Rigorous Computer Solutions of Infinite-Dimensional Inverse Problems

By

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

Technical Report No. 228
December 1989

Department of Statistics
University of California
Berkeley, California

Rigorous Computer Solutions of Infinite-Dimensional Inverse Problems

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

Abstract

It is safe to say that almost all practical inverse problems are plagued by nonuniqueness. Regularization and other techniques for picking a particular solution fitting the data are useful for construction, but not usually for inference: the "true" model may bear little or no resemblance to the particular model the technique constructs. "Confidence set inference" or "strict bounds" works by looking at properties shared by all *a priori* acceptable models that fit the data adequately; *i.e.* it constructs confidence intervals for functionals of the unknown model. Constructing the confidence intervals involves solving certain infinite-dimensional constrained optimization problems in the model space. The usual way of solving these problems approximates the infinite-dimensional problems in a finite-dimensional subspace of the model space, so that the resulting problem can be solved by computer. Clearly, this can produce confidence intervals that are too short. This report uses Lagrangian duality to show how some of these optimization problems may be solved on a computer to get confidence intervals of the correct length.

Acknowledgements. I am grateful to George E. Backus, Peter J. Bickel and David L. Donoho for helpful and encouraging discussions. This work was supported by National Science Foundation Fellowship DMS-8705843, NSF PYI award DMS-8957573, NSF Grant DMS-8810192, and a grant from Schlumberger-Doll Research.

1 Introduction

Please see Backus [2], Cuer [3], Lang [5], and Stark [10] for other approaches to the same or similar problems. Here we treat only linear inverse problems, where we know *a priori* that the unknown model x_0 lies in a cone or a ball, and we wish to make inferences about a linear functional of x_0. The model for the experiment is that we observe n data δ_j, $j = 1, \cdots, n$ that are linearly independent linear functionals f_j^* of x_0, plus additive noise ϵ_j. We assume that the joint probability distribution of the noise terms is known, and that all components of the noise have mean zero. In vector form:

$$\boldsymbol{\delta} = \mathbf{f}^*[x_0] + \boldsymbol{\epsilon}. \tag{1}$$

The model x_0 belongs to the real linear vector space X; the functionals f_j^* belong to its algebraic dual space X^*. We know *a priori* that $x_0 \in \mathsf{C}$, where C is a special convex subset of X: either $\mathsf{C} = \mathsf{X}$ (no special prior information); $\mathsf{C} = \mathsf{P}$, a convex cone with vertex at the origin; or $\mathsf{C} = \mathsf{B}$, a ball in some norm in the space X, which will then be assumed to be a Banach space. In the first two cases, no topology on X is needed. Examples of the constraint $x_0 \in \mathsf{P}$ occur in gravimetry, resistivity and magnetotellurics, and seismology; as mass density, electrical conductivity, and seismic velocities must all be nonnegative. The constraint $x_0 \in \mathsf{B}$ may also arise in these fields if an upper bound on the unknown is available (density is less than that of neutronium, velocity less than the speed of light); it also arises naturally in geomagnetism where one may use *a priori* bounds on the energy stored in the main field, or the rate of dissipation of energy in the core (see Backus [2]). The first examples involve infinity-norm bounds, the last a quadratic norm. We wish to find a confidence interval for $h^*[x_0]$, a linear functional of x_0, using the prior information $x_0 \in \mathsf{C}$ and our knowledge of the probability distribution of the errors. That is, for fixed $\alpha \in (0, 1)$, we wish to find μ_-, μ_+ such that

$$\wp\{[\mu_-, \mu_+] \ni h^*[x_0]\} \geq 1 - \alpha,$$

where $\wp\{\cdot\}$ is the probability that the event in braces occurs.

