal averaging kernels by finding the minimum, $s(\epsilon^2)$, of $\boldsymbol{a}.\boldsymbol{S}.\boldsymbol{a}$ in $\mathscr{E}(\boldsymbol{u}, \boldsymbol{E}, \epsilon^2)$ just as well as by finding the minimum, $\epsilon(s)^2$, of $\boldsymbol{a}.\boldsymbol{E}.\boldsymbol{a}$ in $\mathscr{E}(\boldsymbol{u}, \boldsymbol{S}, s)$. When $s_{\min} \leqslant s \leqslant s_{\max}$ and $\epsilon_{\min} \leqslant \epsilon \leqslant \epsilon_{\max}$ the two functions $\epsilon(s)^2$ and $s(\epsilon^2)$ are inverse to one another. When $s \geqslant s_{\max}$, then $\epsilon(s) = \epsilon_{\min}$, while when $\epsilon \geqslant \epsilon_{\max}$ then $s(\epsilon^2) = s_{\min}$. Therefore we include the dashed vertical line in figure 3 as part of the graph of $\epsilon(s)^2$.

For any particular θ in $0 \leqslant \theta \leqslant \frac{1}{2}\pi$, the N components $a_{r_0, i}(\theta)$ of $\boldsymbol{a}_{r_0}(\theta)$ generate, via (2.5), that unimodular linear combination $A_{r_0, \theta}(r)$ of the data kernels which permits $\langle m_{\mathrm{E}}, A_{r_0, \theta}\rangle$ to be calculated from the data with a smaller absolute error variance than is possible for any other averaging kernel (2.5) whose spread from r_0 is as small as that of $A_{r_0, \theta}$. The spread of $A_{r_0, \theta}$ from r_0 is $s_{r_0}(\theta)$, and the error variance in $\langle m_{\mathrm{E}}, A_{r_0, \theta}\rangle$ is $\epsilon_{r_0}(\theta)^2$.

It is important to note that $\boldsymbol{E}$ and $\boldsymbol{u}$ and hence $\epsilon_{\min}$ and $\boldsymbol{a}_E$ are independent of r_0. Therefore at $\theta = \frac{1}{2}\pi$, $A_{r_0, \theta}$ is independent of r_0; in fact $A_{r_0, \frac{1}{2}\pi} = A_E$, the averaging kernel (2.5) obtained from the components of $\boldsymbol{a}_E$. The averaging kernel A_E is that unimodular linear combination A of the data kernels which minimizes the error variance of $\langle m_{\mathrm{E}}, A\rangle$. Among all averages $\langle m_{\mathrm{E}}, A\rangle$ which can be calculated from the given gross Earth data, $\langle m_{\mathrm{E}}, A_E\rangle$ is the one which is most accurately known. If we define

$$\boldsymbol{q} = (q_1, \ldots, q_N), \tag{5.33}$$

where q_i is given by (2.16), then

$$\langle m_{\mathrm{E}}, A_E\rangle = \boldsymbol{q}.\boldsymbol{a}_E. \tag{5.34}$$

Although this particular average may be very accurately known, usually it will be not at all localized. As we decrease θ to obtain more localized averages $\langle m_{\mathrm{E}}, A_{r_0, \theta}\rangle$, we are forced to accept an increase in the error variance of those averages.

C. RELATIVE ERRORS

6. The geometry of relative errors

The Earth model describing the real Earth is $m_{\mathrm{E}}(r)$. As remarked in § 3*c*, it is conceivable that we would be more concerned about the relative than the absolute error in our estimate of the average $\langle m_{\mathrm{E}}, A\rangle$. Under certain circumstances we would expect that our optimal averaging kernels would be very much the same, whether we obtained them by minimizing the absolute or the relative error at a given spread. In appendix C we describe the circumstances under which the optimal averaging kernels for relative and for absolute error are significantly different. In §§ 6 and 7, we simply discuss the problem of minimizing relative error for a given spread.

(*a*) *The geometrical statement of the problem*

Suppose that m is a $\mathscr{G}$-acceptable Earth model and $G_1, \ldots, G_N$ are the data kernels at m. If $A(r) = \Sigma a_i G_i(r)$ and $\boldsymbol{a} = (a_1, \ldots, a_N)$, and if $\boldsymbol{q}$ is defined by (5.33), then

$$\langle m_{\mathrm{E}}, A\rangle = \boldsymbol{q}.\boldsymbol{a}. \tag{6.1}$$

If all the gross Earth functionals in $\mathscr{G}$ are linear, the only error in the assertion (6.1) is that produced by the errors in the gross Earth data. If some functionals in $\mathscr{G}$ are non-linear, then (6.1) is also in error by a term of order $(m_{\mathrm{E}} - m)^2$ which we have agreed to ignore in the present

paper. Thus the absolute error we commit be using (6.1) to calculate $\langle m_E, A\rangle$ has variance $\epsilon(\boldsymbol{a})^2 = \boldsymbol{a}.\boldsymbol{E}.\boldsymbol{a}$. We can take as a measure of the relative error $\rho(\boldsymbol{a})$ the expression $\epsilon(\boldsymbol{a})/|\boldsymbol{q}.\boldsymbol{a}|$. Then we have

$$\rho(\boldsymbol{a})^2 = \frac{\boldsymbol{a}.\boldsymbol{E}.\boldsymbol{a}}{(\boldsymbol{q}.\boldsymbol{a})^2}. \tag{6.2}$$

For a given s, we seek the averaging kernel A of form (2.5) which minimizes ρ subject to the constraints (2.22) and (3.6). Thus we seek the point $\boldsymbol{a}$ at which $\rho(\boldsymbol{a})^2$ attains its least value in $\mathscr{E}(\boldsymbol{u}, \boldsymbol{S}, s)$. It is clear from (6.2) that this problem is unaltered if we replace $\boldsymbol{q}$ by $-\boldsymbol{q}$; without loss of generality we may and will assume that

$$\boldsymbol{q}.\boldsymbol{a}_E \geqslant 0. \tag{6.3}$$

Equation (5.34) makes clear that in (6.3) we are simply agreeing to use a sign convention for Earth models which gives a non-negative value to that average of m which is known most accurately.

(*b*) *The error cones*

For a fixed positive ρ we are interested in the set of all points $\boldsymbol{a}$ for which $\rho(\boldsymbol{a}) \leqslant \rho$. We denote by $\mathscr{C}(\boldsymbol{q}, \boldsymbol{E}, \rho^2)$ the set of all points $\boldsymbol{a}$ in $\mathscr{R}^N$ which satisfy

$$\boldsymbol{a}.(\boldsymbol{E} - \rho^2\boldsymbol{q}\boldsymbol{q}).\boldsymbol{a} \leqslant 0. \tag{6.4}$$

We call this set the 'error double cone for relative error ρ'. It plays the same role in the theory of relative errors as does the solid error ellipsoid $\mathscr{E}(\boldsymbol{E}, \epsilon^2)$ in the theory of absolute errors. The set of points $\boldsymbol{a}$ for which the inequality (6.4) is strict we call the interior of $\mathscr{C}(\boldsymbol{q}, \boldsymbol{E}, \rho^2)$, denoted $\mathscr{C}^0(\boldsymbol{q}, \boldsymbol{E}, \rho^2)$. The set of points $\boldsymbol{a}$ for which equality holds in (6.4) we call the boundary of $\mathscr{C}(\boldsymbol{q}, \boldsymbol{E}, \rho^2)$, denoted $\partial\mathscr{C}(\boldsymbol{q}, \boldsymbol{E}, \rho^2)$.

Study of the error cone is facilitated by introducing a new geometry into $\mathscr{R}^N$. Lengths and angles in this new geometry are to be calculated from the new inner product

$$(\boldsymbol{f}, \boldsymbol{g})_E = \boldsymbol{f}.\boldsymbol{E}.\boldsymbol{g}, \tag{6.5}$$

where $\boldsymbol{f}$ and $\boldsymbol{g}$ are arbitrary vectors in $\mathscr{R}^N$. Since $\boldsymbol{E}$ is positive definite, (6.5) does define an inner product on $\mathscr{R}^N$. In this new geometry, the length of a vector $\boldsymbol{f}$ is defined as

$$\|\boldsymbol{f}\|_E = (\boldsymbol{f}, \boldsymbol{f})_E^{\frac{1}{2}},$$

while the angle between $\boldsymbol{f}$ and $\boldsymbol{g}$, written $\angle_E(\boldsymbol{f}, \boldsymbol{g})$, is that angle between 0 and π such that

$$\cos\angle_E(\boldsymbol{f}, \boldsymbol{g}) = \frac{(\boldsymbol{f}, \boldsymbol{g})_E}{\|\boldsymbol{f}\|_E \|\boldsymbol{g}\|_E}.$$

Then $\mathscr{C}(\boldsymbol{q}, \boldsymbol{E}, \rho^2)$ consists of those points $\boldsymbol{a}$ such that

$$\|\boldsymbol{a}\|_E^2 \leqslant \rho^2(\boldsymbol{a}, \boldsymbol{E}^{-1}.\boldsymbol{q})_E^2.$$

When $\boldsymbol{a} \neq \boldsymbol{0}$ we can write this inequality as

$$[\cos\angle_E(\boldsymbol{a}, \boldsymbol{E}^{-1}.\boldsymbol{q})]^2 \geqslant \rho^{-2}\|\boldsymbol{E}^{-1}.\boldsymbol{q}\|_E^{-2}.$$

We define

$$\rho_{\min}^2 = (\boldsymbol{q}.\boldsymbol{E}^{-1}.\boldsymbol{q})^{-1}, \tag{6.6}$$

and we define $\psi(\rho)$ as that angle between 0 and $\frac{1}{2}\pi$ such that

$$\cos\psi(\rho) = \frac{\rho_{\min}}{\rho}. \tag{6.7}$$

Then $\mathscr{C}(\boldsymbol{q}, \boldsymbol{E}, \rho^2)$ consists of the origin and all non-zero vectors $\boldsymbol{a}$ such that $\angle_E(\boldsymbol{a}, \boldsymbol{E}^{-1}.\boldsymbol{q})$ is between 0 and $\psi(\rho)$ or between $\pi - \psi(\rho)$ and π. In the geometry defined on $\mathscr{R}^N$ by (6.5), $\mathscr{C}(\boldsymbol{q}, \boldsymbol{E}, \rho^2)$ is a solid, N-dimensional, right-circular double cone with vertex $\mathbf{0}$, axis along $\boldsymbol{E}^{-1}.\boldsymbol{q}$, and vertex half angle $\psi(\rho)$.

Now we fix $\boldsymbol{q}$ and $\boldsymbol{E}$ and consider the family of double cones $\mathscr{C}(\boldsymbol{q}, \boldsymbol{E}, \rho^2)$ as ρ varies. If $\rho < \rho_{\min}$ then $\mathscr{C}(\boldsymbol{q}, \boldsymbol{E}, \rho^2)$ contains only the origin. If $\rho = \rho_{\min}$, $\mathscr{C}(\boldsymbol{q}, \boldsymbol{E}, \rho^2)$ is the axis of the family of double cones, i.e. the straight line through $\mathbf{0}$ which consists of all scalar multiples of $\boldsymbol{E}^{-1}.\boldsymbol{q}$. If $\rho > \rho_{\min}$ then $\mathscr{C}^0(q, E, \rho^2)$ is non-empty. As ρ approaches infinity, $\partial\mathscr{C}(\boldsymbol{q}, \boldsymbol{E}, \rho^2)$ approaches (from both sides) the hyperplane perpendicular to the axis $\boldsymbol{E}^{-1}.\boldsymbol{q}$ and passing through $\mathbf{0}$. This hyperplane, which we denote by $\mathscr{H}_0(\boldsymbol{q})$, consists of all points $\boldsymbol{a}$ such that $(\boldsymbol{a}, \boldsymbol{E}^{-1}.\boldsymbol{q})_E = 0$, or, equivalently $\boldsymbol{a}.\boldsymbol{q} = 0$.

Now suppose $\rho > \rho_{\min}$ so that $\partial\mathscr{C}(\boldsymbol{q}, \boldsymbol{E}, \rho^2)$ is an $(N-1)$-dimensional hypersurface. Let $\boldsymbol{a}$ be any non-zero point on $\partial\mathscr{C}(\boldsymbol{q}, \boldsymbol{E}, \rho^2)$. We would like to have a non-zero vector $\boldsymbol{n}_{q,E}(\boldsymbol{a})$ which is normal to $\partial\mathscr{C}(\boldsymbol{q}, \boldsymbol{E}, \rho^2)$ at $\boldsymbol{a}$ and points out of $\mathscr{C}(\boldsymbol{q}, \boldsymbol{E}, \rho^2)$. Since the function $\boldsymbol{a}.(\boldsymbol{E} - \rho^2\boldsymbol{q}\boldsymbol{q}).\boldsymbol{a}$ is negative in $\mathscr{C}^0(\boldsymbol{q}, \boldsymbol{E}, \rho^2)$ and positive outside $\mathscr{C}(\boldsymbol{q}, \boldsymbol{E}, \rho^2)$, we can take half its gradient with respect to $\boldsymbol{a}$ as the desired outward normal:

$$\boldsymbol{n}_{q,E}(\boldsymbol{a}) = (\boldsymbol{E} - \rho^2\boldsymbol{q}\boldsymbol{q}).\boldsymbol{a}. \tag{6.8}$$

But we must verify that $\boldsymbol{n}_{q,E}(\boldsymbol{a}) \neq \mathbf{0}$. In the contrary case we would have

$$\boldsymbol{a} = \rho^2(\boldsymbol{q}.\boldsymbol{a})\,\boldsymbol{E}^{-1}.\boldsymbol{q}.$$

Since $\boldsymbol{a} \neq \mathbf{0}$, it would follow that $\boldsymbol{q}.\boldsymbol{a} \neq \mathbf{0}$. Then dotting $\boldsymbol{q}$ into the foregoing equation and cancelling $\boldsymbol{q}.\boldsymbol{a}$ gives $\rho = \rho_{\min}$, contrary to hypothesis. (*Note.* The geometry in which (6.8) is normal to $\partial\mathscr{C}(\boldsymbol{q}, \boldsymbol{E}, \rho^2)$ is that of the inner product (4.1), not (6.5). In the geometry of (6.5) the normal is $\boldsymbol{E}^{-1}.\boldsymbol{n}_{q,E}(\boldsymbol{a})$. We will later need the normal at $\boldsymbol{a}$ only to define the tangent hyperplane to $\partial\mathscr{C}(\boldsymbol{q}, \boldsymbol{E}, \rho^2)$ at $\boldsymbol{a}$. This hyperplane is independent of the geometry, so we will use the simpler geometry to discuss it.)

In our discussion of absolute errors a crucial role was played by the strict convexity of the error ellipsoid $\mathscr{E}(\boldsymbol{E}, e^2)$. The error double cone $\mathscr{C}(\boldsymbol{q}, \boldsymbol{E}, \rho^2)$ is not even convex, but it consists of two cones which are. By the 'positive error cone', $\mathscr{C}_+(\boldsymbol{q}, \boldsymbol{E}, \rho^2)$, we will mean the set of all points $\boldsymbol{a}$ in $\mathscr{C}(\boldsymbol{q}, \boldsymbol{E}, \rho^2)$ such that $\boldsymbol{a}.\boldsymbol{q} \geqslant 0$. The interior of the positive error cone, $\mathscr{C}^0_+(\boldsymbol{q}, \boldsymbol{E}, \rho^2)$ consists of all points $\boldsymbol{a}$ in $\mathscr{C}^0(\boldsymbol{q}, \boldsymbol{E}, \rho^2)$ such that $\boldsymbol{a}.\boldsymbol{q} \geqslant 0$; and the boundary of the positive error cone, $\partial\mathscr{C}_+(\boldsymbol{q}, \boldsymbol{E}, \rho^2)$, consists of all points $\boldsymbol{a}$ in $\partial\mathscr{C}(\boldsymbol{q}, \boldsymbol{E}, \rho^2)$ such that $\boldsymbol{a}.\boldsymbol{q} \geqslant 0$. The negative error cone, its interior, and its boundary are defined by replacing '$\boldsymbol{a}.\boldsymbol{q} \geqslant 0$' with '$\boldsymbol{a}.\boldsymbol{q} \leqslant 0$' in the foregoing statements.

Evidently $\mathscr{C}^0_+(\boldsymbol{q}, \boldsymbol{E}, \rho_1^2)$ and $\mathscr{C}^0_-(\boldsymbol{q}, \boldsymbol{E}, \rho_2^2)$ have no common point as long as ρ_1 and ρ_2 are finite. The set

$$\mathscr{C}_+(\boldsymbol{q}, \boldsymbol{E}, \rho_1^2) \cap \mathscr{C}_-(\boldsymbol{q}, \boldsymbol{E}, \rho_2^2)$$

contains only $\mathbf{0}$, and the same is true of

$$\partial\mathscr{C}_+(\boldsymbol{q}, \boldsymbol{E}, \rho_1^2) \cap \partial\mathscr{C}_-(\boldsymbol{q}, \boldsymbol{E}, \rho_2^2).$$

Moreover, $\mathscr{C}(\boldsymbol{q}, \boldsymbol{E}, \rho^2)$ consists of $\mathscr{C}_+(\boldsymbol{q}, \boldsymbol{E}, \rho^2)$ and $\mathscr{C}_-(\boldsymbol{q}, \boldsymbol{E}, \rho^2)$, while $\mathscr{C}^0(\boldsymbol{q}, \boldsymbol{E}, \rho^2)$ consists of $\mathscr{C}^0_+(\boldsymbol{q}, \boldsymbol{E}, \rho^2)$ and $\mathscr{C}^0_-(\boldsymbol{q}, \boldsymbol{E}, \rho^2)$, and $\partial\mathscr{C}(\boldsymbol{q}, \boldsymbol{E}, \rho^2)$ consists of $\partial\mathscr{C}_+(\boldsymbol{q}, \boldsymbol{E}, \rho^2)$ and $\partial\mathscr{C}_-(\boldsymbol{q}, \boldsymbol{E}, \rho^2)$. In the geometry induced by (6.5) both $\mathscr{C}_+(\boldsymbol{q}, \boldsymbol{E}, \rho^2)$ and $\mathscr{C}_-(\boldsymbol{q}, \boldsymbol{E}, \rho^2)$ are solid N-dimensional right-circular cones with axes $\boldsymbol{E}^{-1}.\boldsymbol{q}$ and $-\boldsymbol{E}^{-1}.\boldsymbol{q}$ respectively, with a common vertex at the origin,

and with vertex half angle $\psi(\rho)$. The boundaries $\partial\mathscr{C}_+$ and $\partial\mathscr{C}_-$ are the $(N-1)$-dimensional surfaces of $\mathscr{C}_+$ and $\mathscr{C}_-$, while $\mathscr{C}_+^0$ and $\mathscr{C}_-^0$ are their interiors. We note that

$$\left.\begin{aligned}\mathscr{C}_+(-\boldsymbol{q}, \boldsymbol{E}, \rho^2) &= \mathscr{C}_-(\boldsymbol{q}, \boldsymbol{E}, \rho^2),\\ \mathscr{C}_+^0(-\boldsymbol{q}, \boldsymbol{E}, \rho^2) &= \mathscr{C}_-^0(\boldsymbol{q}, \boldsymbol{E}, \rho^2),\\ \partial\mathscr{C}_+(-\boldsymbol{q}, \boldsymbol{E}, \rho^2) &= \partial\mathscr{C}_-(\boldsymbol{q}, \boldsymbol{E}, \rho^2).\end{aligned}\right\} \tag{6.9}$$

(c) Convexity of the positive and negative error cones

The purpose of this subsection is to prove lemma 6, which states that the positive and negative error cones are almost strictly convex. The proof requires

LEMMA 5. *Suppose* $\boldsymbol{E}$ *is positive definite and* $\boldsymbol{q}$ *is not* $\boldsymbol{0}$ *and* $\rho \geqslant \rho_{\min}$ *as defined by* (6.6). *Then* $\mathscr{C}_+(\boldsymbol{q}, \boldsymbol{E}, \rho^2)$ *consists of all* $\boldsymbol{a}$ *of the form* $\boldsymbol{a} = t\boldsymbol{b}$ *with* $t \geqslant 0$ *and* $\boldsymbol{b}$ *in* $\mathscr{E}(\boldsymbol{q}, \boldsymbol{E}, \rho^2)$; *and* $\partial\mathscr{C}_+(\boldsymbol{q}, \boldsymbol{E}, \rho^2)$ *consists of all* $\boldsymbol{a}$ *of the form* $t\boldsymbol{b}$ *with* $t \geqslant 0$ *and* $\boldsymbol{b}$ *in* $\partial\mathscr{E}(\boldsymbol{q}, \boldsymbol{E}, \rho^2)$; *and* $\mathscr{C}_+^0(\boldsymbol{q}, \boldsymbol{E}, \rho^2)$ *consists of all* $\boldsymbol{a}$ *of the form* $t\boldsymbol{b}$ *with* $t > 0$ *and* $\boldsymbol{b}$ *in* $\mathscr{E}^0(\boldsymbol{q}, \boldsymbol{E}, \rho^2)$. *The foregoing statement remains true if we replace* $+$ *by* $-$, $\geqslant$ *by* $\leqslant$ *and* $>$ *by* $<$ *throughout.*

Proof. We recall that $\mathscr{E}(\boldsymbol{q}, \boldsymbol{E}, \rho^2)$ consists of all points $\boldsymbol{b}$ such that

$$\boldsymbol{q}.\boldsymbol{b} = 1 \quad \text{and} \quad \boldsymbol{b}.\boldsymbol{E}.\boldsymbol{b} < \rho^2.$$

If $\boldsymbol{b}$ is in $\mathscr{E}(\boldsymbol{q}, \boldsymbol{E}, \rho^2)$ and $\boldsymbol{a} = t\boldsymbol{b}$ with $t \geqslant 0$, then $\boldsymbol{q}.\boldsymbol{a} = t$ and $\boldsymbol{a}.\boldsymbol{E}.\boldsymbol{a} \leqslant t^2\rho^2$, so $\boldsymbol{a}.\boldsymbol{E}.\boldsymbol{a} \leqslant \rho^2(\boldsymbol{q}.\boldsymbol{a})^2$. Therefore $\boldsymbol{a}$ is in $\mathscr{C}(\boldsymbol{q}, \boldsymbol{E}, \rho^2)$. Moreover, evidently $\boldsymbol{q}.\boldsymbol{a} \geqslant 0$, so $\boldsymbol{a}$ is in $\mathscr{C}_+(\boldsymbol{q}, \boldsymbol{E}, \rho^2)$. Conversely, if $\boldsymbol{a}$ is in $\mathscr{C}_+(\boldsymbol{q}, \boldsymbol{E}, \rho^2)$ and is not $\boldsymbol{0}$ let $t = \boldsymbol{a}.\boldsymbol{q} > 0$ and let $\boldsymbol{b} = t^{-1}\boldsymbol{a}$. Then $\boldsymbol{a} = t\boldsymbol{b}$ and $\boldsymbol{q}.\boldsymbol{b} = 1$, and from $\boldsymbol{a}.\boldsymbol{E}.\boldsymbol{a} \leqslant \rho^2(\boldsymbol{q}.\boldsymbol{a})^2$ we can infer $\boldsymbol{b}.\boldsymbol{E}.\boldsymbol{b} \leqslant \rho^2$, so $\boldsymbol{b}$ is in $\mathscr{E}(\boldsymbol{q}, \boldsymbol{E}, \rho^2)$. The assertions in lemma 5 concerning $\mathscr{C}_+^0$ and $\partial\mathscr{C}_+$ are proved in the same way. The assertions about $\mathscr{C}_-$, $\mathscr{C}_-^0$ and $\partial\mathscr{C}_-$ follow immediately from (6.9).

Now we can prove

LEMMA 6. *If* $\boldsymbol{a}_1$ *and* $\boldsymbol{a}_2$ *are any two distinct points in* $\mathscr{C}_+(\boldsymbol{q}, \boldsymbol{E}, \rho^2)$ *and* $\boldsymbol{a}'$ *is any point between* $\boldsymbol{a}_1$ *and* $\boldsymbol{a}_2$ *on the straight line segment joining* $\boldsymbol{a}_1$ *and* $\boldsymbol{a}_2$, *then* $\boldsymbol{a}'$ *is in* $\mathscr{C}_+^0(\boldsymbol{q}, \boldsymbol{E}, \rho^2)$ *unless* $\boldsymbol{a}_1$ *and* $\boldsymbol{a}_2$ *are linearly dependent and both lie in* $\partial\mathscr{C}_+(\boldsymbol{q}, \boldsymbol{E}, \rho^2)$. *The foregoing statement remains true if* $+$ *is replaced by* $-$ *throughout.*

Proof. Since $\mathscr{C}_+(\boldsymbol{q}, \boldsymbol{E}, \rho^2)$ contains two distinct points which, if linearly dependent, do not both lie in $\partial\mathscr{C}_+(\boldsymbol{q}, \boldsymbol{E}, \rho^2)$, we infer that $\rho > \rho_{\min}$. If $\boldsymbol{a}_1 = \boldsymbol{0}$, then by hypothesis $\boldsymbol{a}_2$ is in $\mathscr{C}_+^0(\boldsymbol{q}, \boldsymbol{E}, \rho^2)$. Since $\boldsymbol{a}'$ must be a positive multiple of $\boldsymbol{a}_2$, lemma 5 puts $\boldsymbol{a}'$ in $\mathscr{C}_+^0(\boldsymbol{q}, \boldsymbol{E}, \rho^2)$. In the remainder of the proof we can assume that neither $\boldsymbol{a}_1$ nor $\boldsymbol{a}_2$ is $\boldsymbol{0}$. Then by lemma 5 there exist positive numbers t_1 and t_2 and points $\boldsymbol{b}_1$ and $\boldsymbol{b}_2$ in $\mathscr{E}(\boldsymbol{q}, \boldsymbol{E}, \rho^2)$ such that $\boldsymbol{a}_1 = t_1\boldsymbol{b}_1$ and $\boldsymbol{a}_2 = t_2\boldsymbol{b}_2$. Moreover, there exist positive numbers α_1 and α_2 such that

$$\alpha_1 + \alpha_2 = 1 \quad \text{and} \quad \boldsymbol{a}' = \alpha_1\boldsymbol{a}_1 + \alpha_2\boldsymbol{a}_2.$$

If we define $t' = \alpha_1 t_1 + \alpha_2 t_2$, then $t' > 0$, and we can define $\alpha_1' = \alpha_1 t_1/t'$, $\alpha_2' = \alpha_2 t_2/t'$. We have $\alpha_1' > 0$, $\alpha_2' > 0$, and $\alpha_1' + \alpha_2' = 1$. Moreover, $\boldsymbol{a}' = t'\boldsymbol{b}'$ where $\boldsymbol{b}' = \alpha_1'\boldsymbol{b}_1 + \alpha_2'\boldsymbol{b}_2$. If $\boldsymbol{b}_1$ and $\boldsymbol{b}_2$ are distinct, then lemma 2 tells us that $\boldsymbol{b}'$ is in $\mathscr{E}^0(\boldsymbol{q}, \boldsymbol{E}, \rho^2)$, so lemma 5 tells us that $\boldsymbol{a}'$ is in $\mathscr{C}_+^0(\boldsymbol{q}, \boldsymbol{E}, \rho^2)$. If $\boldsymbol{b}_1 = \boldsymbol{b}_2$ then $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ are linearly dependent, so by hypothesis they are in $\mathscr{C}_+^0(\boldsymbol{q}, \boldsymbol{E}, \rho^2)$. But then by lemma 5, $\boldsymbol{b}_1$ and $\boldsymbol{b}_2$ are in $\mathscr{E}^0(\boldsymbol{q}, \boldsymbol{E}, \rho^2)$. Since $\boldsymbol{b}' = \boldsymbol{b}_1 = \boldsymbol{b}_2$, it follows from lemma 5 that $\boldsymbol{a}'$ is in $\mathscr{C}_+^0(\boldsymbol{q}, \boldsymbol{E}, \rho^2)$.

(d) Hyperplane sections of error cones

In the study of absolute errors, an important role was played by the intersection of the hyperplane $\mathscr{H}(\boldsymbol{u})$ with the error ellipsoid $\mathscr{E}(\boldsymbol{E}, \epsilon^2)$. For relative errors, the same role is played by the

intersection of $\mathscr{H}(\boldsymbol{u})$ with the error double cone $\mathscr{C}(\boldsymbol{q}, \boldsymbol{E}, \rho^2)$. Now we examine that intersection. We define

$$\mathscr{C}_{\pm}(\boldsymbol{u}, \boldsymbol{q}, \boldsymbol{E}, \rho^2) = \mathscr{H}(\boldsymbol{u}) \cap \mathscr{C}_{\pm}(\boldsymbol{q}, \boldsymbol{E}, \rho^2),$$

$$\mathscr{C}^0_{\pm}(\boldsymbol{u}, \boldsymbol{q}, \boldsymbol{E}, \rho^2) = \mathscr{H}(\boldsymbol{u}) \cap \mathscr{C}\ (\boldsymbol{q}, \boldsymbol{E}, \rho^2),$$

$$\partial\mathscr{C}_{\pm}(\boldsymbol{u}, \boldsymbol{q}, \boldsymbol{E}, \rho^2) = \mathscr{H}(\boldsymbol{u}) \cap \partial\mathscr{C}_{\pm}(\boldsymbol{q}, \boldsymbol{E}, \rho^2).$$

We also make the corresponding definitions for the whole double cone $\mathscr{C}(\boldsymbol{q}, \boldsymbol{E}, \rho^2)$ by deleting $\pm$ in the three foregoing equations.

If $\boldsymbol{a}$ is a point on $\partial\mathscr{C}(\boldsymbol{u}, \boldsymbol{q}, \boldsymbol{E}, \rho^2)$, we would like to have an expression for a vector normal to $\partial\mathscr{C}(\boldsymbol{u}, \boldsymbol{q}, \boldsymbol{E}, \rho^2)$ at $\boldsymbol{a}$ lying in (i.e. tangent to) $\mathscr{H}(\boldsymbol{u})$, and pointing out of $\mathscr{C}(\boldsymbol{u}, \boldsymbol{q}, \boldsymbol{E}, \rho^2)$. One way to obtain such a vector is to project the normal $\boldsymbol{n}_{q,E}(\boldsymbol{a})$ of (6.8) orthogonally onto $\mathscr{H}(\boldsymbol{u})$. The result is

$$(\boldsymbol{I} - \hat{\boldsymbol{u}}\hat{\boldsymbol{u}}) \,.\, \boldsymbol{n}_{q,E}(\boldsymbol{a}) = (\boldsymbol{I} - \hat{\boldsymbol{u}}\hat{\boldsymbol{u}}) \,.\, (\boldsymbol{E} - \rho^2 \boldsymbol{q}\boldsymbol{q}) \,.\, \boldsymbol{a}. \tag{6.10}$$

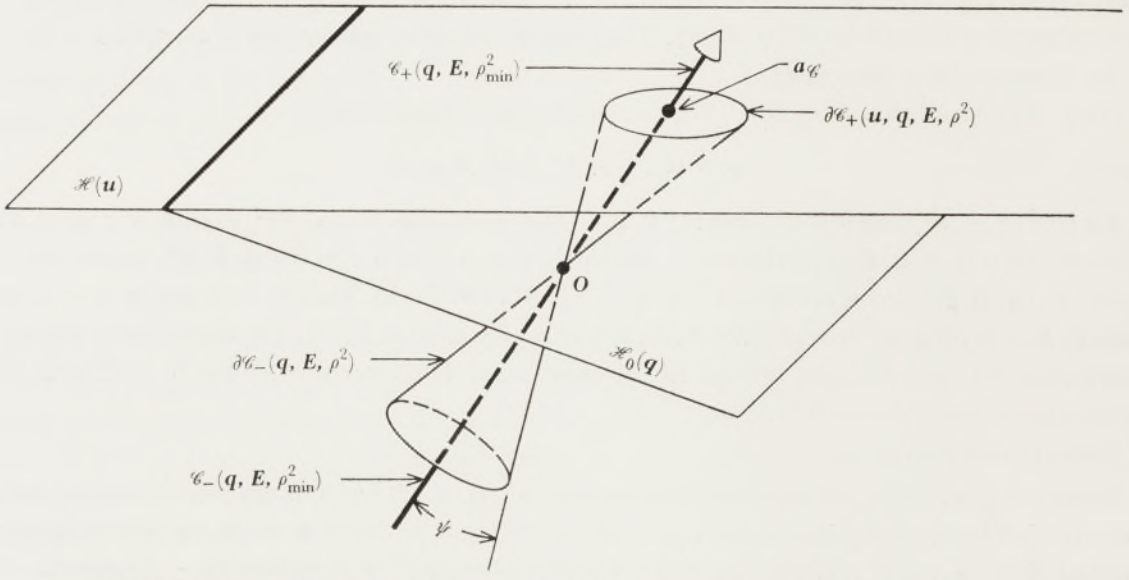

FIGURE 4. Illustration of the geometry and notation for an error cone in the case $N = 3$.

As in (6.8) and for the same reason, the geometry in which the vector (6.10) is normal to $\partial\mathscr{C}(\boldsymbol{u}, \boldsymbol{q}, \boldsymbol{E}, \rho^2)$ is the geometry induced on $\mathscr{R}^N$ by (4.1), not (6.5). To obtain a vector normal to $\partial\mathscr{C}(\boldsymbol{u}, \boldsymbol{q}, \boldsymbol{E}, \rho^2)$ in the geometry of (6.5), one must dot (6.10) on the left with $\boldsymbol{E}^{-1}$.

The whole purpose of introducing $\mathscr{C}_+$ and $\mathscr{C}_-$ is to obtain the following lemma:

LEMMA 7. *If* $\boldsymbol{u}$ *and* $\boldsymbol{q}$ *are non-zero vectors and* $\boldsymbol{E}: \mathscr{R}^N \to \mathscr{R}^N$ *is symmetric and positive definite then* $\mathscr{C}_+(\boldsymbol{u}, \boldsymbol{q}, \boldsymbol{E}, \rho^2)$ *and* $\mathscr{C}_-(\boldsymbol{u}, \boldsymbol{q}, \boldsymbol{E}, \rho^2)$ *are strictly convex.*

The proof is simply an application of lemma 6 and the observation that two linearly dependent vectors in $\mathscr{H}(\boldsymbol{u})$ are equal.

