IOP PUBLISHING INVERSE PROBLEMS

Inverse Problems **24** (2008) 034014 (17pp) doi:10.1088/0266-5611/24/3/034014

Generalizing resolution

Philip B Stark

Department of Statistics, #3860 University of California, Berkeley, CA 94720-3860, USA

E-mail: stark@stat.berkeley.edu

Received 11 October 2007, in final form 20 April 2008
Published 23 May 2008
Online at stacks.iop.org/IP/24/034014

Abstract

Of those things that can be estimated well in an inverse problem, which is best to estimate? Backus–Gilbert resolution theory answers a version of this question for linear (or linearized) inverse problems in Hilbert spaces with additive zero-mean errors with known, finite covariance, and no constraints on the unknown other than the data. This paper generalizes resolution: it defines the resolution and Bayes resolution of an estimator, intrinsic minimax and Bayes resolution, and intrinsic minimax and Bayes design resolution. Intrinsic resolution is the smallest value of a penalty across parameters that can be estimated with controlled (minimax or Bayes) risk. Intrinsic minimax resolution includes Backus–Gilbert resolution and subtractive optimally localized averages (SOLA) as special cases. Intrinsic design resolution is the smallest value of a penalty among parameters that can be estimated with controlled (minimax or Bayes) risk using observations with controlled acquisition cost. Intrinsic resolution wraps the classical problem of choosing an optimal estimator of an abstract parameter inside the problem of choosing an optimal parameter to estimate. Intrinsic design resolution adds another layer: optimizing what to observe. Equivalently, it wraps a problem in information-based complexity inside the problem of choosing an optimal parameter. The definitions apply to inverse problems with constraints, to nonlinear inverse problems, to nonlinear and biased estimators and estimators with controlled computational cost, to general definitions of risk (not just the variance of unbiased estimators), to confidence set estimators as well as point estimators, and to abstract penalties not necessarily related to 'spread'. Simple examples are given, including a definition of the resolution of 'strict bounds' confidence intervals.

1. Introduction

In a seminal series of papers, George Backus and Freeman Gilbert [1, 2, 3, 4, 7] studied the problem of estimating an element $\theta = \theta(r)$ of a separable Hilbert space $\mathcal{H}$ of functions of

position r from a set of $n < \infty$ noisy observations $\{X_j\}_{j=1}^n$ related to θ through bounded linear functionals. (They also addressed local linearizations of Fréchet-differentiable problems.) Backus and Gilbert showed that no finite set of linear functionals determines θ, even if the data are free of error. Hence, they advocated estimating properties of the model that the data do constrain. Backus and Gilbert [7, section 5] showed how to select from the linear functionals that can be estimated linearly with zero bias and controlled variance the linear functional that comes as close as possible to measuring $\theta(r_0)$, the value of the model at the point r_0. They quantify 'as close as possible' using several measures of spread.

The Backus and Gilbert approach to measuring resolution has been applied to many problems, including geomagnetism [28], gravimetry [15], heat flow [17], helioseismology [13], magnetotellurics [23, 27], optics [39], seismology [7] and signal processing [22, 24]. Backus–Gilbert theory has been generalized in several ways, including definitions of optimality other than minimal spread [14, 18, 24, 29, 30] and the effect of nonlinearity [34]. Parts of the theory have been generalized to Banach spaces [12]. And its symptotic properties have been studied [16, 33].

This paper generalizes resolution in a number of different directions: the new definition encompasses nonlinear problems; model constraints; nonlinear, biased and controlled-cost estimators; set estimators as well as point estimators; new definitions of concentration; loss functions other than squared error; and minimax and Bayes risk. It also introduces the notion of *design resolution* to optimize the observations to make; this includes some problems in information-based complexity as special cases.

Section 2 lays out the notation. Section 3 contains the new ideas. Section 3.1 reviews Backus–Gilbert resolution. Section 3.2 defines two new generalizations of resolution: the intrinsic minimax resolution and the minimax resolution of an estimator. Section 3.3 defines two more new generalizations: the intrinsic Bayes resolution and the Bayes resolution of an estimator. Section 3.4 shows how the definitions in sections 3.2 and 3.3 can be used to quantify the resolution of confidence set estimators, a new notion. Section 3.5 defines two more new quantities: the minimax and Bayes intrinsic design resolution. Section 4 has some simple examples. Section 4.1 gives an example of computing the intrinsic resolution in a constrained inverse problem. Section 4.2 gives an example of computing the resolution of the 'strict bounds' approach to finding confidence intervals in constrained problems. Section 4.3 illustrates optimal design resolution using a problem related to optimal quadrature. Sections 5 and 6 contain discussion and conclusions.

