Minimax Confidence Intervals in Geomagnetism

by

Philip B. Stark

Technical Report No. 292
February 1991

Department of Statistics
University of California
Berkeley, California 94720

Minimax Confidence Intervals in Geomagnetism

Philip B. Stark
Department of Statistics
University of California
Berkeley, CA 94720

February 26, 1991

Abstract

Backus [1] uses the prior information that the rest mass of Earth's magnetic field is less than the mass of Earth, or that the Ohmic heat liberated in the core by the currents giving rise to the main magnetic field is less than the surface heat flow, to compute the lengths of confidence intervals for low-degree Gauss coefficients of the magnetic field. His technique for producing confidence intervals yields intervals that are longer than necessary to guarantee the nominal coverage probability. The present paper uses theory of Donoho [2] to find lower bounds on the lengths of optimally short fixed-length confidence intervals (*minimax confidence intervals*) for Gauss coefficients of the field of degree $1 \leq l \leq 12$ using the heat flow constraint. The bounds on optimal minimax intervals are about 40% shorter than Backus' intervals: no procedure for producing fixed-length confidence intervals, linear or nonlinear, can give intervals shorter than about 60% the length of Backus' in this problem. Procedures that examine the data before determining the length of the interval can do arbitrarily better for some data sets.

Keywords: Inverse Problems, Confidence Intervals, Minimax Estimates, Confidence Set Inference, Core Magnetic Field, Modulus of Continuity.

Acknowledgements: I am very grateful to David L. Donoho for helpful conversations. This work was supported by NASA Grant NAG-883, NSF Presidential Young Investigator Award DMS-8957573, NSF Grant DMS-8810192 and donations from the AT&T Foundation and IBM.

1 Introduction

Consider the following model of a linear inverse problem. We observe an n-vector of data $\boldsymbol{\delta}$ related by the linear operator K to the unknown model $\mathbf{x}_0$, with additive noise $\boldsymbol{\epsilon}$:

$$\boldsymbol{\delta} = K\mathbf{x}_0 + \boldsymbol{\epsilon}, \tag{1}$$

where $\boldsymbol{\delta}, \boldsymbol{\epsilon} \in \mathbf{R}^n$, $\mathbf{x}_0 \in \mathsf{X} = l_2(w)$, the space of weighted square-summable sequences

$$l_2(w) \equiv \{\mathbf{x} = (x_1, x_2, \cdots) : \sum_{j=1}^{\infty} w_j x_j^2 < \infty\}, \tag{2}$$

$w_j > 0$, with norm

$$||\mathbf{x}|| \equiv \left(\sum_{j=1}^{\infty} w_j x_j^2 \right)^{1/2}; \tag{3}$$

the linear data mapping

$$\begin{aligned} K : \quad l_2 \rightarrow & \quad \mathbf{R}^n \\ \mathbf{x} \mapsto & \quad K\mathbf{x}; \end{aligned} \tag{4}$$

and the components of $\boldsymbol{\epsilon}$ are independent, identically distributed errors with zero mean and variance σ^2. We assume we know *a priori* that

$$\mathbf{x}_0 \in \mathsf{C} \equiv \{\mathbf{x} \in l_2(w) : ||\mathbf{x}|| \leq 1\}. \tag{5}$$

Let L be a linear functional

$$\begin{aligned} L : \quad l_2(w) \rightarrow & \quad \mathbf{R} \\ \mathbf{x} \mapsto & \quad L\mathbf{x}. \end{aligned} \tag{6}$$

We wish to use the data 1 and the prior information 5 to find a $1 - \alpha$ confidence interval for $L\mathbf{x}_0$. We refer to the definitions and relations 1-6 as the problem $\mathcal{P}$.

Backus [1] develops a procedure for producing confidence intervals for this problem. Donoho [2] develops theory for finding the lengths of optimally short confidence intervals in a class larger than that considered by Backus, for the case the errors ϵ_j are iid Gaussian. The present paper is bounds the improvement one could hope to obtain over Backus' result via nonlinear methods, in the geomagnetic problem of estimating low-degree Gauss coefficients of the magnetic field from satellite observations, using the constraint that the rate of heat production by currents in the core is less than the surface heat flow rate.

Table 1 gives the lengths of confidence intervals for Gauss coefficients of Earth's magnetic field derived for this problem by Backus [1], and bounds on optimally short nonlinearly based confidence intervals computed using Donoho's theory [2]. The bounds on optimal intervals are about 59-62% of the length of Backus' intervals.

Section 2 of the paper outlines Backus' technique and Donoho's minimax results. Section 3 applies Donoho's results to the geomagnetic problem, giving the approximations necessary to make direct comparisons with Backus' confidence intervals. Section 4 concludes with a short discussion of variable-length confidence intervals.

2 Confidence Intervals for Linear Functionals

There are two principal types of procedures for constructing confidence intervals. In the first type, "fixed-length," the length of the confidence interval is determined without reference to the measurements. The measurements are then used to determine where to center the interval. This is the most common sort of procedure for constructing confidence intervals.

In the second type, "variable-length," both the location and length of the confidence interval depend on the observed data. This sort is less common (at least among frequentists: Bayesian confidence intervals are often in this class). It is discussed in the context of inverse problems by Stark [6], and a brief indication of the improvement one might obtain over fixed-length procedures is given in section 4 below. The present paper primarily concerns fixed-length procedures.