Let Ξ be a $1 - \alpha$ confidence region for $\boldsymbol{\epsilon}$. Then $\wp\{\boldsymbol{\delta} - \mathbf{f}^*[x_0] \in \Xi\} \geq 1 - \alpha$. Let

$$\mathsf{D} \equiv \{x \in \mathsf{X} : \boldsymbol{\delta} - \mathbf{f}^*[x_0] \in \Xi\}.$$

Then we also have

$$\wp\{\mathsf{D} \ni x_0\} \geq 1 - \alpha,$$

and since $x_0 \in \mathsf{C}$,

$$\wp\{\mathsf{C} \cap \mathsf{D} \ni x_0\} \geq 1 - \alpha. \tag{2}$$

Let

$$\mu_- \equiv \inf_{x \in \mathsf{C} \cap \mathsf{D}} h^*[x]$$

and

$$\mu_+ \equiv \sup_{x \in \mathsf{C} \cap \mathsf{D}} h^*[x]$$

then μ_-, μ_+ satisfy 1. It remains to choose Ξ and solve the resulting optimization problems.

It has been traditional to choose Ξ to be a ball in some norm in $\mathbf{R}^n$; often an infinity norm. Backus [2] advocates a "slab," the region between parallel hyperplanes in $\mathbf{R}^n$. In the case $\mathsf{C} = \mathsf{X}$, it is well known that the only linear functionals that can be bounded by the data are linear combinations of the data functionals f_j^* (for example, [1]). As we have assumed that the f_j^* are linearly independent, there exists at most one representation of h^* as a linear combination of $\{f_j^*\}_{j=1}^n$. Say

$$h^* = \boldsymbol{\lambda} \cdot \mathbf{f}^* \equiv \sum_{j=1}^{n} \lambda_j f_j^*.$$

It is intuitively clear that in this case, if the distribution of errors is symmetric and unimodal, the optimal choice of confidence regions Ξ to get the narrowest possible confidence interval for $h^*[x_0]$ is

$$\Xi = \{\boldsymbol{\beta} \in \mathbf{R}^n : |\boldsymbol{\lambda} \cdot \boldsymbol{\beta}| \leq \chi\}, \tag{3}$$

for an appropriate choice of χ which may be computed from the known distribution of $\boldsymbol{\epsilon}$. This is the "slab" (not his terminology) of Backus. This choice allows D to be unbounded in every direction other than h^*; it concentrates all the statistical power of the data in the relevant direction.

For other sorts of prior information, the optimal choice of sets Ξ to get the shortest confidence intervals in expectation seem to be unsolved problems. In this report, we will use slabs as the confidence regions in data space. The orientation of the slab will be given by the vector γ, which need bear no relation to h^*. In [10] I develop results for general balls in l_p norms, for other functionals, and for slightly more general prior information.

2 Primal Approaches

We specialize to the minimization problem only since μ_+ can be found by minimizing $-h^*$ (also a linear functional). For Ξ as in 3,

$$\mathsf{D} = \{x \in \mathsf{X} : |\gamma \cdot (\mathbf{f}^*[x] - \boldsymbol{\delta})| \leq \chi\}.$$

Define $g^* \equiv \gamma \cdot \mathbf{f}^*$ and $\zeta \equiv \gamma \cdot \boldsymbol{\delta}$. Then

$$\mathsf{D} = \{x \in \mathsf{X} : |g^*[x] - \zeta| \leq \chi\} \tag{4}$$

In case $\mathsf{C} = \mathsf{X}$, it is well known that the only linear functionals with nontrivial confidence intervals are of the form $h^* = \lambda g^*$, $\lambda \in \mathbf{R}$. In case $\mathsf{C} = \mathsf{P}$ or $\mathsf{C} = \mathsf{B}$, the traditional means of solving the *primal* problem

$$\mu_- = \inf_{x \in \mathsf{C} \cap \mathsf{D}} H[x]$$

is to discretize the problem by approximating x in a finite-dimensional subspace of X. Given $\{x_j\}_{j=1}^m \subset \mathsf{X}$, let $\tilde{\mathsf{X}}$ be the subspace spanned by $\{x_j\}_{j=1}^m$. Then $x \in \tilde{\mathsf{X}}$ can be written $\sum_{k=1^m} \beta_j x_j$, and $x \in \mathsf{D} \cap \tilde{\mathsf{X}}$ if $|\boldsymbol{\phi} \cdot \boldsymbol{\beta} - \zeta| \le \chi$, where the m-vector $\boldsymbol{\phi}$ has elements $\phi_k = g^*[x_k]$ and $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_m)$. Similarly, $h^*[x] = \boldsymbol{\eta} \cdot \boldsymbol{\beta}$, where $\boldsymbol{\eta}$ is the m-vector with components $\eta_j \equiv h^*[x_j]$. One completes the picture by expressing $x \in \mathsf{C}$ in terms of the coefficients β_j. In my experience, it appears to be possible to pick the $\{x_k\}$ so that this is not too difficult. The result is a finite-dimensional optimization problem:

$$\tilde{\mu}_- = \min_{\{\boldsymbol{\beta}: \sum_k \beta_k x_k \in \mathsf{C} \cap \mathsf{D} \cap \tilde{\mathsf{X}}\}} \boldsymbol{\eta} \cdot \boldsymbol{\beta}.$$

It is clear that since the minimization is over the more restrictive set $\mathsf{C} \cap \mathsf{D} \cap \tilde{\mathsf{X}}$, $\tilde{\mu}_- \ge \mu_-$ (and correspondingly $\tilde{\mu}_+ \le \mu_+$), and thus the resulting confidence interval may be too short to achieve the nominal coverage probability. We turn now to another method of solution.

3 Lagrangian Duality

See the excellent book by Luenberger [6]. If (A) there exists $x_1 \in \mathsf{C}$ such that $|g^*[x_1] - \zeta| < \chi$, and (B) $\mu_- > -\infty$, Lagrangian duality says

$$\mu_- = \max_{\lambda \ge 0} \inf_{x \in \mathsf{C}} \{h^*[x] + \lambda(|g^*[x] - \zeta| - \chi)\}. \tag{5}$$

Note that condition (A) can be verified empirically by finding a feasable point for the discretized primal problem; condition (B) can be verified algebraically, as we see below. We now trace the implications of Lagrangian duality for the various choices of C.

3.1 $\mathsf{C} = \mathsf{X}$

We have

$$\begin{aligned} \mu_- &= \sup_{\lambda \ge 0}\{ \inf_{x \in \mathsf{X}} (h^*[x] + \lambda(|g^*[x] - \zeta| - \chi))\} \\ &= \sup_{\lambda \ge 0}\{-\lambda\chi + \inf_{x \in \mathsf{X}} \max[(h^* + \lambda g^*)[x] - \lambda\zeta, (h^* - \lambda g^*)[x] + \lambda\zeta]\}. \end{aligned}$$

Name the second term τ:

$$\tau \equiv \inf_{x \in \mathsf{C}} \max[(h^* + \lambda g^*)[x] - \lambda\zeta, (h^* - \lambda g^*)[x] + \lambda\zeta]. \tag{6}$$

For $h^* = \pm\lambda g^*$,

$$\tau = \inf_{x \in \mathsf{X}} \max[\pm\lambda\zeta, \pm 2\lambda g^*[x] \mp \lambda\zeta] = \pm\lambda\zeta.$$

If $h^* \neq \lambda g^*$ and $h^* \neq -\lambda g^*$, write

$$h^* = e^* \pm \lambda g^*,$$

where $e^* \neq 0$ and the sign will be fixed presently. Then

$$\tau \leq \inf_{x \in \mathsf{X}} (e^*[x] \pm \lambda g^*[x] + \lambda |g^*[x]| + \lambda \zeta).$$

Pick the sign so that there exists $y \in \mathsf{X}$ with both $e^*[y] < 0$ and $\pm g^*[y] < 0$ (this is possible since both functionals are nonzero). Then

$$\tau \leq \inf_{x \in \mathsf{X}} (e^*[x] + \lambda \zeta) = -\infty.$$

Thus

$$\begin{aligned} \mu_- &= \sup_{\substack{\lambda \geq 0 \\ h^* = \pm \lambda g^*}} -\lambda \chi \pm \lambda \zeta \\ &= \lambda \zeta - |\lambda| \chi, \;\; h^* = \lambda g^*, \;\lambda \in \mathbf{R}. \end{aligned}$$

The only functionals one may make inferences about in this case are multiples of h^*. Note that since g^* is a nonzero linear functional, there always exists x_1 such that $g^*[x_1] = \zeta$, so condition (A) is met. We now have no optimization to perform at all, just the (easy) infinite-dimensional constraint that h^* is proportional to g^*.