Now we examine the families $\mathscr{C}_{\pm}(\boldsymbol{u}, \boldsymbol{q}, \boldsymbol{E}, \rho^2)$ in some detail as ρ varies. The axis of all the double error cones $\mathscr{C}(\boldsymbol{q}, \boldsymbol{E}, \rho^2)$ for various ρ is $\mathscr{C}(\boldsymbol{q}, \boldsymbol{E}, \rho^2_{\min})$; it consists of all vectors $\boldsymbol{a} = t\boldsymbol{E}^{-1} \,.\, \boldsymbol{q}$. The axis intersects $\mathscr{H}(\boldsymbol{u})$ at all such vectors which have $\boldsymbol{u} \,.\, \boldsymbol{a} = 1$, or $t(\boldsymbol{u} \,.\, \boldsymbol{E}^{-1} \,.\, \boldsymbol{q}) = 1$. Thus if $\boldsymbol{u} \,.\, \boldsymbol{E}^{-1} \,.\, \boldsymbol{q} = 0$ the axis does not intersect $\mathscr{H}(\boldsymbol{u})$, while if $\boldsymbol{u} \,.\, \boldsymbol{E}^{-1} \,.\, \boldsymbol{q} > 0$ there is exactly one point of intersection,

$$\boldsymbol{a}_{\mathscr{C}} = \frac{\boldsymbol{E}^{-1} \,.\, \boldsymbol{q}}{\boldsymbol{u} \,.\, \boldsymbol{E}^{-1} \,.\, \boldsymbol{q}}, \tag{6.11}$$

and $\boldsymbol{a}_{\mathscr{C}}$ lies in all the positive error cones $\mathscr{C}_+(\boldsymbol{q}, \boldsymbol{E}, \rho^2)$ which have $\rho \geqslant \rho_{\min}$. Thus if $\rho \geqslant \rho_{\min}$, $\boldsymbol{a}_{\mathscr{C}}$ is in $\mathscr{C}_+(\boldsymbol{u}, \boldsymbol{q}, \boldsymbol{E}, \rho^2)$, and that set is non-empty. The possibility that $\boldsymbol{u}.\boldsymbol{E}^{-1}.\boldsymbol{q} < 0$ is excluded by (6.3). The effect of (6.3) is to insure that if the axis $\mathscr{C}(\boldsymbol{q}, \boldsymbol{E}, \rho^2_{\min})$ is not parallel to $\mathscr{H}(\boldsymbol{u})$ then $\mathscr{C}_+(\boldsymbol{q}, \boldsymbol{E}, \rho^2_{\min})$ intersects $\mathscr{H}(\boldsymbol{u})$ but $\mathscr{C}_-(\boldsymbol{q}, \boldsymbol{E}, \rho^2_{\min})$ does not. When $\boldsymbol{u}.\boldsymbol{E}^{-1}.\boldsymbol{q} = 0$, i.e. when $\mathscr{C}(\boldsymbol{q}, \boldsymbol{E}, \rho^2_{\min})$ is parallel to $\mathscr{H}(\boldsymbol{u})$, it is heuristically useful to think of $\boldsymbol{a}_{\mathscr{C}}$ as $(+\infty)\ \boldsymbol{E}^{-1}.\boldsymbol{q}$, a point which has receded to infinity on $\mathscr{C}_+(\boldsymbol{q}, \boldsymbol{E}, \rho^2_{\min})$. For $N = 3$ and $\boldsymbol{u}.\boldsymbol{E}^{-1}.\boldsymbol{q} > 0$, the geometry of the error cones and their relation to $\mathscr{H}(\boldsymbol{u})$ is shown in figure 4.

From figure 4 it is clear that $\mathscr{C}(\boldsymbol{u}, \boldsymbol{q}, \boldsymbol{E}, \rho^2)$ is the N-dimensional generalization of a conic section. As one would expect, the three-dimensional theory generalizes without difficulty. Depending on the value of ρ, the intersection of the hyperplane $\mathscr{H}(\boldsymbol{u})$ with the double cone $\mathscr{C}(\boldsymbol{q}, \boldsymbol{E}, \rho^2)$ can be of elliptic, parabolic, or hyperbolic type. The critical value of ρ which separates the elliptic from the hyperbolic intersections, and itself yields the parabolic intersection, is the positive number ρ_{par} such that

$$\rho^2_{\text{par}} = \frac{(\boldsymbol{a}_{\mathscr{C}} - \boldsymbol{a}_E).\boldsymbol{E}.(\boldsymbol{a}_{\mathscr{C}} - \boldsymbol{a}_E)}{[\boldsymbol{q}.(\boldsymbol{a}_{\mathscr{C}} - \boldsymbol{a}_E)]^2}. \tag{6.12}$$

Of course if the axis $\mathscr{C}(\boldsymbol{q}, \boldsymbol{E}, \rho^2_{\min})$ is parallel to $\mathscr{H}(\boldsymbol{u})$ there is only the hyperbolic case, and $\rho_{\text{par}} = \rho_{\min}$. If the axis intersects $\mathscr{H}(\boldsymbol{u})$ then necessarily $\rho_{\text{par}} > \rho_{\min}$. The whole situation is summarized in

THEOREM 2. *Let $\boldsymbol{u}$ and $\boldsymbol{q}$ be non-zero vectors in $\mathscr{R}^N$ while $\boldsymbol{E}$: $\mathscr{R}^N \to \mathscr{R}^N$ is symmetric and positive definite. Define $\boldsymbol{a}_E$ by (4.24) and when $\boldsymbol{u}.\boldsymbol{E}^{-1}.\boldsymbol{q} > 0$ define $\boldsymbol{a}_{\mathscr{C}}$ by (6.11) and ρ_{par} by (6.12). If $\boldsymbol{u}.\boldsymbol{E}^{-1}.\boldsymbol{q} = 0$, then $\mathscr{C}(\boldsymbol{u}, \boldsymbol{q}, \boldsymbol{E}, \rho^2)$ is empty as long as $\rho \leqslant \rho_{\min}$, while for $\rho > \rho_{\min}$, $\mathscr{C}_\pm(\boldsymbol{u}, \boldsymbol{q}, \boldsymbol{E}, \rho^2)$ are the two halves of a non-empty, unbounded, solid $(N-1)$-dimensional hyperboloid of two sheets in $\mathscr{H}(\boldsymbol{u})$. If $\boldsymbol{u}.\boldsymbol{E}^{-1}.\boldsymbol{q} > 0$ then*

(i) *when $\rho < \rho_{\min}$, $\mathscr{C}(\boldsymbol{u}, \boldsymbol{q}, \boldsymbol{E}, \rho^2)$ is empty;*

(ii) *when $\rho = \rho_{\min}$, $\mathscr{C}_-(\boldsymbol{u}, \boldsymbol{q}, \boldsymbol{E}, \rho^2)$ is empty while $\mathscr{C}_+(\boldsymbol{u}, \boldsymbol{q}, \boldsymbol{E}, \rho^2)$ consists of the single point $\boldsymbol{a}_{\mathscr{C}}$;*

(iii) *when $\rho_{\min} < \rho < \rho_{\text{par}}$, then $\mathscr{C}_-(\boldsymbol{u}, \boldsymbol{q}, \boldsymbol{E}, \rho^2)$ is empty while $\mathscr{C}_+(\boldsymbol{u}, \boldsymbol{q}, \boldsymbol{E}, \rho^2)$ is a compact, solid, $(N-1)$-dimensional ellipsoid in $\mathscr{H}(\boldsymbol{u})$ containing $\boldsymbol{a}_{\mathscr{C}}$;*

(iv) *when $\rho = \rho_{\text{par}}$, then $\mathscr{C}_-(\boldsymbol{u}, \boldsymbol{q}, \boldsymbol{E}, \rho^2)$ is empty while $\mathscr{C}_+(\boldsymbol{u}, \boldsymbol{q}, \boldsymbol{E}, \rho^2)$ is an unbounded, solid, $(N-1)$-dimensional paraboloid of revolution in the geometry (6.5), its axis consisting of the points $\boldsymbol{a}_E + t(\boldsymbol{a}_{\mathscr{C}} - \boldsymbol{a}_E)$ for real t;*

(v) *when $\rho_{\text{par}} < \rho < \infty$, $\mathscr{C}_-(\boldsymbol{u}, \boldsymbol{q}, \boldsymbol{E}, \rho^2)$ and $\mathscr{C}_+(\boldsymbol{u}, \boldsymbol{q}, \boldsymbol{E}, \rho^2)$ are the two unbounded halves of a solid $(N-1)$-dimensional hyperboloid of two sheets in $\mathscr{H}(\boldsymbol{u})$.*

When $N = 3$, theorem 2 is a well-known part of the theory of conic sections. Figure 5 illustrates the case $N = 3$, $\boldsymbol{u}.\boldsymbol{E}^{-1}.\boldsymbol{q} > 0$. A rigorous proof of the theorem for all N is simply an exercise in N-dimensional geometry. One such proof among the many available is given in appendix D.

Even for $N = 3$, the expression (6.12) for ρ_{par} is not trivial, so here we will describe briefly how that expression is obtained. First, (6.6) and (6.11) imply that

$$\boldsymbol{u}.\boldsymbol{E}^{-1}.\boldsymbol{q} = \frac{1}{\rho^2_{\min}(\boldsymbol{q}.\boldsymbol{a}_{\mathscr{C}})}, \tag{6.13}$$

and that

$$(\boldsymbol{q}.\boldsymbol{a}_E)\,(\boldsymbol{q}.\boldsymbol{a}_{\mathscr{C}}) = \frac{\epsilon^2_{\min}}{\rho^2_{\min}}. \tag{6.14}$$

Then from (4.23), (4.24), (6.13) and (6.14) we can show with a little algebra that (6.12) is equivalent to

$$\frac{\rho^2_{\min}}{\rho^2_{\text{par}}} = 1 - \frac{\boldsymbol{q}.\boldsymbol{a}_E}{\boldsymbol{q}.\boldsymbol{a}_{\mathscr{C}}}. \tag{6.15}$$

Further manipulation of (4.23), (4.24), (6.6), (6.13) and (6.14) transforms (6.15) into the equation

$$\frac{\rho^2_{\min}}{\rho^2_{\text{par}}} = 1 - \frac{(\boldsymbol{u}.\boldsymbol{E}^{-1}.\boldsymbol{q})^2}{(\boldsymbol{u}.\boldsymbol{E}^{-1}.\boldsymbol{u})\,(\boldsymbol{q}.\boldsymbol{E}^{-1}.\boldsymbol{q})}. \tag{6.16}$$

Equation (6.16) is very easy to interpret. In the geometry induced on $\mathscr{R}^N$ by (6.5), let ψ_{par} denote the vertex half-angle of $\mathscr{C}_-(\boldsymbol{q}, \boldsymbol{E}, \rho^2_{\text{par}})$. Then according to (6.7), $\cos^2\psi_{\text{par}} = \rho^2_{\min}/\rho^2_{\text{par}}$. Therefore equation (6.16) can be written

$$\cos^2\psi_{\text{par}} + \cos^2\angle_E(\boldsymbol{E}^{-1}.\boldsymbol{q}, \boldsymbol{E}^{-1}.\boldsymbol{u}) = 1.$$

Since both ψ_{par} and $\angle_E(\boldsymbol{E}^{-1}.\boldsymbol{q}, \boldsymbol{E}^{-1}.\boldsymbol{u})$ are between 0 and $\frac{1}{2}\pi$, this last equation is equivalent to

$$\psi_{\text{par}} = \tfrac{1}{2}\pi - \angle_E(\boldsymbol{E}^{-1}.\boldsymbol{q}, \boldsymbol{E}^{-1}.\boldsymbol{u}). \tag{6.17}$$

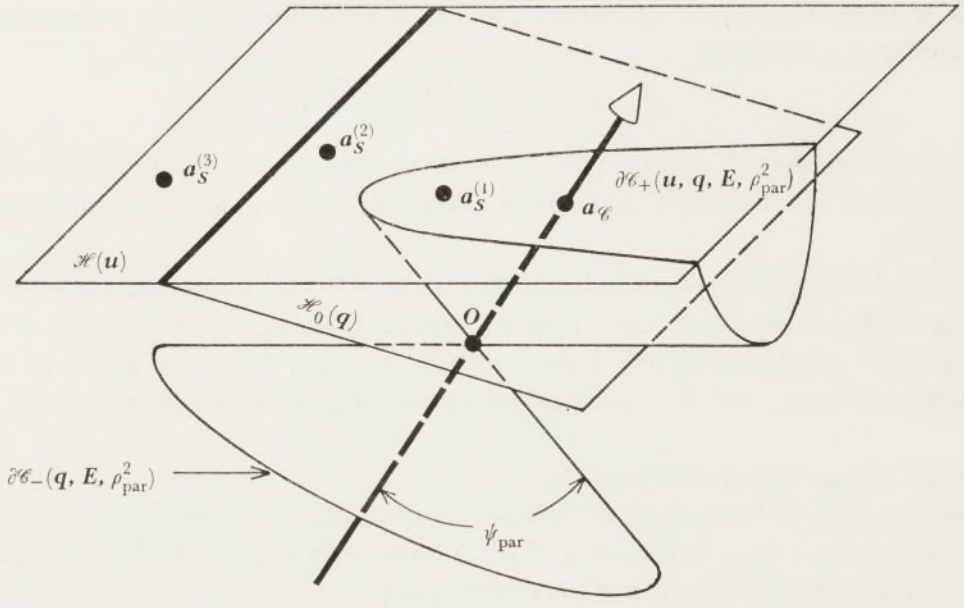

FIGURE 5. Illustration, for $N = 3$, of the geometrical significance of ρ_{par}. The three possible positions of $\boldsymbol{a}_S$ generate three cases: $\boldsymbol{a}_S^{(1)}$ has $\boldsymbol{q}.\boldsymbol{a}_S > 0$ and $\rho_{\text{par}} > \rho^+_{\max} = \rho_{\max}$; $\boldsymbol{a}_S^{(2)}$ has $\boldsymbol{q}.\boldsymbol{a}_S > 0$ and $\rho_{\text{par}} < \rho^+_{\max} = \rho_{\max}$; $\boldsymbol{a}_S^{(3)}$ has $\boldsymbol{q}.\boldsymbol{a}_S < 0$ and $\rho_{\max} = \rho^-_{\max}$.

The angle on the right in (6.17) is the angle between the cone axis $\boldsymbol{E}^{-1}.\boldsymbol{q}$ and the hyperplane $\mathscr{H}(\boldsymbol{u})$, since $\boldsymbol{E}^{-1}.\boldsymbol{u}$ is the normal to $\mathscr{H}(\boldsymbol{u})$ in the geometry of (6.5). Thus (6.17) says simply that if the hyperplane $\mathscr{H}(\boldsymbol{u})$ is to intersect both $\mathscr{C}_+(\boldsymbol{q}, \boldsymbol{E}, \rho^2)$ and $\mathscr{C}_-(\boldsymbol{q}, \boldsymbol{E}, \rho^2)$ then $\psi(\rho)$, the vertex half-angle of those cones, must be greater than the angle between their axis and $\mathscr{H}(\boldsymbol{u})$.

(e) *Elementary remarks about $\rho(s)$, the tradeoff curve between relative error and spread*

As pointed out in § 6*a*, to minimize the relative error in $\langle m_E, A\rangle$ for a given spread s of the averaging kernel A from r_0, we must minimize $\rho(\boldsymbol{a})$, defined by (6.2), when $\boldsymbol{a}$ is restricted to $\mathscr{E}(\boldsymbol{u}, \boldsymbol{S}, s)$. We define $\rho(s)$ as the greatest lower bound of the values of $\rho(\boldsymbol{a})$ when $\boldsymbol{a}$ is in $\mathscr{E}(\boldsymbol{u}, \boldsymbol{S}, s)$. If $s < s_{\min}$, then $\mathscr{E}(\boldsymbol{u}, \boldsymbol{S}, s)$ is empty, so we agree that $\rho(s) = +\infty$.

If $s_1 < s_2$, then $\mathscr{E}(\boldsymbol{u}, \boldsymbol{S}, s_1)$ is contained in $\mathscr{E}(\boldsymbol{u}, \boldsymbol{S}, s_2)$, so evidently $\rho(s_1) \geqslant \rho(s_2)$. If $s = s_{\min}$, then $\mathscr{E}(\boldsymbol{u}, \boldsymbol{S}, s)$ consists of the single point $\boldsymbol{a}_S$, while $\rho(s)$ takes its largest finite value, $\rho_{\max}$, given by

$$\rho_{\max} = \frac{\boldsymbol{a}_S.\boldsymbol{E}.\boldsymbol{a}_S}{(\boldsymbol{q}.\boldsymbol{a}_S)^2} = \frac{\boldsymbol{u}.\boldsymbol{S}^{-1}.\boldsymbol{E}.\boldsymbol{S}^{-1}.\boldsymbol{u}}{(\boldsymbol{u}.\boldsymbol{S}^{-1}.\boldsymbol{q})^2}. \tag{6.18}$$

As s increases, $\mathscr{E}(\boldsymbol{u}, \boldsymbol{S}, s)$ expands in $\mathscr{H}(\boldsymbol{u})$ until finally a value $\tilde{s}_{\max}$ is reached at which $\partial\mathscr{E}(\boldsymbol{u}, \boldsymbol{S}, s)$ contains $\boldsymbol{a}_{\mathscr{C}}$. Clearly

$$\tilde{s}_{\max} = \boldsymbol{a}_{\mathscr{C}} . \boldsymbol{S} . \boldsymbol{a}_{\mathscr{C}} = \frac{\boldsymbol{q} . \boldsymbol{E}^{-1} . \boldsymbol{S} . \boldsymbol{E}^{-1} . \boldsymbol{q}}{(\boldsymbol{u} . \boldsymbol{E}^{-1} . \boldsymbol{q})^2} . \tag{6.19}$$

(Note that $\tilde{s}_{\max}$ is different from the $s_{\max}$ which appears in the study of the tradeoff curve for absolute errors.) If $s \geqslant \tilde{s}_{\max}$ then $\mathscr{E}(\boldsymbol{u}, \boldsymbol{S}, s)$ contains $\boldsymbol{a}_{\mathscr{C}}$. But at $\boldsymbol{a}_{\mathscr{C}}$, $\rho(\boldsymbol{a})$ takes the value $\rho_{\min}$, which is the least value $\rho(\boldsymbol{a})$ can have in $\mathscr{H}(\boldsymbol{u})$. Hence if $s \geqslant \tilde{s}_{\max}$, then $\rho(s) = \rho_{\min}$.

We see that $\rho(s)$ is $+\infty$ for $s < s_{\min}$, monotonic non-increasing from $\rho_{\max}$ to $\rho_{\min}$ in $s_{\min} \leqslant s \leqslant \tilde{s}_{\max}$, and constant if $s \geqslant \tilde{s}_{\max}$. Evidently the only interesting domain for s is $s_{\min} \leqslant s \leqslant \tilde{s}_{\max}$.

What if $s_{\min} = \tilde{s}_{\max}$? To deal with this possibility, we need more geometry. Suppose $\boldsymbol{K}: \mathscr{R}^N \to \mathscr{R}^N$ is symmetric and positive definite. For any $\boldsymbol{f}$ and $\boldsymbol{g}$ in $\mathscr{R}^N$, let $(\boldsymbol{f}, \boldsymbol{g})_{\boldsymbol{K}}$ stand for $\boldsymbol{f} . \boldsymbol{K} . \boldsymbol{g}$. Then $(\boldsymbol{f}, \boldsymbol{g})_{\boldsymbol{K}}$ defines an inner product with a Schwarz inequality on $\mathscr{R}^N$. Now

$$\frac{\tilde{s}_{\max}}{s_{\min}} = \frac{(\boldsymbol{a}_{\boldsymbol{S}}, \boldsymbol{a}_{\boldsymbol{S}})_{\boldsymbol{S}} (\boldsymbol{a}_{\mathscr{C}}, \boldsymbol{a}_{\mathscr{C}})_{\boldsymbol{S}}}{(\boldsymbol{a}_{\boldsymbol{S}}, \boldsymbol{a}_{\mathscr{C}})_{\boldsymbol{S}}^2} \quad \text{and} \quad \left(\frac{\rho_{\max}}{\rho_{\min}}\right)^2 = \frac{(\boldsymbol{a}_{\boldsymbol{S}}, \boldsymbol{a}_{\boldsymbol{S}})_{\boldsymbol{E}} (\boldsymbol{a}_{\mathscr{C}}, \boldsymbol{a}_{\mathscr{C}})_{\boldsymbol{E}}}{(\boldsymbol{a}_{\boldsymbol{S}}, \boldsymbol{a}_{\mathscr{C}})_{\boldsymbol{E}}^2} .$$

Hence $s_{\min} < \tilde{s}_{\max}$ and $\rho_{\min} < \rho_{\max}$ unless $\boldsymbol{a}_{\mathscr{C}}$ and $\boldsymbol{a}_{\boldsymbol{S}}$ are linearly dependent. But since both vectors are in $\mathscr{H}(\boldsymbol{u})$, if they are linearly dependent they are equal. In this trivial case, $\rho(s) = +\infty$ when $s < s_{\min}$ and $\rho(s) = \rho_{\min}$ when $s \geqslant s_{\min}$. Henceforth we will assume that $\boldsymbol{a}_{\mathscr{C}} \neq \boldsymbol{a}_{\boldsymbol{S}}$, so that $s_{\min} < \tilde{s}_{\max}$ and $\rho_{\min} < \rho_{\max}$.

We must also deal with a second trivial case. If $\boldsymbol{u}$ and $\boldsymbol{q}$ are linearly dependent, then we can write $\boldsymbol{q} = \kappa \boldsymbol{u}$, and from (6.3) we infer that $\kappa > 0$. For any $\boldsymbol{a}$ in $\mathscr{H}(\boldsymbol{u})$ we will have $\boldsymbol{q} . \boldsymbol{a} = \kappa$, so $\rho(\boldsymbol{a}) = \epsilon(\boldsymbol{a})/\kappa$. Then to minimize the relative error $\rho(\boldsymbol{a})$ in $\mathscr{E}(\boldsymbol{u}, \boldsymbol{S}, s)$ we simply minimize the absolute error $\epsilon(\boldsymbol{a})$. This problem has already been treated in §§ 4 and 5. Therefore we will assume that $\boldsymbol{u}$ and $\boldsymbol{q}$ are linearly independent. Then $\boldsymbol{a}_{\boldsymbol{E}}$ and $\boldsymbol{a}_{\mathscr{C}}$ are linearly independent, so $\boldsymbol{a}_{\boldsymbol{E}} \neq \boldsymbol{a}_{\mathscr{C}}$. In fact, from (4.23), (6.6) and (6.13),

$$\boldsymbol{q} . \boldsymbol{a}_{\mathscr{C}} - \boldsymbol{q} . \boldsymbol{a}_{\boldsymbol{E}} = \frac{\boldsymbol{q} . \boldsymbol{E}^{-1} . \boldsymbol{q}}{\boldsymbol{u} . \boldsymbol{E}^{-1} . \boldsymbol{q}} - \frac{\boldsymbol{u} . \boldsymbol{E}^{-1} . \boldsymbol{q}}{\boldsymbol{u} . \boldsymbol{E}^{-1} . \boldsymbol{u}} .$$

Then Schwarz's inequality and (6.3) imply that

$$0 \leqslant \boldsymbol{q} . \boldsymbol{a}_{\boldsymbol{E}} < \boldsymbol{q} . \boldsymbol{a}_{\mathscr{C}} . \tag{6.20}$$

In our discussion of relative errors, we shall ignore both trivial cases. We shall assume that $\boldsymbol{a}_{\mathscr{C}} \neq \boldsymbol{a}_{\boldsymbol{S}}$ and $\boldsymbol{a}_{\mathscr{C}} \neq \boldsymbol{a}_{\boldsymbol{E}}$, so that $s_{\min} < \tilde{s}_{\max}$, $\rho_{\min} < \rho_{\max}$, and we have (6.20). These assumptions still admit a singular case, $\boldsymbol{u} . \boldsymbol{E}^{-1} . \boldsymbol{q} = 0$, for which $\boldsymbol{a}_{\mathscr{C}}$ is at $(+\infty)\, \boldsymbol{E}^{-1} . \boldsymbol{q}$ and $\boldsymbol{q} . \boldsymbol{a}_{\mathscr{C}} = +\infty$, and $\tilde{s}_{\max} = +\infty$. Except in this singular case, $\rho_{\min} < \rho_{\text{par}}$.

(*f*) *The complete geometrical theory of* $\rho(s)$

In § 4*e* the discussion of the tradeoff curve for absolute error was based on the fact that the level surfaces of both $\epsilon(\boldsymbol{a})$ and $s(\boldsymbol{a})$ in $\mathscr{H}(\boldsymbol{u})$ were the boundaries of strictly convex sets. This is not true of $\rho(\boldsymbol{a})$; if $\rho > \rho_{\text{par}}$ then $\mathscr{C}(\boldsymbol{u}, \boldsymbol{q}, \boldsymbol{E}, \rho^2)$ consists of the two separate pieces $\mathscr{C}_{\pm}(\boldsymbol{u}, \boldsymbol{q}, \boldsymbol{E}, \rho^2)$. Evidently we want to study those pieces separately, so we define two functions:

$$\left.\begin{aligned} \rho_+(\boldsymbol{a}) &= \rho(\boldsymbol{a}) \quad \text{if} \quad \boldsymbol{q} . \boldsymbol{a} > 0 \\ &= +\infty \quad \text{if} \quad \boldsymbol{q} . \boldsymbol{a} \leqslant 0; \\ \rho_-(\boldsymbol{a}) &= \rho(\boldsymbol{a}) \quad \text{if} \quad \boldsymbol{q} . \boldsymbol{a} < 0 \\ &= +\infty \quad \text{if} \quad \boldsymbol{q} . \boldsymbol{a} \geqslant 0. \end{aligned}\right\} \tag{6.21}$$

If we take the branch of arctan which lies between 0 and $\frac{1}{2}\pi$, $\arctan \rho_+(\boldsymbol{a})$ and $\arctan \rho_-(\boldsymbol{a})$ are continuous functions of $\boldsymbol{a}$; therefore we will say that $\rho_\pm(\boldsymbol{a})$ are continuous in an extended sense, even though they take on the value $+\infty$.

For any $\boldsymbol{a}$, $\rho(\boldsymbol{a})$ is the least of $\rho_+(\boldsymbol{a})$ and $\rho_-(\boldsymbol{a})$: $\rho(\boldsymbol{a}) = \min\{\rho_+(\boldsymbol{a}), \rho_-(\boldsymbol{a})\}$. If ρ is finite, $\mathscr{C}_+(\boldsymbol{u}, \boldsymbol{q}, \boldsymbol{E}, \rho^2)$ is the set of all $\boldsymbol{a}$ in $\mathscr{H}(\boldsymbol{u})$ where $\rho_+(\boldsymbol{a}) \leqslant \rho$, while $\mathscr{C}_-(\boldsymbol{u}, \boldsymbol{q}, \boldsymbol{E}, \rho^2)$ is the set of all $\boldsymbol{a}$ in $\mathscr{H}(\boldsymbol{u})$ where $\rho_-(\boldsymbol{a}) \leqslant \rho$. We define $\rho_+(s)$ to be the greatest lower bound of the values of $\rho_+(\boldsymbol{a})$ for $\boldsymbol{a}$ in $\mathscr{E}(\boldsymbol{u}, \boldsymbol{S}, s)$, while $\rho_-(s)$ is the greatest lower bound of the values of $\rho_-(\boldsymbol{a})$ for $\boldsymbol{a}$ in $\mathscr{E}(\boldsymbol{u}, \boldsymbol{S}, s)$. For any s,

$$\rho(s) = \min\{\rho_+(s), \rho_-(s)\}. \tag{6.22}$$

Both $\rho_+(s)$ and $\rho_-(s)$ are monotonic non-increasing functions of s and both are finite only if s is so large that $\mathscr{E}(\boldsymbol{u}, \boldsymbol{S}, s)$ includes points $\boldsymbol{a}$ with $\boldsymbol{q}.\boldsymbol{a} > 0$ and also points with $\boldsymbol{q}.\boldsymbol{a} < 0$. This can happen only if $s > s_\infty$, where $\partial\mathscr{E}(\boldsymbol{u}, \boldsymbol{S}, s_\infty)$ is tangent to $\mathscr{H}_0(\boldsymbol{q}) \cap \mathscr{H}(\boldsymbol{u})$ in $\mathscr{H}(\boldsymbol{u})$. We must find s_∞ and the point of tangency, $\boldsymbol{a}_\infty$. The subscript ∞ refers to the fact that $\boldsymbol{q}.\boldsymbol{a}_\infty = 0$, so that $\rho(\boldsymbol{a}_\infty) = \infty$.

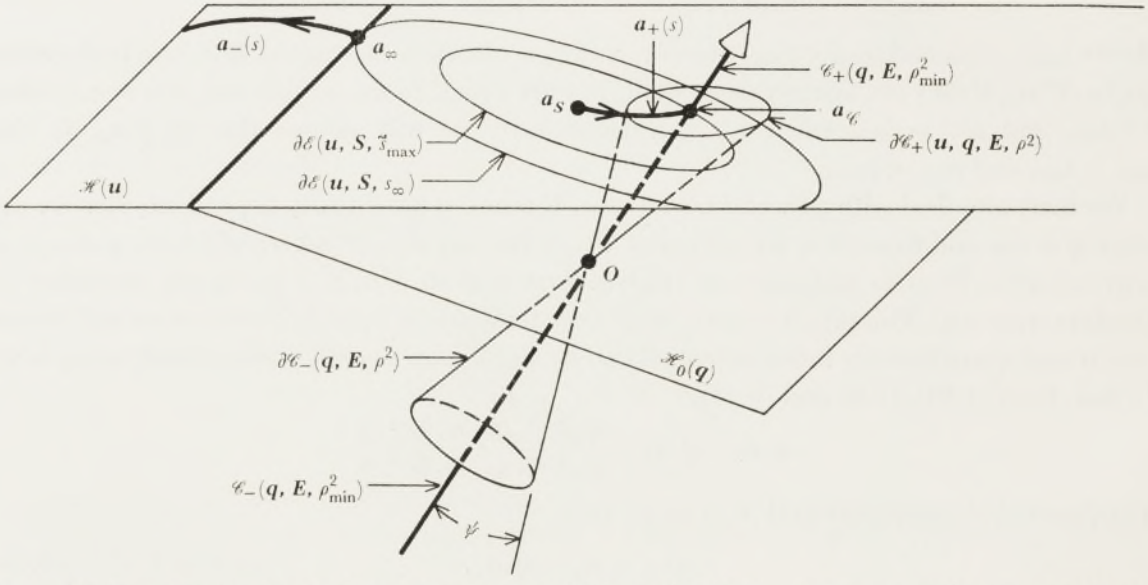

FIGURE 6. Illustration for $N = 3$ of the case $s_\infty > \vec{s}_{\max}$, which requires $\boldsymbol{q}.\boldsymbol{a}_S > 0$. Arrows on the curves $\boldsymbol{a}_+(s)$ and $\boldsymbol{a}_-(s)$ in $\mathscr{H}(\boldsymbol{u})$ point in the direction of increasing θ.

The normal to $\mathscr{H}_0(\boldsymbol{q})$ in $\mathscr{H}(\boldsymbol{u})$ is $(\boldsymbol{I} - \hat{\boldsymbol{u}}\hat{\boldsymbol{u}}).\boldsymbol{q}$ if we work in the geometry induced on $\mathscr{R}^N$ by (4.1). The normal to $\partial\mathscr{E}(\boldsymbol{u}, \boldsymbol{S}, s_\infty)$ at $\boldsymbol{a}_\infty$ is given by (4.22). Thus there is a constant α, such that

$$(\boldsymbol{I} - \hat{\boldsymbol{u}}\hat{\boldsymbol{u}}).\boldsymbol{S}.\boldsymbol{a}_\infty = \alpha(\boldsymbol{I} - \hat{\boldsymbol{u}}\hat{\boldsymbol{u}}).\boldsymbol{q}.$$

Then there are constants α and β such that

$$\boldsymbol{S}.\boldsymbol{a}_\infty = \alpha\boldsymbol{q} + \beta\boldsymbol{u},$$

or

$$\boldsymbol{a}_\infty = \alpha\boldsymbol{S}^{-1}.\boldsymbol{q} + \beta\boldsymbol{S}^{-1}.\boldsymbol{u}.$$

Now $\boldsymbol{q}.\boldsymbol{a}_\infty = 0$ and $\boldsymbol{u}.\boldsymbol{a}_\infty = 1$, so we can evaluate α and β as the solution of a pair of inhomogeneous equations. The result is

$$\boldsymbol{a}_\infty = \frac{(\boldsymbol{S}^{-1}.\boldsymbol{u})(\boldsymbol{q}.\boldsymbol{S}^{-1}.\boldsymbol{q}) - (\boldsymbol{S}^{-1}.\boldsymbol{q})(\boldsymbol{u}.\boldsymbol{S}^{-1}.\boldsymbol{q})}{(\boldsymbol{u}.\boldsymbol{S}^{-1}.\boldsymbol{u})(\boldsymbol{q}.\boldsymbol{S}^{-1}.\boldsymbol{q}) - (\boldsymbol{u}.\boldsymbol{S}^{-1}.\boldsymbol{q})^2}. \tag{6.23}$$

Since $s_\infty = \boldsymbol{a}_\infty . \boldsymbol{S} . \boldsymbol{a}_\infty$, we have

$$s_\infty = \frac{\boldsymbol{q} . \boldsymbol{S}^{-1} . \boldsymbol{q}}{(\boldsymbol{u} . \boldsymbol{S}^{-1} . \boldsymbol{u})(\boldsymbol{q} . \boldsymbol{S}^{-1} . \boldsymbol{q}) - (\boldsymbol{u} . \boldsymbol{S}^{-1} . \boldsymbol{q})^2}. \tag{6.24}$$

From the definition of $s_{\min}$ we always have $s_{\min} \leqslant s_\infty$, but it is perfectly possible to have $s_\infty > \tilde{s}_{\max}$. Figure 6 shows how this can occur; the relative sizes of s_∞ and $\tilde{s}_{\max}$ are determined by the location of $\boldsymbol{a}_S$ in $\mathscr{H}(\boldsymbol{u})$.