Among fixed-length procedures, the dependence of the center of the interval on the measured data may be linear or nonlinear. If we are clever at centering the interval, we may use a shorter interval and still have the nominal coverage probability. This leads us to the questions of what procedures allow us to use the shortest intervals, and what are their lengths?

The prior information $\mathbf{x}_0 \in \mathsf{C}$ is obviously key in constructing short intervals. Let $\hat{L}(\boldsymbol{\delta})$ be a procedure for determining the center of a confidence interval for $L\mathbf{x}_0$ from data $\boldsymbol{\delta}$. The minimum fixed length of a $1-\alpha$ confidence interval for $L\mathbf{x}_0$ centered at $\hat{L}(\boldsymbol{\delta})$, valid whatever be $\mathbf{x}_0 \in \mathsf{C}$ is

$$C_\alpha(\hat{L}, \sigma) \equiv 2 \inf \left\{ \chi : Prob\{|\hat{L}(\boldsymbol{\delta}) - L\mathbf{x}| \leq \chi\} \geq 1 - \alpha, \forall \mathbf{x} \in \mathsf{C} \right\}. \tag{7}$$

Backus [1] considers a finite set of linear functionals $\{\hat{L}_k\}_{k=0}^N$ at which to center confidence intervals, and finds $C_\alpha(\hat{L}_k, \sigma)$. He then picks as his procedure that with smallest C_α. For the case the errors are Gaussian, Donoho [2] shows how to construct the affine functional $\hat{L}_A$ with smallest C_α among all affine functionals (the *affine minimax procedure*), and gives exact expressions for the resulting confidence interval length. Further, he finds upper and lower bounds on the length of the shortest fixed-length confidence interval based on nonlinear procedures (the *nonlinear minimax* confidence interval length). Below we use Donoho's theory to bound the improvement one could hope to obtain over Backus' results in the most favorable circumstances.

2.1 Backus' Procedure

We shall state Backus' method for constructing confidence intervals in slightly more detail, to compare and contrast with minimax procedures. Let $\mathbf{N}(K)$ denote the null-space of the linear function K, and let $\mathbf{D}(K) \equiv \mathbf{N}^{\perp}(K)$ be its orthogonal complement in X. Since K has finite-dimensional range, $\mathbf{D}(K)$ is closed, and any $\mathbf{x} \in \mathsf{X}$ can be written uniquely as the sum

$$\mathbf{x} = \mathbf{x}_{\mathbf{N}} + \mathbf{x}_{\mathbf{D}}$$

where $\mathbf{x}_{\mathbf{N}} \in \mathbf{N}(K)$ and $\mathbf{x}_{\mathbf{D}} \in \mathbf{D}(K)$; *i.e.*

$$\mathsf{X} = \mathbf{N}(K) \oplus \mathbf{D}(K).$$

We assume from here on that $n = dim\{\mathbf{D}(K)\}$. Consider the restriction $K_{\mathbf{D}}$ of K to $\mathbf{D}(K)$. The linear function $K_{\mathbf{D}} : \mathbf{D} \to \mathbf{R}^n$ is one-to-one, and hence invertable. Let $\{\mathbf{D}_0, \mathbf{D}_1, \mathbf{D}_2, \cdots \mathbf{D}_n\}$ be a nested sequence of subspaces of $\mathbf{D}(K)$ such that:

$$\begin{aligned} \dim\{\mathbf{D}_k\} &= k && (8)\\ \mathbf{D}_k &\subset \mathbf{D}_{k+1} && (9)\\ \mathbf{D}_0 &= \{0\} && (10)\\ \mathbf{D}_n &= \mathbf{D}(K); && (11) \end{aligned}$$

and let $K_{\mathbf{D}_k}$ be the restriction of K to $\mathbf{D}_k$. Let $P_{\mathbf{D}_k}$ be the orthogonal projection operator from X to $\mathbf{D}_k$. Let $\mathbf{R}_k$ be the range of $K_{\mathbf{D}_k}$. Then $\mathbf{R}_k$ is isomorphic to $\mathbf{R}^k$. Let $P_{\mathbf{R}_k}$ be the orthogonal projection operator from $\mathbf{R}^n$ to $\mathbf{R}_k$. Backus considers the following set of estimators of $L[\mathbf{x}_0]$:

$$\hat{L}_k(\boldsymbol{\delta}) \equiv L K_{\mathbf{D}_k}^{-1} P_{\mathbf{R}_k} \boldsymbol{\delta}. \tag{12}$$

(It follows from the invertability of $K_{\mathbf{D}}$ on $\mathbf{R}^n$ and the definitions that $K_{\mathbf{D}_k}$ is invertable on $P_{\mathbf{R}_k}\mathbf{R}^n$ for a unique element of $\mathbf{D}_k$.) Note that $\hat{L}_k(\boldsymbol{\delta})$ is simply a linear functional of the data $\boldsymbol{\delta}$ (the composition of two linear functions and a linear functional), and so can be written

$$\hat{L}_k(\boldsymbol{\delta}) = \mathbf{c}(k) \cdot \boldsymbol{\delta}, \tag{13}$$

where $\mathbf{c}(k) = (c_j(k))_{j=1}^n \in \mathbf{R}^n$.