3.2 $\mathsf{C} = \mathsf{P}$

The algebra leads us to consider τ (equation 6) again, but for $\mathsf{C} = \mathsf{P}$. Let $\mathsf{P}^{\oplus}$ be the conjugate cone of P in X^*:

$$\mathsf{P}^{\oplus} \equiv \{x^* \in \mathsf{X}^* : x^*[x] \geq 0, \; \forall x \in \mathsf{P}\}.$$

Then if $h^* \pm \lambda g^* \in \mathsf{P}^{\oplus}$,

$$\tau = \inf_{x \in \mathsf{P}} \max [(h^* + \lambda g^*)[x] - \lambda \zeta, (h^* - \lambda g^*)[x] + \lambda \zeta] = \mp \lambda \zeta,$$

while if $h^* + \lambda g^* \notin \mathsf{P}^{\oplus}$ and $h^* - \lambda g^* \notin \mathsf{P}^{\oplus}$, then using the convexity of P it is easy to show that there exists $y \in \mathsf{P}$ such that $(h^* \pm \lambda g^*)[y] < 0$ for *both* choices of sign. Since P is a cone with vertex at the origin, $y \in \mathsf{P}$ implies $\beta y \in \mathsf{P}$, $\beta \geq 0$, and so

$$\tau = \inf_{x \in \mathsf{P}} \max [(h^* + \lambda g^*)[x] - \lambda \zeta, (h^* - \lambda g^*)[x] + \lambda \zeta] = -\infty.$$

We conclude that

$$\begin{aligned}
\mu_- &= \max_{\substack{\lambda \geq 0 \\ h^* \pm \lambda g^* \in \mathsf{P}^{\oplus}}} \{-\lambda\chi \mp \lambda\zeta\} \\
&= \max_{\lambda : h^* + \lambda g^* \in \mathsf{P}^{\oplus}} \{-|\lambda|\chi - \lambda\zeta\} \qquad (7)
\end{aligned}$$

Here we have a one-dimensional optimization problem, with the infinite-dimensional constraint on λ that $h^* + \lambda g^* \in \mathsf{P}^{\oplus}$. In applications, this usually amounts to something like requiring that the sum of a function and a multiple of another is nonnegative everywhere. This sort of constraint can readily be imposed since the two functions are *known*, and so their continuity properties can be exploited to impose the constraint conservatively using only values on a finite grid.

3.3 C=B

Here X is normed; we will assume it is a Banach space and that the functionals h^* and f_j^* (and thus g^*) are continuous. Let B be the ball with radius ν. For this problem, we find

$$\begin{aligned}
\tau &= \inf_{x \in \mathsf{B}} \max[(h^* + \lambda g^*)[x] - \lambda\zeta, (h^* - \lambda g^*)[x] + \lambda\zeta] \\
&\geq \max\left[\inf_{x \in \mathsf{B}} \{(h^* + \lambda g^*)[x] - \lambda\zeta\}, \inf_{x \in \mathsf{B}} \{(h^* - \lambda g^*)[x] + \lambda\zeta\}\right] \\
&= \max[-\nu\|h^* + \lambda g^*\| - \lambda\zeta, -\nu\|h^* - \lambda g^*\| + \lambda\zeta] \\
&\equiv \sigma,
\end{aligned}$$

so

$$\begin{aligned}
\mu_- &\geq \max_{\lambda \geq 0} \{-\lambda\chi + \max[-\nu\|h^* + \lambda g^*\| - \lambda\zeta, -\nu\|h^* - \lambda g^*\| + \lambda\zeta]\} \\
&= \max_{\lambda \in \mathbf{R}} \{-|\lambda|\chi - \nu\|h^* + \lambda g^*\| - \lambda\zeta\}. \qquad (8)
\end{aligned}$$

We shall see that there exists $y \in \mathsf{B}$ such that

$$h^*[y] + \lambda|g^*[y] - \zeta| = \sigma,$$

and thus equality holds in 8.