It is clear from figure 6 that the structure of the problem will depend heavily on the sign of $\boldsymbol{q} . \boldsymbol{a}_S$. An appropriate notation is required. If $\boldsymbol{q} . \boldsymbol{a}_S \geqslant 0$, we define

$$s^+_{\min} = s_{\min}, \quad s^-_{\min} = s_\infty; \qquad \rho^+_{\max} = \rho_{\max}, \quad \rho^-_{\max} = +\infty; \tag{6.25}$$

while if $\boldsymbol{q} . \boldsymbol{a}_S \leqslant 0$ we define

$$s^+_{\min} = s_\infty, \quad s^-_{\min} = s_{\min}; \qquad \rho^+_{\max} = +\infty, \quad \rho^-_{\max} = \rho_{\max}. \tag{6.26}$$

Also, whatever the sign of $\boldsymbol{q} . \boldsymbol{a}_S$, we define

$$s^+_{\max} = s_{\max}, \quad s^-_{\max} = +\infty; \qquad \rho^+_{\min} = \rho_{\min}, \quad \rho^-_{\min} = \rho_{\text{par}}. \tag{6.27}$$

Reading either + or − consistently throughout, we observe that if $s < s^{+;-}_{\min}$ then $\rho_\pm(s) = +\infty$. If $s^{+;-}_{\min} \leqslant s < s^{+;-}_{\max}$ then $\rho_\pm(s)$ decreases monotonically (we have proved only that it does not increase) from $\rho^{+;-}_{\max}$ to $\rho^{+;-}_{\min}$. If $s^{+;-}_{\max} \leqslant s \leqslant \infty$ then $\rho_\pm(s) = \rho^{+;-}_{\min}$.

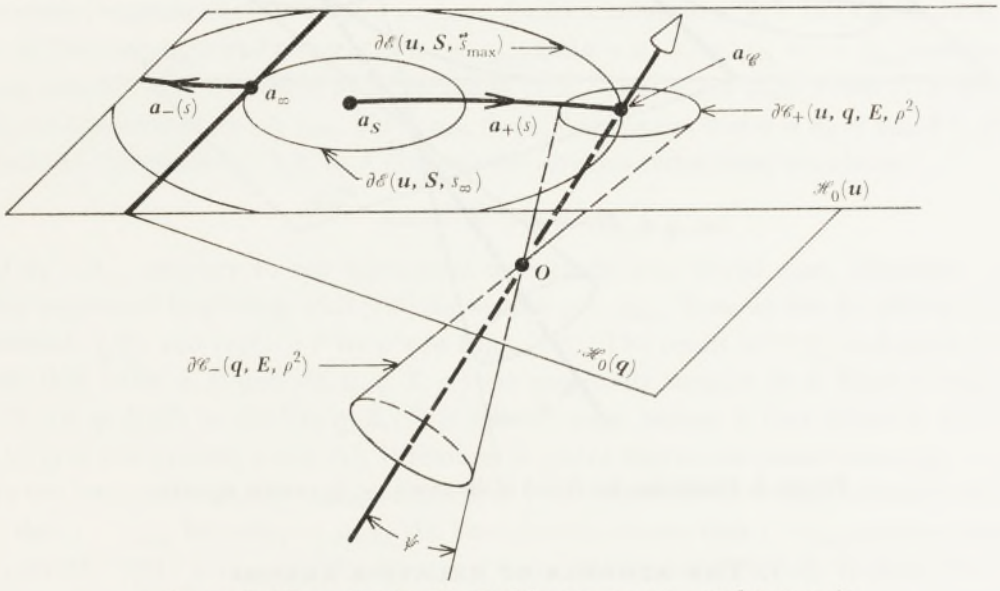

FIGURE 7. Illustration for $N = 3$ of the case $s_\infty < \tilde{s}_{\max}$ and $\boldsymbol{q} . \boldsymbol{a}_S > 0$.

By analogy with lemma 3 we have (reading + or − consistently throughout):

LEMMA 8. *If $s^{+;-}_{\min} \leqslant s$ then $\mathscr{E}(\boldsymbol{u}, \boldsymbol{S}, s) \cap \mathscr{C}_\pm(\boldsymbol{u}, \boldsymbol{q}, \boldsymbol{E}, \rho_\pm(s)^2)$ contains exactly one point, which we denote by $\boldsymbol{a}_\pm(s)$. This point lies on $\partial\mathscr{C}_\pm(\boldsymbol{u}, \boldsymbol{q}, \boldsymbol{E}, \rho_\pm(s)^2)$. When $s^{+;-}_{\min} \leqslant s \leqslant s^{+;-}_{\max}$, $\boldsymbol{a}_\pm(s)$ also lies on $\partial\mathscr{E}(\boldsymbol{u}, \boldsymbol{S}, s)$, and consequently is a point of external tangency of $\partial\mathscr{E}(\boldsymbol{u}, \boldsymbol{S}, s)$ and $\partial\mathscr{C}_\pm(\boldsymbol{u}, \boldsymbol{q}, \boldsymbol{E}, \rho_\pm(s)^2)$.*

The proof of lemma 8 is based on the strict convexity of $\mathscr{C}_\pm(\boldsymbol{u}, \boldsymbol{q}, \boldsymbol{E}, \rho^2)$ and $\mathscr{E}(\boldsymbol{u}, \boldsymbol{S}, s)$, and proceeds almost exactly like the deduction of lemma 3 from lemma 2. We omit the details. A generalization which includes both lemmas 3 and 8 is given in appendix A.

Continuing the same line of argument, we obtain an analogue of lemma 4 (read + or − consistently throughout):

LEMMA 9. *Suppose that* $\mathscr{E}(\boldsymbol{u}, \boldsymbol{S}, s)$ *is externally tangent to* $\mathscr{C}_{\pm}(\boldsymbol{u}, \boldsymbol{q}, \boldsymbol{E}, \rho^2)$. *Then* $s_{\min}^{+,-} \leqslant s \leqslant s_{\max}^{+,-}$ *and* $\rho_{\min}^{+,-} \leqslant \rho \leqslant \rho_{\max}^{+,-}$, *and* $\rho = \rho_{\pm}(s)$, *so the point of external tangency is* $\boldsymbol{a}_{\pm}(s)$.

Again we omit the proof, since it is so like the deduction of lemma 4 from lemmas 2 and 3. The general theory in appendix A includes lemmas 4 and 9.

The argument following lemma 4 can be repeated almost verbatim. If $\rho_{\min}^{+,-} \leqslant \rho \leqslant \rho_{\max}^{+,-}$ then there is precisely one s in $s_{\min}^{+,-} \leqslant s \leqslant s_{\max}^{+,-}$ such that $\rho = \rho_{\pm}(s)$. Therefore, since $\rho_{\pm}(s)$ is non-increasing, it must be strictly decreasing and continuous on $s_{\min}^{+,-} \leqslant s \leqslant s_{\max}^{+,-}$. (The continuity is in the extended sense if $\rho_{\max}^{+,-} = +\infty$.)

Figures 6 to 8 show the relation between $\mathscr{E}(\boldsymbol{u}, \boldsymbol{S}, s)$ and $\mathscr{C}_{+}(\boldsymbol{u}, \boldsymbol{q}, \boldsymbol{E}, s)$ in the three possible cases. Figure 6 refers to the case $\boldsymbol{q}.\boldsymbol{a}_{\boldsymbol{S}} > 0$, $s_{\infty} > \tilde{s}_{\max}$; figure 7 refers to the case $\boldsymbol{q}.\boldsymbol{a}_{\boldsymbol{S}} > 0$, $s_{\infty} < \tilde{s}_{\max}$; and figure 8 has $\boldsymbol{q}.\boldsymbol{a}_{\boldsymbol{S}} < 0$. When $\boldsymbol{q}.\boldsymbol{a}_{\boldsymbol{S}} < 0$, we must always have $s_{\infty} < \tilde{s}_{\max}$, because $\mathscr{E}^0(\boldsymbol{u}, \boldsymbol{S}, s_{\infty})$ contains only points $\boldsymbol{a}$ with $\boldsymbol{q}.\boldsymbol{a} < 0$ while $\mathscr{E}^0(\boldsymbol{u}, \boldsymbol{S}, \tilde{s}_{\max})$ contains points $\boldsymbol{a}$ with $\boldsymbol{q}.\boldsymbol{a} > 0$ and therefore must be the larger ellipsoid.

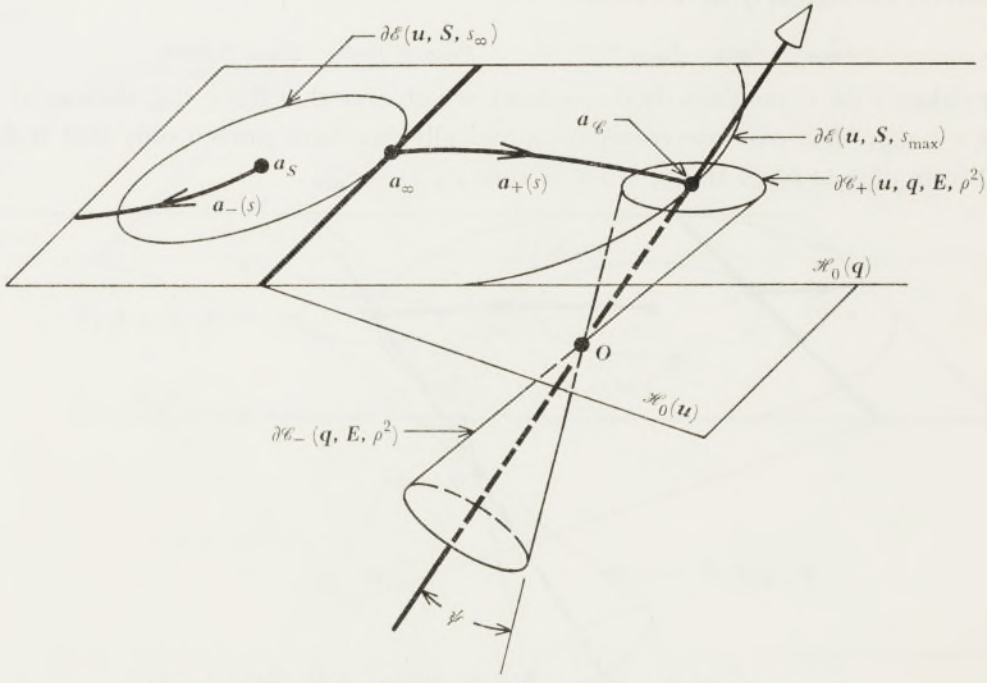

FIGURE 8. Illustration for $N = 3$ of the case $s_{\infty} < \tilde{s}_{\max}$ and $\boldsymbol{q}.\boldsymbol{a}_{\boldsymbol{S}} < 0$.

7. THE ALGEBRA OF RELATIVE ERRORS

The external tangency of $\partial\mathscr{E}(\boldsymbol{u}, \boldsymbol{S}, s)$ and $\partial\mathscr{C}_{\pm}(\boldsymbol{u}, \boldsymbol{q}, \boldsymbol{E}, \rho_{\pm}(s)^2)$ at $\boldsymbol{a}_{\pm}(s)$ gives us a means of calculating $\boldsymbol{a}_{\pm}(s)$ and hence $\rho_{\pm}(s)$.

(a) *Algebraic statement of the geometric problem*

Suppose that $\partial\mathscr{E}(\boldsymbol{u}, \boldsymbol{S}, s)$ and $\partial\mathscr{C}(\boldsymbol{u}, \boldsymbol{q}, \boldsymbol{E}, \rho^2)$ are externally tangent at $\boldsymbol{a}$. Then in the hyperplane $\mathscr{H}(\boldsymbol{u})$ the outward normals to those two hypersurfaces must be antiparallel. These outward normals are given by (4.22) and (6.10). Therefore there is a positive number λ such that

$$\lambda(\boldsymbol{I} - \hat{\boldsymbol{u}}\hat{\boldsymbol{u}}).\boldsymbol{S}.\boldsymbol{a} + (\boldsymbol{I} - \hat{\boldsymbol{u}}\hat{\boldsymbol{u}}).(\boldsymbol{E} - \rho^2\boldsymbol{q}\boldsymbol{q}).\boldsymbol{a} = \boldsymbol{0}. \tag{7.1}$$

If we define $\beta' = [\lambda\boldsymbol{u}.\boldsymbol{S}.\boldsymbol{a} + \boldsymbol{u}.(\boldsymbol{E} - \rho^2\boldsymbol{q}\boldsymbol{q}).\boldsymbol{a}]/(\boldsymbol{u}.\boldsymbol{u})$ and $t' = \rho^2(\boldsymbol{q}.\boldsymbol{a})$ then (7.1) is

$$(\lambda\boldsymbol{S} + \boldsymbol{E}).\boldsymbol{a} = t'\boldsymbol{q} + \beta'\boldsymbol{u}.$$

By dotting this equation on the left with $\boldsymbol{a}$ and using the facts that $\boldsymbol{u}.\boldsymbol{a} = 1$, $\boldsymbol{a}.\boldsymbol{S}.\boldsymbol{a} = s$, and $\boldsymbol{a}.\boldsymbol{E}.\boldsymbol{a} = \rho^2(\boldsymbol{q}.\boldsymbol{a})^2$, we infer that $\beta' = \lambda s$. Thus there is a number t' and a positive number λ such that

$$(\lambda \boldsymbol{S} + \boldsymbol{E}).\boldsymbol{a} = t'\boldsymbol{q} + \lambda s \boldsymbol{u}. \tag{7.2}$$

In addition we have

$$\boldsymbol{u}.\boldsymbol{a} = 1, \tag{7.3}$$

$$\boldsymbol{a}.\boldsymbol{S}.\boldsymbol{a} = s, \tag{7.4}$$

$$\rho^2 = (\boldsymbol{a}.\boldsymbol{E}.\boldsymbol{a})/(\boldsymbol{q}.\boldsymbol{a})^2. \tag{7.5}$$

Equations (7.2) and (7.5) imply a useful expression for ρ:

$$\rho^2 = t'/(\boldsymbol{q}.\boldsymbol{a}).$$

If s is given, then equations (7.2), (7.3), (7.4) and (7.5) are a vector equation and three scalar equations for the vector unknown $\boldsymbol{a}$ and the three scalar unknowns λ, t' and ρ. The first three equations do not involve ρ, so we would expect to find $\boldsymbol{a}$, λ, t' by solving (7.2), (7.3), (7.4), and then to find ρ^2 from (7.5).

If $s^+_{\min} < s < s^+_{\max}$, then according to lemma 8, equations (7.2) to (7.5) have a solution $\boldsymbol{a}$, λ, t', ρ with $\boldsymbol{a} = \boldsymbol{a}_+(s)$, $\lambda > 0$, and $\rho = \rho_+(s)$. And if $s^-_{\min} < s < s^-_{\max}$, then equations (7.2) to (7.5) have a solution $\boldsymbol{a}$, λ, t', ρ with $\boldsymbol{a} = \boldsymbol{a}_-(s)$, $\lambda > 0$, and $\rho = \rho_-(s)$.

Conversely, suppose that for some s we have found a solution $\boldsymbol{a}$, λ, t' of (7.2) to (7.4) which has $\lambda > 0$. Then we claim that either $s^-_{\min} < s < s^-_{\max}$ and $\boldsymbol{a} = \boldsymbol{a}_-(s)$ or $s^+_{\min} < s < s^+_{\max}$ and $\boldsymbol{a} = \boldsymbol{a}_+(s)$. To prove this converse we define ρ^2 by means of (7.5). Clearly $\rho \geqslant \rho_{\min}$. From (7.3) and (7.4), $\mathscr{E}(\boldsymbol{u}, \boldsymbol{S}, s)$ is non-empty, so $s \geqslant s_{\min}$. If $s = s_{\min}$ then (7.4) implies that $\boldsymbol{a} = \boldsymbol{a}_S = s_{\min}\boldsymbol{S}^{-1}.\boldsymbol{u}$; when we substitute this result in (7.2) and perform the obvious reductions, we obtain

$$s_{\min}\boldsymbol{E}.\boldsymbol{S}^{-1}.\boldsymbol{u} = t'\boldsymbol{q},$$

whence $\boldsymbol{a}_S = \boldsymbol{a}_{\mathscr{C}}$, contrary to our agreement to exclude that trivial case. Therefore $s > s_{\min}$. A similar argument beginning with (7.5) shows that $\rho > \rho_{\min}$. Now we dot $\boldsymbol{I} - \hat{\boldsymbol{u}}\hat{\boldsymbol{u}}$ on the left of both sides of (7.2), and replace t' by $\rho^2\boldsymbol{q}.\boldsymbol{a}$ from (7.5). The result is (7.1), and since $\lambda > 0$ we conclude that $\partial\mathscr{E}(\boldsymbol{u}, \boldsymbol{S}, s)$ and $\partial\mathscr{C}(\boldsymbol{u}, \boldsymbol{q}, \boldsymbol{E}, \rho^2)$ are externally tangent at $\boldsymbol{a}$. Since $\boldsymbol{a}$ must be on either $\partial\mathscr{C}_+(\boldsymbol{u}, \boldsymbol{q}, \boldsymbol{E}, \rho^2)$ or $\partial\mathscr{C}_-(\boldsymbol{u}, \boldsymbol{q}, \boldsymbol{E}, \rho^2)$ it follows from lemma 8 that either $\boldsymbol{a} = \boldsymbol{a}_+(s)$ and $\rho = \rho_+(s)$ or $\boldsymbol{a} = \boldsymbol{a}_-(s)$ and $\rho = \rho_-(s)$. It remains to prove that in the former case $s^+_{\min} < s < s^+_{\max}$ while in the latter case $s^-_{\min} < s < s^-_{\max}$. Equations (6.27) show that $s < s^-_{\max}$ is trivial while if $\boldsymbol{a}$ is on $\partial\mathscr{C}_+$ then $s < s^+_{\max}$ because $\rho > \rho_{\min}$. We have already shown that $s > s_{\min}$ and we must have $s \geqslant s^+_{\min}$ on $\partial\mathscr{C}_+$ and $s \geqslant s^-_{\min}$ on $\partial\mathscr{C}_-$, since ρ is finite. It remains then only to show that $s \neq s_\infty$. But if $s = s_\infty$ then $\boldsymbol{a} = \boldsymbol{a}_\infty$ and $\boldsymbol{q}.\boldsymbol{a} = 0$, while $\boldsymbol{a}.\boldsymbol{E}.\boldsymbol{a} > 0$, a situation which forces (7.5) to produce the result $t' = \infty$.

We have proved the analogue of theorem 1:

THEOREM 3. *Suppose* $s_{\min} < \tilde{s}_{\max}$. *Then for any* s *in* $s^{+,-}_{\min} < s < s^{+,-}_{\max}$ *equations* (7.2) *to* (7.5) *have a solution* $\boldsymbol{a}, \lambda, t', \rho$ *with* $\lambda > 0$, $\boldsymbol{a} = \boldsymbol{a}_\pm(s)$ *and* $\rho = \rho_\pm(s)$. *Conversely, if for some* s *those equations have a solution* $\boldsymbol{a}, \lambda, t', \rho$ *with* $\lambda > 0$, *then either* $s^-_{\min} < s < s^-_{\max}$, $\boldsymbol{a} = \boldsymbol{a}_-(s)$ *and* $\rho = \rho_-(s)$, *or* $s^+_{\min} < s < s^+_{\max}$, $\boldsymbol{a} = \boldsymbol{a}_+(s)$, *and* $\rho = \rho_+(s)$.

Again, we could easily have deduced (7.2) by using the method of Lagrange multipliers to minimize $\boldsymbol{a}.\boldsymbol{E}.\boldsymbol{a}/(\boldsymbol{q}.\boldsymbol{a})^2$ subject to the constraints $\boldsymbol{u}.\boldsymbol{a} = 1$ and $\boldsymbol{a}.\boldsymbol{S}.\boldsymbol{a} = s$, but we would not then know that $\lambda > 0$, and we would not have the converse assertion in theorem 3. We will make heavy use of these two extra items of information.

(b) Solving the algebraic problem

For any s, equations (7.2) to (7.5) have at most two solutions with $\lambda > 0$, namely $\boldsymbol{a}_\pm(s)$, $\lambda_\pm(s)$, $t_\pm(s)$, $\rho_\pm(s)$. As with (5.2), the structure of (7.2) suggests that the computation will be simpler if we regard λ rather than s as the independent variable and try to find solutions $\boldsymbol{a}_\pm(\lambda)$, $t'_\pm(\lambda)$, $s_\pm(\lambda)$, $\rho_\pm(\lambda)$ to (7.2) to (7.5). We know from theorem 3 that we will find at most two solutions, one on $\partial\mathscr{C}_+$ and one on $\partial\mathscr{C}_-$, and that if they exist then $\boldsymbol{a}_\pm(\lambda) = \boldsymbol{a}_\pm(s_\pm(\lambda))$. Thus $\boldsymbol{a}_+(\lambda)$ will be the point of external tangency of $\partial\mathscr{E}(\boldsymbol{u}, \boldsymbol{S}, s_+(\lambda))$ and $\partial\mathscr{C}_+(\boldsymbol{u}, \boldsymbol{q}, \boldsymbol{E}, \rho_+(\lambda)^2)$ while $\boldsymbol{a}_-(\lambda)$ will be the point of external tangency of $\partial\mathscr{E}(\boldsymbol{u}, \boldsymbol{S}, s(\lambda))$ and $\partial\mathscr{C}_-(\boldsymbol{u}, \boldsymbol{q}, \boldsymbol{E}, \rho_-(\lambda)^2)$.

In the spirit of equation (5.6), we choose a positive constant w so that, roughly speaking, $\boldsymbol{S}$ and $w\boldsymbol{E}$ are of comparable numerical size. Then we replace λ by a new independent variable θ, defined thus:

$$\lambda = (w \tan\theta)^{-1}, \tag{7.6}$$

The domain of the independent variable θ is the finite interval $0 < \theta < \frac{1}{2}\pi$, corresponding to the infinite interval $0 < \lambda < \infty$. If we define $\boldsymbol{W}(\theta)$ by (5.8) and define $t = t' w \sin\theta$ then equations (7.2) to (7.5) can be rewritten as

$$\boldsymbol{W}(\theta).\boldsymbol{a} = t\boldsymbol{q} + s\cos\theta\boldsymbol{u} \tag{7.7}$$

$$\boldsymbol{u}.\boldsymbol{a} = 1, \tag{7.8}$$

$$s = \boldsymbol{a}.\boldsymbol{S}.\boldsymbol{a}, \tag{7.9}$$

$$\rho^2 = \frac{\boldsymbol{a}.\boldsymbol{E}.\boldsymbol{a}}{(\boldsymbol{q}.\boldsymbol{a})^2}. \tag{7.10}$$

As one consequence of these equations we deduce that (7.10) can be replaced by

$$\rho^2 = \frac{t}{(\boldsymbol{q}.\boldsymbol{a})\, w \sin\theta}. \tag{7.11}$$

Since we are regarding θ as the independent variable, the problem is to pick a value for θ in the interval $0 \leqslant \theta \leqslant \frac{1}{2}\pi$ and then try to solve the four equations (7.7) to (7.10) for the four unknowns $\boldsymbol{a}(\theta)$, $t(\theta)$, $s(\theta)$, $\rho(\theta)$. Evidently the heart of the problem is to solve (7.7) to (7.9) for $\boldsymbol{a}(\theta)$, $t(\theta)$, $s(\theta)$; then $\rho(\theta)$ can be obtained immediately from either (7.10) or (7.11).

To give the solution of (7.7) to (7.9) in compact form we will need some notation. For any vector $\boldsymbol{f}$ in $\mathscr{R}_N$ we define

$$\tilde{\boldsymbol{f}}(\theta) = \boldsymbol{W}(\theta)^{-1}.\boldsymbol{f}.$$

Then for any vectors $\boldsymbol{f}$ and $\boldsymbol{g}$ we have $\boldsymbol{f}.\tilde{\boldsymbol{g}} = \tilde{\boldsymbol{f}}.\boldsymbol{g} = \boldsymbol{f}.\boldsymbol{W}^{-1}.\boldsymbol{g} = \tilde{\boldsymbol{f}}.\boldsymbol{W}.\tilde{\boldsymbol{g}}$. Since $\boldsymbol{W}^{-1}$ is positive definite, $\boldsymbol{f}.\tilde{\boldsymbol{f}} > 0$ unless $\boldsymbol{f} = \boldsymbol{0}$. In terms of the particular vectors $\boldsymbol{u}$ and $\boldsymbol{q}$ given by (2.19) and (2.16), we define

$$\boldsymbol{h}(\theta) = (\boldsymbol{q}\boldsymbol{u} - \boldsymbol{u}\boldsymbol{q}).\tilde{\boldsymbol{u}}(\theta), \tag{7.12}$$

so that

$$\tilde{\boldsymbol{h}}(\theta) = (\tilde{\boldsymbol{q}}\tilde{\boldsymbol{u}} - \tilde{\boldsymbol{u}}\tilde{\boldsymbol{q}}).\boldsymbol{u}. \tag{7.13}$$

Two properties of $\tilde{\boldsymbol{h}}$ will be useful. Evidently

$$\boldsymbol{u}.\tilde{\boldsymbol{h}} = 0. \tag{7.14}$$

Then if we dot $\tilde{\boldsymbol{h}}$ into (7.12) we obtain

$$\boldsymbol{q}.\tilde{\boldsymbol{h}} = \frac{\boldsymbol{h}.\tilde{\boldsymbol{h}}}{\boldsymbol{u}.\tilde{\boldsymbol{u}}} > 0. \tag{7.15}$$

In (7.15) the inequality is strict because if $\boldsymbol{h}.\tilde{\boldsymbol{h}} = 0$ then $\boldsymbol{h} = \boldsymbol{0}$, so, from (7.12), $\boldsymbol{u}$ and $\boldsymbol{q}$ are linearly dependent.

Now we can solve (7.7) to (7.9). If for the moment we regard t as given, then (7.7) and (7.8) can be solved for $\boldsymbol{a}$ and s. The result is

$$\boldsymbol{a}(\theta) = \frac{\tilde{\boldsymbol{u}} + t\tilde{\boldsymbol{h}}}{\boldsymbol{u}.\tilde{\boldsymbol{u}}}, \tag{7.16}$$

$$s(\theta)\cos\theta = \frac{1 - t\boldsymbol{u}.\tilde{\boldsymbol{q}}}{\boldsymbol{u}.\tilde{\boldsymbol{u}}}. \tag{7.17}$$

When we substitute these expressions for $\boldsymbol{a}$ and s in (7.9) we obtain a quadratic equation for t:

$$at^2 + bt - c = 0, \tag{7.18}$$

The coefficients in this equation are calculated as follows: we define $\boldsymbol{R} = \boldsymbol{S}\cos\theta$ and $\boldsymbol{D} = w\boldsymbol{E}\sin\theta$, so that $\boldsymbol{W} = \boldsymbol{R} + \boldsymbol{D}$. Then

$$\left.\begin{aligned} a &= \tilde{\boldsymbol{h}}.\boldsymbol{R}.\tilde{\boldsymbol{h}}, \\ b &= (\boldsymbol{u}.\tilde{\boldsymbol{u}})\,(\boldsymbol{q}.\tilde{\boldsymbol{u}}) + 2\tilde{\boldsymbol{u}}.\boldsymbol{R}.\tilde{\boldsymbol{h}}, \\ &= (\boldsymbol{u}.\tilde{\boldsymbol{u}})\,(\boldsymbol{q}.\tilde{\boldsymbol{u}}) - 2\tilde{\boldsymbol{u}}.\boldsymbol{D}.\tilde{\boldsymbol{h}}, \\ c &= \tilde{\boldsymbol{u}}.\boldsymbol{D}.\tilde{\boldsymbol{u}}. \end{aligned}\right\} \tag{7.19}$$

In deducing the second expression for b, we used (7.14).

The two solutions of (7.18) are

$$t_{\pm} = \frac{-b \pm (b^2 + 4ac)^{\frac{1}{2}}}{2a}, \tag{7.20}$$

where, at this stage, the subscript + or − on t refers to nothing but the choice of signs on the right-hand side of (7.20). If $0 \leqslant \theta \leqslant \frac{1}{2}\pi$ then $a \geqslant 0$ and $c \geqslant 0$, so both roots (7.20) are real. In fact if $0 < \theta < \frac{1}{2}\pi$ then $a > 0$ and $c > 0$, so in that open interval t_+ and t_- are distinct. At $\theta = 0$, we have $b = s_{\min}^{-2}\,\boldsymbol{q}.\boldsymbol{a}_{\boldsymbol{S}}$, so t_+ and t_- are distinct unless $\boldsymbol{q}.\boldsymbol{a}_{\boldsymbol{S}} = 0$. At $\theta = \frac{1}{2}\pi$, we have $b = w^{-1}e_{\min}^{-4}\,\boldsymbol{q}.\boldsymbol{a}_{\boldsymbol{E}}$, so t_+ and t_- are distinct unless $\boldsymbol{q}.\boldsymbol{a}_{\boldsymbol{E}} = 0$.

If we insert any solution t of (7.18) into (7.16) and (7.17) the result is a solution $\boldsymbol{a}$, t, s of (7.7) to (7.9). Then we can calculate ρ^2 from (7.10), and by theorem 3 we know that $\boldsymbol{a}$ is a point of external tangency of $\partial\mathscr{E}(\boldsymbol{u}, \boldsymbol{S}, s)$ and $\partial\mathscr{C}(\boldsymbol{u}, \boldsymbol{q}, \boldsymbol{E}, \rho^2)$. But is $\boldsymbol{a}$ on $\partial\mathscr{C}_+(\boldsymbol{u}, \boldsymbol{q}, \boldsymbol{E}, \rho^2)$ or $\partial\mathscr{C}_-(\boldsymbol{u}, \boldsymbol{q}, \boldsymbol{E}, \rho^2)$? From (7.15) and (7.16) we see that $\boldsymbol{q}.\boldsymbol{a}$ is positive or negative, and $\boldsymbol{a}$ is on $\partial\mathscr{C}_+$ or $\partial\mathscr{C}_-$, according as $t - t_\infty$ is positive or negative, where

$$t_\infty = -\frac{\boldsymbol{q}.\tilde{\boldsymbol{u}}}{\boldsymbol{q}.\tilde{\boldsymbol{h}}}. \tag{7.21}$$

Now we claim that if $0 < \theta < \frac{1}{2}\pi$ then t_+ produces via (7.16) an $\boldsymbol{a}$ which is on $\partial\mathscr{C}_+$ while t_- produces an $\boldsymbol{a}$ on $\partial\mathscr{C}_-$. That is, we claim that $t_- < t_\infty < t_+$. Since (7.19) shows that $a > 0$, it suffices to prove that $P(t_\infty) < 0$ where $P(t) = at^2 + bt - c$. From (7.14), (7.19) and (7.21) we deduce that

$$\begin{aligned} -(\boldsymbol{q}.\tilde{\boldsymbol{h}})^2 P(t_\infty) &= (\boldsymbol{q}.\tilde{\boldsymbol{u}})^2(\tilde{\boldsymbol{h}}.\boldsymbol{D}.\tilde{\boldsymbol{h}}) - 2(\boldsymbol{q}.\tilde{\boldsymbol{u}})\,(\boldsymbol{q}.\tilde{\boldsymbol{h}})\,(\tilde{\boldsymbol{u}}.\boldsymbol{D}.\tilde{\boldsymbol{h}}) + (\boldsymbol{q}.\tilde{\boldsymbol{h}})^2(\tilde{\boldsymbol{u}}.\boldsymbol{D}.\tilde{\boldsymbol{u}}) \\ &\geqslant |\boldsymbol{q}.\tilde{\boldsymbol{u}}|^2(\tilde{\boldsymbol{h}}.\boldsymbol{D}.\tilde{\boldsymbol{h}}) - 2|\boldsymbol{q}.\tilde{\boldsymbol{u}}|\,|\boldsymbol{q}.\tilde{\boldsymbol{h}}|\,|\tilde{\boldsymbol{u}}.\boldsymbol{D}.\tilde{\boldsymbol{h}}| + |\boldsymbol{q}.\tilde{\boldsymbol{h}}|^2(\tilde{\boldsymbol{u}}.\boldsymbol{D}.\tilde{\boldsymbol{u}}). \end{aligned}$$

Since $\boldsymbol{u}$ and $\boldsymbol{q}$ are assumed linearly independent, (7.12) implies that $\boldsymbol{u}$ and $\boldsymbol{h}$ must be. Then $\tilde{\boldsymbol{u}}$ and $\tilde{\boldsymbol{h}}$ are linearly independent. When $0 < \theta < \frac{1}{2}\pi$, $\boldsymbol{D}$ is positive definite, so by Schwarz's inequality

$$|\tilde{\boldsymbol{u}}.\boldsymbol{D}.\tilde{\boldsymbol{h}}| < (\tilde{\boldsymbol{u}}.\boldsymbol{D}.\tilde{\boldsymbol{u}})^{\frac{1}{2}}(\tilde{\boldsymbol{h}}.\boldsymbol{D}.\tilde{\boldsymbol{h}})^{\frac{1}{2}}.$$

But then

$$-(\boldsymbol{q}.\tilde{\boldsymbol{h}})^2 P(t_\infty) > [|\boldsymbol{q}.\tilde{\boldsymbol{u}}|\,(\tilde{\boldsymbol{h}}.\boldsymbol{D}.\tilde{\boldsymbol{h}})^{\frac{1}{2}} - |\boldsymbol{q}.\tilde{\boldsymbol{h}}|\,(\tilde{\boldsymbol{u}}.\boldsymbol{D}.\tilde{\boldsymbol{u}})^{\frac{1}{2}}]^2.$$

Thus $P(t_\infty) < 0$. We conclude that the subscript + or − on the left side of (7.20) gives not merely the choice of sign on the right hand side of that equation but also the sign of $\boldsymbol{q}.\boldsymbol{a}$ when $\boldsymbol{a}$ is obtained by substituting t_+ or t_- from (7.20) into (7.16).