Let us now compare the estimator 12 with $L\mathbf{x}_0$:

$$\begin{aligned} L\mathbf{x}_0 - \hat{L}_k(\boldsymbol{\delta}) &= L\mathbf{x}_0 - L K_{\mathbf{D}_k}^{-1} P_{\mathbf{R}_k}(K\mathbf{x}_0 + \boldsymbol{\epsilon})\\ &= L\mathbf{x}_0 - L P_{\mathbf{D}_k}\mathbf{x}_0 - L K_{\mathbf{D}_k}^{-1} P_{\mathbf{R}_k}\boldsymbol{\epsilon}\\ &= L P_{\mathbf{D}_k^{\perp}}\mathbf{x}_0 - L K_{\mathbf{D}_k}^{-1} P_{\mathbf{R}_k}\boldsymbol{\epsilon}\\ &= L P_{\mathbf{D}_k^{\perp}}\mathbf{x}_0 - \mathbf{c}(k)\cdot\boldsymbol{\epsilon}, \end{aligned} \tag{14}$$

where $P_{\mathbf{D}_k^\perp}$ is the orthogonal projection operator onto the orthogonal complement $\mathbf{D}_k^\perp$ of $\mathbf{D}_k$. Since the error vector ϵ is assumed to have zero mean, the first term in 14 is the bias of the estimator; the second is a random variable with zero mean. The distribution of the second term can be deduced from the distribution of the error vector ϵ, since it is merely a linear functional of ϵ. Thus one can compute a number γ_k such that

$$Prob\{|\mathbf{c}(k) \cdot \epsilon| \leq \gamma_k\} \geq 1 - \alpha. \tag{15}$$

The bias term can be bounded using the prior information that $||\mathbf{x}_0|| \leq 1$:

$$\begin{aligned} |LP_{\mathbf{D}_k^\perp}\mathbf{x}_0| &\leq ||LP_{\mathbf{D}_k^\perp}||||\mathbf{x}_0|| \\ &\leq ||LP_{\mathbf{D}_k^\perp}|| \\ &= ||L_{\mathbf{D}_k^\perp}||, \end{aligned} \tag{16}$$

where $L_{\mathbf{D}_k^\perp}$ is the restriction of L to the orthogonal complement of $\mathbf{D}_k$. Thus the following intervals are $1 - \alpha$ confidence intervals for $L\mathbf{x}_0$:

$$[\hat{L}_k(\delta) - ||L_{\mathbf{D}_k^\perp}|| - \gamma_k, \hat{L}_k(\delta) + ||L_{\mathbf{D}_k^\perp}|| + \gamma_k].$$

These intervals have lengths

$$C_{\alpha,k}(\sigma) = 2(\gamma_k + ||L_{\mathbf{D}_k^\perp}||). \tag{17}$$

Backus then takes as the confidence interval the one with shortest length.

The point here is that Backus picks the best of but $n + 1$ linear estimators $\hat{L}_k$ of $L\mathbf{x}_0$ at which to center his confidence interval, where "best" means that giving the shortest interval. Instead of only the estimators $\hat{L}_k$, we could consider *all* linear estimators; all *affine* estimators of Lx_0 of the form

$$\hat{L}_A(\delta) \equiv c_0 + \mathbf{c} \cdot \delta,$$

where $\mathbf{c}$ is the n-vector $(c_1, \cdots c_n)$; or even all *nonlinear* estimators $\hat{L}_N$ with arbitrary (measurable) dependence on the data; then find the best estimators in this wider classes. Donoho's results [2] bound the length of the shortest confidence intervals based on general affine and nonlinear estimates.

2.2 Donoho's Minimax Results

Donoho [2] shows how to find the length of the shortest affine-based confidence interval for a large class of inverse problems. He shows further that the length of the shortest confidence interval based on *nonlinear* functions of the data is bounded by a fraction of the length of the shortest affine-derived interval.

A few definitions are required in order to proceed. The length of the *minimax affine confidence interval* is

$$C^*_{\alpha,A}(\sigma) \equiv 2\inf\left\{\chi : \exists \hat{L}_A \text{ affine } \ni Prob\{|\hat{L}_A(\delta) - L\mathbf{x}| \le \chi\} \ge 1-\alpha,\ \forall \mathbf{x} \in \mathsf{C}\right\}. \tag{18}$$

The length of the *minimax nonlinear confidence interval* is

$$C^*_{\alpha,N}(\sigma) \equiv 2\inf\left\{\chi : \exists \hat{L}_N \ni Prob\{|\hat{L}_N(\delta) - L\mathbf{x}| \le \chi\} \ge 1-\alpha,\ \forall \mathbf{x} \in \mathsf{C}\right\}. \tag{19}$$

(Note that these definitions differ from those of Donoho by a factor of two: he gives the half-lengths of the intervals.)