2. Notation

The possible states of Nature (models) are elements θ of a set Θ, a nonempty subset of a separable Banach space $\mathcal{T}$. The set Θ is known, but the value of θ is not. The data X take values x in a separable Banach space $\mathcal{X}_K$. For the moment, K is an abstract index and is considered to be fixed. When we consider design resolution in section 3.5, K will index forward operators. Typically, $\mathcal{X}_K = \mathbb{R}^n$, but n can depend on K. The probabability distribution of X depends on θ and on K: $X \sim \mathbb{P}_{\theta,K}$. Assume that for fixed K all the probability distributions $\mathcal{P}_K \equiv \{\mathbb{P}_{\theta,K} : \theta \in \Theta\}$ are defined on a common σ-algebra $\mathcal{F}$ of subsets of $\mathcal{X}_K$. Throughout the paper, I use the convention that $\inf_{\emptyset}(\cdot) = \infty$ and $\sup_{\emptyset}(\cdot) = -\infty$.

A *parameter* is the value at θ of a mapping $g : \Theta \to \mathcal{B}$, where $\mathcal{B}$ is a separable Banach space. Sometimes the mapping g, rather than its value at θ, is called the parameter. A parameter is real valued if $\mathcal{B} = \mathbb{R}$. The parameter $g[\theta]$ is *linear* if g is a linear mapping. For example, if Θ is a collection of mass density distributions within the Earth that satisfy bounds $a \leqslant \theta(r) \leqslant b$, and g is the average density in a volume within the Earth, then $g[\theta]$ is a

real-valued linear parameter. Similarly, if Θ is the collection of scalar potentials of magnetic fields at Earth's core–mantle boundary corresponding to fields with total energy less than some bound, and $g[\theta]$ is a single Gauss coefficient of the potential, then $g[\theta]$ is a real-valued linear parameter [5, 6, 36].

A parameter g is *identifiable* (*for forward mapping* K) if $g[\eta] \neq g[\zeta]$, $\eta, \zeta \in \Theta$, implies $\mathbb{P}_{\eta,K} \neq \mathbb{P}_{\zeta,K}$. If a parameter is not identifiable, two models can have different values of the parameter and yet produce the same observations with the same probabilities, so data cannot discriminate between them.

A forward problem is *linear* if Θ is a subset of a separable Banach space $\mathcal{T}$ with normed dual $\mathcal{T}^*$, and for some fixed n-tuple $(\kappa_j)_{j=1}^n$ of elements of $\mathcal{T}^*$, the datum is $X = (X_j)_{j=1}^n$ where

$$X_j = \kappa_j[\theta] + \epsilon_j, \qquad \theta \in \Theta, \tag{1}$$

and $\epsilon = (\epsilon_j)_{j=1}^n$ is a vector of random errors with a probability distribution that does not depend on θ. The functionals $(\kappa_j)_{j=1}^n$ are sometimes called 'data kernels' or 'representers'. A more compact notation is

$$X = K[\theta] + \epsilon, \tag{2}$$

where $K : \Theta \to \mathbb{R}^n; \theta \mapsto (\kappa_j[\theta])_{j=1}^n$. Backus–Gilbert theory [1, 2, 3, 4, 7] addresses estimating identifiable real-valued linear parameters when the forward problem is linear (or linearized), the model space $\mathcal{T}$ is a Hilbert space, and there are no constraints other than the data, i.e., $\Theta = \mathcal{T}$. Essentially, Backus and Gilbert show that a linear parameter $g[\theta]$ is identifiable if and only $g[\cdot] = \sum_{j=1}^n \beta_j \kappa_j[\cdot]$. See [12] for a generalization of this result.