We conclude that for any θ in $0 < \theta < \frac{1}{2}\pi$, equations (7.7) to (7.10) have two different solutions, $\boldsymbol{a}_+(\theta)$, $t_+(\theta)$, $s_+(\theta)$, $\rho_+(\theta)$ and $\boldsymbol{a}_-(\theta)$, $t_-(\theta)$, $s_-(\theta)$, $\rho_-(\theta)$. To obtain these two solutions, we calculate $a(\theta)$, $b(\theta)$, $c(\theta)$ from (7.19), and then obtain $t_+(\theta)$ and $t_-(\theta)$ from (7.20). We find $\boldsymbol{a}_\pm(\theta)$ by substituting $t_\pm(\theta)$ in (7.16). Then we can find $s_\pm(\theta)$ from either (7.17) or (7.9), and we can find $\rho_\pm(\theta)$ from either (7.11) or (7.10). Then $\boldsymbol{a}_+(\theta)$ is the unique point of external tangency of $\partial\mathscr{E}(\boldsymbol{u}, \boldsymbol{S}, s_+(\theta))$ and $\partial\mathscr{C}_+(\boldsymbol{u}, \boldsymbol{q}, \boldsymbol{E}, \rho_+(\theta)^2)$ while $\boldsymbol{a}_-(\theta)$ is the unique point of external tangency of $\partial\mathscr{E}(\boldsymbol{u}, \boldsymbol{S}, s_-(\theta))$ and $\partial\mathscr{C}(\boldsymbol{u}, \boldsymbol{q}, \boldsymbol{E}, \rho_-(\theta)^2)$. The functions $s_+(\theta)$ and $\rho_+(\theta)^2$ give a parametric representation of $\rho_+(s)$, while $s_-(\theta)$ and $\rho_-(\theta)^2$ give a parametric representation of $\rho_-(s)$. It is clear from (7.20), (7.16), (7.17) and (7.11) that if $0 < \theta < \frac{1}{2}\pi$ then $\boldsymbol{a}_\pm(\theta)$, $t_\pm(\theta)$, $s_\pm(\theta)$ and $\rho_\pm(\theta)$ are continuous (in fact continuously differentiable) functions of θ.

It follows from theorem 3 that $s^-_{\min} < s_-(\theta) < s^-_{\max}$, $s^+_{\min} < s_+(\theta) < s^+_{\max}$, $\rho^-_{\min} < \rho_-(\theta) < \rho^-_{\max}$ and $\rho^+_{\min} < \rho_+(\theta) < \rho^+_{\max}$ when $0 < \theta < \frac{1}{2}\pi$. Theorem 3 also implies that if $s^-_{\min} < s < s^-_{\max}$ then there is exactly one θ in $0 < \theta < \frac{1}{2}\pi$ such that $s = s_-(\theta)$; and if $s^+_{\min} < s < s^+_{\max}$ then there is exactly one θ in $0 < \theta < \frac{1}{2}\pi$ such that $s = s_+(\theta)$. Thus $s_\pm(\theta)$ must be strictly monotone functions of θ in $0 < \theta < \frac{1}{2}\pi$. Since $\rho_\pm(s)$ are strictly monotone in $s^{+,-}_{\min} < s < s^{+,-}_{\max}$, therefore $\rho_\pm(\theta)$ must also be strictly monotone functions of θ in $0 < \theta < \frac{1}{2}\pi$. The interesting (non-constant) parts of the curves $\rho_\pm(s)$ are traced out by the pairs $s_\pm(\theta)$, $\rho_\pm(\theta)$ as θ increases from 0 to $\frac{1}{2}\pi$, but the uninteresting (constant) parts of those curves fall outside the parametrization.

(c) *The shape of $\rho(s)$, the tradeoff curve for relative error against spread*

We can now describe in general terms the appearance of the two curves $\rho_\pm(s)$ and their parametrization by θ.

First, what happens as θ approaches 0 or $\frac{1}{2}\pi$? If $\theta \to \frac{1}{2}\pi$ then $\boldsymbol{R} = \boldsymbol{S}\cos\theta \to \boldsymbol{0}$ while $\boldsymbol{D} = w\boldsymbol{E}\sin\theta \to w\boldsymbol{E}$ and $\boldsymbol{W} \to w\boldsymbol{E}$. Thus

$$\lim_{\theta\to\frac{1}{2}\pi} \tilde{\boldsymbol{h}} = \frac{\boldsymbol{q}.\boldsymbol{a}_E}{w^2\epsilon^4_{\min}}(\boldsymbol{a}_{\mathscr{C}} - \boldsymbol{a}_E), \tag{7.22}$$

and, from (7.19) and (7.22),

$$\lim_{\theta\to\frac{1}{2}\pi}\left(\frac{a}{\frac{1}{2}\pi-\theta}\right) = \left(\frac{\boldsymbol{q}.\boldsymbol{a}_E}{w^2\epsilon^4_{\min}}\right)^2 (\boldsymbol{a}-\boldsymbol{a}_E).\boldsymbol{S}.(\boldsymbol{a}_{\mathscr{C}}-\boldsymbol{a}_E)$$

$$\lim_{\theta\to\pi} b = \left(\frac{\boldsymbol{q}.\boldsymbol{a}_E}{w^2\epsilon^4_{\min}}\right),$$

$$\lim_{\theta\to\frac{1}{2}\pi} c = \frac{1}{w\epsilon^2_{\min}}.$$

Then

$$\lim_{\theta\to\frac{1}{2}\pi} t_+(\theta) = \frac{w\epsilon^2_{\min}}{\boldsymbol{q}.\boldsymbol{a}_E}$$

and

$$\lim_{\theta\to\frac{1}{2}\pi} (\tfrac{1}{2}\pi-\theta)\, t_-(\theta) = -\left(\frac{w^2\epsilon^4_{\min}}{\boldsymbol{q}.\boldsymbol{a}_E}\right)\frac{1}{(\boldsymbol{a}_{\mathscr{C}}-\boldsymbol{a}_E).\boldsymbol{S}.(\boldsymbol{a}_{\mathscr{C}}-\boldsymbol{a}_E)},$$

so

$$\lim_{\theta\to\pi} \boldsymbol{a}_+(\theta) = \boldsymbol{a}_{\mathscr{C}} \tag{7.23}$$

and

$$\lim_{\theta\to\pi} (\tfrac{1}{2}\pi-\theta)\,\boldsymbol{a}(\theta) = -\left[\frac{w\epsilon^2_{\min}}{(\boldsymbol{a}_{\mathscr{C}}-\boldsymbol{a}_E).\boldsymbol{S}.(\boldsymbol{a}_{\mathscr{C}}-\boldsymbol{a}_E)}\right](\boldsymbol{a}_{\mathscr{C}}-\boldsymbol{a}_E). \tag{7.24}$$

It follows that
$$\lim_{\theta\to\frac{1}{2}\pi} s_+(\theta) = \tilde{s}_{\max}, \quad \lim_{\theta\to\frac{1}{2}\pi} \rho_+(\theta) = \rho_{\min}, \tag{7.25}$$

and
$$\left.\begin{aligned} \lim_{\theta\to\frac{1}{2}\pi} (\tfrac{1}{2}\pi-\theta)^2 s_-(\theta) &= \frac{w^2 e_{\min}^4}{(\boldsymbol{a}_{\mathscr{C}}-\boldsymbol{a}_E).\boldsymbol{S}.(\boldsymbol{a}_{\mathscr{C}}-\boldsymbol{a}_E)}, \\ \lim_{\theta\to\frac{1}{2}\pi} \rho_-(\theta) &= \rho_{\mathrm{par}}. \end{aligned}\right\} \tag{7.26}$$

As $\theta\to 0$, we have $\boldsymbol{R}\to\boldsymbol{S}$, $\boldsymbol{D}\to 0$, and $\boldsymbol{W}\to\boldsymbol{S}$. Thus

$$\lim_{\theta\to 0} \tilde{\boldsymbol{h}} = \left[\frac{\boldsymbol{q}.\boldsymbol{a}_S}{s_{\min}(s_\infty - s_{\min})}\right](\boldsymbol{a}_S - \boldsymbol{a}_\infty). \tag{7.27}$$

In deducing (7.27), we have used (6.24) to infer that

$$\boldsymbol{q}.\boldsymbol{S}^{-1}.\boldsymbol{q} = \frac{s_\infty(\boldsymbol{q}.\boldsymbol{a}_S)^2}{s_{\min}(s_\infty - s_{\min})},$$

and (6.23) to infer that
$$\boldsymbol{S}^{-1}.\boldsymbol{q} = \left(\frac{s_\infty \boldsymbol{q}.\boldsymbol{a}_S}{s_\infty - s_{\min}}\right)\left(\frac{\boldsymbol{a}_S}{s_{\min}} - \frac{\boldsymbol{a}_\infty}{s_\infty}\right).$$

Also, as $\theta\to 0$, from (7.19) and (7.27),

$$\lim_{\theta\to 0} a = \frac{(\boldsymbol{q}.\boldsymbol{a}_S)^2}{s_{\min}^2(s_\infty - s_{\min})}, \quad \lim_{\theta\to 0} b = \frac{\boldsymbol{q}.\boldsymbol{a}_S}{s_{\min}^2}, \quad \lim_{\theta\to 0}\left(\frac{c}{\theta}\right) = \frac{w e_{\max}^2}{s_{\min}^2}.$$

It is clear from (7.20) that as $\theta\to 0$ and $c\to 0$ the behaviour of t_+ and t_- depends on the sign of b, i.e. the sign of $\boldsymbol{q}.\boldsymbol{a}_S$. We will ignore the special case $\boldsymbol{q}.\boldsymbol{a}_S = 0$. If $\boldsymbol{q}.\boldsymbol{a}_S > 0$ we have

$$\lim_{\theta\to 0} t_+(\theta) = 0, \quad \lim_{\theta\to 0} t_-(\theta) = \frac{s_\infty - s_{\min}}{\boldsymbol{q}.\boldsymbol{a}_S}, \tag{7.28}$$

and
$$\lim_{\theta\to 0} \boldsymbol{a}_+(\theta) = \boldsymbol{a}_S, \tag{7.29}$$

$$\lim_{\theta\to 0} \boldsymbol{a}_-(\theta) = \boldsymbol{a}_\infty. \tag{7.30}$$

Then, for $\boldsymbol{q}.\boldsymbol{a}_S > 0$,
$$\lim_{\theta\to 0} s_+(\theta) = s_{\min}, \quad \lim_{\theta\to 0} \rho_+(\theta) = \rho_{\max}, \tag{7.31}$$

and
$$\lim_{\theta\to 0} s_-(\theta) = s_\infty, \quad \lim_{\theta\to 0} \theta\rho_-(\theta) = \frac{(\boldsymbol{a}_\infty.\boldsymbol{E}.\boldsymbol{a}_\infty)^{\frac{1}{2}}}{|\boldsymbol{q}.\partial_\theta \boldsymbol{a}_-(0)|}. \tag{7.32}$$

In (7.32) and henceforth, if $f(\theta)$ is a function of θ, its derivative with respect to θ is written $\partial_\theta f$. Formulae (7.28) to (7.32) all refer to the case $\boldsymbol{q}.\boldsymbol{a}_S > 0$. When $\boldsymbol{q}.\boldsymbol{a}_S < 0$, the subscripts + and − must be interchanged everywhere in (7.28) to (7.32).

At this point we are finally able to state that $s_+(\theta)$ and $s_-(\theta)$ *increase* monotonically as θ increases from 0 to $\frac{1}{2}\pi$. We already knew that $s_+(\theta)$ depended monotonically on θ, and now it is clear that $s_-(0) < s_-(\frac{1}{2}\pi)$ and $s_+(0) < s_+(\frac{1}{2}\pi)$. But since $\rho_\pm(s)$ is monotonic decreasing in s, it follows that $\rho_+(\theta)$ and $\rho_-(\theta)$ decrease monotonically as θ increases from 0 to $\frac{1}{2}\pi$.

For $N = 3$, figures 6 to 8 show the curves $\boldsymbol{a}_+(s)$ and $\boldsymbol{a}_-(s)$ and their behaviour as $\theta\to 0$ or $\theta\to\frac{1}{2}\pi$. The arrows on the curves $\boldsymbol{a}_\pm(s)$ in those figures point in the direction of increasing θ.

Now we are able to sketch the general appearance of the two curves $\rho_+(s)^2$ and $\rho_-(s)^2$. A number of cases must be considered. Figure 9 shows the two possibilities when $\boldsymbol{q}.\boldsymbol{a}_S > 0$ and $\tilde{s}_{\max} < s_\infty$. Since we are only interested in $\rho(s)$ for $s_{\min} \leqslant s \leqslant \tilde{s}_{\max}$, and since $\rho(s) = \min\{\rho_+(s), \rho_-(s)\}$, evidently when $\boldsymbol{q}.\boldsymbol{a}_S > 0$ and $\tilde{s}_{\max} < s_\infty$ we can ignore $\rho_-(s)$ altogether, whatever the sign of $\rho_{\mathrm{par}} - \rho_{\max}$.

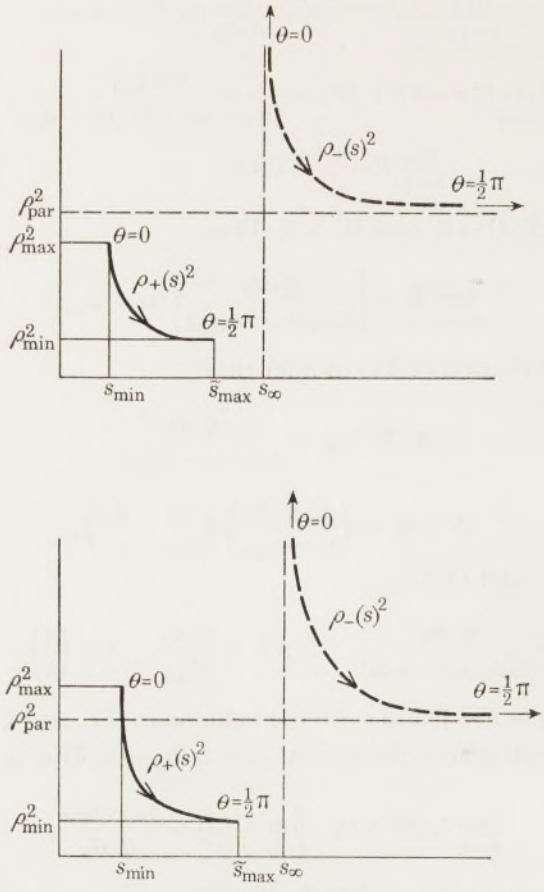

FIGURE 9. A schematic picture of the two curves, $\rho_+(s)$ solid and $\rho_-(s)$ dashed, when $s_\infty > \bar{s}_{\max}$. The tradeoff curve of relative error against spread in $\rho(s) = \min\{\rho_+(s), \rho_-(s)\}$. In the upper graph we assume $\rho_{\text{par}} > \rho_{\max}$, and in the lower $\rho_{\text{par}} < \rho_{\max}$.

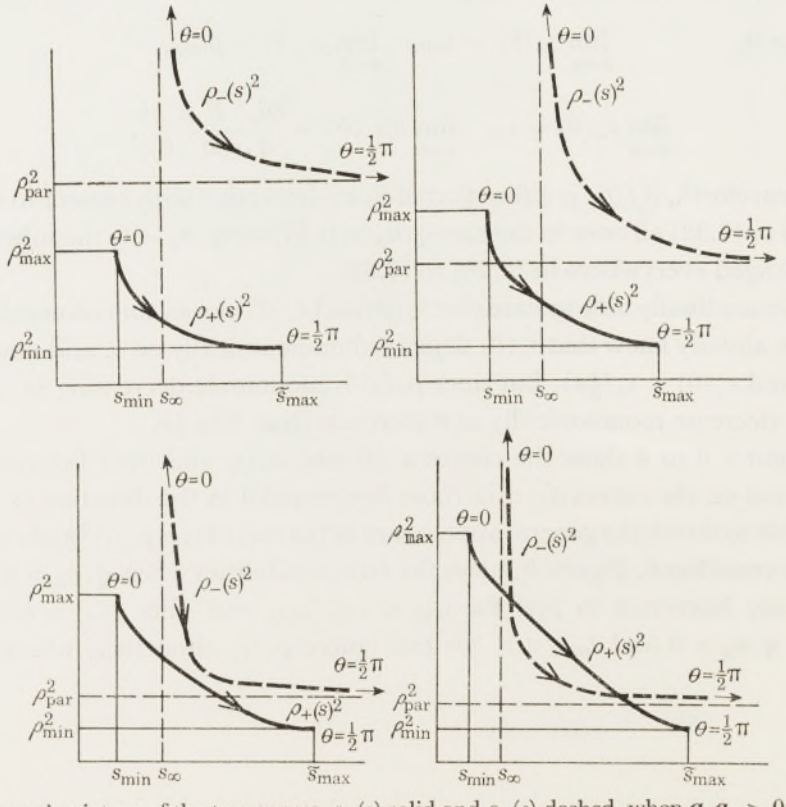

FIGURE 10. A schematic picture of the two curves, $\rho_+(s)$ solid and $\rho_-(s)$ dashed, when $\boldsymbol{q} \cdot \boldsymbol{a}_S > 0$ and $s_\infty < \bar{s}_{\max}$. The tradeoff curve of relative error against spread is $\rho(s) = \min\{\rho_+(s), \rho_-(s)\}$. All four cases are possible, and examples of each have been computed.

In figure 10 we suppose that $\boldsymbol{q}.\boldsymbol{a}_S > 0$ and $s_\infty < \tilde{s}_{\max}$. In figure 10*a* we have $\rho_{\text{par}} \geqslant \rho_{\max}$, while in figure 10*b* the point $(s_\infty, \rho_{\text{par}}^2)$ lies above $(s_\infty, \rho_+(s_\infty)^2)$, even though $\rho_{\text{par}} < \rho_{\max}$. In figure 10*c* the former point has dropped below the latter, but not far enough to produce an intersection of $\rho_+(s)$ and $\rho_-(s)$. In figure 10*d* such an intersection has appeared. Since $\rho(s) = \min\{\rho_+(s), \rho_-(s)\}$, it is clear that $\rho_-(s)$ is of interest only in figure 10*d*. When $\boldsymbol{q}.\boldsymbol{a}_S > 0$ and $s_\infty < \tilde{s}_{\max}$ the most efficient scheme for calculating the curve $\rho(s)$ is evidently as follows: first calculate the whole curve $\rho_+(s)$. Then if $\rho_{\text{par}} \geqslant \rho_{\max}$, set $\rho(s) = \rho_+(s)$ and ignore $\rho_-(s)$. If $\rho_{\max} > \rho_{\text{par}} \geqslant \rho_+(s_\infty)$, set $\rho(s) = \rho_+(s)$ and ignore $\rho_-(s)$. If $\rho_+(s_\infty) > \rho_{\text{par}}$, then calculate that part of the curve $\rho_-(s)$ which has $\rho_-(s) < \rho_+(s_\infty)$ and $\rho_+(s) > \rho_{\text{par}}$. In this range of s, set $\rho(s) = \min\{\rho_+(s), \rho_-(s)\}$, and elsewhere set $\rho(s) = \rho_+(s)$.

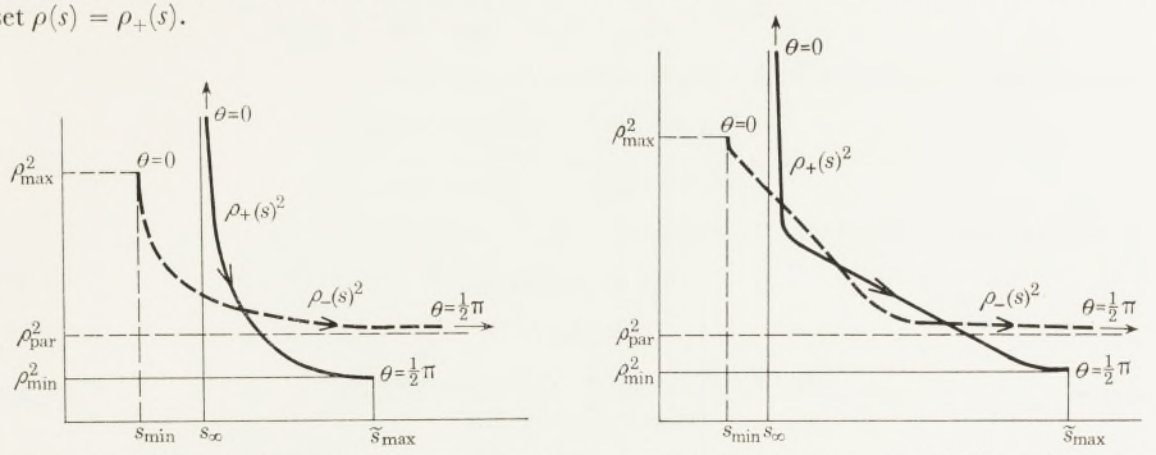

FIGURE 11. A schematic picture of the two curves, $\rho_+(s)$ solid and $\rho_-(s)$ dashed, when $\boldsymbol{q}.\boldsymbol{a}_S < 0$. The tradeoff curve of relative error against spread is $\rho(s) = \min\{\rho_+(s), \rho_-(s)\}$. Multiple crossings of the ρ_+ and ρ_- curves, as in the right figure, are believed to be possible, but no examples have yet been encountered.

In figure 11 we suppose that $\boldsymbol{q}.\boldsymbol{a}_S < 0$. In that case we always have $s_\infty < \tilde{s}_{\max}$ and $\rho_{\text{par}} < \rho_{\max}$, so the two curves $\rho_+(s)$ and $\rho_-(s)$ always cross at least once, as in figure 11*a*. It appears possible in principle that they can cross three times, as in figure 11*b*, or perhaps more often when N is large. When s is close to $s_{\min}$, $\rho(s) = \rho_-(s)$, and when s is close to $s_{\max}$, $\rho(s) = \rho_+(s)$. To calculate $\rho(s)$, we calculate $\rho_-(s)$ from $s_{\min}$ to the s_2 such that $\rho_+(s_2) = \rho_{\text{par}}$, and we calculate $\rho_+(s)$ from the s_1 such that $\rho_+(s_1) = \rho_-(s_\infty)$ to $\tilde{s}_{\max}$. Then $\rho(s) = \rho_-(s)$ in $s_{\min} \leqslant s \leqslant s_1$, $\rho(s) = \min\{\rho_+(s), \rho_-(s)\}$ in $s_1 \leqslant s \leqslant s_2$, and $\rho(s) = \rho_+(s)$ in $s_2 \leqslant s \leqslant \tilde{s}_{\max}$.

A word of caution about figures 9 to 11 is required. We have proved that $\rho_+(s)$ and $\rho_-(s)$ are monotone decreasing, as shown in those figures. We will show in appendix C that they have vertical tangents when $\theta = 0$ and horizontal tangents when $\theta = \frac{1}{2}\pi$. We have a numerical example, not given here, which looks like figure 10*d*, so that figure shows a real possibility. We have no numerical examples like figure 11*b*, but suspect that such examples will be found.

We will not prove, and suspect it is not always true, that $\rho_+(s)^2$ and $\rho_-(s)^2$ are convex. The convexity of $\epsilon(s)^2$, the tradeoff curve for absolute error against spread, followed from (5.29). The analogous equation for relative errors, proved in appendix B, is

$$[\boldsymbol{q}.\boldsymbol{a}_\pm(s)]^2 \frac{\mathrm{d}[w_\pm \rho_\pm(s)^2]}{\mathrm{d}s} = -\cot\theta. \tag{7.33}$$

Because $\partial_\theta s_\pm(\theta) \geqslant 0$, we can deduce from (7.33) that convexity of $\rho_\pm(s)^2$ is equivalent to the inequality

$$\partial_\theta\{\tan\theta[\boldsymbol{q}.\boldsymbol{a}_\pm(\theta)]^2\} \geqslant 0. \tag{7.34}$$

We have not been able either to prove or to disprove (7.34).

Equation (7.33) gives us an interpretation of the parameter θ. Unlike (5.29), (7.33) contains $\boldsymbol{a}_{\pm}(s)$ explicitly, so the value of θ at a point on the curve $\rho_{\pm}(s)^2$ cannot be deduced from that curve alone. Nevertheless, (7.33) does resemble (5.29) rather closely, in that for any $\boldsymbol{a}$ in $\mathscr{R}^N$ we have $(\boldsymbol{q}.\boldsymbol{a})^2\rho(\boldsymbol{a})^2 = \epsilon(\boldsymbol{a})^2$.

(d) *The choice of* $w_{\pm}$

In calculating $\rho_+(s)$ and $\rho_-(s)$, the two branches of the tradeoff curve for relative error against spread, it will be convenient to choose the parameters w_+ and w_- to be different.

When $\boldsymbol{q}.\boldsymbol{a}_S > 0$, we choose w_+ so that there is a single number L_+ such that

$$L_+ = -\frac{\partial_\theta[\rho_+(0)^2]}{\rho^2_{\max}-\rho^2_{\min}} = \frac{\partial_\theta s(\tfrac{1}{2}\pi)}{\tilde{s}_{\max}-s_{\min}}, \tag{7.35}$$

and we choose w_- so that there is a single number L_- such that

$$L_- = \frac{\lim_{\theta\to 0}\theta^2\rho_-(\theta)^2}{\rho^2_{\max}-\rho^2_{\min}} = \frac{\lim_{\theta\to\frac{1}{2}\pi}(\tfrac{1}{2}\pi-\theta)^2 s(\theta)}{\tilde{s}_{\max}-s_{\min}}. \tag{7.36}$$

It is shown in appendix B that the result is

$$w_{\pm} = (N_{\pm}/M_{\pm})^{\frac{1}{2}}, \tag{7.37}$$

$$L_{\pm} = (M_{\pm}N_{\pm})^{\frac{1}{2}}, \tag{7.38}$$

where

$$M_+ = \frac{\boldsymbol{a}_S.(\boldsymbol{E}-\rho^2_{\max}\boldsymbol{q}\boldsymbol{q}).\boldsymbol{S}^{-1}.(\boldsymbol{E}-\rho^2_{\max}\boldsymbol{q}\boldsymbol{q}).\boldsymbol{a}_S}{(\rho^2_{\max}-\rho^2_{\min})(\boldsymbol{q}.\boldsymbol{a}_S)^2}, \tag{7.39}$$

$$N_+ = \frac{\boldsymbol{a}_{\mathscr{C}}.\boldsymbol{S}.\boldsymbol{E}^{-1}.\boldsymbol{S}.\boldsymbol{a}_{\mathscr{C}} + (\tilde{s}^2_{\max} - 2\tilde{s}_{\max}\boldsymbol{a}_{\mathscr{C}}.\boldsymbol{S}.\boldsymbol{a}_E)/\epsilon^2_{\min}}{\tilde{s}_{\max}-s_{\min}}, \tag{7.40}$$

$$M_- = \frac{\epsilon^4_{\min}}{(s_{\max}-s_{\min})[(\boldsymbol{a}_{\mathscr{C}}-\boldsymbol{a}_E).\boldsymbol{S}.(\boldsymbol{a}_{\mathscr{C}}-\boldsymbol{a}_E)]}, \tag{7.41}$$

and

$$N_- = \frac{(s_\infty - s_{\min})^2}{(\rho^2_{\max}-\rho^2_{\min})(\boldsymbol{q}.\boldsymbol{a}_S)^2(\boldsymbol{a}_\infty.\boldsymbol{E}.\boldsymbol{a}_\infty)}. \tag{7.42}$$

When $\boldsymbol{q}.\boldsymbol{a}_S < 0$, we choose w_+ so that

$$\partial_\theta s_+(\tfrac{1}{2}\pi) = \tilde{s}_{\max} - s_\infty \tag{7.43}$$

and w_- so that

$$-\partial_\theta[\rho_-(0)^2] = \rho^2_{\max}-\rho^2_{\text{par}}. \tag{7.44}$$

In appendix B it is shown that these choices imply

$$w_- = \frac{1}{2M_+}\left(\frac{\rho^2_{\max}-\rho^2_{\text{par}}}{\rho^2_{\max}-\rho^2_{\min}}\right) \tag{7.45}$$

and

$$w_+ = 2N_+\left(\frac{\tilde{s}_{\max}-s_{\min}}{\tilde{s}_{\max}-s_\infty}\right) \tag{7.46}$$

where M_+ is given by (7.39) and N_+ by (7.40).

D. NUMERICAL ILLUSTRATIONS

For purposes of illustration we now discuss two inverse problems in some detail: in the first problem the density and seismic velocities of a spherical Earth are assumed known, dissipation is assumed small, and the radial dependence of the dissipation function is sought from the observed damping rates of a finite number of normal modes of elastic–gravitational oscillation.

In this problem the relevant gross Earth functionals are linear. As a second, nonlinear, example we try to determine the radial dependence of the density in a spherical Earth whose seismic velocities are known; the gross Earth data are the Earth's total mass and moment and the squared circular frequencies of oscillation of a finite number of identified normal modes.

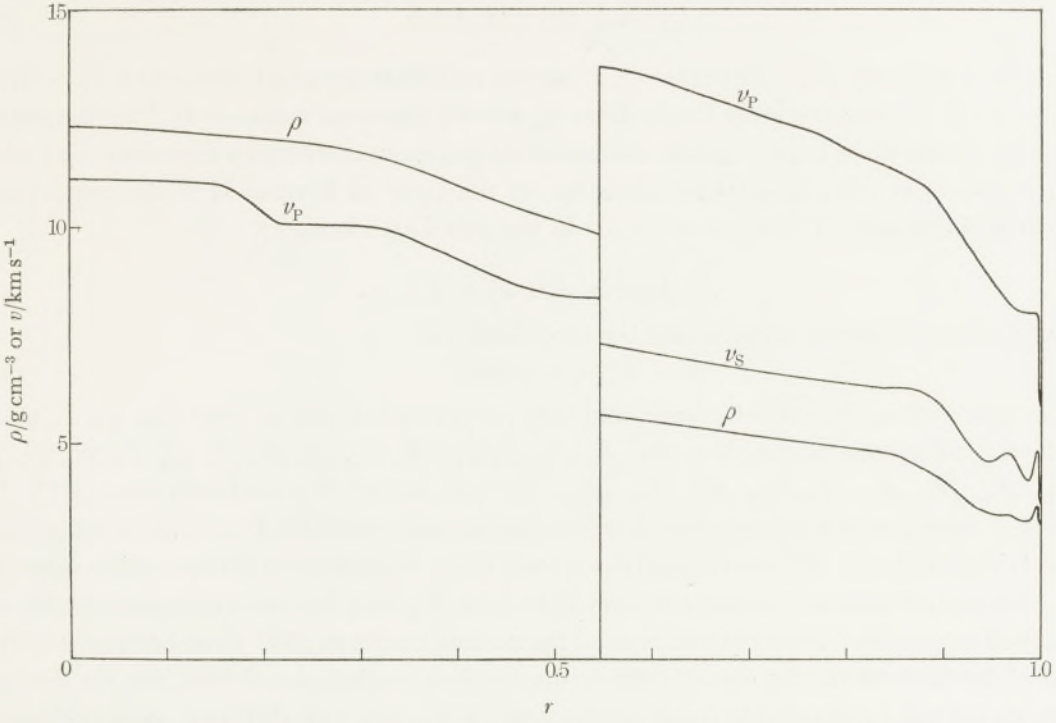

FIGURE 12. The density and seismic velocities used in all the calculations in § D. It is shown in § 9 that the small hump in density just below the crust is not resolvable with the data used here.

8. A LINEAR EXAMPLE: DISSIPATION

If the dissipation is small ($Q \gg 1$) we can calculate it by first-order perturbation theory. We assume in this section that $\rho(r)$, $v_P(r)$ and $v_S(r)$, the density and seismic velocities, are known functions of the radius r. Figure 12 shows the functions we have actually used. In the absence of dissipation we can calculate the displacement field for any normal mode of oscillation of the Earth. For the ith normal mode we denote by $S_i(r)\,\mathrm{d}r$ the maximum elastic shear energy stored in the spherical shell between r and $r+\mathrm{d}r$, and by E_i the total energy of oscillation in that normal mode. Both $S_i(r)$ and E_i are proportional to the square of the amplitude of the oscillation, and their ratio,

$$G_i(r) = S_i(r)/E_i, \tag{8.1}$$

is completely determined by $\rho(r)$, $v_P(r)$, $v_S(r)$ and i, the index identifying the oscillation. An explicit expression for $G_i(r)$ appears in Inverse II.

We denote by $2\pi/Q_i$ the fraction of the energy of oscillation of the ith normal mode which is dissipated as heat in one cycle of the oscillation. We denote by $2\pi/Q(r)$ the fraction of the maximum shear-energy density at radius r which is dissipated as heat in one cycle of oscillation. As the notation indicates, we assume that this fraction is independent of the amplitude of the oscillation and the period required to execute it, that the dissipation process is isotropic, and that pure compression is a non-dissipative process. If any of these assumptions fails, then the radial

dependence of dissipation requires more than one function of r for its description, and the appropriate Earth model is n-dimensional with $n > 1$. This situation is discussed in Inverse II but will not be pursued here. We neglect viscous dissipation in the core (Backus 1968).

According to first-order perturbation theory we have

$$Q_i^{-1} = \int_0^1 Q(r)^{-1} G_i(r)\, \mathrm{d}r, \tag{8.2}$$

where $G_i(r)$ is given by (8.1). Equation (8.2) has the standard form (2.1) if we set $m(r) = Q(r)^{-1}$ and $g_i(m) = Q_i^{-1}$. Then the gross Earth data, γ_i, are the observed values of Q_i^{-1} for the normal modes. The space $\mathfrak{M}$ of Earth models consists of all piecewise continuous functions $m(r)$ which vanish in the fluid core, and 'reasonableness' in the sense of Inverse II is the requirement $m(r) \geqslant 0$ in the mantle. The value of $Q(r)^{-1}$ in the real Earth is $m_{\mathrm{E}}(r)$.