Donoho's results for the affine minimax confidence interval and his bounds on nonlinear minimax confidence interval lengths, $C^*_{\alpha,A}(\sigma)$ and $C^*_{\alpha,N}(\sigma)$, involve the *modulus of continuity* $\omega(\rho, L, K, \mathsf{C})$ of the functional L with respect to the data mapping K and the prior information set C:

$$\omega(\rho, L, K, \mathsf{C}) \equiv \sup\left\{|L\mathbf{x}_1 - L\mathbf{x}_{-1}| : \|K(\mathbf{x}_1 - \mathbf{x}_{-1})\| \le \rho \text{ and } \mathbf{x}_i \in \mathsf{C}\right\}. \tag{20}$$

The modulus measures how much the functional L can vary among members of the *a priori* class C if their predicted data differ by no more than ρ in the norm. The arguments L, K, and C will be suppressed in the sequel. Donoho shows that

$$2\omega(2\mathcal{Z}_{1-\alpha}\sigma) \le C^*_{\alpha,N}(\sigma) \le C^*_{\alpha,A}(\sigma) \le 2\omega(2\mathcal{Z}_{1-\alpha/2}\sigma) \tag{21}$$

$$C^*_{\alpha,A}(\sigma) \le \frac{\mathcal{Z}_{1-\alpha/2}}{\mathcal{Z}_{1-\alpha}} C^*_{\alpha,N}(\sigma), \tag{22}$$

and, if $\omega(\rho) \to 0$ as $\rho \to 0$,

$$C^*_{\alpha,A}(\sigma) = 2\sup_{\rho \ge 0}\left(\frac{\omega(\rho)}{\rho}\right)\chi_{\alpha,A}(\rho/2, \sigma). \tag{23}$$

(See also Donoho and Low [3] for further discussion and applications.) Here $\mathcal{Z}_\alpha$ is the 100α percentage point of the standard normal distribution, and $\chi_{\alpha,A}$ is the length of the minimax affine confidence interval for a bounded normal mean:

Suppose we wish to estimate θ from the single observation $Y = \theta + Z$ where θ is known *a priori* to lie in $[-\tau, \tau]$, and Z is a zero-mean Gaussian random variable with variance σ^2. The half-length of the minimax affine confidence interval for θ is

$$\chi_{\alpha,A}(\tau, \sigma) \equiv \inf\left\{\chi : \sup_c Prob\{|cY - \theta| \le \chi\} \ge 1-\alpha \ \ \forall \theta \in [-\tau, \tau]\right\}. \tag{24}$$

That is, $\chi_{\alpha,A}(\tau, \sigma)$ is the smallest number χ such that for some $c \in \mathbf{R}$ the interval $[cY - \chi, cY + \chi]$ is a valid $1-\alpha$ confidence interval for θ, whatever be

$\theta \in [-\tau, \tau]$. Conservative and optimistic values of $\chi_{\alpha,A}(\tau, 1)$ are given by Stark [5].

The equation 23 is not useful for the present problem as the functional L that evaluates a single Gauss coefficient has a component in the null-space of K whenever the number of data n is finite; hence $\omega(\rho)$ does not go to zero as ρ does.

To use Donoho's results to compute or bound the lengths of the minimax affine or nonlinear confidence intervals, we must evaluate the modulus. We do this for the problem of inferring Gauss coefficients of the geomagnetic field in the next section.

3 Geomagnetism

3.1 The Geomagnetic Problem

Following Backus [1], we represent the magnetic field $\mathbf{B}$ outside Earth's core due to currents within the core as the gradient of a scalar field Ψ:

$$\mathbf{B} = -\nabla\Psi,$$

where Ψ has the spherical harmonic expansion

$$\Psi(\mathbf{r}) = a\sum_{l=1}^{\infty}(a/r)^{l+1}\sum_{m=-l}^{l} x_l^m(a)Y_l^m(\hat{\mathbf{r}}). \tag{25}$$

In this expansion, a is the core radius, $\mathbf{r}$ is the position vector with origin at Earth's center, r is the Euclidean length of $\mathbf{r}$, $\hat{\mathbf{r}}$ is the unit vector in the direction $\mathbf{r}$, and Y_l^m are spherical harmonics normalized so that

$$\frac{1}{4\pi}\int_{r=a}|Y_l^m|^2 d^2\mathbf{r} = (2l+1)^{-1}. \tag{26}$$

In the previous equation, the core-mantle boundary is idealized as a sphere of radius a. We will define all operations on magnetic field models in terms of such spherical harmonic expansions. Essentially, the model space for the inverse problem is a weighted space $l_2(w)$ of coefficients x_l^m in these expansions. The norms considered here have weights w_l that depend only on l and not m—they are invariant under rotations of the field $\mathbf{B}$ in $\mathbf{R}^3$. We will denote the magnetic field of the model whose Gauss coefficients are $\{x_l^m\}$ by $\mathbf{B}_x$.

The weights w_l we shall use are induced by either the "energy bound" that the rest mass of the energy of $\mathbf{B}$ is less than the total mass of Earth (equation 28, see [1]), or the "heat flow bound" that the rate of Ohmic dissipation in the core is less than the Earth's rate of surface heat flow (equation 29, see [1]). In terms of the coefficients x_l^m, this implies that

$$\sum_{l=1}^{\infty} w_l \sum_{m=-l}^{l} |x_l^m(a)|^2 \le q, \tag{27}$$

where for the energy bound

$$w_l = (2l+1)(l+1)^{-1} \tag{28}$$

and $q = 2 \times 10^{33} nT^2$; or for the heat flow bound

$$w_l = (l+1)(2l+1)(2l+3)l^{-1} \tag{29}$$

and $q = 3 \times 10^{17} nT^2$, when the units of x_l^m are nanoTesla. Let $\mathbf{x}$ and $\mathbf{y}$ be two field models, with vectors of coefficients $\{x_l^m\}$ and $\{y_l^m\}$, respectively. We define the inner product of the two models to be

$$\mathbf{x} \cdot \mathbf{y} \equiv q^{-1} \sum_{l=1}^{\infty} w_l \sum_{m=-l}^{l} x_l^m y_l^m. \tag{30}$$