Define y_+ such that $-y_+$ is aligned with $h^* + \lambda g^*$, and $\|y_+\| = \nu$. Define y_- such that $-y_-$ is aligned with $h^* - \lambda g^*$, and $\|y_-\| = \nu$. Then

$$(h^* + \lambda g^*)[y_+] - \lambda\zeta = -\nu\|h^* + \lambda g^*\| - \lambda\zeta,$$

and

$$(h^* - \lambda g^*)[y_-] + \lambda\zeta = -\nu\|h^* - \lambda g^*\| + \lambda\zeta.$$

A bit of algebra shows that if $\sigma = -\nu||h^* + \lambda g^*|| - \lambda\zeta$, then $(h^* + \lambda g^*)[y_+] - \lambda\zeta \geq (h^* - \lambda g^*)[y_+] + \lambda\zeta$, and y_+ is the required vector. On the other hand, if $\sigma = -\nu||h^* - \lambda g^*|| + \lambda\zeta$, then $(h^* - \lambda g^*)[y_-] + \lambda\zeta \geq (h^* + \lambda g^*)[y_-] - \lambda\zeta$, and y_- is the required vector. □

Note that in this problem, the dual is the unconstrained maximization of a concave function of λ—the infinite-dimensional constrained optimization problem is certainly simplified. However, the evaluation of $||h^* - \lambda g^*||$ can involve infinite-dimensional operations such as summation of series or integration. Fortunately, the summation or integration involves the ***known*** vectors h^* and g^*, and so their properties can be used to bound the accuracy of a truncated sum, or of numerical quadrature. There are many error bounds one may use, depending on how much information about the functionals is incorporated. The error bounds are convex functions of λ. If one of these error bounds is subtracted from the original objective function, and the difference is maximized numerically, the result is a rigorous lower bound on μ_-.

4 Discretizing the Dual Problem

We will look at an example of discretizing the dual problem for the case $\mathsf{C} = \mathsf{B}$ related to the "ideal bodies" problem in gravimetry [7, 8, 9]. We make n noisy observations δ_j of the vertical component of gravity at the points $\{\vec{r_j}\}_{j=1}^n$ at a height β above Earth's surface. Let E denote the region of $\mathbf{R}^3$ occupied by Earth, and $x_0(\vec{r})$ be the mass density of Earth as a function of position $\vec{r}$. Then

$$\delta_j = f_j^*[x_0] + \epsilon_j, \quad j = 1, \cdots, n$$

where

$$f_j^*[x] \equiv \int_E \mathcal{G} \frac{\hat{z} \cdot (\vec{r_j} - \vec{r})}{|\vec{r_j} - \vec{r}|^3} x(\vec{r}) d^3\vec{r}.$$

Here $\hat{z}$ is the vertical unit vector in $\mathbf{R}^3$. We shall assume that we know ***a priori*** that the density is bounded below by zero and above by $v(\vec{r})$. We assume the joint distribution of the noise components is known. We wish to find a 99% confidence interval for the mass in the region $D \subset E$, which is given by the linear functional h^*:

$$h^*[x] = \int_E 1_D(\vec{r}) x(\vec{r}) d^3\vec{r}.$$

Here $1_D(\vec{r}) = \{1, \ \vec{r} \in D; \ 0, \ \text{otherwise}\}$. This problem may be reduced to the form in section 3.3 as follows. We pose the problem in the Banach space $L^\infty(E)$ of bounded functions on the domain E, with the "sup" norm

$$||x|| \equiv \sup_{\vec{r} \in E} |x(\vec{r})|.$$

Let $x'(\vec{r}) \equiv \frac{1}{2}v(\vec{r})$. Pick $\gamma \in \mathbf{R}^n$ (for example, let $\gamma \equiv \arg\min_{\rho \in \mathbf{R}^n} \|h^* - \sum_{j=1}^n \rho_j f_j^*\|$, or an approximation to it). Let χ be such that $\wp\{\gamma \cdot \epsilon \le \chi\} = 0.99$. Then the optimization problem we must solve has the form:

$$\mu_- = \inf_{\substack{x \in \mathbf{L}_\infty : \|\frac{x-x'}{x'}\| \le 1 \\ |\gamma \cdot \mathbf{f}^*[x] - \gamma \cdot \delta| \le \chi}} h^*[x].$$