(*a*) *Artificial data, absolute errors*

As a first illustration we suppose that the true Earth has

$$m_{\mathrm{E}}(r) = 0.004r \tag{8.3}$$

in the mantle and $m_{\mathrm{E}}(r) = 0$ in the core, and that we have succeeded in observing $Q_1, \ldots, Q_N$ for the following 24 normal modes ($N = 24$): ${}_0S_0$, ${}_1S_0$, ${}_2S_0$, ${}_3S_0$, ${}_1S_1$, ${}_2S_1$, ${}_0S_2$, ${}_2S_2$, ${}_1S_3$, ${}_0S_4$, ${}_1S_4$, ${}_2S_4$, ${}_4S_4$, ${}_0S_7$, ${}_1S_8$, ${}_0S_{25}$, ${}_0S_{49}$, ${}_0S_{73}$, ${}_0S_{97}$, ${}_0T_7$, ${}_0T_{14}$, ${}_0T_{27}$, ${}_0T_{53}$, ${}_0T_{105}$. We call this set of gross Earth data ${}_n\mathscr{G}_l^{ST}$. The labelling of normal modes by type, radial order n and angular order l is described, for example, in Inverse I. The modes in ${}_n\mathscr{G}_l^{ST}$ are chosen to represent Q_i for surface waves with periods longer than about 100 s and, in addition, to include some data from the deep mantle which conceivably will be observed eventually. The frequencies of all the normal modes in ${}_n\mathscr{G}_l^{ST}$ have been observed in records of the Chilean earthquake of 1960 or the Alaskan earthquake of 1964, but for the most of them Q_i has not been reliably measured because of the spin and ellipticity gaps (Gilbert & Backus 1965; Dahlen 1969). The problems raised by these gaps now seem to be under control (Dahlen 1969), but this fact has not yet been exploited.

We assume that all the $\gamma_i = Q_i^{-1}$ in ${}_n\mathscr{G}_l^{ST}$ have been measured with errors (standard deviations) of 5 % and that these errors are uncorrelated. This last assumption is based on the fact that the different γ_i are measured from a single time series by band-passing different parts of its frequency spectrum. Because of our assumptions, the 24×24 matrix E_{ij} defined by (3.4) is diagonal:

$$E_{ij} = 2.5 \times 10^{-3} (\gamma_i)^2 \delta_{ij}, \tag{8.4}$$

where no sum on i is intended.

We proceed as in § B. At a given r_0 in the mantle we calculate the tradeoff curve for absolute error against spread, in order to see what combinations of accuracy and resolution the data make available to us in our attempt to calculate local averages of $m(r)$ near r_0. Figure 13 shows these trade-off curves for five different values of r_0, namely 0.55, 0.65, 0.75, 0.85 and 0.95. It is noteworthy how rapidly the error ϵ drops when the spread s is increased only slightly above $s_{\min}$ on these curves. A very slight loss of resolution in the local averages permits a very large increase in their accuracy. This feature of the curves is somewhat obscured by our plotting $\lg \epsilon$ against $\lg s$.* The graphs of ϵ against s look almost like a right angle, and do not show the details of the 'corner'.

Figure 14 is a contour map of the width $w(A)$ of the optimal averaging kernel $A(r_0, r)$ as a function of r_0 and the absolute error ϵ committed in calculating the local average $\langle m_{\mathrm{E}}, A \rangle$ from the given inaccurate data. The oblique nearly straight line in figure 14 is $\lg \{m_{\mathrm{E}}(r_0)\}$ calculated from (8.3). The region above this oblique line is uninteresting; there the error ϵ is larger than the

* $\lg \equiv \log_{10}$.

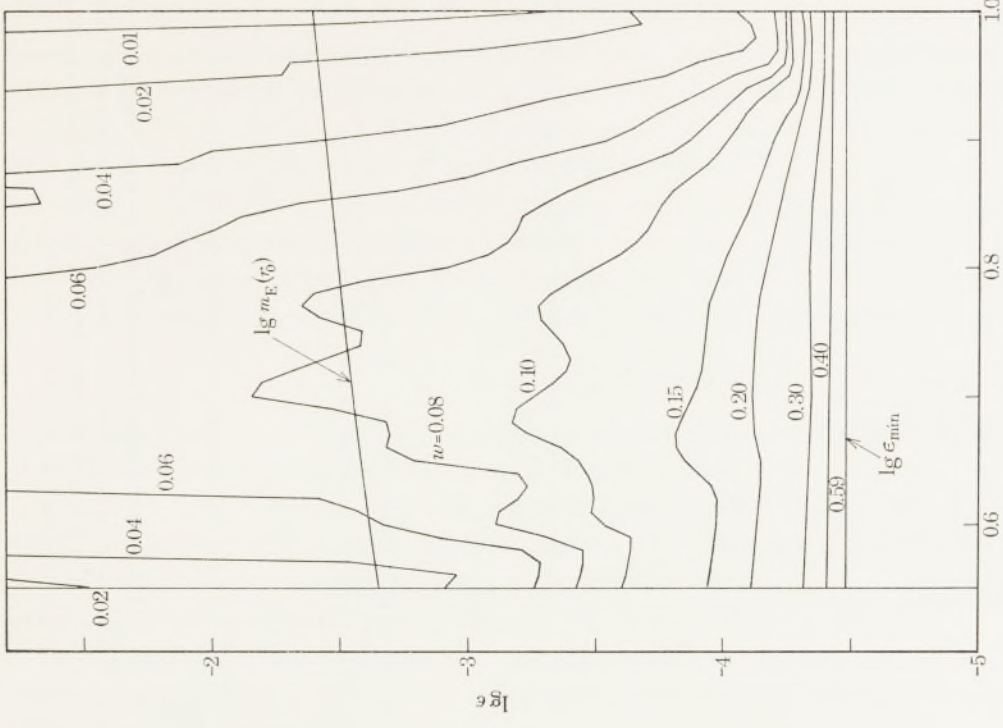

FIGURE 13. Tradeoff curves of absolute error ϵ against spread s at $r_0 = 0.55, 0.65, 0.75, 0.85$ and 0.95. The gross Earth data are ${}_n\mathscr{G}_l^{\mathrm{ST}}$, the damping rates of 24 normal modes calculated from the artificial assumption that $m_{\mathrm{E}}(r)$ is given by equation (8.3). The function to be determined from the data is $m(r) = Q(r)^{-1}$, the dissipation function.

FIGURE 14. A contour map of the width w of the optimal averaging kernel $A(r_0, r)$ as a function of r_0 and the absolute error ϵ in $\langle m_{\mathrm{E}}, A\rangle$ or $\boldsymbol{q}.\boldsymbol{a}$. The oblique nearly straight line is $\lg\{m_{\mathrm{E}}(r_0)\}$ with $m_{\mathrm{E}}(r_0)$ given by equation (8.3). The gross Earth data and the function to be determined are as in figure 13.

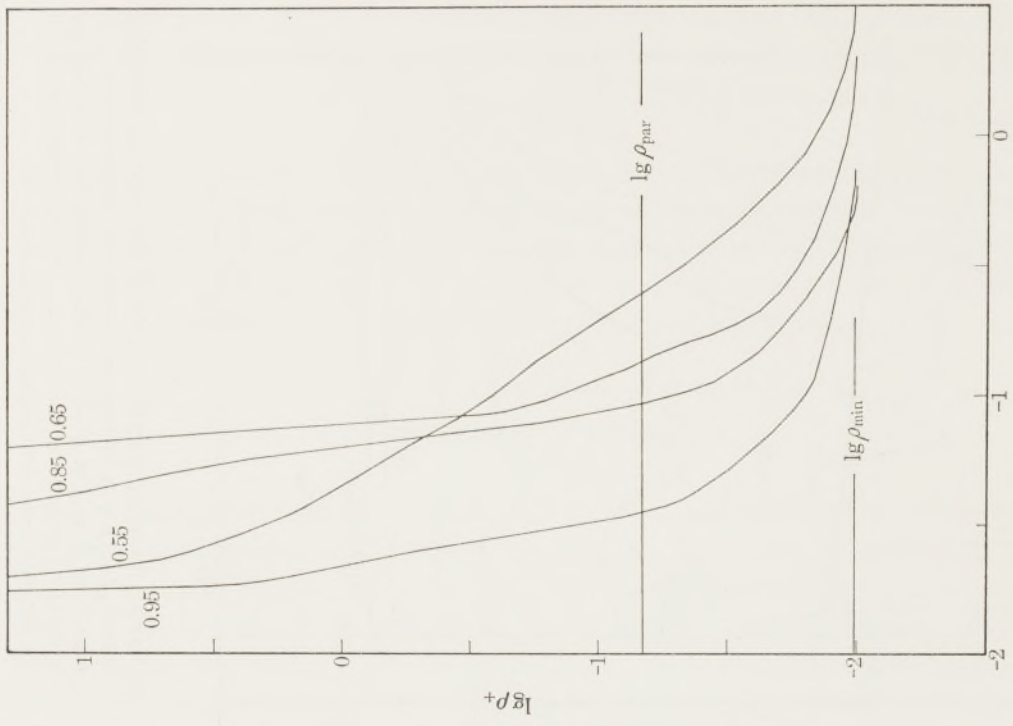

FIGURE 16. Tradeoff curves of relative error ρ against spread s at $r_0 = 0.55$, 0.65, 0.85 and 0.95. The gross Earth data and the function to be determined are as in figure 13; with these data $\rho_-(s) > \rho_+(s)$ so ρ is always ρ_+.

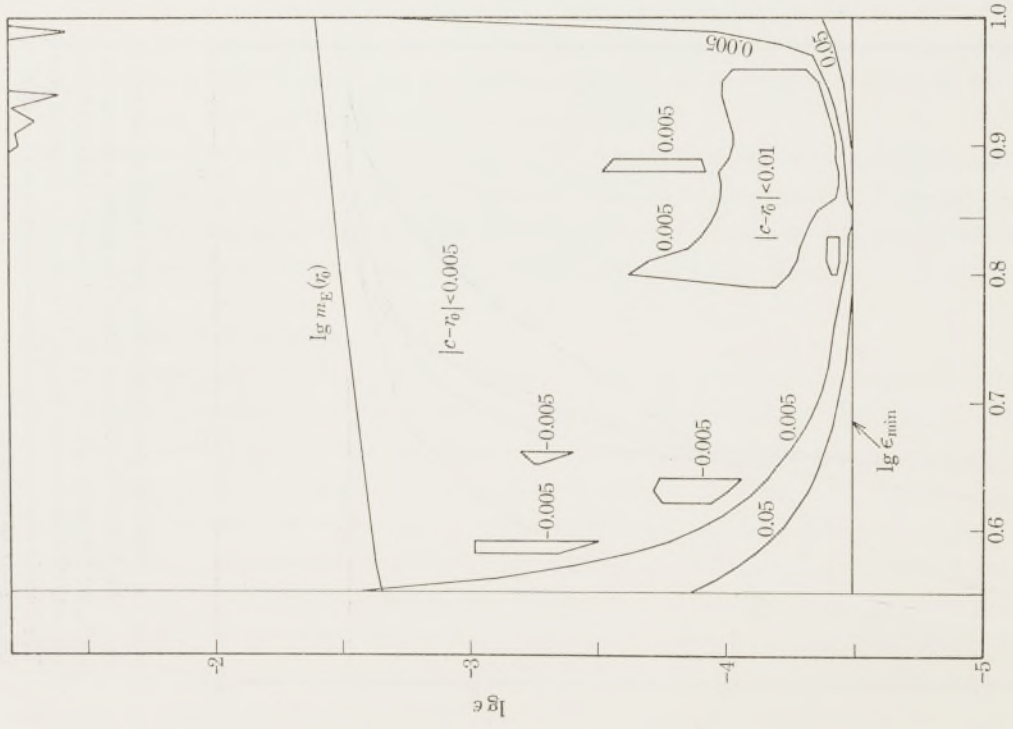

FIGURE 15. A contour map of $c(A) - r_0$ as a function of r_0 and the absolute error ϵ in $\langle m_E, A\rangle$. Here $c(A)$ is the centre of $A(r_0, r)$, the optimal averaging kernel for ϵ at r_0. The gross Earth data and the function to be determined are as in figure 13. This map is the companion to figure 14.

quantity being calculated. It is clear from figure 14 that if errors in our estimate of the local average $\langle m_E, A\rangle$ are acceptable when no larger than 10 % of m_E, then the width w need never be greater than about 0.11 (700 km) and if $0.9 < r < 1$ (the upper 600 km of the mantle) the width can be between 0.06 and 0.01 (between 360 and 60 km). Thus figure 14 makes clear at a glance what resolutions are available at various radii r_0 and how these resolutions depend on the absolute error ϵ which we will tolerate in our estimate of the local average of m_E at r_0.

Figure 15 is a contour map of $c(A) - r_0$ as a function of r_0 and ϵ. Here as in figure 14, for any ϵ and r_0 the function $A(r_0, r)$ is that linear combination of $G_1(r), \ldots, G_N(r)$ which minimizes the spread $s(r_0, A)$ subject to the constraint that the variance of the error in $\langle m_E, A\rangle$ shall be no greater than ϵ^2. It is clear from figure 15 that the centre $c(A(r_0, r))$ is quite close to r_0 except in the far lower left corner of the figure. There the resolution is so poor (see figure 14) that it is of no importance that the optimal averaging kernels are not centred on their r_0.

(*b*) *Artificial data, relative errors*

Here we repeat the calculations of § 8*a* verbatim except that we use the relative error ρ rather than the absolute error ϵ, so that § C rather than § B is illustrated.

Figure 16 gives the tradeoff curves between spread s and relative error ρ at four different radii r_0, namely 0.55, 0.65, 0.85 and 0.95. Again a very small sacrifice of resolution provides an enormous improvement in the accuracy of the local averages $\langle m_E, A\rangle$. In figure 16 the error ρ is always ρ_+ because $\rho_-(s) > \rho_+(s)$ for all s if the data are ${}_n\mathscr{G}_l^{ST}$.

In figure 17 we give the optimal averaging kernel at $r_0 = 0.90$ for eight different values of the relative error ρ (in this case always ρ_+) starting with $\rho_{\max}$ and ending with $\rho_{\min}$. This figure illustrates again how small a change is required in $A(r_0, r)$ to decrease ρ from a ridiculous to a usable value when s is near $s_{\min}$, and how little we can decrease ρ by relaxing our demands on resolution when s is near $\tilde{s}_{\max}$.

Figure 18 is a contour map of the width w of the optimal averaging kernel $A(r_0, r)$ as a function of r_0 and the relative error ρ. It leads to essentially the same conclusions as figure 14. In the present instance little is gained by using relative rather than absolute errors except that of course the contour map for relative errors is slightly easier to interpret than the map for absolute errors.

Figure 19, a contour map of $c(A) - r_0$ as a function of r_0 and the relative error ρ, is almost indistinguishable from the corresponding map (figure 15), for absolute errors, even though the $A(r_0, r)$ in figure 19 is chosen to minimize $s(r_0, A)$ subject to a constraining bound on the relative rather than the absolute error in $\langle m_E, A\rangle$.

(*c*) *Real data, relative errors*

Now we repeat the calculations of § 8*b* with a different set of gross Earth data, only ten in number. These values of $Q_i^{-1} = \gamma_i$ for ten normal modes have actually been measured and reported in the literature. The data, γ_i, and the estimated standard deviations of their errors, $[\overline{(\delta\gamma_i)^2}]^{\frac{1}{2}}$, are given in table 1. The values for $i = 1$ and 2 are from Knopoff (1965), while γ_3 through γ_{10} are from Ben-Menahem (1965) and the errors are from Toksöz & A. Ben-Menahem (personal communication). Other workers have used more extensive tables of Q_i^{-1}, but without recognizing the very serious systematic errors produced by the spin and ellipticity gaps (Gilbert & Backus 1965; Dahlen 1969). Again we assume that the error variance matrix E_{ij} defined by (3.4) is diagonal; the ith diagonal entry, $\overline{(\delta\gamma_i)^2}$, is obtained from table 1. The set of 10 gross Earth data in table 1 we denote by ${}_0\mathscr{G}_l^{ST}$.

Figure 20 gives the tradeoff curve between spread s and relative error ρ at $r_0 = 0.83$. Evidently $\rho(s) = \rho_-(s)$ when s is to the left of the crossing point of the curves for $\rho_+(s)$ and $\rho_-(s)$, while $\rho(s) = \rho_+(s)$ to the right of that point. The curve $\rho(s)$ is of interest only inside the rectangle bounded by the dashed lines. Outside that rectangle either the spread or the relative error is greater than unity. At $r_0 = 0.83$ the resolution is poor; if we want an error less than 10 % in our estimate of a local average of $Q(r)^{-1}$, the spread of that average from r_0 must be at least 0.3, or 2000 km.

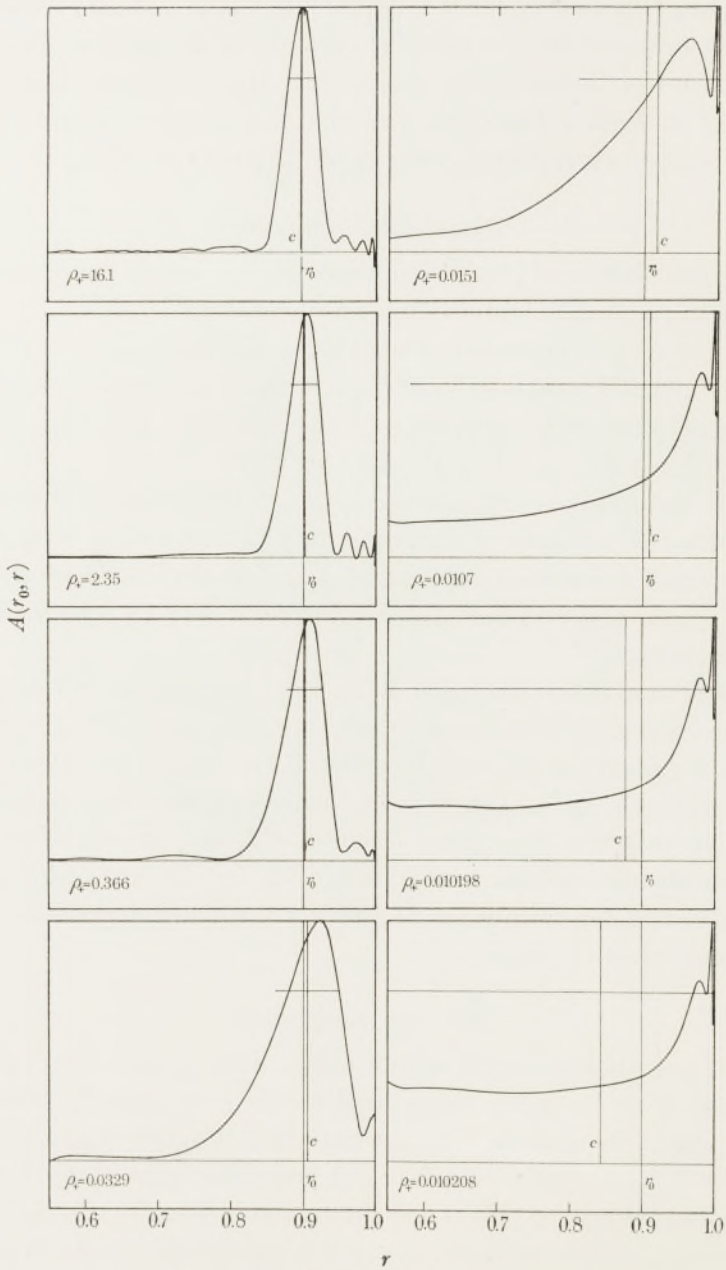

FIGURE 17. The optimal averaging kernel $A(0.90, r)$ for various values of relative error ρ (always ρ_+) when the gross Earth data and the function to be determined are as in figure 13. The vertical line extending above and below the r_0 axis is at $r_0 = 0.90$, and the vertical line lying wholly above the axis is at $r_0 = c(A)$, the centre of A. The horizontal line gives the width of A except for the two widest kernels, whose widths are greater than 0.45 and will not fit in the mantle.

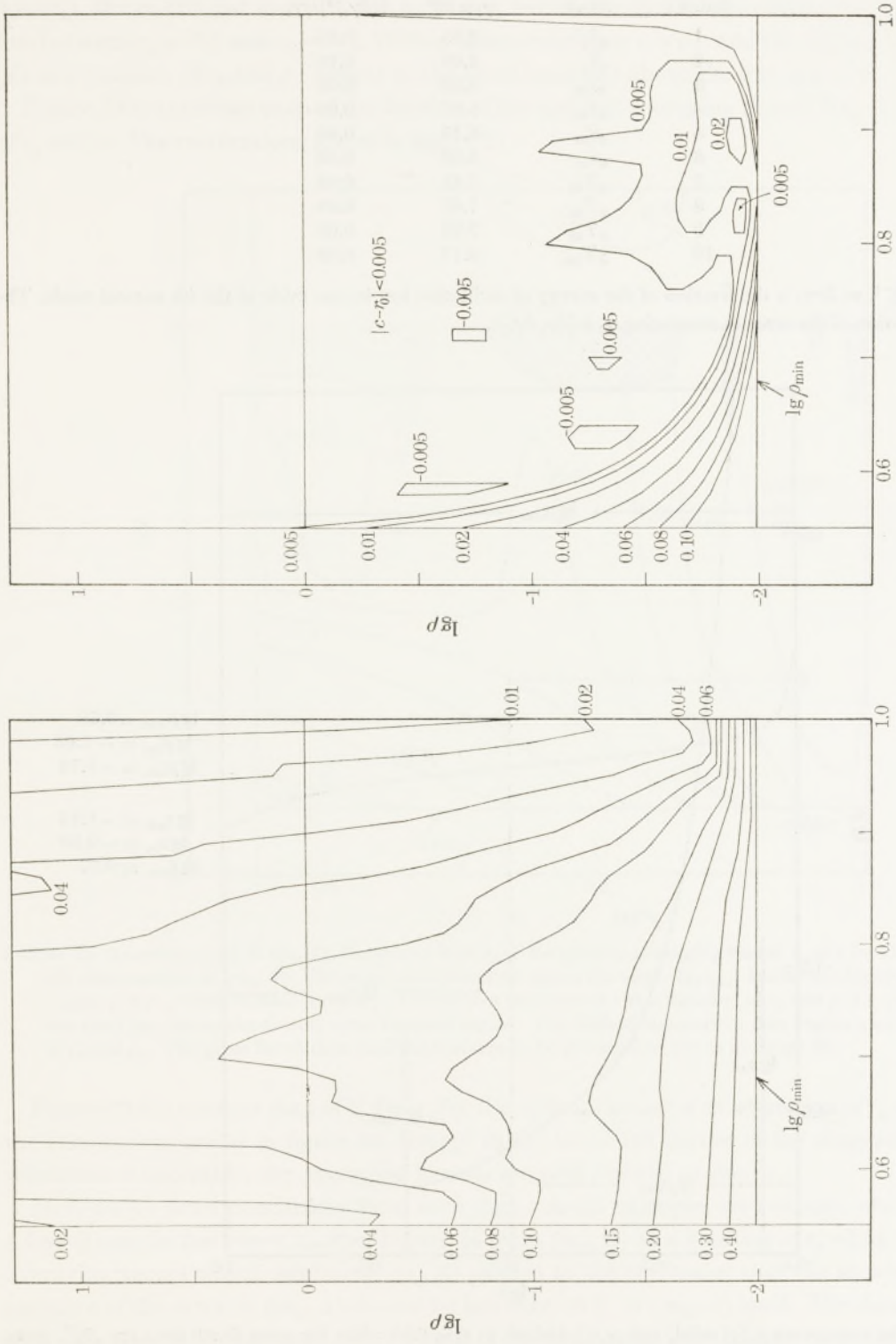

FIGURE 18. A contour map of the width w of the optimal averaging kernel $A(r_0, r)$ as a function of r_0 and the relative error ρ in $\langle m_E, A \rangle$. The gross Earth data and the function to be determined are as in figure 13.

FIGURE 19. A contour map of $c(A) - r_0$ as a function of r_0 and the relative error ρ. The gross Earth data and the function to be determined are as in figure 13. This map is a companion to figure 18.

TABLE 1

mode index i	mode	$\gamma_i \times 10^3$	$[(\overline{\delta\gamma_i})^2]^{\frac{1}{2}}/\gamma_i$
1	${}_0S_2$	2.85	0.15
2	${}_0S_3$	2.63	0.15
3	${}_0S_{22}$	3.89	0.05
4	${}_0S_{42}$	5.67	0.05
5	${}_0S_{60}$	6.10	0.05
6	${}_0S_{97}$	6.90	0.05
7	${}_0T_{23}$	7.41	0.05
8	${}_0T_{39}$	7.63	0.05
9	${}_0T_{64}$	7.87	0.05
10	${}_0T_{105}$	9.17	0.05

Here γ_i is Q_i^{-1}, so $2\pi\gamma_i$ is the fraction of the energy of oscillation lost in one cycle of the ith normal mode. The standard deviation of the error in measuring γ_i is $[\overline{\delta\gamma_i\,\delta\gamma_i}]^{\frac{1}{2}}$.

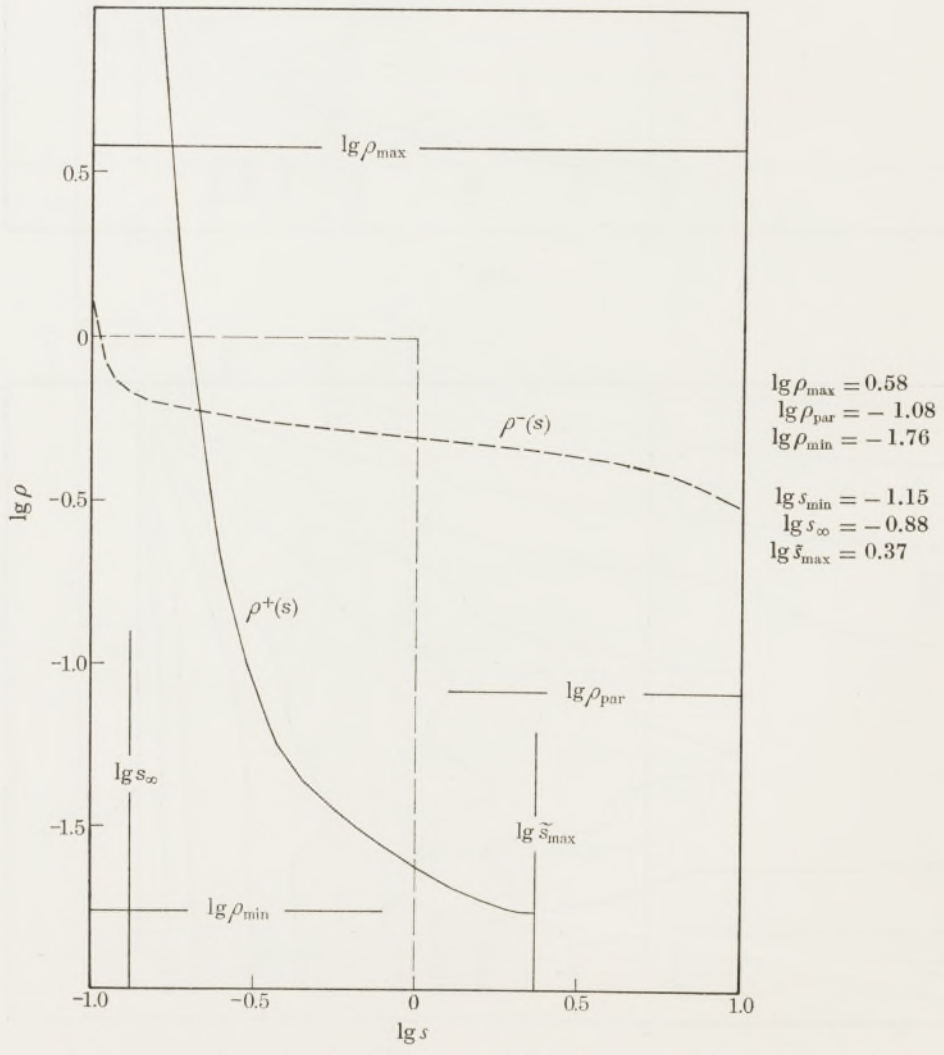

FIGURE 20. The two curves $\rho_+(s)$ solid, and $\rho_-(s)$ dashed, at $r_0 = 0.83$ when the gross Earth data are ${}_0\mathscr{G}_i^{sT}$, given in table 1, and the function to be determined is $m(r) = Q(r)^{-1}$, the dissipation function. The tradeoff curve between relative error and spread is $\rho(s) = \min\{\rho_+(s), \rho_-(s)\}$.

Figure 21 gives a contour map of the spread $s(r_0, A)$ of the optimal averaging kernel $A(r_0, r)$ as a function of r_0 and the relative error ρ. The bottom boundary of the cross-hatched region is $\rho_{\max}(r_0)$. Below the line $\lg \rho = 0$, ρ is ρ_+ except within the small semi-ellipse pendant from that line between $r_0 = 0.7$ and $r_0 = 0.9$. Within that semi-ellipse ρ is ρ_-, and the appropriate contours of s as a function of r_0 and ρ_- appear in the small inset just above the line $\lg \rho = 0$.

Figure 22 is a contour map of the width w of the optimal averaging kernel $A(r_0, r)$ as a function of r_0 and ρ. The conventions are as in figure 21.

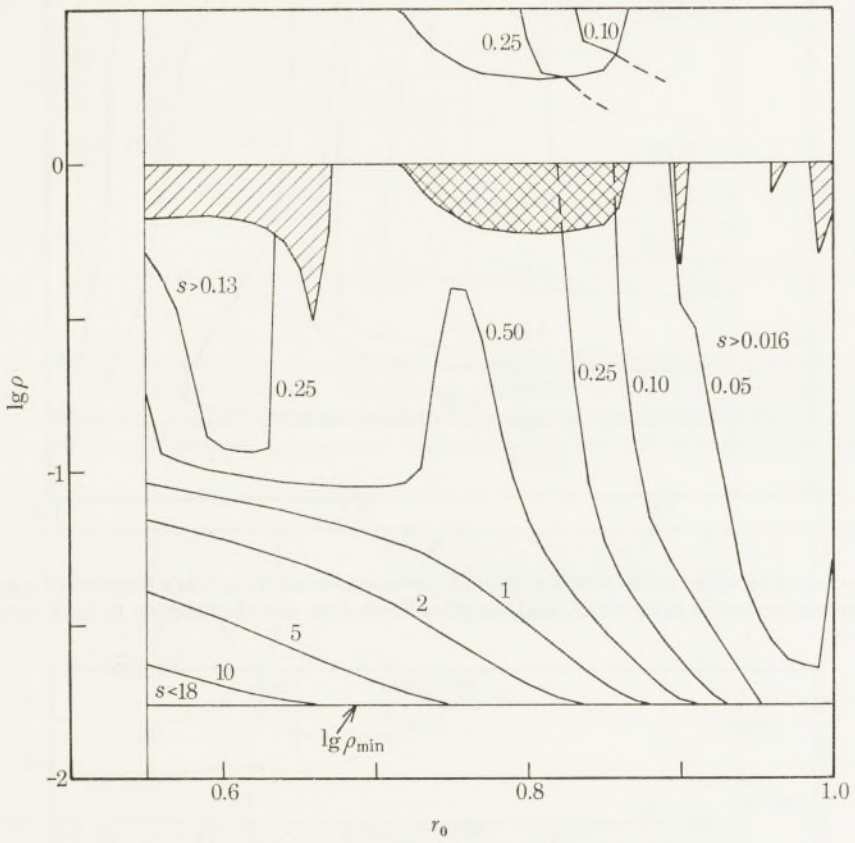

FIGURE 21. A contour map of $s(r_0, A)$, the spread from r_0 of the optimal averaging kernel A, as a function of r_0 and ρ, the relative error in $\langle m_E, A \rangle$. The singly cross-hatched region lies above $\lg \rho_{\max}$. Inside the doubly cross-hatched region ρ is ρ_-, while outside it ρ is ρ_+. The correct contours of s as a function of r_0 and ρ (i.e. ρ_-) are given in the inset just above the doubly cross-hatched region. The dashed contours in that region give s as a function of r_0 and ρ_+. The gross Earth data and the function to be determined are as in figure 20.

Figure 23 is a contour map of $c(A) - r_0$ for the optimal kernel A as a function of r_0 and ρ. Again the conventions are as in figure 21. Except in the lower left corner of the diagram, where the resolution is execrable, the averaging kernels are well centred at their r_0.

How do we draw conclusions from such data? As an example, we consider whether there is a low-Q zone in the upper mantle. Figure 24 gives $\langle m_E, A \rangle$ as a function of r_0 when $A(r_0, r)$ is the averaging kernel which minimizes $s(r_0, A)$ subject to the constraint that the square root of the variance of the error in $\langle m_E, A \rangle$ should be less than 10 % of $\langle m_E, A \rangle$ itself. The dots in figure 24 give $\langle m_E, A(r_0, r) \rangle$ at various r_0. The length of the horizontal line through each dot is the width of the corresponding optimal averaging kernel, while the length of the vertical line is the error (standard deviation) in $\langle m_E, A \rangle$ at r_0. The dots show a hump in $\langle m_E, A \rangle$ with a maximum near

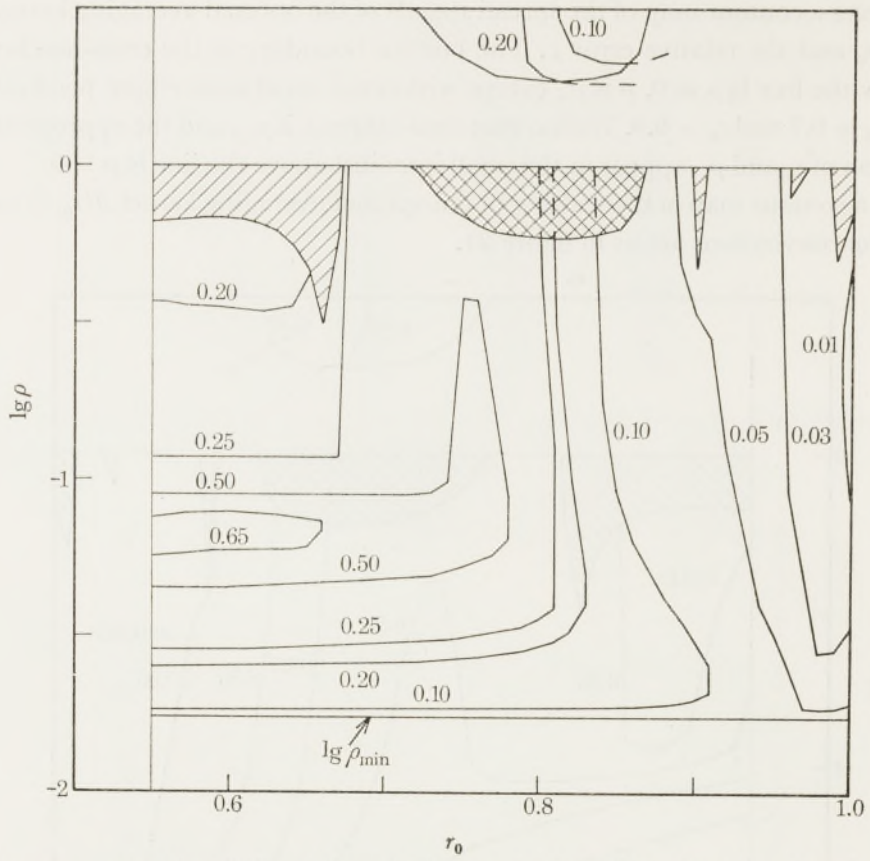

FIGURE 22. A contour map of the width w of the optimal averaging kernel $A(r_0, r)$ as a function of r_0 and the relative error ρ. Conventions are as in figure 21, and the gross Earth data and the function to be determined are as in figure 20.