With the induced norm

$$||\mathbf{x}|| = \left(q^{-1} \sum_{l=1}^{\infty} w_l \sum_{m=-l}^{l} |x_l^m|^2 \right)^{1/2},$$

the constraint that the rest mass of the energy of $\mathbf{B}$ is less than the mass of Earth, or that the Ohmic heat production is less than the surface heat flow take the desired form

$$\mathbf{x} \in \mathsf{C} \equiv \{\mathbf{x} \in l_2(w) : ||\mathbf{x}|| \le 1\}.$$

We are interested in estimating the linear functional L that evaluates a single Gauss coefficient x_l^m. From the definition of the inner product 30 we may deduce the form of L:

$$L = q w_l^{-1} \nabla Y_l^m. \tag{31}$$

Then we have

$$Lx = q^{-1} \sum_{l'=1}^{\infty} w_{l'} \sum_{m'=-l'}^{l'} q w_l^{-1} \delta_l^{l'} \delta_m^{m'} x_{l'}^{m'} \tag{32}$$

$$= x_l^m. \tag{33}$$

(Here $\delta_j^k \equiv \{1, j = k; \;\; 0, j \ne k\}$.) Backus [1] shows that these functionals L are bounded in the norms induced by the energy bound and by the heat flow bound; in fact, by inspection we find that

$$||L||^2 = q/w_l. \tag{34}$$

The data will be assumed to be measurements of the three Cartesian components of $\mathbf{B}$ at locations $\{\mathbf{r}_j\}_{j=1}^{N/3}$ on or above Earth's surface $r = c$. We have (from Backus eq. 9.9)

$$\mathbf{B}_x(\mathbf{r}_j) = \sum_{l=1}^{\infty}(a/r_j)^{l+2} \sum_{m=-l}^{l} x_l^m \nabla[r^{-l-1}Y_l^m(\hat{\mathbf{r}_j})]_{r=1}. \tag{35}$$

The observational errors ϵ consist of random errors of measurement, plus contributions to the magnetic field from the crust, mantle, ionosphere and magnetosphere. We will ignore all these except the crust, which we shall treat as independent, identically distributed Gaussian random errors, statistically independent of the measurement errors, following Backus' "most favorable treatment" in equation 9.15 of [1]. Backus takes the measurement errors to be Gaussian with zero mean and standard deviation $\sigma_m = 6nT$, and the crustal signal to contribute independent Gaussian errors with zero mean and standard deviation $\sigma_c = 12nT$. The combined errors are then Gaussian with zero mean and standard deviation

$$\sigma = \sqrt{\sigma_m^2 + \sigma_c^2} = 13.416. \tag{36}$$

If we make the usual correspondence between sets of spherical harmonic coefficients $(x_l^m)_{l=1}^{\infty}{}_{m=-l}^{l}$ and singly-indexed sets of coefficients $(x_j)_{j=1}^{\infty}$, then, with K defined implicitly through 35, the problem of estimating a Gauss coefficient of the field from surface and satellite observations of $\mathbf{B}$ takes the form of the original estimation problem $\mathcal{P}$ of the introduction. To apply the theory of Donoho, we must evaluate the modulus of continuity of L.

3.2 Modulus of Continuity of a Gauss Coefficient

This subsection develops the expressions needed to compute the modulus of continuity of the functional that evaluates a Gauss coefficient of the magnetic field at the core-mantle boundary. Let L stand for the linear functional that evaluates a particular Gauss coefficient x_l^m at the CMB, as defined in equation 31. We have by the definition 20

$$\begin{aligned}\omega(\rho) &\equiv \sup\{|L\mathbf{x}_1 - L\mathbf{x}_{-1}| : \|K(\mathbf{x}_1 - \mathbf{x}_{-1})\| \le \rho \text{ and } \mathbf{x}_i \in \mathsf{C}\} \\ &= \sup\{L\mathbf{x} : \|K\mathbf{x}\| \le \rho \text{ and } \|\mathbf{x}\| \le 2\}, \end{aligned} \tag{37}$$

where we have used the linearity of L and K, the symmetry of C, and the fact that

$$\mathsf{C} \setminus \mathsf{C} \equiv \{\mathbf{x} = \mathbf{x}_1 - \mathbf{x}_{-1} : \|\mathbf{x}_1\|, \|\mathbf{x}_{-1}\| \le 1\} \tag{38}$$

$$= \{\mathbf{x} : \|\mathbf{x}\| \le 2\}. \tag{39}$$

Let $\boldsymbol{\lambda}$ denote the ordered pair (λ_1, λ_2) and let

$$P = \{\boldsymbol{\lambda} = (\lambda_1, \lambda_2) : \lambda_1, \lambda_2 \ge 0\} = \{\boldsymbol{\lambda} \ge 0\}$$

denote the positive cone in $\mathbf{R}^2$. Using Lagrangian duality [4], and the fact that $\sup L\mathbf{x} = -\inf\{-L\mathbf{x}\}$, we may rewrite the optimization problem 37 as follows:

$$\omega(\rho) \leq -\sup_{\boldsymbol{\lambda}\geq 0} \inf_{\mathbf{x}\in l_2(w)} \{-L\mathbf{x} + \lambda_1(||\mathbf{x}||^2 - 4) + \lambda_2(||K\mathbf{x}||^2 - \rho^2)\}. \tag{40}$$

The conditions for equality here are

1. There exists an $\mathbf{x} \in l_2(w)$ such that

$$||\mathbf{x}||^2 < 4 \text{ and } ||Kx||^2 < \rho^2,$$

   ($\mathbf{x} = 0$ will do),

2. the functional

$$\begin{aligned} G : l_2(w) &\rightarrow \mathbf{R}^2 \\ \mathbf{x} &\mapsto (||\mathbf{x}||^2 - 4, ||K\mathbf{x}||^2 - \rho^2) \end{aligned}$$

   is convex with respect to the partial order on $\mathbf{R}^2$ induced by the positive cone P (this follows from the convexity of norms), and

3. $\omega(\rho) < \infty$, which follows from the fact that L is a bounded linear functional, and that we are constrained to the ball $||\mathbf{x}|| \leq 2$.