By defining $g^* = \gamma \cdot \mathbf{f}^*$ and $\zeta = \gamma \cdot (\delta - \mathbf{f}^*[x'])$; and renorming the space to take into account the weight $\frac{1}{x'}$, we reduce the optimization problem to the canonical form

$$\mu_- = \inf_{\substack{\|x\| \le 1 \\ |g^*[x] - \zeta| \le \chi}} h^*[x].$$

A common approach to solving this problem is to divide E into m regions and approximate X by the m-dimensional subspace of functions that are constant in each of the regions; the resulting discretized problem can be solved by linear programming (see [9, 4]). The linear programming solution will almost certainly be larger than the true solution to the original infinite-dimensional problem. Similarly, the value for μ_+ determined by the finite-dimensional approximation may be too small. In either case the confidence interval may not be large enough to have 99% coverage probability.

Instead, we use duality to get a different finite (one) dimensional problem. The solution is given in equation 8 to be

$$\max_\lambda\{-|\lambda|\chi - \lambda\zeta - \|h^* + \lambda g^*\|\}$$

which is the maximization of a one-dimensional convex function. In practice, it is necessary to approximate numerically the integral

$$\|h^* - \lambda g^*\| = \int_E \left| 1_D(\vec{r}) - \lambda \sum_{j=1}^n \gamma_j \mathcal{G} \frac{\hat{z} \cdot (\vec{r_j} - \vec{r})}{|\vec{r_j} - \vec{r}|^3} \right| d^3\vec{r},$$

However, since $\{f_j^*\}_{j=1}^n$ and 1_D are *known* functions, their continuity properties can be used to determine the maximum possible error in the quadrature. As noted in section 3.3, the maximum quadrature error is a convex function of λ (and depends on the quadrature scheme). If the error bound is subtracted from $-|\lambda|\chi - \lambda\zeta - \|h^* - \lambda g^*\|$, where the norm is to be approximated by quadrature, and the difference, which is a concave function of λ, is maximized numerically, the resulting number yields a confidence interval of at least the correct length. Note that the dual space of X is the (non-separable) space of finitely additive finite signed measures, elements of which would, in general, be rather difficult to approximate even with a countable number of approximating functions. Fortunately, in the dual approach, we need only consider particular known elements of the dual space. This is true for $\mathsf{C} = \mathsf{X}$ and $\mathsf{C} = \mathsf{P}$ as well.

5 Conclusion

A common approach to computing confidence intervals for functionals of a model in linear inverse problems involves approximating the infinite-dimensional model space with a finite-dimensional subspace. The result can be confidence intervals that are too short. Duality provides a tool to formulate finite-dimensional problems that exactly solve the original infinite-dimensional problems. These finite-dimensional problems may have infinite-dimensional constraints, or involve integration or some such infinite-dimensional operation. However, the functions that are involved in the dual problem are known, and so the accuracy of methods to approximate the infinite-dimensional part of the dual problem can be calculated or bounded explicitly. The result is that one may construct confidence intervals for functionals that, under the assumptions of the observation model, are certain have their nominal coverage probability.

References

[1] G. Backus. Inference from inadequate and inaccurate data, I. *Proc. Natl. Acad. Sci.*, 65:1–7, 1970.

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

[3] M. Cuer. Discretisation en $o(h^p)$, $p > 2$ de problèmes de programmation lineare dans $L^\infty(0,1)$. *Manuscript*, 1989.

[4] S.P. Huestis and M.E. Ander. IDB2—a Fortran program for computing extremal bounds in gravity data interpretation. *Geophysics*, 48:999–1010, 1983.

[5] S.W. Lang. Bounds from noisy linear measurements. *IEEE Trans. Info. Th.*, IT-31:498–508, 1985.

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

[7] R.L. Parker. The theory of ideal bodies for gravity interpretation. *Geophys. J. Roy. astr. Soc.*, 42:315–334, 1975.

[8] P.C. Sabatier. Positivity constraints in linear inverse problems—I. General theory, II. Applications. *Geophys. J. Roy. astr. Soc.*, 48:415–459, 1977.

[9] C. Safon, G. Vasseur and M. Cuer. Some applications of linear programming to the inverse gravity problem. *Geophysics*, 42:1215–1229, 1977.

[10] P.B. Stark. Duality and discretization in linear inverse problems. *Technical Report, Dept. of Statistics, University of California, Berkeley*, 1989.