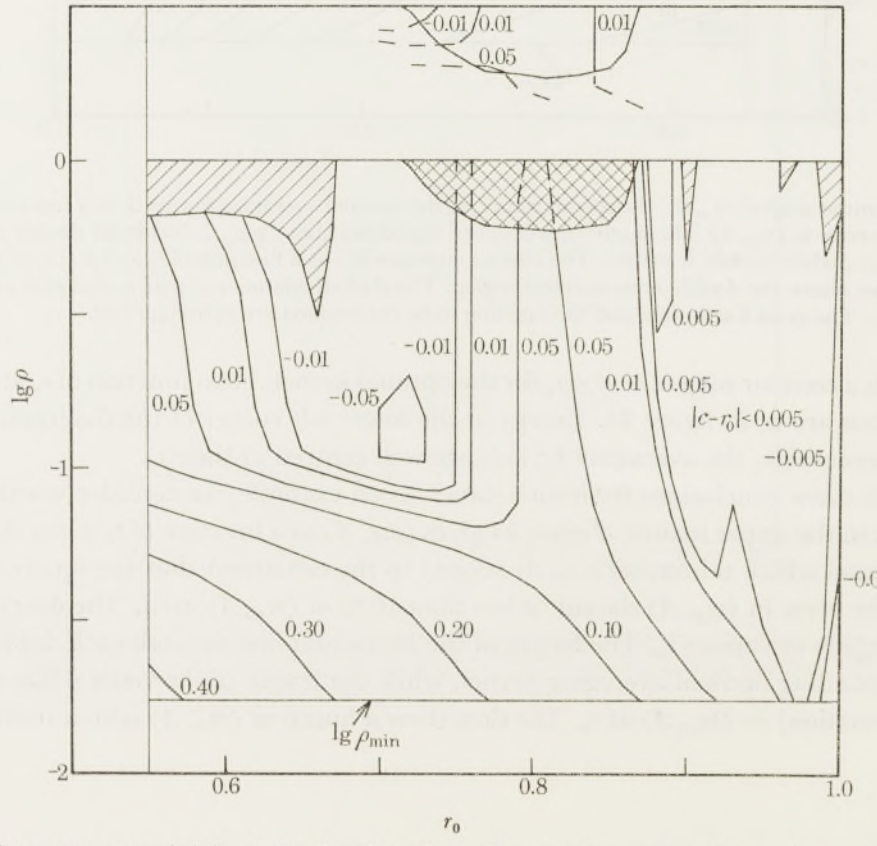

FIGURE 23. A contour map of $c(A) - r_0$ as a function of r_0 and the relative error ρ. Conventions are as in figure 21, and the gross Earth data and the function to be determined are as in figure 20. This map is the companion to figure 22.

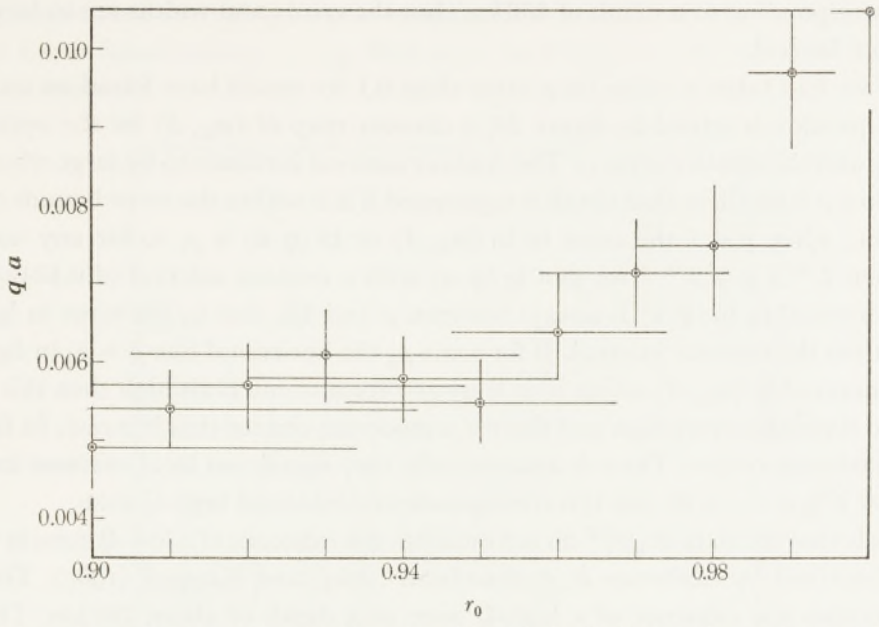

FIGURE 24. Values of $\langle m_E, A\rangle$ as a function of r when $A(r_0, r)$ is optimized subject to the requirement that the relative error in $\langle m_E, A\rangle$ be no greater than 0.1. The dots are values of $\langle m_E, A\rangle$; the vertical line through each dot represents a 10 % error in $\langle m_E, A\rangle$; the horizontal line through each dot gives the width of the corresponding A. Gross Earth data are ${}_0\mathcal{G}_l^{ST}$ in table 1, as in figure 20, and $\langle m_E, A\rangle$ is a local average of the dissipation function $Q(r)^{-1}$.

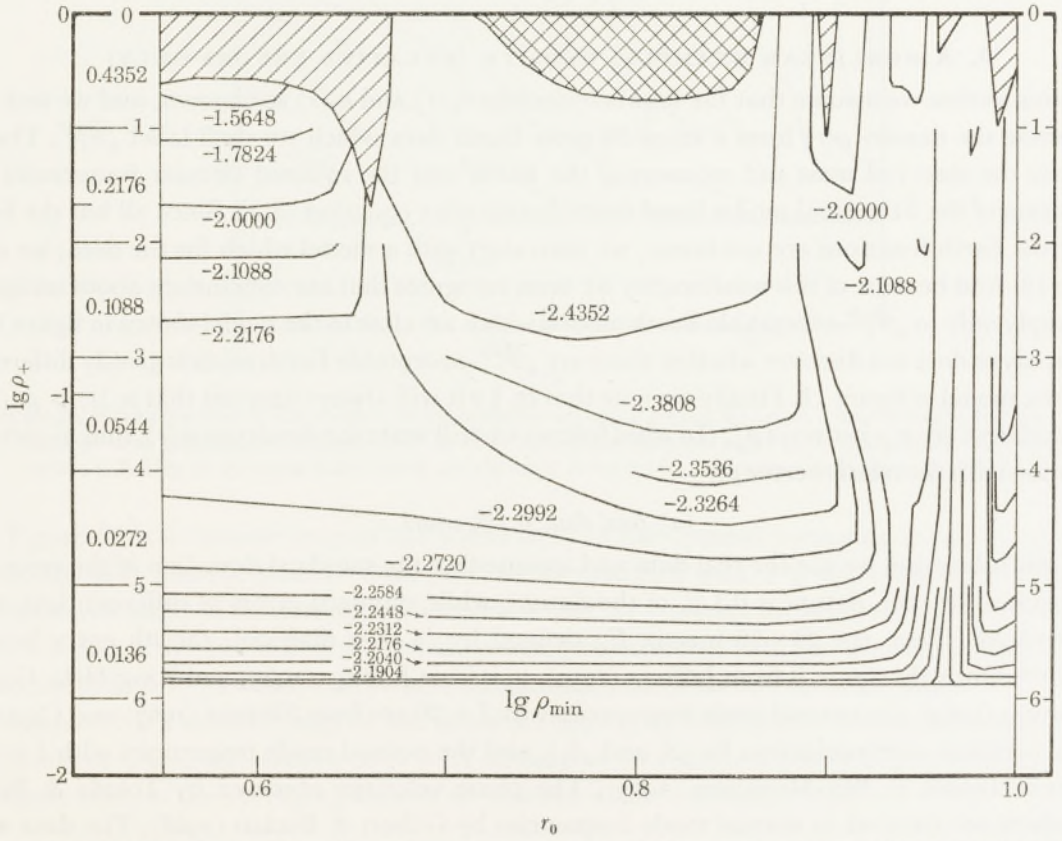

FIGURE 25. A contour map of $\langle m_E, A\rangle$ as a function of r_0 and the relative error ρ. Conventions are as in figure 21, except that the contour interval varies with ρ; when $2^{-k} \geqslant \rho \geqslant 2^{-k-1}$ for any non-negative integer k the contour interval of $\lg \langle m_E, A\rangle$ is $0.4343/2^k$. Thus the contour interval is always between ρ and 2ρ where ρ is the error in the quantity being contoured. Gross Earth data and m_E are as in figure 24 or figure 20.

$r_0 = 0.93$ (corresponding to a depth of 450 km) but the errors and widths are so large that this hump may not be real.

Perhaps if we had taken a value for ρ other than 0.1 we would have found an unambiguous hump. This question is settled by figure 25, a contour map of $\langle m_E, A \rangle$ for the optimal A as a function of r_0 and the relative error ρ. The contour interval is chosen to be large when ρ is large and small when ρ is small, so that detail is suppressed if it is within the error bounds on $\langle m_E, A \rangle$. To be specific, when $\rho \ll 1$ the error in $\ln \langle m_E, A \rangle$ or $\ln (\boldsymbol{q}.\boldsymbol{a})$ is ρ, so for any non-negative integer k when $2^{-k} \geqslant \rho \geqslant 2^{-k-1}$ we plot $\lg (\boldsymbol{q}.\boldsymbol{a})$ with a contour interval of $0.4343/2^{k+1}$. Thus the contour interval in $\ln (\boldsymbol{q}.\boldsymbol{a})$ is always between ρ and 2ρ; that is, the error in $\lg \langle m_E, A \rangle$ is never more than the contour interval. If for some ρ_0 the horizontal line $\rho = \rho_0$ in figure 25 has a local maximum of $\lg \langle m_E, A \rangle$ which is at least two contour intervals high then this maximum is at least two standard errors high and there is a moderate chance that it is real. In figure 25 no such local maximum occurs. There is a statistically very significant local *minimum* in $\lg \langle m_E, A \rangle$ near $r_0 = 0.97$ if $\lg \rho < -1.25$, but this corresponds to a localized high-Q zone.

We conclude that the data in ${}_0\mathscr{G}_l^{ST}$ do not establish the existence of a low-Q zone in the mantle of the sort described by Anderson & Archambeau (1964) and Knopoff (1965). The data do, however, establish the existence of a high-Q zone at a depth of about 200 km. This positive conclusion, of course, presupposes that we are looking at the right space of Earth models and the right gross Earth functionals; i.e. that $Q(r)$ is frequency independent and that the density and seismic velocities which we use to compute the data kernels are correct. The conclusion also presupposes that our estimates of the errors in the data are not excessively optimistic.

9. A nonlinear example: density (relative errors only)

In this section we assume that the seismic velocities $v_P(r)$ and $v_S(r)$ are known, and we seek to determine the density $\rho(r)$ from a set of 26 gross Earth data which we shall label ${}_n\overline{\mathscr{G}}_l^{ST}$. These data are the observed mass and moment of the Earth and the squared circular frequencies of oscillation of the 24 normal modes listed immediately after equation (8.3). Since all but the first two gross Earth functions are nonlinear, we must start with a model which fits the data; we use figure 12. And because of this nonlinearity we must recognize that our conclusions about uniqueness apply only to ${}_n\overline{\mathscr{G}}_l^{ST}$-acceptable Earth models which are close to the model shown in figure 12. Our analysis does not discover whether there are ${}_n\overline{\mathscr{G}}_l^{ST}$-acceptable Earth models grossly different from the model of figure 12. Finally we note that in § 9 it will always turn out that $\rho_-(s) > \rho_+(s)$, so the relative error ρ is always ρ_+. (In what follows we will write the density as $m(r)$, thus avoiding confusion with the relative error ρ.)

(a) *Real data, 0.1% errors*

In this subsection we use the real data and assume that the standard deviation of the error of observation for each datum is 0.1 % of the datum, while standard errors of different data are uncorrelated. Then the 26×26 matrix E_{ij} defined by (3.4) is diagonal, the ith entry being $\gamma_i^2 \times 10^{-6}$. Our value for γ_1 is from Jeffreys (1959), γ_2 is from Jeffreys (1963) and King-Hele, Cook & Watson (1964), the normal mode frequencies with $l < 20$ are from Slichter (1967) and Caputo (1967, personal communication for ${}_1S_1$ and ${}_2S_1$), and the normal mode frequencies with $l \geqslant 20$ are from Toksöz & Ben-Menahem (1963). The phase velocities observed by Toksöz & Ben-Menahem are reduced to normal mode frequencies by Gilbert & Backus (1968). The data are listed in table 2.

In order to calculate averaging kernels we must have the Fréchet derivatives $G_1(r), \ldots, G_{26}(r)$ of the gross Earth functionals $g_1, \ldots, g_{26}$. Both $G_1(r)$ and $G_2(r)$ are trivial, while $G_3(r), \ldots, G_{26}(r)$ are given in Inverse I.

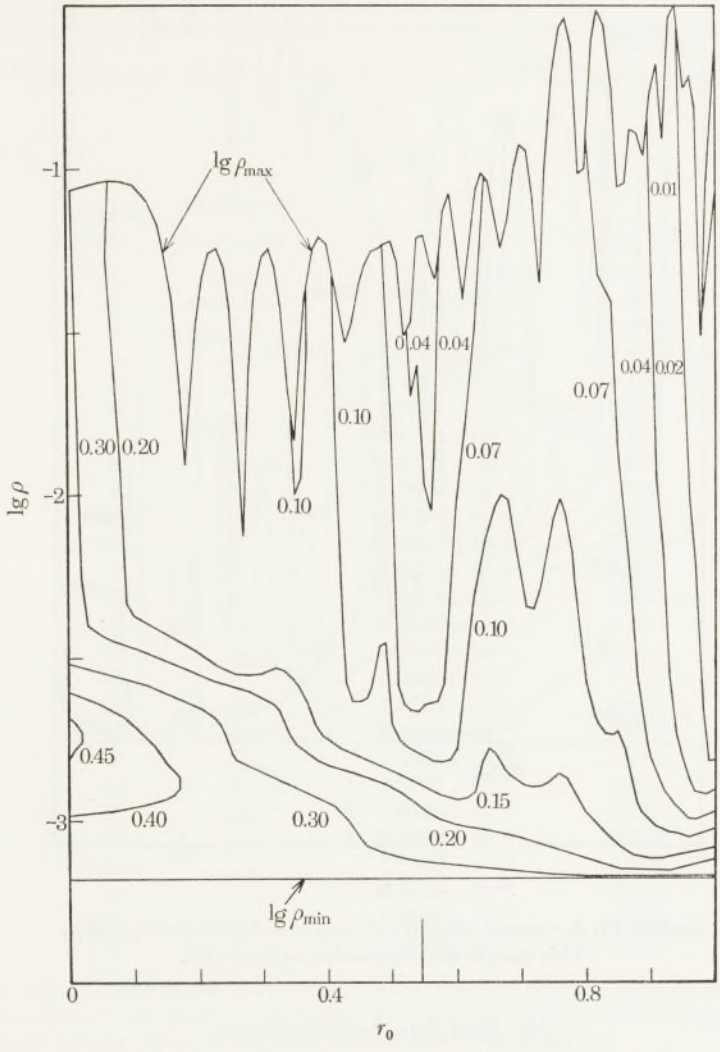

FIGURE 26. A contour map of the width w of the optimal averaging kernel $A(r_0, r)$ as a function of r_0 and the relative error ρ. The gross Earth data are ${}_n\bar{\mathscr{G}}_l^{ST}$ in table 2, and the function $m_E(r)$ to be determined is the density at radius r. Errors in the gross Earth data are assumed to be 0.1 %. Throughout the diagram ρ is ρ_+.

Figure 26 is a contour map of the width $w(A)$ of the optimal averaging kernel $A(r_0, r)$ as a function of r_0 and the relative error ρ. If we are satisfied with 1 % errors in our estimates of localized averages of the density, then it is clear that the resolving power is moderate. The widths are less than 0.07 (450 km) in the upper mantle, less than 0.10 (640 km) throughout the mantle, and less than 0.15 (950 km) everywhere except in the deepest 1000 km of the core.

Figure 27 is a contour map of the value of $c(A) - r_0$ for the optimal averaging kernel A as a function of r_0 and the relative error ρ. If we set $\rho = 10^{-2}$ then clearly $|c - r_0| \ll w$ except in the deepest core, so the averaging kernels are centred where we want them.

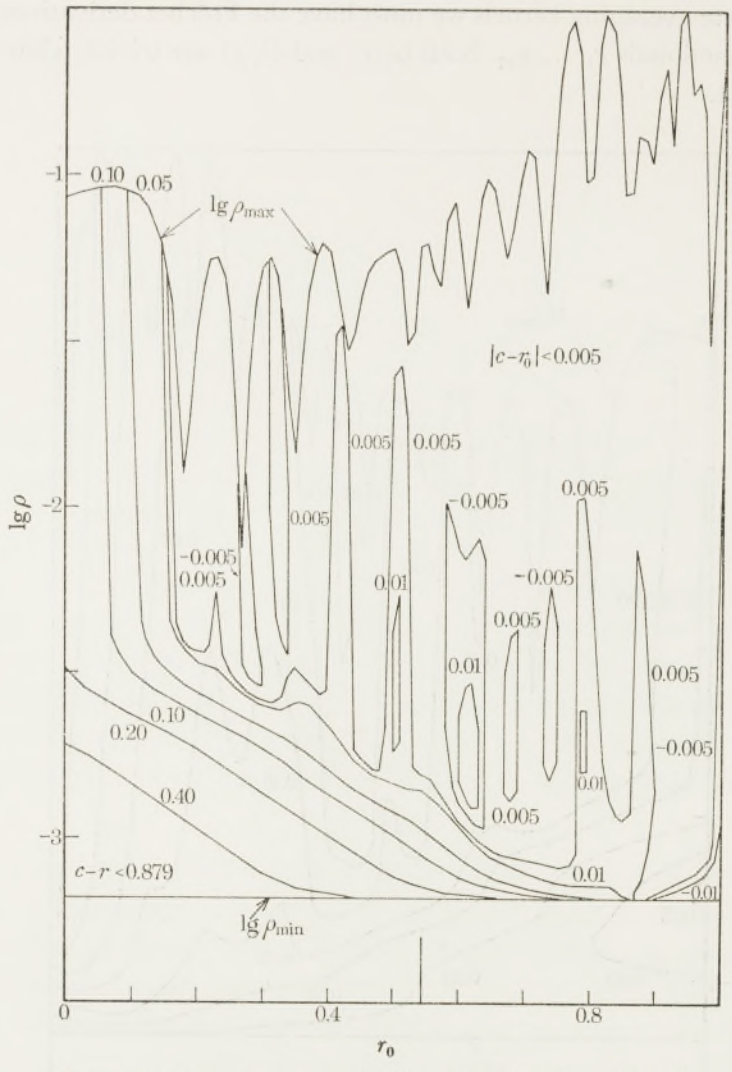

FIGURE 27. A contour map of $c(A) - r_0$ as a function of r_0 and ρ. This map is the companion to figure 26.

(b) Real data, published errors

In this subsection we repeat §9*a*, except that we use the published errors for the gross Earth data instead of assuming errors of 0.1 %. The data and errors are listed in table 2. Again we suppose that $\overline{\delta\gamma_i \delta\gamma_j} = 0$ if $i \neq j$.

Figure 28 is a contour map of the width $w(A)$ of the optimal averaging kernel $A(r_0, r)$ as a function of r_0 and the relative error ρ. For a given ρ, the values of w are larger than in figure 26 because most of the published errors are larger than 0.1 %.

Figure 29 gives $c(A) - r_0$ for the optimal averaging kernel A as a function of r_0 and ρ. As in figure 27, the averaging kernels are centred where we want them, except in the deepest core.

In Inverse II we gave estimates for the density just above and just below the core–mantle boundary, based on the hypothesis that the lower-mantle density and the upper-core density varied linearly with r over the interval where the optimal averaging kernel had appreciable weight. If the real Earth resembles figure 12, inspection of that figure shows that the procedure in Inverse II is applicable only when the averaging kernels have widths less than about 0.1.

From figure 28, if this is to hold in $0.45 < r_0 < 0.65$ then we must have $\lg\rho \geqslant -1.5$ or $\rho > 0.03$ in $0.55 < r_0 < 0.65$ and $\lg\rho \geqslant -2.0$ or $\rho > 0.01$ in $0.45 < r_0 < 0.55$. Therefore the error in our estimate of the density $m(r_c+)$ just above the core–mantle boundary is 3 % while the error in our estimate of the density $m(r_c-)$ just below that boundary is 1 %, when the uncertainties in the observations are taken into account. The result is

$$m(r_c+) = 5.72 \pm 0.17\,\mathrm{g\,cm^{-3}}, \quad m(r_c-) = 10.0 \pm 0.1\,\mathrm{g\,cm^{-3}}. \tag{9.1}$$

TABLE 2

datum index i	functional	γ_i	$[\overline{(\delta\gamma_i)^2}]^{\frac{1}{2}}/\gamma_i$
1	mass	$5.517\,\mathrm{g\,cm^{-3}}$	0.0006
2	moment	0.00389	0.0005
3	${}_0S_0$	2.6175×10^{-5}	0.0004
4	${}_1S_0$	1.0822×10^{-4}	0.0004
5	${}_2S_0$	2.4790×10^{-4}	0.004
6	${}_3S_0$	4.2994×10^{-4}	0.004
7	${}_1S_1$	6.4892×10^{-6}	0.008
8	${}_2S_1$	3.5202×10^{-5}	0.008
9	${}_0S_2$	3.7679×10^{-6}	0.004
10	${}_2S_2$	4.7294×10^{-5}	0.004
11	${}_1S_3$	3.4467×10^{-5}	0.012
12	${}_0S_4$	1.6471×10^{-5}	0.012
13	${}_1S_4$	5.4014×10^{-5}	0.012
14	${}_2S_4$	7.5094×10^{-5}	0.012
15	${}_4S_4$	2.2326×10^{-4}	0.016
16	${}_0S_7$	5.9659×10^{-5}	0.012
17	${}_1S_8$	1.2764×10^{-4}	0.012
18	${}_0S_{25}$	4.4310×10^{-4}	0.012
19	${}_0S_{49}$	1.1981×10^{-3}	0.012
20	${}_0S_{73}$	2.3731×10^{-3}	0.012
21	${}_0S_{97}$	3.9239×10^{-3}	0.012
22	${}_0T_7$	5.8334×10^{-5}	0.012
23	${}_0T_{14}$	1.7300×10^{-4}	0.012
24	${}_0T_{27}$	4.9320×10^{-4}	0.012
25	${}_0T_{53}$	1.6004×10^{-3}	0.012
26	${}_0T_{105}$	5.6920×10^{-3}	0.012

γ_1 is the observed mean density of the Earth and γ_2 is the observed dimensionless ratio of its moment of inertia to the product of its mass and the square of its radius. The data $\gamma_3, \ldots, \gamma_{26}$ are the squared circular frequencies of oscillation of the normal modes listed in the 'functional column', units are $\mathrm{s^{-2}}$.

In Inverse II the gross Earth data were assumed perfectly accurate and the much smaller quoted errors in the two densities referred only to the effect of deviations from linearity in the graph of the density in figure 12. Of course equations (9.1) still assume that the $v_P(r)$ and $v_S(r)$ in figure 12 exactly describe the real Earth, and they assume that the density of the real Earth is not grossly different from that of figure 12. To eliminate the former assumption requires a tripling of the number of gross Earth data we use, and requires a larger computer than has so far been available to us. To eliminate the latter assumption requires a theory for solving the nonlinear uniqueness problem in the large.

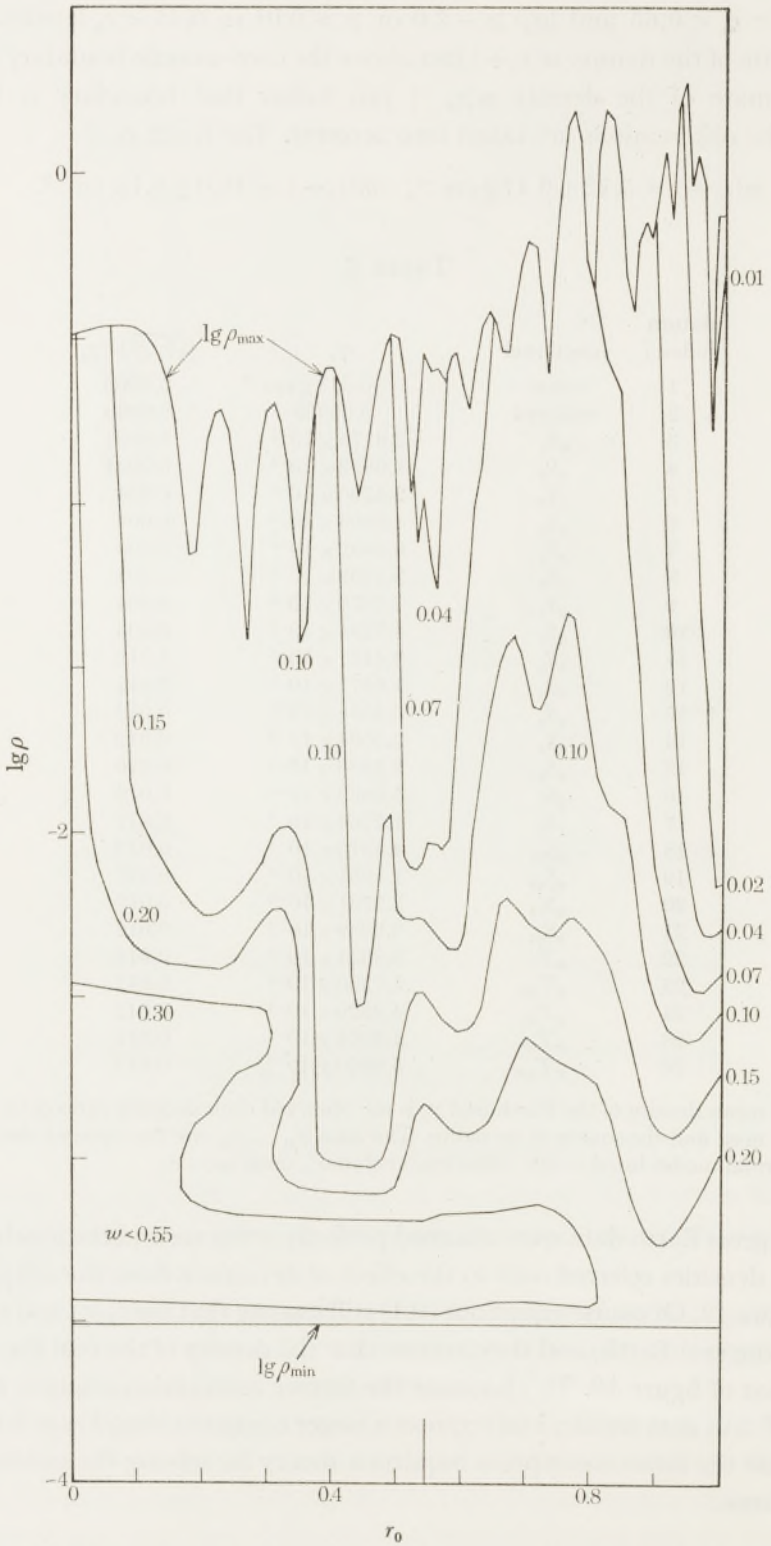

FIGURE 28. The same as figure 26 except that the errors in the gross Earth data are those reported in the literature and listed in table 2.

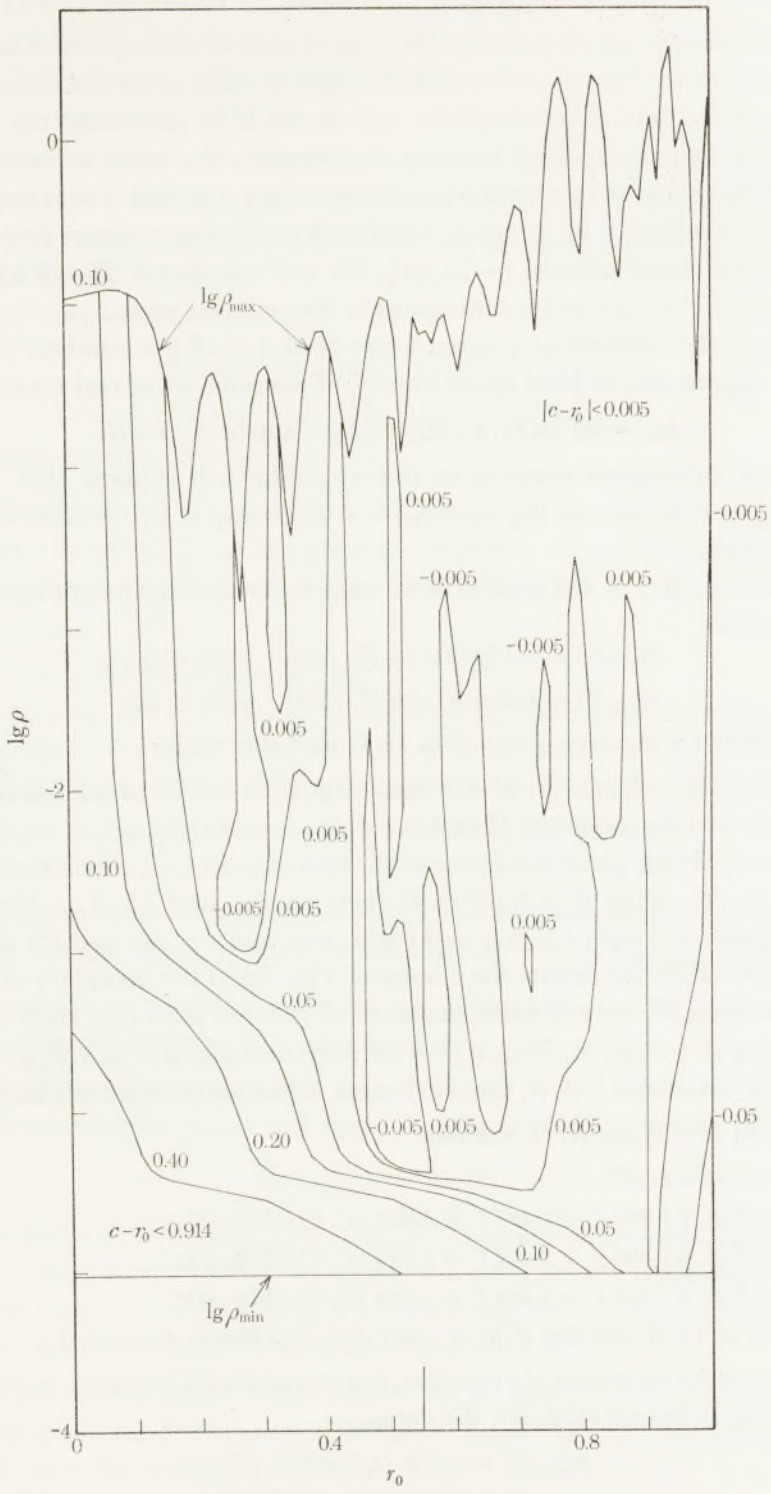

FIGURE 29. The same as figure 27 except that the errors in the gross Earth data are those reported in the literature and listed in table 2. This map is the companion to figure 28.

Appendix A. Constrained infimums of pairs of functions

The purpose of this appendix is to isolate the critical parts of the hypothesis in lemmas 3 and 8 and thus to prove a general result which includes those lemmas as special cases.

We denote by $\mathscr{R}$ the field of real numbers, and we let H be any connected Hausdorff space containing at least two points (and hence a continuum). We want to examine real-valued functions ϕ on H, but we want to permit ϕ to take the values $+\infty$ and $-\infty$ at certain points of H. We can do this if we replace ϕ by $\arctan\phi$, which will then be an ordinary real-valued function mapping H into the closed interval $[-\frac{1}{2}\pi, \frac{1}{2}\pi]$. We will say that $\phi: H \to \mathscr{R} \cup \{\infty, -\infty\}$ is continuous if $\arctan\phi: H \to [-\frac{1}{2}\pi, \frac{1}{2}\pi]$ is continuous in the ordinary sense.

We define $\phi_{\inf}$ as the infimum or greatest lower bound of all the numbers $\phi(h)$ with h in H, while $\phi_{\sup}$ is the supremum, or least upper bound, of the same set of real numbers:

$$\phi_{\inf} = \inf\{\phi(h) : h \in H\}, \quad \phi_{\sup} = \sup\{\phi(h) : h \in H\}.$$

The proofs of our subsequent remarks in this appendix will suppose that $-\infty < \phi_{\inf}$ and $\phi_{\sup} < +\infty$, but the extension to the case $\phi_{\inf} = -\infty$ or $\phi_{\sup} = +\infty$ is immediate: we simply replace ϕ by $\arctan\phi$.

For two functions $\phi: H \to \mathscr{R}$ and $\psi: H \to \mathscr{R}$ we want to consider the monotonic, non-increasing real-valued functions

$$\left.\begin{aligned} m_{\psi,\phi}(s) &= \inf\{\psi(h): h \in H \quad \text{and} \quad \phi(h) \leqslant s\}, \\ m_{\phi,\psi}(t) &= \inf\{\phi(h): h \in H \quad \text{and} \quad \psi(h) \leqslant t\}. \end{aligned}\right\} \tag{A 1}$$

To exploit the symmetry between ϕ and ψ in (A 1) we begin with:

Definition. If $\phi: H \to \mathscr{R}$ and $\psi: H \to \mathscr{R}$ then $S(\phi, \psi)$ is the set of all ordered pairs (s, t) of real numbers such that for some h in H we have $\phi(h) < s$ and $\psi(h) < t$.