Thus equality obtains in 40.

Define

$$F[\mathbf{x}; \boldsymbol{\lambda}] \equiv -L\mathbf{x} + \lambda_1(||\mathbf{x}||^2 - 4) + \lambda_2(||K\mathbf{x}||^2 - \rho^2). \tag{41}$$

Then

$$\omega(\rho) = -\sup_{\boldsymbol{\lambda}\geq 0} \inf_{\mathbf{x}\in l_2(w)} F[\mathbf{x}; \boldsymbol{\lambda}]. \tag{42}$$

For fixed $\boldsymbol{\lambda}$, F is Fréchet differentiable with respect to $\mathbf{x}$, and so, since F is convex, necessary and sufficient conditions for the unconstrained minimum of $F[\cdot; \boldsymbol{\lambda}]$ are given by $\partial_{\mathbf{x}} F = 0$. Now

$$\partial_{\mathbf{x}} F = -L + 2\lambda_1 \mathbf{x} + 2\lambda_2 K^T K\mathbf{x}, \tag{43}$$

so the optimal $\mathbf{x}$, $\mathbf{x}^*$ satisfies

$$2(\lambda_1 I + \lambda_2 K^T K)\mathbf{x}^* = L, \tag{44}$$

where I is the identity operator on $l_2(w)$.

At this point we begin to make approximations to have results directly comparable with those of Backus [1]. In order to find the length of confidence intervals without reference to the actual locations at which observations were made, Backus takes the observations to be approximately evenly distributed at $n/3$ points on a sphere 400km larger than Earth (the approximate altitude of the MAGSAT satellite). If the observations are sufficiently numerous, the inner product in data space is well-approximated by an integral over the sphere. Now we shall use the felicitous fact that L is the functional that evaluates a single Gauss coefficient x_l^m. Backus [1] shows that if we restrict K and L to a finite-dimensional subspace spanned by spherical harmonics of degree at most $l_{max} > l$ where l_{max} is much less than the square-root of the number of data n, then

$$K^T K L \approx \phi_l L, \tag{45}$$

where

$$\phi_l \equiv \sigma^{-2} qn/3(a/c)^{2(l+2)}(l+1)w_l^{-1}, \tag{46}$$

and q and w_l are given by equations 27 or 29, depending upon the prior information norm we are using.

This will not hold in the case we consider here where l_{max} is infinite; however, we shall see presently that the effect of the approximation is to make the confidence intervals *shorter*, so the results are still a lower bound on the shortest valid confidence intervals, and bound the improvement one can hope to obtain through methods more sophisticated than that of Backus [1]—at least for methods that pick the length of the interval prior to observing the data (see section 4). We have from 45,

$$(\lambda_1 I + \lambda_2 K^T K)L \approx (\lambda_1 + \lambda_2 \phi_l)L, \tag{47}$$

and hence

$$2(\lambda_1 + \lambda_2 K^T K)\frac{L}{2(\lambda_1 + \lambda_2 \phi_l)} \approx L. \tag{48}$$

Let

$$\tilde{\mathbf{x}}(\boldsymbol{\lambda}) = \tilde{\mathbf{x}} \equiv \frac{L}{2(\lambda_1 + \lambda_2 \phi_l)}. \tag{49}$$

What is the effect of approximating $\mathbf{x}^*$ by $\tilde{\mathbf{x}}$? Since $F[\tilde{\mathbf{x}};\boldsymbol{\lambda}] \geq F[\mathbf{x}^*;\boldsymbol{\lambda}]$, we have

$$\sup_{\boldsymbol{\lambda}} F[\tilde{\mathbf{x}};\boldsymbol{\lambda}] \geq \sup_{\boldsymbol{\lambda}} F[\mathbf{x}^*;\boldsymbol{\lambda}], \tag{50}$$

and

$$-\sup_{\boldsymbol{\lambda}} F[\tilde{\mathbf{x}};\boldsymbol{\lambda}] \leq -\sup_{\boldsymbol{\lambda}} F[\mathbf{x}^*;\boldsymbol{\lambda}] = \omega(\rho). \tag{51}$$

Thus the approximate modulus will be too small, and the bound on the length of the minimax confidence interval 21 using the approximate modulus will still lower-bound the length of the nonlinear minimax confidence interval, as desired.