We will denote by $\mathscr{R}^2$ the plane consisting of all ordered pairs (s, t) of real numbers. Evidently $S(\phi, \psi)$ is open in $\mathscr{R}^2$, while if $(s, t) \in S(\phi, \psi)$ then $s > \phi_{\inf}$ and $t > \psi_{\inf}$. Moreover, $S(\phi, \psi)$ contains all the pairs (s, t) with $s \geqslant \phi_{\sup}$ and $t > \psi_{\inf}$ or with $s > \phi_{\inf}$ and $t \geqslant \psi_{\sup}$.

If S is any subset of $\mathscr{R}^2$, we denote the closure of S by $\bar{S}$ and the boundary of S by ∂S. By the transpose of S, written S^T, we will mean the set of all ordered pairs (t, s) such that (s, t) is in S. Then clearly $S(\phi, \psi)^T = S(\psi, \phi)$, $\partial S(\phi, \psi)^T = \partial S(\psi, \phi)$ and $\overline{S(\phi, \psi)}^T = \overline{S(\psi, \phi)}$.

The elementary facts about $S(\phi, \psi)$ listed in lemma A 1 follow immediately from the definition:

Lemma A 1. *If $\phi: H \to \mathscr{R}$ and $\psi: H \to \mathscr{R}$ then*

(i) *$S(\phi, \psi)$ is open in $\mathscr{R}^2$;*
(ii) *if $(s, t) \in S(\phi, \psi)$ and $s' \geqslant s$ and $t' \geqslant t$ then $(s', t') \in S(\phi, \psi)$;*
(iii) *if $(s, t) \in \overline{S(\phi, \psi)}$ and $s' \geqslant s$ and $t' \geqslant t$ then $(s', t') \in \overline{S(\phi, \psi)}$;*
(iv) *if $(s, t) \in \overline{S(\phi, \psi)}$ and $s' > s$ and $t' > t$ then $(s', t') \in S(\phi, \psi)$;*
(v) *if there is an $h \in H$ such that $\phi(h) \leqslant s$ and $\psi(h) \leqslant t$ then $(s, t) \in \overline{S(\phi, \psi)}$.*

Clearly $\partial S(\phi, \psi)$ is not the graph of a function, but to study it we introduce two functions whose graphs will turn out to bound $\partial S(\phi, \psi)$. We define

$$\left.\begin{aligned} m^+_{\psi,\phi}(s) &= \inf\{t: (s, t) \in S(\phi, \psi)\}, \\ m^-_{\psi,\phi}(s) &= \inf\{t: (s, t) \in \overline{S(\phi, \psi)}\}. \end{aligned}\right\} \tag{A 2}$$

We agree that the infimum of the empty set is $+\infty$, so $m^+_{\psi,\phi}(s) = +\infty$ if $s \leqslant \phi_{\inf}$ and $m^-_{\psi,\phi}(s) = +\infty$ if $s < \phi_{\inf}$. We will need a number of properties of $m^+_{\psi,\phi}$ and $m^-_{\psi,\phi}$. Obviously $m^+_{\psi,\phi}(s) \geqslant m^-_{\psi,\phi}(s)$. Then we have

LEMMA A 2.

(i) $t > m^+_{\psi,\phi}(s)$ *implies* $(s,t) \in S(\phi,\psi)$;

(ii) $m^-_{\psi,\phi}(s) \leqslant t \leqslant m^+_{\psi,\phi}(s)$ *implies* $(s,t) \in \partial S(\phi,\psi)$;

(iii) $t < m^-_{\psi,\phi}(s)$ *implies* $(s,t) \notin \overline{S(\phi,\psi)}$.

Proof. The first and second of equations (A 2) imply (i) and (iii) immediately. Moreover, from the first of equations (A 2), if $t < m^+_{\psi,\phi}(s)$ then $(s,t) \notin S(\phi,\psi)$, while from the second of those equations if $t > m^-_{\psi,\phi}(s)$ then $(s,t) \in \overline{S(\phi,\psi)}$. Thus if $m^-_{\psi,\phi}(s) < t < m^+_{\psi,\phi}(s)$ then $(s,t) \in \partial S(\phi,\psi)$. Then (ii) follows because $\partial S(\phi,\psi)$ is closed.

As a corollary of lemma A 2 we have

$$\left.\begin{aligned} m^+_{\psi,\phi}(s) &= \sup\{t\colon (s,t) \in \partial S(\phi,\psi)\},\\ m^-_{\psi,\phi}(s) &= \inf\{t\colon (s,t) \in \partial S(\phi,\psi)\}. \end{aligned}\right\} \tag{A 3}$$

As another corollary of lemma A 2 we have

LEMMA A 3. *If* $h \in H$ *and* $\phi(h) < s$ *then* $\psi(h) \geqslant m^+_{\psi,\phi}(s)$.

Proof. By lemma A 2, $(\phi(h), m^+_{\psi,\phi}(\phi(h))) \in \overline{S(\phi,\psi)}$. If $\psi(h) < m^+_{\psi,\phi}(s)$ then by part (iv) of lemma A 1, $(s, m^+_{\psi,\phi}(s)) \in S(\phi,\psi)$. This contradicts part (ii) of lemma A 2.

We also have

LEMMA A 4. *If* $s_1 < s_2$ *then* $m^-_{\psi,\phi}(s_1) \geqslant m^+_{\psi,\phi}(s_2)$.

Proof. Let $t_1 = m^-_{\psi,\phi}(s_1)$ and $t_2 = m^+_{\psi,\phi}(s_2)$. From part (ii) of lemma A 2, both (s_1,t_1) and (s_2,t_2) are in $\partial S(\phi,\psi)$. Then from part (iv) of lemma A 1 we must have $t_2 \leqslant t_1$.

Because $m^-_{\psi,\phi}(s) \leqslant m^+_{\psi,\phi}(s)$, lemma A 4 implies without further argument that

LEMMA A 5. $m^-_{\psi,\phi}(s)$ *and* $m^+_{\psi,\phi}(s)$ *are monotonic non-increasing functions of* s.

It is also easy to prove

LEMMA A 6. $m^-_{\psi,\phi}(s)$ *is continuous on the right and* $m^+_{\psi,\phi}(s)$ *is continuous on the left for all* s.

Proof. We may assume $\phi_{\inf} \leqslant s_0$. Let $t = \lim_{s \to s_0+} m^+_{\psi,\phi}(s)$. If $t < m^-_{\psi,\phi}(s_0)$ then the fact that $(s_0,t) \in \overline{S(\phi,\psi)}$ contradicts the definition of $m^-_{\psi,\phi}(s_0)$. Therefore $t \geqslant m^-_{\psi,\phi}(s_0)$, and lemma A 4 requires equality. The same argument applies to $m^+_{\psi,\phi}(s_0)$.

LEMMA A 7. $m^-_{\psi,\phi}(s) \leqslant m_{\psi,\phi}(s) \leqslant m^+_{\psi,\phi}(s)$.

Proof. Again we may assume $\phi_{\inf} \leqslant s$. If $m^+_{\psi,\phi}(s) < m_{\psi,\phi}(s)$ then, by lemma A 2,

$$(s, m_{\psi,\phi}(s)) \in S(\phi,\psi)$$

so there is an h such that $\phi(h) < s$ and $\psi(h) < m_{\psi,\phi}(s)$. This contradicts the definition of $m_{\psi,\phi}(s)$. Again, if $m_{\psi,\phi}(s) < m^-_{\psi,\phi}(s)$ then for any t in $m_{\psi,\phi}(s) < t < m^-_{\psi,\phi}(s)$ we know from lemma A 2 that $(s,t) \notin \overline{S(\phi,\psi)}$. But then part (v) of lemma A 1 implies that for no such t is there an $h \in H$ with $\phi(h) \leqslant s$ and $\psi(h) \leqslant t$. Again we contradict the definition of $m_{\psi,\phi}(s)$.

LEMMA A 8. *Let* s *be any real number* $\geqslant \phi_{\inf}$. *Suppose that for every* $s' < s$ *we have* $m^-_{\psi,\phi}(s') > m^-_{\psi,\phi}(s)$. *(This is always true if* $s = \phi_{\inf}$.*) Then there is an infinite sequence* $\{h_1, h_2, \ldots\}$ *of points in* H *such that* $\lim_{n\to\infty} \phi(h_n) = s$ *and* $\lim_{n\to\infty} \psi(h_n) = m^-_{\psi,\phi}(s)$.

Proof. Let $t = m^-_{\psi,\phi}(s)$. Since $(s,t) \in \overline{S(\phi,\psi)}$ there is an infinite sequence $(s_n,t_n) \in S(\phi,\psi)$ such that $\lim_{n\to\infty}(s_n,t_n) = (s,t)$. Then, by the definition of $S(\phi,\psi)$, for each n there is an h_n in H such that $\phi(h_n) < s_n$ and $\psi(h_n) < t_n$. Thus $\limsup_{n\to\infty} \phi(h_n) \leqslant s$ and $\limsup_{n\to\infty} \psi(h_n) \leqslant t$. Let $s' = \liminf_{n\to\infty} \phi(h_n)$.

If $s' < s$ then there are infinitely many n such that $\phi(h_n) < \frac{1}{2}(s'+s)$. By lemma A 3 these all have $\psi(h_n) \geqslant m^+_{\psi,\phi}\frac{1}{2}(s'+s) \geqslant m^-_{\psi,\phi}\frac{1}{2}(s'+s)$, so $\limsup_{n\to\infty} \psi(h_n) \geqslant m^-_{\psi,\phi}\frac{1}{2}(s'+s) > m^-_{\psi,\phi}(s) = t$. This contradiction forces us to conclude that $s' = s$, so $\lim_{n\to\infty} \phi(h_n) = s$. Let $t' = \liminf_{n\to\infty} \psi(h_n)$. Clearly $(s,t') \in \overline{S(\phi,\psi)}$ so by part (iii) of lemma (A 2) $t' \geqslant t$. Thus $\lim_{n\to\infty} \psi(h_n) = t$.

Now we define $$\psi_{\max}(\phi) = m^-_{\psi,\phi}(\phi_{\inf}), \quad \phi_{\max}(\psi) = m^-_{\phi,\psi}(\psi_{\inf}). \tag{A 4}$$

We have $\phi_{\inf} \leqslant \phi_{\max}(\psi) \leqslant \phi_{\sup}$ and $\psi_{\inf} \leqslant \psi_{\max}(\phi) \leqslant \psi_{\sup}$. Also, if $s \geqslant \phi_{\max}(\psi)$ then $m^-_{\psi,\phi}(s) = \psi_{\inf}$, while if $s > \phi_{\max}(\psi)$ then $m^+_{\psi,\phi}(s) = \psi_{\inf}$. Similarly, if $t \geqslant \psi_{\max}(\phi)$ then $m^-_{\phi,\psi}(t) = \phi_{\inf}$, while if $t > \psi_{\max}(\phi)$ then $m^+_{\phi,\psi}(t) = \phi_{\inf}$. Moreover, if $\phi_{\inf} < s < \phi_{\max}(\psi)$ then

$$\psi_{\inf} < m^-_{\psi,\phi}(s) \leqslant m^+_{\psi,\phi}(s) \leqslant \psi_{\max}(\phi). \tag{A 5}$$

The fact that $\psi_{\inf} < m^-_{\psi,\phi}(s)$ when $s < \phi_{\max}(\psi)$ follows from the definition of $m^-_{\phi,\psi}(\psi_{\inf})$, equation (A 1). The fact that $m^+_{\psi,\phi}(s) \leqslant \psi_{\max}(\phi)$ when $s > \phi_{\inf}$ follows from the definition of $\psi_{\max}(\phi)$ and lemma A 4. By taking transposes of all the sets in question we immediately infer from (A 5) that when $\psi_{\inf} < t < \psi_{\max}(\phi)$ then

$$\phi_{\inf} < m^-_{\phi,\psi}(t) \leqslant m^+_{\phi,\psi}(t) \leqslant \phi_{\max}(\psi). \tag{A 6}$$

It follows that $\psi_{\inf} = \psi_{\max}(\phi)$ if and only if $\phi_{\inf} = \phi_{\max}(\psi)$, in which case

$$S(\phi,\psi) = \{(s,t): \phi_{\inf} < s \text{ and } \psi_{\inf} < t\}.$$

We define $$G(\phi,\psi) = \partial S(\phi,\psi) \cap \{(s,t): s > \phi_{\inf} \text{ and } t > \psi_{\inf}\}.$$

From what has been said, $\overline{G(\phi,\psi)}$ consists of $G(\phi,\psi)$ and the two points $(\phi_{\inf}, \psi_{\max}(\phi))$ and $(\phi_{\max}(\psi), \psi_{\inf})$, so

$$\overline{G(\phi,\psi)} = \partial S(\phi,\psi) \cap \{(s,t): s \leqslant \phi_{\max}(\psi) \text{ and } t \leqslant \psi_{\max}(\phi)\}.$$

From lemma A 3 we infer that there are at most denumerably many values of s where $m^-_{\psi,\phi}(s) < m^+_{\psi,\phi}(s)$, i.e. where $\overline{G(\phi,\psi)}$ has straight segments parallel to the t axis. We want to know when these occur.

Definition. *If $\{h_1, h_2, \ldots\}$ is a sequence of points in H such that $\phi(h_n) < s$ for all n and $\lim_{n\to\infty} \psi(h_n) = t$, and if $(s,t) \in \overline{G(\phi,\psi)}$, then the sequence $\{h_n\}$ is said to '(ϕ,ψ)-evoke the point (s,t)', and (s,t) is 'an evoked point in $\overline{G(\phi,\psi)}$'.*

We have

Lemma A 9. *If $(s, m^-_{\psi,\phi}(s))$ is an evoked point in $\overline{G(\phi,\psi)}$ then $m^-_{\psi,\phi}(s) = m^+_{\psi,\phi}(s)$.*

Proof. Let $t = m^-_{\psi,\phi}(s)$ and let $\{h_n\}$ be a sequence in H which (ϕ,ψ)-evokes (s,t). Let $s_n = \phi(h_n)$ and $t_n = \psi(h_n)$. According to

Lemma A 3. $t_n \geqslant m^+_{\psi,\phi}(s)$. *But* $\lim_{n\to\infty} t_n = t$, *whence the conclusion.*

To proceed further with the discussion, we need some limitations on the functions ϕ and ψ. We introduce

Definition. *If $\phi: H \to \mathscr{R}$ then a point $h \in H$ is a weak local minimum of ϕ if there is a neighbourhood U of h such that $\phi(h') \geqslant \phi(h)$ for all $h' \in U$.*

Definition. *If $\phi: H \to \mathscr{R}$ then ϕ° is the restriction of ϕ to the domain $\{h: \phi_{\inf} < \phi(h) < \phi_{\sup}\} \cap H$.*

Definition. *We say $\phi: H \to \mathscr{R}$ is ψ-compactly defined if for any real $s < \phi_{\max}(\psi)$ the set $\{h: \phi(h) \leqslant s\}$ is compact in H.*

Using these definitions we can prove

LEMMA A 10. *Suppose $\phi: H \to \mathscr{R}$ and $\psi: H \to \mathscr{R}$ are both continuous, that either ϕ is ψ-compactly defined or ψ is ϕ-compactly defined, and that ϕ^0 has no weak local minima. Then if $\phi_{\inf} < s < \phi_{\max}(\psi)$ we have $m^-_{\psi,\phi}(s) = m^+_{\psi,\phi}(s)$.*

Proof. Let $t = m^-_{\psi,\phi}(s)$ and suppose $t < m^+_{\psi,\phi}(s)$. Then $t < \psi_{\max}(\phi)$. According to lemma A 4, $m^-_{\psi,\phi}(s') > m^-_{\psi,\phi}(s)$ for all $s' < s$. Then lemma A 8 gives us a sequence $\{h_n\}$ with $\lim_{n\to\infty} \phi(h_n) = s$ and $\lim_{n\to\infty} \psi(h_n) = t$. Since $s < \phi_{\max}(\psi)$ and $t < \psi_{\max}(\phi)$, the fact that either ϕ is ψ-compactly defined or ψ is ϕ-compactly defined assures the existence of a limit point h of the sequence $\{h_n\}$. The continuity of ϕ and ψ then implies $\phi(h) = s$ and $\psi(h) = t$. But $\phi^0(h) = s$ and ϕ^0 has no weak local minima. Hence there exists a second sequence $\{h'_n\}$ such that $\lim_{n\to\infty} h'_n = h$ and $\phi(h_n) < s$. According to lemma A 3, $\psi(h'_n) \geqslant m^+_{\psi,\phi}(s) > t$, which contradicts the continuity of ψ. Hence the conclusion.

In consequence of lemma A 6 the hypotheses of lemma A 10 also imply that $m^-_{\psi,\phi}(s)$ is continuous in $\phi_{\inf} \leqslant s < \phi_{\max}(\psi)$. If $(\phi_{\max}(\psi), \psi_{\inf})$ is an evoked point in $\overline{G(\phi, \psi)}$, then lemma A 9 shows that $m^-_{\psi,\phi}(s) = m^+_{\psi,\phi}(s)$ if $s > \phi_{\inf}$, and that $m^-_{\psi,\phi}(s)$ is continuous if $s \geqslant \phi_{\inf}$.

Now we summarize our observations as theorems.

THEOREM A 1. *Suppose $\phi: H \to \mathscr{R}$ and $\psi: H \to \mathscr{R}$ are continuous and that ϕ^0 and ψ^0 have no weak local minima. Suppose either that ϕ is ψ-compactly defined or ψ is ϕ-compactly defined. Suppose that $(\phi_{\max}(\psi), \psi_{\inf})$ is an evoked point in $\overline{G(\phi, \psi)}$ and $(\psi_{\max}(\phi), \phi_{\inf})$ is an evoked point in $\overline{G(\psi, \phi)}$. Then $m^-_{\psi,\phi}(s)$ is a continuous, monotonic decreasing function in $\phi_{\inf} \leqslant s \leqslant \phi_{\max}(\psi)$ which maps that closed interval in one-to-one fashion onto the closed interval $[\psi_{\inf}, \psi_{\max}(\phi)]$. The function on $\psi_{\inf} \leqslant t \leqslant \psi_{\max}(\phi)$ which is inverse to $m^-_{\psi,\phi}(s)$ is $m^-_{\phi,\psi}(t)$. Furthermore, if $\phi_{\inf} < s$ then*

$$m^-_{\psi,\phi}(s) = m_{\psi,\phi}(s) = m^+_{\psi,\phi}(s),$$

and if $\psi_{\inf} < t$ then

$$m^-_{\phi,\psi}(t) = m_{\phi,\psi}(t) = m^+_{\phi,\psi}(t).$$

The proof of theorem A 1 is a straightforward application of the preceding lemmas to $S(\phi, \psi)$ and $S(\psi, \phi)$, so the details will be omitted. Equally obvious is

THEOREM A 2. *Under the hypotheses of theorem A 1, if $(s, t) \in G(\phi, \psi)$ then there is a point h in H such that $\phi(h) = s$ and $\psi(h) = t$.*

Proof. The point h was constructed in the proof of lemma A 10.

THEOREM A 3 *Under the hypotheses of theorem A 1, if $\phi_{\inf} < s < \phi_{\max}(\psi)$ then the set*

$$\{h: \phi(h) \leqslant s\} \cap \{h: \psi(h) \leqslant t\} \cap H$$

(i) *is empty if $t < m_{\psi,\phi}(s)$;*

(ii) *is non-empty with empty interior if $t = m_{\psi,\phi}(s)$, and contains only points h with $\phi(h) = s$ and $\psi(h) = t$;*

(iii) *has non-empty interior if $t > m_{\psi,\phi}(s)$.*

Proof. Assertion (i) follows immediately from part (v) of lemma A 1, while assertion (iii) follows from part (i) of lemma A 2. In assertion (ii) the existence of an h such that $\phi(h) = s$ and $\psi(h) = t$ is theorem A 2. Next we show that if $t = m^-_{\psi,\phi}(s)$ then there is no open set U in H where $\phi(h) \leqslant s$ and $\psi(h) \leqslant t$. If such a set U exists, then the open set $U \cap \{h: \phi(h) < s\}$ cannot be empty, because ϕ^0 has no weak local minimum. Then the open set $U \cap \{h: \phi(h) < s\} \cap \{h: \psi(h) < s\}$ cannot be empty, because ψ^0 has no weak local minimum. But then $(s, t) \in S(\phi, \psi)$, contrary to the hypothesis of assertion (ii). It remains to show that if $t = m_{\psi,\phi}(s)$ and $\phi(h) \leqslant s$ and $\psi(h) \leqslant t$ then

$\phi(h) = s$ and $\psi(h) = t$. We cannot have both $\phi(h) < s$ and $\psi(h) < t$ because then $(s, t) \in S(\phi, \psi)$, contrary to part (ii) of lemma A 2. But if one of the two equations $\phi(h) = s$ or $\psi(h) = t$ holds, then according to theorem A 1 and lemma A 2 so does the other.

Now we must show that the foregoing results are applicable to convexly defined functions.

Definition. Suppose H is a convex subset of a real vector space V and $\phi: H \to \mathscr{R}$. We say ϕ is 'strictly convexly defined' if any $s < \phi_{\text{sup}}$ has the following property: if $\boldsymbol{h}_i \in H$ and $\phi(\boldsymbol{h}_i) \leqslant s$, with $i = 1$ or 2, then $\phi(\boldsymbol{h}) < s$ whenever $\boldsymbol{h}$ is a point between $\boldsymbol{h}_1$ and $\boldsymbol{h}_2$ on the straight line segment joining them (i.e. whenever there are positive numbers α_1 and α_2 such that $\alpha_1 + \alpha_2 = 1$ and $\boldsymbol{h} = \alpha_1 \boldsymbol{h}_1 + \alpha_2 \boldsymbol{h}_2$).

Now we have

LEMMA A 11. *If $\phi: H \to \mathscr{R}$ is strictly convexly defined then ϕ^0 has no weak local minima.*

Proof. Let $\boldsymbol{h}_1$ be in the domain of ϕ^0. Since $\phi(\boldsymbol{h}_1) > \phi_{\text{inf}}$, there is a point $\boldsymbol{h}_2$ in H where $\phi(\boldsymbol{h}_2) < \phi(\boldsymbol{h}_1)$. Then the straight line segment joining h_1 and h_2 consists of points h where $\phi(\boldsymbol{h}) < \phi(\boldsymbol{h}_1)$, so $\boldsymbol{h}_1$ is not a weak local minimum.

If $\phi: H \to \mathscr{R}$ and $\psi: H \to \mathscr{R}$ and the domain H is a convex subset of a real vector space, then evidently in theorems A 1, A 2 and A 3 we can replace the hypothesis that ϕ and ψ have no weak local minima by the hypothesis that ϕ and ψ are strictly convexly defined. This stronger hypothesis gives us some additional information:

LEMMA A 12. *Suppose that H is a convex subset of a real vector space and that $\phi: H \to \mathscr{R}$ and $\psi: H \to \mathscr{R}$ are strictly convexly defined. Suppose also that they are both continuous and their either ϕ is ψ-compactly defined or ψ is ϕ-compactly defined. Then the hypothesis and conclusions of theorems* A 1, A 2 *and* A 3 *apply to ϕ and ψ. In addition, if $\phi_{\text{inf}} < s < \phi_{\max}(\psi)$ and $t = m_{\psi,\phi}(s)$ then the set*

$$\{\boldsymbol{h}: \phi(\boldsymbol{h}) \leqslant s\} \cap \{\boldsymbol{h}: \psi(\boldsymbol{h}) \leqslant t\} \cap H$$

contains exactly one point if $t = m_{\psi,\phi}(s)$, and this point has $\phi(\boldsymbol{h}) = s$, $\psi(\boldsymbol{h}) = t$.

Proof. Everything in lemma A 12 follows immediately from lemma A 11 and theorems A 1, A 2 and A 3 except for the assertion that when $\phi_{\text{inf}} < s < \phi_{\max}(\psi)$ and $t = m_{\psi,\phi}(s)$ then there is exactly one point in H which satisfies $\phi(\boldsymbol{h}) \leqslant s$ and $\psi(\boldsymbol{h}) \leqslant t$. From theorem A 2 we are assured of the existence of such a point, and from part (ii) of theorem A 3 we are sure that for any such point $\phi(\boldsymbol{h}) = s$ and $\psi(\boldsymbol{h}) = t$. It remains only to prove that there cannot be two such points. In fact if we had $\phi(\boldsymbol{h}_i) \leqslant s$ and $\psi(\boldsymbol{h}_i) \leqslant t$ for $i = 1$ and 2, we could define $\boldsymbol{h} = \frac{1}{2}(\boldsymbol{h}_1 + \boldsymbol{h}_2)$ and conclude that $\phi(\boldsymbol{h}) < s$ and $\psi(\boldsymbol{h}) < t$. This contradiction implies $\boldsymbol{h}_1 = \boldsymbol{h}_2$.

APPENDIX B. DERIVATIVES ALONG THE TRADEOFF CURVES FOR RELATIVE ERROR

In order to obtain the expressions (7.37) to (7.42) for w_+ and w_- and the rates at which $s_\pm(\theta)$ and $\rho_\pm(\theta)$ approach their limits as θ approaches 0 or $\frac{1}{2}\pi$, we must calculate the derivatives $\partial_\theta s_\pm$, $\partial_\theta \rho_\pm$ and $\partial_\theta \boldsymbol{a}_\pm$. We use the notation of § 7.

If we differentiate (7.7) with respect to θ and solve for $\partial_\theta \boldsymbol{a}$, we obtain

$$\partial_\theta \boldsymbol{a} = \tilde{\boldsymbol{q}}(\partial_\theta t) + \tilde{\boldsymbol{u}}[(\partial_\theta s)\cos\theta - s\sin\theta] - \tilde{\boldsymbol{y}} \tag{B 1}$$

where

$$\tilde{\boldsymbol{y}} = \boldsymbol{W}^{-1}.\partial_\theta \boldsymbol{W}.\boldsymbol{a}. \tag{B 2}$$

Next we differentiate (7.8) and (7.9) with respect to θ, obtaining

$$\boldsymbol{u}.\partial_\theta \boldsymbol{a} = 0, \tag{B 3}$$

$$\tfrac{1}{2}\partial_\theta s = \boldsymbol{a}.\boldsymbol{S}.\partial_\theta \boldsymbol{a}. \tag{B 4}$$

When we substitute (B 1) in these two equations the result is a pair of inhomogeneous linear equations for $\partial_\theta s$ and $\partial_\theta t$, whose solution is

$$\Delta\partial_\theta s = \boldsymbol{a}.\boldsymbol{S}.(\tilde{\boldsymbol{q}}\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{y}}\tilde{\boldsymbol{q}}).\boldsymbol{u}+s(\sin\theta)\boldsymbol{a}.\boldsymbol{S}.\tilde{\boldsymbol{h}}, \tag{B 5}$$

$$\Delta\partial_\theta t = (\cos\theta)\boldsymbol{a}.\boldsymbol{S}.(\tilde{\boldsymbol{y}}\tilde{\boldsymbol{u}}-\tilde{\boldsymbol{u}}\tilde{\boldsymbol{y}}).\boldsymbol{u}+\tfrac{1}{2}\boldsymbol{u}.\tilde{\boldsymbol{y}}+\tfrac{1}{2}s(\sin\theta)\boldsymbol{u}.\tilde{\boldsymbol{u}}, \tag{B 6}$$

where
$$\Delta = \cos\theta(\boldsymbol{a}.\boldsymbol{S}.\tilde{\boldsymbol{h}})+\tfrac{1}{2}\boldsymbol{q}.\tilde{\boldsymbol{u}} \tag{B 7}$$

and
$$\boldsymbol{a} = t\tilde{\boldsymbol{q}}+s\cos\theta\tilde{\boldsymbol{u}}. \tag{B 8}$$

If we substitute the expressions (B 5) and (B 6) for $\partial_\theta s$ and $\partial_\theta t$ back in (B 1), we obtain an explicit expression for $\partial_\theta \boldsymbol{a}(\theta)$. Then differentiating (7.10) gives

$$\tfrac{1}{2}\partial_\theta\rho(\theta)^2 = \frac{\boldsymbol{a}.(\boldsymbol{E}-\rho^2\boldsymbol{q}\boldsymbol{q}).\partial_\theta\boldsymbol{a}}{(\boldsymbol{q}.\boldsymbol{a})^2}. \tag{B 9}$$

The explicit expressions for $\partial_\theta\boldsymbol{a}$ and $\partial_\theta\rho(\theta)^2$ in general seem to be extremely complicated, and are not given here. We do not need them to deduce the identity (7.33). We simply multiply equation (B 4) by $2\cos\theta$, multiply (B 9) by $2(\boldsymbol{q}.\boldsymbol{a})^2 w\sin\theta$, and add the products. The result is

$$(\boldsymbol{q}.\boldsymbol{a})^2\sin\theta\,\partial_\theta[w\rho(\theta)^2]+\cos\theta\,\partial_\theta s(\theta) = 2\boldsymbol{a}.[\boldsymbol{W}-w(\sin\theta)\rho^2\boldsymbol{q}\boldsymbol{q}].\partial_\theta\boldsymbol{a}.$$

According to equations (7.7) and (7.11), the right-hand side of this equation is $2s(\cos\theta)\boldsymbol{u}.\partial_\theta\boldsymbol{a}$. Then (B 3) yields

$$(\boldsymbol{q}.\boldsymbol{a})^2\sin\theta\,\partial_\theta[w\rho(\theta)^2]+\cos\theta\,\partial_\theta s(\theta) = 0, \tag{B 10}$$

which is equivalent to (7.33).

A more explicit expression for $\partial_\theta s$ can be obtained by using the facts that

$$\sin\theta\,\partial_\theta\boldsymbol{W} = \boldsymbol{W}\cos\theta-\boldsymbol{S}, \quad \cos\theta\,\partial_\theta\boldsymbol{W} = w\boldsymbol{E}-\boldsymbol{W}\sin\theta,$$

and the identity
$$\boldsymbol{a}.\boldsymbol{S}.\tilde{\boldsymbol{q}}-s\cos\theta(\boldsymbol{a}.\boldsymbol{S}.\tilde{\boldsymbol{h}}) = s\tilde{\boldsymbol{u}}.\boldsymbol{q}.$$

This identity can be deduced from (B 8), (7.8), (7.9) and (7.13). The results for $\partial_\theta s$ are these:

$$(\sin\theta)\Delta\partial_\theta s = s\boldsymbol{a}.\boldsymbol{S}.\tilde{\boldsymbol{h}}+(\boldsymbol{q}.\tilde{\boldsymbol{u}})(\boldsymbol{a}.\boldsymbol{S}.\boldsymbol{W}.\boldsymbol{S}.\boldsymbol{a})-(\boldsymbol{a}.\boldsymbol{S}.\tilde{\boldsymbol{u}})(\boldsymbol{a}.\boldsymbol{S}.\tilde{\boldsymbol{q}}), \tag{B 11}$$

$$(\cos\theta)\Delta\partial_\theta s = (\boldsymbol{a}.\boldsymbol{S}.\tilde{\boldsymbol{q}})(\boldsymbol{a}.w\boldsymbol{E}.\tilde{\boldsymbol{u}})-(\boldsymbol{q}.\tilde{\boldsymbol{u}})(\boldsymbol{a}.\boldsymbol{S}.\boldsymbol{W}^{-1}.w\boldsymbol{E}.\boldsymbol{a}). \tag{B 12}$$

Similar but more complicated expressions for $\partial_\theta t$, $\partial_\theta\boldsymbol{a}$ and $\partial_\theta\rho^2$ can be deduced, but we omit them.

Now we can examine the rates of approach to the four limits we need. These four limits are

(i) $\theta\to 0$, $\boldsymbol{a}\to\boldsymbol{a}_S$, $s\to s_{\min}$, $\rho^2\to\rho^2_{\max}$;
(ii) $\theta\to 0$, $\boldsymbol{a}\to\boldsymbol{a}_\infty$, $s\to s_\infty$, $\rho^2\to\infty$;
(iii) $\theta\to\frac{1}{2}\pi$, $\boldsymbol{a}\to\boldsymbol{a}_\mathscr{C}$, $s\to\tilde{s}_{\max}$, $\rho^2\to\rho^2_{\min}$;
(iv) $\theta\to\frac{1}{2}\pi$, $\boldsymbol{a}\to\infty(\boldsymbol{a}_{\mathrm{E}}-\boldsymbol{a}_\mathscr{C})$, $s\to\infty$, $\rho^2\to\rho^2_{\mathrm{par}}$.

Which of cases (i) and (ii) is assigned to $\rho_+(s)$ and which is assigned to $\rho_-(s)$ depends on the sign of $\boldsymbol{q}.\boldsymbol{a}_S$, but the limiting behaviour can be calculated independently of that assignment. Case (iii) is always $\rho_+(s)$ and case (iv) is always $\rho_-(s)$.

Case (i): $\theta\to 0$, $\boldsymbol{a}\to\boldsymbol{a}_S$.

In this case, $\boldsymbol{W}\to\boldsymbol{S}$, $\partial_\theta\boldsymbol{W}\to w\boldsymbol{E}$, $\tilde{\boldsymbol{y}}\to\boldsymbol{S}^{-1}.w\boldsymbol{E}.\boldsymbol{a}_S$, $s\to s_{\min}$, $\rho\to\rho_{\max}$, $\tilde{\boldsymbol{u}}\to\boldsymbol{a}_S/s_{\min}$ and $\boldsymbol{a}_S.\boldsymbol{S}.\tilde{\boldsymbol{h}} = 0$ because of (7.14). Thus

$$\Delta = \frac{\boldsymbol{q}.\boldsymbol{a}_S}{2s_{\min}}. \tag{B 13}$$

Then $$\Delta\partial_\theta s(0) = 0, \quad \Delta\partial_\theta t(0) = \frac{\boldsymbol{a}_S . w\boldsymbol{E} . \boldsymbol{a}_S}{2s_{\min}},$$

so $$\partial_\theta \boldsymbol{a}(0) = -w\boldsymbol{S}^{-1} . (\boldsymbol{E} - \rho_{\max}^2 \boldsymbol{qq}) . \boldsymbol{a}_S. \tag{B 14}$$

Therefore $$\tfrac{1}{2}\partial_\theta \rho(0)^2 = -w \frac{\boldsymbol{a}_S . (\boldsymbol{E} - \rho_{\max}^2 \boldsymbol{qq}) . \boldsymbol{S}^{-1} . (\boldsymbol{E} - \rho_{\max}^2 \boldsymbol{qq}) . \boldsymbol{a}_S}{(\boldsymbol{q} . \boldsymbol{a}_S)^2}. \tag{B 15}$$

It follows from (B 14) and (B 15) that $\rho(s)$ has a vertical tangent at $s = s_{\min}$ when $\boldsymbol{a} = \boldsymbol{a}_S$.