Let $\gamma \equiv \frac{1}{2}(\lambda_1 + \lambda_2 \phi_l)^{-1}$. Then $\tilde{\mathbf{x}} = \gamma L$, and

$$L\tilde{\mathbf{x}} = \gamma ||L||^2, \tag{52}$$

$$||\tilde{\mathbf{x}}||^2 = \gamma^2 ||L||^2, \tag{53}$$

and

$$\|K\tilde{\mathbf{x}}\|^2 = \gamma^2\phi_l\|L\|^2. \tag{54}$$

Combining these and substituting into 41 we have

$$\begin{aligned} F[\tilde{\mathbf{x}};\boldsymbol{\lambda}] &= -\gamma\|L\|^2 + \lambda_1\left(\gamma^2\|L\|^2 - 4\right) + \\ &\quad \lambda_2\left(\gamma^2\phi_l\|L\|^2 - \rho^2\right) \\ &= -\frac{\|L\|^2}{4}(\lambda_1 + \lambda_2\phi_l)^{-1} - 4\lambda_1 - \rho^2\lambda_2. \end{aligned} \tag{55}$$

Define

$$f(\boldsymbol{\lambda}) \equiv F[\tilde{\mathbf{x}}(\boldsymbol{\lambda});\boldsymbol{\lambda}]. \tag{56}$$

(It follows from Lagrangian Duality that f is concave in $\boldsymbol{\lambda}$.) The modulus is then bounded by

$$\begin{aligned} \omega(\rho) &\geq \min_{\boldsymbol{\lambda}\geq 0} -f(\boldsymbol{\lambda}) \\ &= \min_{\boldsymbol{\lambda}\geq 0}\left\{\frac{q}{4w_l}(\lambda_1 + \lambda_2\phi_l)^{-1} + 4\lambda_1 + \rho^2\lambda_2\right\}. \end{aligned} \tag{57}$$

This maximization was performed numerically using Stanford Systems Optimization Laboratory code NPSOL, given the analytic derivatives

$$\partial_{\lambda_1}(-f) = -\frac{q}{4w_l}(\lambda_1 + \lambda_2\phi_l)^{-2} + 4 \tag{58}$$

$$\partial_{\lambda_2}(-f) = -\phi_l\frac{q}{4w_l}(\lambda_1 + \lambda_2\phi_l)^{-2} + \rho^2. \tag{59}$$

Note that f has no stationary point unless $\phi_l/\rho^2 = 4$, so the maximum occurs at infinity or on the boundary of the positive cone $\boldsymbol{\lambda} \geq 0$. By inspection of f, the minimum does not occur at $\boldsymbol{\lambda} = 0$; also, for $\boldsymbol{\lambda} \geq 0$, $-f$ grows for sufficiently large $\|\boldsymbol{\lambda}\|$, so the minimum occurs on the boundary where exactly one of λ_1, λ_2 is zero. For sufficiently small ρ, the minimum has $\lambda_1 = 0$; eventually, as ρ grows, the minimum has $\lambda_2 = 0$. This is an artifact of the approximation 49, since $\tilde{\mathbf{x}}$ is an eigenfunction of K^TK— if $\mathbf{x}^*$ were found exactly using the actual positions at which measurements were made, in general the minimum would lie in the interior of the cone. Another consequence of this approximation is that intervals based on the heat flow bound and energy bound have the same length up to some critical value of ρ: for small ρ only the data constraint is active.

The numerical minimization of $-f$ directly bounds $C^*_{\alpha,N}$ via equation 21. Table 1 compares bounds on the half-lengths of nonlinear minimax confidence intervals with Backus' [1] results in microtesla for the heat flow bound, with the total error standard deviation $\sigma = 13.416$, $n = 26500$, $\alpha = 0.0001$, $1 \leq l \leq 12$. The bounds on optimal intervals are all about 60% of the lengths of Backus' intervals.

4 Discussion

We have seen that no fixed-length method for finding confidence intervals for Gauss coefficients of degree $1 \leq l \leq 12$ can produce confidence intervals shorter than about 60% of the length of Backus' intervals, and still guarantee 99.99% coverage probability for all models satisfying the heat flow bound. What about variable-length methods that use the data not only to center the confidence interval, but also to determine its length?

It can be shown [6] that variable-length procedures exist that do no worse than the best affine fixed-length procedure in the worst case, and do arbitrarily better than both the affine minimax and nonlinear minimax procedures for some data sets. Here we sketch why that is so.

Suppose first of all that both

$$L^+ \equiv \sup_{\mathbf{x}\in\mathsf{C}} Lx < \infty \tag{60}$$

and

$$L^- \equiv \inf_{\mathbf{x}\in\mathsf{C}} Lx > -\infty, \tag{61}$$

as they are in the geomagnetic problem since L is a bounded linear functional, and C is the unit ball. Define the interval

$$J \equiv [L^-, L^+].$$

Then with probability 1, $J \ni L\mathbf{x}_0$. For any interval $I \subset \mathbf{R}$, let $|I|$ denote the length of I. Let $I(\boldsymbol{\delta})$ be any procedure for producing a $1-\alpha$ confidence interval for the problem $\mathcal{P}$.