Case (ii): $\theta \to 0$, $\boldsymbol{a} \to \boldsymbol{a}_\infty$

As in case (i), $\boldsymbol{W} \to \boldsymbol{S}$, $\partial_\theta \boldsymbol{W} \to w\boldsymbol{E}$ and $\tilde{\boldsymbol{u}} \to \boldsymbol{a}_S / s_{\min}$. However, $\tilde{\boldsymbol{y}} \to \boldsymbol{S}^{-1} . w\boldsymbol{E} . \boldsymbol{a}_\infty$, $s \to s_\infty$, and $\rho \to \infty$. If we define

$$\epsilon_\infty^2 = \boldsymbol{a}_\infty . \boldsymbol{E} . \boldsymbol{a}_\infty, \tag{B 16}$$

then from (7.32) we have $$\lim_{\theta \to 0} \theta\rho(\theta) = \frac{\epsilon_\infty}{|\boldsymbol{q} . \partial_\theta \boldsymbol{a}(0)|}. \tag{B 17}$$

The problem is to evaluate $\boldsymbol{q} . \partial_\theta \boldsymbol{a}(0)$ in case (ii).

Because $\boldsymbol{q} . \boldsymbol{a}_\infty = 0$, we have $$\Delta(0) = -\frac{\boldsymbol{q} . \boldsymbol{a}_S}{2s_{\min}}, \tag{B 18}$$

so, from (B 5), $$\partial_\theta s(0) = 2w\epsilon_\infty^2. \tag{B 19}$$

From (B 6), $$\partial_\theta t(0) = -\frac{w(\boldsymbol{a}_S . \boldsymbol{E} . \boldsymbol{a}_\infty - 2\epsilon_\infty^2)}{\boldsymbol{q} . \boldsymbol{a}_S}. \tag{B 20}$$

Then, from (B 1) and the two equations immediately following (7.27) we infer that when $\boldsymbol{a}(0) = \boldsymbol{a}_\infty$,

$$\boldsymbol{q} . \partial_\theta \boldsymbol{a}(0) = -\frac{w\epsilon_\infty^2 (\boldsymbol{q} . \boldsymbol{a}_S)}{s_\infty - s_{\min}}. \tag{B 21}$$

Thus in case (ii) $$\lim_{\theta \to 0} \theta\rho(\theta) = \frac{s_\infty - s_{\min}}{w\epsilon_\infty |\boldsymbol{q} . \boldsymbol{a}_S|}. \tag{B 22}$$

Case (iii): $\theta \to \frac{1}{2}\pi$, $\boldsymbol{a}_+ \to \boldsymbol{a}_{\mathscr{C}}$

In this case,

$$\boldsymbol{W} \to w\boldsymbol{E}, \quad \partial_\theta \boldsymbol{W} \to -\boldsymbol{S}, \quad \tilde{\boldsymbol{u}} \to \boldsymbol{a}_E / w_+ \epsilon_{\min}^2, \quad s \to \tilde{s}_{\max}, \quad \rho \to \rho_{\min},$$

and $$\tilde{\boldsymbol{q}} \to \boldsymbol{E}^{-1} . \boldsymbol{q} / w_+ = \boldsymbol{a}_{\mathscr{C}} (\boldsymbol{q} . \boldsymbol{a}_E) / w_+ \epsilon_{\min}^2.$$

From (B 10) we infer immediately that

$$\partial_\theta [\rho_+ (\tfrac{1}{2}\pi)^2] = 0. \tag{B 23}$$

From (B 7), $$\Delta_+ (\tfrac{1}{2}\pi) = \frac{\boldsymbol{q} . \boldsymbol{a}_E}{2w_+ \epsilon_{\min}^2}, \tag{B 24}$$

while $$w_+ \tilde{\boldsymbol{y}} = -\boldsymbol{E}^{-1} . \boldsymbol{S} . \boldsymbol{a}_{\mathscr{C}}.$$

Then a straightforward calculation reduces (B 5) to

$$\tfrac{1}{2} w_+ \partial_\theta s_+ (\tfrac{1}{2}\pi) = \boldsymbol{a}_{\mathscr{C}} . \boldsymbol{S} . \boldsymbol{E}^{-1} . \boldsymbol{S} . \boldsymbol{a}_{\mathscr{C}} + \frac{\tilde{s}_{\max}}{\epsilon_{\min}^2} (\tilde{s}_{\max} - 2\boldsymbol{a}_E . \boldsymbol{S} . \boldsymbol{a}_{\mathscr{C}}). \tag{B 25}$$

It follows from (B 23) and (B 25) that $\rho_+(s)$ has a horizontal tangent at $s = \tilde{s}_{\max}$.

Case (iv): $\theta \to \frac{1}{2}\pi$, $\boldsymbol{a} \to \infty(\boldsymbol{a}_E - \boldsymbol{a}_{\mathscr{C}})$.

This case has already been treated completely in equation (7.26).

Appendix C. Comparison of relative with absolute errors

Absolute errors are so much easier to use than relative errors that we prefer the former when both lead to essentially the same information about $\mathscr{G}$-acceptable Earth models. In the present appendix we study conditions which assure that minimizing either relative or absolute errors will indeed yield essentially the same information.

Any $\boldsymbol{a}$ in $\mathscr{R}^N$ which satisfies $\boldsymbol{u}.\boldsymbol{a} = 1$ produces, via (2.5), an averaging kernel A. With $\boldsymbol{u}$ and $\boldsymbol{q}$ defined in § 2 we have

$$\langle m_{\mathrm{E}}, A\rangle = \boldsymbol{q}.\boldsymbol{a}, \tag{C 1}$$

$$\epsilon(\boldsymbol{a})^2 = \boldsymbol{a}.\boldsymbol{E}.\boldsymbol{a}, \quad \rho(\boldsymbol{a})^2 = \epsilon(\boldsymbol{a})^2/(\boldsymbol{q}.\boldsymbol{a})^2. \tag{C 2}$$

By definition, $\epsilon(s)$ is the minimum of $\epsilon(\boldsymbol{a})$ and $\rho(s)$ is the minimum of $\rho(\boldsymbol{a})$ when $\boldsymbol{a}$ is constrained to lie in $\mathscr{E}(\boldsymbol{u}, \boldsymbol{S}, s)$. Let $\boldsymbol{a}(s)$ be the point in $\mathscr{E}(\boldsymbol{u}, \boldsymbol{S}, s)$ such that $\epsilon(s) = \epsilon(\boldsymbol{a}(s))$, and let $\tilde{\boldsymbol{a}}(s)$ be the point in $\mathscr{E}(\boldsymbol{u}, \boldsymbol{S}, s)$ such that $\rho(s) = \rho(\tilde{\boldsymbol{a}}(s))$. Let A_s and $\tilde{A}_s$ be the averaging kernels (2.5) produced by $\boldsymbol{a}(s)$ and $\tilde{\boldsymbol{a}}(s)$. Define

$$\rho^*(s) = \frac{\epsilon(\boldsymbol{a}(s))}{|\boldsymbol{q}.\boldsymbol{a}(s)|}. \tag{C 3}$$

Both $\langle m_{\mathrm{E}}, A_s\rangle$ and $\langle m_{\mathrm{E}}, \tilde{A}_s\rangle$ are measureable properties of the Earth; both are localized averages of m_{E} with the same spread from r_0, namely s. The average $\langle m_{\mathrm{E}}, A_s\rangle$ is known with the smaller absolute error and the average $\langle m_{\mathrm{E}}, \tilde{A}_s\rangle$ is known with the smaller relative error.

A measurement of the velocity of light *in vacuo* with an error of $\pm 100\,\mathrm{cm\,s^{-1}}$ would be regarded by almost everyone as 'more accurate' than a measurement of the velocity of sound in sea water with an error of $\pm 1\,\mathrm{cm\,s^{-1}}$, because the former error is three parts in 10^9 while the latter is one part in 10^5. The figure of merit for the accuracy of a measurement of a physical quantity p is not the absolute error Δp but the relative error $\Delta p/p$. From this point of view, $\langle m_{\mathrm{E}}, \tilde{A}_s\rangle$ is more 'accurately' known than $\langle m_{\mathrm{E}}, A_s\rangle$. Since both are estimates of a local average of m_{E} at the same position r_0 and with the same spread s, in principle $\tilde{A}_s$ is preferable to A_s as an averaging kernel. However, the added labour of computing $\tilde{A}_s$ is justified only if the relative error $\rho(s)$ in $\langle m_{\mathrm{E}}, \tilde{A}_s\rangle$ is significantly smaller than the relative error $\rho^*(s)$ in $\langle m_{\mathrm{E}}, A_s\rangle$.

When $s = s_{\min}$, we have $\tilde{\boldsymbol{a}}(s) = \boldsymbol{a}(s) = \boldsymbol{a}_{\boldsymbol{S}}$, so $\rho(s) = \rho^*(s)$. The question is, how small must $s - s_{\min}$ be to insure that $\rho^*(s) - \rho(s)$ is only a small fraction of $\rho^*(s)$ or $\rho(s)$. It is always possible by accident that $\rho^*(s) = \rho(s)$, so we seek only a sufficient condition to assure the inequality

$$\rho^*(s) - \rho(s) \ll \rho(s).$$

First we verify that

$$0 \leqslant \rho^*(s) - \rho(s). \tag{C 4}$$

This inequality is an immediate consequence of the definition of $\rho(s)$, which requires that $\rho(s) \leqslant \epsilon(\boldsymbol{a})/|\boldsymbol{q}.\boldsymbol{a}|$ for any $\boldsymbol{a}$, and in particular for $\boldsymbol{a} = \boldsymbol{a}(s)$.

Next we need an upper bound on $\rho^*(s) - \rho(s)$. The definition of $\epsilon(s)$ requires that

$$\epsilon(s)^2 \leqslant \tilde{\boldsymbol{a}}(s).\boldsymbol{E}.\tilde{\boldsymbol{a}}(s),$$

and therefore

$$\frac{\epsilon(s)^2}{|\boldsymbol{q}.\boldsymbol{a}(s)|^2} \leqslant \left(\frac{\boldsymbol{q}.\tilde{\boldsymbol{a}}(s)}{\boldsymbol{q}.\boldsymbol{a}(s)}\right)^2 \rho(s)^2. \tag{C 5}$$

If we define $q(s)$ and $Q(s)$ as, respectively, the minimum and maximum values of $|\boldsymbol{q}.\boldsymbol{a}|$ when $\boldsymbol{a}$ is confined to $\mathscr{E}(\boldsymbol{u}, \boldsymbol{S}, s)$, then we have from (C 5)

$$\rho^*(s) \leqslant \frac{Q(s)}{q(s)}\rho(s).$$

Therefore
$$0 \leqslant \rho^*(s) - \rho(s) \leqslant \left[\frac{Q(s) - q(s)}{q(s)}\right]\rho(s). \tag{C 6}$$

It remains to calculate $q(s)$ and $Q(s)$. These two extremal values of $\boldsymbol{q}.\boldsymbol{a}$ in $\mathscr{E}(\boldsymbol{u}, \boldsymbol{S}, s)$ are assumed at two points $\boldsymbol{a}_q$ and $\boldsymbol{a}_Q$ on $\partial\mathscr{E}(\boldsymbol{u}, \boldsymbol{S}, s)$, where the boundary is tangent to the two hyperplanes $\boldsymbol{q}.\boldsymbol{a} = q(s)$ and $\boldsymbol{q}.\boldsymbol{a} = Q(s)$. The argument is the now familiar one based on convexity, and will be omitted. It follows that the equations

$$\boldsymbol{S}.\boldsymbol{a} = \alpha\boldsymbol{q} + \beta\boldsymbol{u}, \quad \boldsymbol{u}.\boldsymbol{a} = 1, \quad s = \boldsymbol{a}.\boldsymbol{S}.\boldsymbol{a} \tag{C 7}$$

have a solution $\boldsymbol{a}_q$, α_q, β_q and another $\boldsymbol{a}_Q$, α_Q, β_Q. If we define $\tilde{\boldsymbol{q}} = \boldsymbol{S}^{-1}.\boldsymbol{q}$, $\tilde{\boldsymbol{u}} = \boldsymbol{S}^{-1}.\boldsymbol{u}$ and $\tilde{\boldsymbol{h}} = (\tilde{\boldsymbol{q}}\tilde{\boldsymbol{u}} - \tilde{\boldsymbol{u}}\tilde{\boldsymbol{q}}).\boldsymbol{u}$, then (C 7) imply that

$$\boldsymbol{a} = \frac{\tilde{\boldsymbol{u}} + \alpha\tilde{\boldsymbol{h}}}{\boldsymbol{u}.\tilde{\boldsymbol{u}}} \tag{C 8}$$

and
$$\alpha^2(\tilde{\boldsymbol{h}}.\boldsymbol{S}.\tilde{\boldsymbol{h}}) + 2\alpha(\tilde{\boldsymbol{u}}.\boldsymbol{S}.\tilde{\boldsymbol{h}}) + \tilde{\boldsymbol{u}}.\boldsymbol{S}.\tilde{\boldsymbol{u}} - s(\boldsymbol{u}.\tilde{\boldsymbol{u}})^2 = 0. \tag{C 9}$$

The two roots of (C 9) are α_q and α_Q, and

$$q(s) = \boldsymbol{q}.\boldsymbol{a}_q = \frac{\boldsymbol{q}.\tilde{\boldsymbol{u}} + \alpha_q\boldsymbol{q}.\tilde{\boldsymbol{h}}}{\boldsymbol{u}.\tilde{\boldsymbol{u}}},$$

$$Q(s) = \boldsymbol{q}.\boldsymbol{a}_Q = \frac{\boldsymbol{q}.\tilde{\boldsymbol{u}} + \alpha_Q\boldsymbol{q}.\tilde{\boldsymbol{h}}}{\boldsymbol{u}.\tilde{\boldsymbol{u}}}.$$

To solve (C 9) we first observe the following facts:

$$\tilde{\boldsymbol{u}}.\boldsymbol{S}.\tilde{\boldsymbol{u}} = \frac{1}{s_{\min}}, \quad \tilde{\boldsymbol{u}}.\boldsymbol{S}.\tilde{\boldsymbol{h}} = 0, \quad \tilde{\boldsymbol{h}}.\boldsymbol{S}.\tilde{\boldsymbol{h}} = \frac{(\boldsymbol{q}.\boldsymbol{a}_S)^2}{s_{\min}^2(s_\infty - s_{\min})}.$$

From the resulting α_q and α_Q we obtain

$$q(s) = \boldsymbol{q}.\boldsymbol{a}_S(1 - \zeta(s)), \quad Q(s) = \boldsymbol{q}.\boldsymbol{a}_S(1 + \zeta(s))$$

where
$$\zeta(s) = \left(\frac{s - s_{\min}}{s_\infty - s_{\min}}\right)^{\frac{1}{2}}. \tag{C 10}$$

Therefore
$$0 \leqslant \frac{\rho^*(s) - \rho(s)}{\rho(s)} \leqslant \frac{2\zeta(s)}{1 - \zeta(s)}. \tag{C 11}$$

For purposes of illustration, suppose we are willing to accept $\langle m_E, A_s\rangle$ rather than $\langle m_E, \tilde{A}_s\rangle$ as long as the relative error of the former is no more than 10 % larger than the relative error of the latter. To be sure that we have $\rho^*(s) \leqslant 1.1\rho(s)$ we must require

$$0 \leqslant \frac{2\zeta(s)}{1 - \zeta(s)} \leqslant 0.1 \quad \text{or} \quad 0 \leqslant \zeta(s) \leqslant 1/21, \quad \text{or} \quad 0 \leqslant \frac{s - s_{\min}}{s_\infty - s_{\min}} \leqslant \frac{1}{441}.$$

Appendix D. Calculation of ρ_{par} and proof of theorem 2

For any vectors $\boldsymbol{f}$ and $\boldsymbol{g}$ in $\mathscr{R}^N$ we define $\boldsymbol{f}.\boldsymbol{g}$ by (4.1). We suppose that $\boldsymbol{E}$: $\mathscr{R}^N \to \mathscr{R}^N$ is symmetric and positive definite and that $\boldsymbol{u}$ and $\boldsymbol{q}$ are two linearly independent vectors in $\mathscr{R}^N$. We seek the number ρ_{par} such that in the $(N-1)$-dimensional space $\mathscr{H}(\boldsymbol{u})$ the $(N-2)$-dimensional surface $\partial\mathscr{C}(\boldsymbol{u}, \boldsymbol{q}, \boldsymbol{E}, \rho^2)$ is an ellipsoid when $\rho < \rho_{\text{par}}$, a paraboloid when $\rho = \rho_{\text{par}}$, and a hyperboloid of two sheets when $\rho > \rho_{\text{par}}$.

We abbreviate the inner product $(\boldsymbol{f}, \boldsymbol{g})_E = \boldsymbol{f}.\boldsymbol{E}.\boldsymbol{g}$ as $(\boldsymbol{f}, \boldsymbol{g})$. The idea is to work entirely in the geometry defined on $\mathscr{R}^N$ by the inner product $(\boldsymbol{f}, \boldsymbol{g})$, because in this geometry $\mathscr{C}(\boldsymbol{q}, \boldsymbol{E}, \rho^2)$ is a right circular cone. We define $\boldsymbol{U} = \boldsymbol{E}^{-1}.\boldsymbol{u}$, $U = (\boldsymbol{U}, \boldsymbol{U})^{\frac{1}{2}}$, $\hat{\boldsymbol{U}} = U^{-1}\boldsymbol{U}$, and $\boldsymbol{Q} = \boldsymbol{E}^{-1}.\boldsymbol{q}$. Then a vector $\boldsymbol{a}$ is in $\mathscr{C}(\boldsymbol{u}, \boldsymbol{q}, \boldsymbol{E}, \rho^2)$ if and only if

$$(\boldsymbol{U}, \boldsymbol{a}) = 1 \tag{D 1}$$

and

$$(\boldsymbol{a}, \boldsymbol{a}) = (\rho\boldsymbol{Q}, \boldsymbol{a})^2. \tag{D 2}$$

Our first step is, roughly speaking, to translate $\mathscr{C}(\boldsymbol{u}, \boldsymbol{q}, \boldsymbol{E}, \rho^2)$ rigidly down from $\mathscr{H}_1(\boldsymbol{u})$ to $\mathscr{H}_0(\boldsymbol{u})$, so as to exploit the fact that $\mathscr{H}_0(\boldsymbol{u})$ is an $(N-1)$-dimensional subspace of $\mathscr{R}^N$. The condition that a vector $\boldsymbol{b}$ lie in $\mathscr{H}_0(\boldsymbol{u})$ is $(\boldsymbol{U}, \boldsymbol{b}) = 0$. Therefore, if $\boldsymbol{a}$ is any vector in $\mathscr{R}^N$ there is a unique vector $\boldsymbol{b}$ in $\mathscr{H}_0(\boldsymbol{u})$ such that

$$\boldsymbol{a} = (\boldsymbol{a}, \hat{\boldsymbol{U}})\,\hat{\boldsymbol{U}} + \boldsymbol{b}.$$

Evidently $\boldsymbol{b}$ is the perpendicular projection of $\boldsymbol{a}$ onto $\mathscr{H}_0(\boldsymbol{u})$. If $\boldsymbol{a}$ is in $\mathscr{H}(\boldsymbol{u})$ then $(\boldsymbol{a}, \hat{\boldsymbol{U}}) = U^{-1}$ so

$$\boldsymbol{a} = U^{-1}\hat{\boldsymbol{U}} + \boldsymbol{b}. \tag{D 3}$$

Equation (D 3) establishes a one-to-one correspondence between the points $\boldsymbol{a}$ in $\mathscr{H}(\boldsymbol{u})$ and the points $\boldsymbol{b}$ in $\mathscr{H}_0(\boldsymbol{u})$. The correspondence really consists in translating $\mathscr{H}_0(\boldsymbol{u})$ perpendicularly to itself through the vector displacement $U^{-1}\hat{\boldsymbol{U}}$.

If we substitute (D 3) in (D 2) we obtain

$$(\boldsymbol{b}, \boldsymbol{b}) + U^{-2} = [U^{-1}(\rho\boldsymbol{Q}, \hat{\boldsymbol{U}}) + (\rho\boldsymbol{Q}, \boldsymbol{b})]^2. \tag{D 4}$$

A point $\boldsymbol{a}$ in $\mathscr{R}^N$ is in $\mathscr{C}(\boldsymbol{u}, \boldsymbol{q}, \boldsymbol{E}, \rho^2)$ if and only if it has the form (D 3) where $\boldsymbol{b}$ is in $\mathscr{H}_0(\boldsymbol{u})$ and satisfies (D 4). Now we define $\boldsymbol{P} = \boldsymbol{Q} - (\boldsymbol{Q}, \hat{\boldsymbol{U}})\,\hat{\boldsymbol{U}}$. Since $\boldsymbol{q}$ and $\boldsymbol{u}$ are linearly independent, so are $\boldsymbol{Q}$ and $\boldsymbol{U}$. Therefore $\boldsymbol{P} \neq \boldsymbol{0}$, and if we define $P = (\boldsymbol{P}, \boldsymbol{P})^{\frac{1}{2}}$ we have $P \neq 0$. Then we can define $\hat{\boldsymbol{P}} = P^{-1}\boldsymbol{P}$. Now $\hat{\boldsymbol{P}}$ lies in $\mathscr{H}_0(\boldsymbol{u})$ and for any vector $\boldsymbol{b}$ in $\mathscr{H}_0(\boldsymbol{u})$ there is a unique vector $\boldsymbol{b}_\perp$ such that

$$\boldsymbol{b} = (\boldsymbol{b}, \hat{\boldsymbol{P}})\,\hat{\boldsymbol{P}} + \boldsymbol{b}_\perp \tag{D 5}$$

and

$$(\hat{\boldsymbol{P}}, \boldsymbol{b}_\perp) = 0. \tag{D 6}$$

With the help of (D 5) and (D 6) we can rewrite (D 4) as

$$(\boldsymbol{b}_\perp, \boldsymbol{b}_\perp) + (1-\rho^2P^2)\,(\hat{\boldsymbol{P}}, \boldsymbol{b})^2 - 2[\rho^2PU^{-1}(\boldsymbol{Q}, \hat{\boldsymbol{U}})]\,(\hat{\boldsymbol{P}}, \boldsymbol{b}) = U^{-2}[(\rho\boldsymbol{Q}, \hat{\boldsymbol{U}})^2 - 1]. \tag{D 7}$$

If $\rho^2 = P^{-2}$, then (D 7) is clearly the equation of a paraboloid of revolution whose vertex is at

$$\boldsymbol{b}_0 = \frac{1}{2U}\left[\frac{P}{(\boldsymbol{Q}, \hat{\boldsymbol{U}})} - \frac{(\boldsymbol{Q}, \hat{\boldsymbol{U}})}{P}\right]\hat{\boldsymbol{P}}$$

and whose axis consists of the vectors $t\hat{\boldsymbol{P}}$ with $t \geqslant (\hat{\boldsymbol{P}}, \boldsymbol{b}_0)$.

When $\rho^2 \neq P^{-2}$, we define

$$J = \frac{\rho^2P(\boldsymbol{Q}, \hat{\boldsymbol{U}})}{U(1-\rho^2P^2)} \tag{D 8}$$

and write (D 7) as

$$(\boldsymbol{b}_\perp, \boldsymbol{b}_\perp) + (1-\rho^2P)\,[(\hat{\boldsymbol{P}}, \boldsymbol{b})^2 - 2J(\hat{\boldsymbol{P}}, \boldsymbol{b})] = U^{-2}[(\rho\boldsymbol{Q}, \hat{\boldsymbol{U}})^2 - 1].$$

Completing the square yields

$$(\boldsymbol{b}_\perp, \boldsymbol{b}_\perp) + (1-\rho^2P^2)\,[(\hat{\boldsymbol{P}}, \boldsymbol{b}) - J]^2 = \frac{\rho^2(\boldsymbol{Q}, \boldsymbol{Q}) - 1}{U^2(1-\rho^2P^2)}, \tag{D 9}$$

where we have used the fact that $P^2 + (\boldsymbol{Q}, \hat{\boldsymbol{U}})^2 = (\boldsymbol{Q}, \boldsymbol{Q})$.

The easiest way to analyse (D 9) is to introduce an orthonormal basis $\hat{\boldsymbol{P}}_1, \ldots, \hat{\boldsymbol{P}}_{N-1}$ in $\mathscr{H}_0(\boldsymbol{u})$ with $\hat{\boldsymbol{P}}_1 = \hat{\boldsymbol{P}}$ and to write $\boldsymbol{b}$ as

$$\sum_{i=1}^{N-1} b_i \hat{\boldsymbol{P}}_i.$$

Then $(\hat{\boldsymbol{P}}, \boldsymbol{b}) = b_1$ and $(\boldsymbol{b}_\perp, \boldsymbol{b}_\perp) = b_2^2 + \ldots + b_{N-1}^2$. Evidently the surface in $\mathscr{H}_0(\boldsymbol{u})$ described by (D 9) contains no real points unless $\rho^2 \geqslant (\boldsymbol{Q}, \boldsymbol{Q})^{-1}$. If $\rho^2 = (\boldsymbol{Q}, \boldsymbol{Q})^{-1}$ then that 'surface' consists of the single point $J\hat{\boldsymbol{P}}$. If

$$\frac{1}{(\boldsymbol{Q}, \boldsymbol{Q})} < \rho^2 < \frac{1}{(\boldsymbol{P}, \boldsymbol{P})}, \tag{D 10}$$

then (D 9) describes an ellipsoid of revolution in $\mathscr{H}_0(\boldsymbol{u})$ with centre at $J\hat{\boldsymbol{P}}$ and axis parallel to $\hat{\boldsymbol{P}}$. If $\rho^2 > (\boldsymbol{P}, \boldsymbol{P})^{-1}$ then (D 9) describes a hyperboloid of revolution with two sheets, having its centre at $J\hat{\boldsymbol{P}}$ and its axis parallel to $\hat{\boldsymbol{P}}$.

We are already familiar with $(\boldsymbol{Q}, \boldsymbol{Q})$; it is $\boldsymbol{q} . \boldsymbol{E}^{-1} . \boldsymbol{q}$, or $\rho_{\min}^{-2}$. It remains to evaluate the axis $\boldsymbol{P}$ and $(\boldsymbol{P}, \boldsymbol{P})$. But $(\boldsymbol{P}, \boldsymbol{P}) = (\boldsymbol{Q}, \boldsymbol{Q}) - (\boldsymbol{Q}, \hat{\boldsymbol{U}})^2$ or

$$(\boldsymbol{P}, \boldsymbol{P}) = \boldsymbol{q} . \boldsymbol{E}^{-1} . \boldsymbol{q} - \frac{(\boldsymbol{u} . \boldsymbol{E}^{-1} . \boldsymbol{q})^2}{\boldsymbol{u} . \boldsymbol{E}^{-1} . \boldsymbol{u}}.$$

Thus $(\boldsymbol{P}, \boldsymbol{P}^{-1})$ is ρ_{par}^2, as given by (6.16).

The axis $\boldsymbol{P}$ is $\boldsymbol{Q} - (\boldsymbol{Q}, \hat{\boldsymbol{U}})\, \hat{\boldsymbol{U}}$, so

$$\boldsymbol{P} = \boldsymbol{E}^{-1} . \boldsymbol{q} - \frac{\boldsymbol{u} . \boldsymbol{E}^{-1} . \boldsymbol{q}}{\boldsymbol{u} . \boldsymbol{E}^{-1} . \boldsymbol{u}} \boldsymbol{E}^{-1} . \boldsymbol{u},$$

or $\boldsymbol{P} = (\boldsymbol{u} . \boldsymbol{E}^{-1} . \boldsymbol{q})\, (\boldsymbol{a}_{\mathscr{C}} - \boldsymbol{a}_E)$. The proof of theorem 2 is complete.

References

Anderson, D. L. & Archambeau, C. B. 1964 *J. Geophys. Res.* **69**, 2071–2084.
Asbel, I. Y., Keilis-Borok, V. I. & Yanovskaya, T. B. 1966 *Geophys. J. R. astr. Soc.* **11**, 25–55.
Backus, G. 1968 *Phil. Trans. Roy. Soc. Lond.* A **263**, 239–266.
Backus, G. & Gilbert, F. 1967 *Geophys. J. R. astr. Soc.* **13**, 247–276 (Inverse I).
Backus, G. & Gilbert, F. 1968 *Geophys. J. R. astr. Soc.* **16**, 169–205 (Inverse II).
Bartle, R. G. 1955 *Proc. Am. math. Soc.* **6**, 827–831.
Ben-Menahem, A. 1965 *J. Geophys. Res.* **70**, 4641–4651.
Bridgeman, P. 1963 *Dimensional analysis*. New Haven, Connecticut: Yale University Press.
Cramér, H. 1946 *Mathematical methods of statistics*. Princeton, New Jersey: Princeton University Press.
Dahlen, A. 1969 *Geophys. J. R. astr. Soc.* (in the Press).
Gibbs, J. W. 1901 *Vector analysis*. New Haven, Connecticut: Yale University Press.
Gilbert, F. & Backus, G. 1965 *Rev. Geophys.* **3**, 1–10.
Gilbert, F. & Backus, G. 1968 *Bull. Seism. Soc. Am.* **58**, 103–131.
Halmos, P. 1958 *Finite dimensional vector spaces*. Princeton, New Jersey: Van Nostrand.
Jeffreys, H. 1959 *The Earth*. Cambridge University Press.
Jeffreys, H. 1963 *Geophys. J. R. astr. Soc.* **8**, 196–202.
Keilis-Borok, V. I. & Yanovskaya, T. B. 1967 *Geophys. J. R. astr. Soc.* **13**, 223–234.
King-Hele, D. C., Cook, G. E. & Watson, H. M. 1964 *Nature, Lond.* **202**, 996.
Knopoff, L. 1965 *Rev. Geophys.* **2**, 625–660.
Levshin, A. L., Sabitova, T. M. & Valus, V. P. 1966 *Geophys. J. R. astr. Soc.* **11**, 57–66.
Parzen, E. 1962 *Stochastic processes*. San Francisco: Holden-Day.
Press, F. 1968 *Science, N.Y.* **160**, 1218–1221.
Slichter, L. B. 1967 *International dictionary of geophysics* (ed. S. K. Runcorn). New York: Pergamon.
Toksöz, M. N. & Ben-Menahem, A. 1963 *Bull. Seism. Soc. Am.* **53**, 741–764.

SOME PUBLICATIONS OF THE ROYAL SOCIETY

PROCEEDINGS OF THE ROYAL SOCIETY, SERIES A AND B

Per part (post extra)	£1 15*s*	(£1.75)	$4.55
Per volume (cloth bound) (post extra)	£8		$20.80
Subscription rates, payable in advance of publication			
Per volume in parts as published (post free)	£6 5*s*	(£6.25)	$16.00
Binding cases are available at	8*s*	(40p)	$1.05
Per volume cloth bound (post free)	£7		$18.00

BIOGRAPHICAL MEMOIRS OF FELLOWS OF THE ROYAL SOCIETY

Volume 15	£2 10*s*	(£2.50)	$6.50

NOTES AND RECORDS OF THE ROYAL SOCIETY

Volumes 15 to 25 *each volume*	£1 10*s*	(£1.50)	$4.00

YEAR BOOK OF THE ROYAL SOCIETY

1970	£1 1*s*	(£1.05)	$2.75

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY

SERIES A: MATHEMATICAL AND PHYSICAL SCIENCES
SERIES B: BIOLOGICAL SCIENCES

The price of each part in either series of the *Philosophical Transactions* is directly related to the number of pages and plates it contains. The price of a volume is the total of the prices of its constituent parts (which are published irregularly).

Volumes are closed when the total price of the parts reaches a minimum of £17 10*s* (£17.50) $45.50.

Subscription Rates

Subscriptions for complete volumes may be placed through booksellers or direct to The Royal Society, 6 Carlton House Terrace, London s.w.1.

The subscription rates to either series, *payable in advance of publication*, are:

Per volume in parts as issued (post free)	£14 5*s*	(£14.25)	$37.00
(Binding cases are available at	8*s*	(40p)	$1.05)
Per volume cloth bound (post free)	£15		$40.00

SYMBOLS, SIGNS, AND ABBREVIATIONS

RECOMMENDED FOR BRITISH SCIENTIFIC PUBLICATIONS

A REPORT BY
THE SYMBOLS COMMITTEE OF
THE ROYAL SOCIETY

1969

Since the publication in 1951 of the last Report of the Symbols Committee of the Royal Society considerable progress has been made towards international agreement on the names and symbols for physical quantities, on the definitions, names, and symbols for units, and on the rules for the expression of relations involving numbers between physical quantities and units.

The recommendations of the following international bodies, on each of which the U.K. is represented, are in virtually complete agreement wherever they overlap:

The General Conference on Weights and Measures
The International Organization for Standardization
The International Union of Pure and Applied Physics
The International Union of Pure and Applied Chemistry
The International Electrochemical Commission

The British Standard B.S. 1991, Part 1, second edition, 1967, is also in virtually complete agreement with the international recommendations.

This new version of the Report of the Symbols Committee thus recommends the procedures and symbols which have been internationally agreed.

Price 7*s.* (35p) per copy, or £6 per 25 copies from The Royal Society, 6 Carlton House Terrace, London, S.W.1, or through booksellers.

Printed in Great Britain
for the Royal Society at the University Printing House, Cambridge