Let $I_A^*(\boldsymbol{\delta})$ be the affine minimax procedure, with corresponding center

$$\hat{L}_A^*(\boldsymbol{\delta}) = c_0^* + \mathbf{c}^* \cdot \boldsymbol{\delta} \tag{62}$$

and length

$$|I_A^*| = C_{\alpha,A}^*,$$

(*i.e.*

$$I_A^*(\boldsymbol{\delta}) \equiv [\hat{L}_A^*(\boldsymbol{\delta}) - C_{\alpha,A}^*/2, \hat{L}_A^*(\boldsymbol{\delta}) + C_{\alpha,A}^*/2])$$

and let $I_N^*(\boldsymbol{\delta})$ be a nonlinear minimax procedure with length

$$|I_N^*| = C_{\alpha,N}^*.$$

Since the errors ϵ_j are iid Gaussian, $||\epsilon||$ has positive probability of being arbitrarily large; hence, by 62 and the symmetry of the n-dimensional Gaussian distribution,

$$Prob\{\hat{L}_A^*(\boldsymbol{\delta}) > \beta\} > 0, \ \ \forall \beta \in \mathbf{R}. \tag{63}$$

In particular,

$$Prob\{\hat{L}_A^* > L^+ + C_{\alpha,A}^*/2\} > 0. \tag{64}$$

Denote by $D_{1/2}$ the set of $\boldsymbol{\delta} \in \mathbf{R}^n$ for which $\hat{L}_A^* > L^+ + C_{\alpha,A}^*/2$. Since $J \ni L\mathbf{x}_0$ and $Prob\{I_A^* \ni L\mathbf{x}_0\} \geq 1-\alpha$,

$$Prob\{I_A^*(\boldsymbol{\delta}) \cap J \ni L\mathbf{x}_0\} \geq 1-\alpha : \tag{65}$$

the variable-length procedure $I_A^*(\boldsymbol{\delta}) \cap J$ has coverage probability at least $1-\alpha$. The length of $I_A^*(\boldsymbol{\delta}) \cap J$ obviously never exceeds that of $I_A^*(\boldsymbol{\delta})$. In fact, for $\boldsymbol{\delta} \in D_{1/2}$,

$$|I_A^*(\boldsymbol{\delta}) \cap J| \leq 1/2|I_A^*(\boldsymbol{\delta})|. \tag{66}$$

Similarly for any $\gamma > 0$ there are sets D_γ such that $Prob\{\boldsymbol{\delta} \in D_\gamma\} > 0$ and whenever $\boldsymbol{\delta} \in D_\gamma$

$$|I_A^*(\boldsymbol{\delta}) \cap J| \leq \gamma|I_A^*(\boldsymbol{\delta})|. \tag{67}$$

As Donoho [2] shows (inequality 22 above) the length of the nonlinear minimax interval is never less than the constant fraction $\mathcal{Z}_{1-\alpha}/\mathcal{Z}_{1-\alpha/2}$ of the length of the affine minimax confidence interval length; thus with strictly positive probability the variable-length procedure $I_A^*(\boldsymbol{\delta}) \cap J$ produces intervals arbitrarily shorter than the the nonlinear minimax length.

For variable-length procedures, we need a new criterion for optimality. For example, for any procedure $I(\boldsymbol{\delta})$ that produces $1-\alpha$ confidence intervals for the problem $\mathcal{P}$, valid whatever be $\mathbf{x}_0 \in \mathsf{C}$, define

$$R(I(\boldsymbol{\delta})) \equiv \sup_{\mathbf{x}_0 \in \mathsf{C}} E|I(\boldsymbol{\delta})|, \tag{68}$$

where E is the expectation operator. This functional is one possibility. Let us say that $I^*(\boldsymbol{\delta})$ is R-optimal for the problem $\mathcal{P}$ if

$$R(I^*) = \inf_{\text{procedures } I(\boldsymbol{\delta})} R(I(\boldsymbol{\delta})). \tag{69}$$

Do there exist variable-length procedures that are R-optimal and have lengths never exceeding $C_{\alpha,N}^*$? Can $R(I^*) < C_{\alpha,N}^*$? What is the R-optimal procedure for producing confidence intervals for a bounded normal mean (the problem described above equation 24), and what is its value of R? These questions are the subject of current research.

References

[1] G.E. Backus. Confidence set inference with a prior quadratic bound. *Geophys. J.*, 97:119–150, 1989.

[2] D.L. Donoho. Statistical estimation and optimal recovery. *Technical Report No. 214, Department of Statistics, UC Berkeley*, 1989.

[3] D.L. Donoho and M.G. Low. Renormalization exponents and optimal pointwise rates of convergence. *Technical Report No. 249, Department of Statistics, UC Berkeley*, 1990.

[4] D.G. Luenberger. *Optimization by Vector Space Methods.* John Wiley and Sons, Inc., New York, 1969.

[5] P.B. Stark. Affine minimax confidence intervals for a bounded normal mean. *Technical Report, Department of Statistics, University of California, Berkeley*, 1991.

[6] P.B. Stark. Inference in infinite-dimensional inverse problems: Discretization and duality. *J. Geophys. Res.*, to appear, 1991.

l	rms x_l^m	Backus Interval	Nonlinear Bound
1	107.77	0.009	0.0055
2	22.59	0.015	0.0087
3	23.01	0.024	0.0148
4	17.86	0.042	0.0258
5	11.96	0.075	0.0458
6	9.36	0.135	0.0824
7	7.93	0.25	0.1500
8	4.73	0.45	0.2752
9	6.78	0.83	0.5080
10	4.59	1.53	0.9423
11	4.34	2.85	1.7555
12	3.62	5.32	3.2816

Table 1: Average estimated Gauss coefficients, length of Backus' confidence intervals, lower bounds on lengths of nonlinearly based confidence intervals in microTesla, at 99.99% confidence, using the heat-flow bound. Columns 2 and 3 from Backus [1989].