A (non-randomized) estimator $\hat{g}(\cdot)$ of a parameter $g[\theta]$ is a measurable mapping from $\mathcal{X}_K$ to $\mathcal{A}$, where $\mathcal{A}$ (the space of *actions*) is a measurable space. For *point estimators* the action space is $\mathcal{B}$, the space of possible parameter values. For *set estimators* the action space is a collection of subsets of $\mathcal{B}$. In general, $\mathcal{A}$ need not have any particular relationship to $\mathcal{B}$. For example, the parameter might be the depth to the top of a reservoir, so that $\mathcal{B} = \mathbb{R}$, and the action space might be $\{0, 1\}$, where 0 means 'do not drill' and 1 means 'drill'. The crucial thing is that the *risk* of an estimator $\hat{g}$ when the true parameter value is $g[\eta]$ and the model is η must be well defined. In this example, the cost of drilling might be polynomial in the depth, and the cost of not drilling might be constant. An estimator would be a rule for deciding whether to drill on the basis of an observation X. The risk would be the expected cost of using that rule to make the decision. See [19] for a rigorous exposition. To each possible observation $x \in \mathcal{X}_K$, $\hat{g}$ assigns an element $\hat{g}(x) \in \mathcal{A}$, in such a way that the pre-image under $\hat{g}$ of measurable subsets of $\mathcal{A}$ are $\mathbb{P}_{\theta,K}$-measurable subsets of $\mathcal{X}_K$ for all $\theta \in \Theta$. In Backus–Gilbert resolution, $\mathcal{A} = \mathcal{B} = \mathbb{R}$.

There are also *randomized estimators*, which assign probability distributions on $\mathcal{A}$ to elements of $\mathcal{X}_K$. See [12] or [11] for more detail and examples; see [19] for rigor. Often the class $\mathcal{E}_K$ of estimators considered is restricted to depend on the data in a specific way or to have limited computational cost. For example, when $\mathcal{A}$ is a linear space, $\mathcal{E}_K$ might contain only estimators with linear or affine dependence on the data, or estimators that require at most a given number of floating point operations. In Backus–Gilbert resolution, $\mathcal{E}_K$ consists of estimators with linear dependence on the data: $\hat{g}(x) = \beta \cdot x$ for some $\beta \in \mathbb{R}^n$.

The expectation operator when the true state of Nature is η is denoted as $\mathbb{E}_\eta$. Generically, $\rho_\eta(\hat{g}; g[\eta])$, the *$\rho$-risk at η of the estimator $\hat{g}$ of the parameter $g[\eta]$*, denotes the expected cost of estimating $g[\eta]$ by $\hat{g}$ when the true model is η. For fixed K, ρ maps $\mathcal{E}_K \times \Theta \times \mathcal{G}$ into $\mathbb{R}^+$, where $\mathcal{G}$ is the set of parameters under consideration.

Mean squared error (MSE) is an example of a risk function for a point estimator. The *mean squared error at η of the point estimator $\hat{g}$ of the real-valued parameter $g[\eta]$* is

$$\mathrm{MSE}_\eta(\hat{g}; g[\eta]) \equiv \mathbb{E}_\eta(\hat{g}(X) - g[\eta])^2. \tag{3}$$

MSE can be decomposed into a sum of the variance of $\hat{g}$, $\mathbf{Var}_\eta(\hat{g}) = \mathbb{E}_\eta(\hat{g}(X) - \mathbb{E}_\eta(\hat{g}(X)))^2$ and the square of the bias of $\hat{g}$, $\mathbf{bias}_\eta(\hat{g}; g[\eta]) \equiv \mathbb{E}_\eta[\hat{g}(X) - g[\eta]]$:

$$\mathrm{MSE}_\eta(\hat{g}; g[\eta]) = \mathbf{Var}_\eta(\hat{g}) + (\mathbf{bias}_\eta(\hat{g}; g[\eta]))^2. \tag{4}$$

The MSE of an unbiased real-valued point estimator is thus equal to its variance. Backus and Gilbert work with unbiased estimators; a bound on the variance of an unbiased estimator is the same as a bound on its MSE. The MSE quantifies the average accuracy of $\hat{g}(X)$ as an estimator of $g[\eta]$. There are countless ways to define the accuracy of point estimators of real-valued parameters; MSE is common.

One way to define the risk of randomized set estimators is the expected measure of the set [11, 32], or a combination of expected size and coverage probability [8].

The *maximum ρ-risk of $\hat{g}$ (for estimating g)* is

$$\rho_\Theta(\hat{g}; g) \equiv \sup_{\eta\in\Theta} \rho_\eta(\hat{g}; g). \tag{5}$$

The *minimax ρ-risk for estimating g (over estimators in $\mathcal{E}_K$)* is

$$\rho_\Theta(\mathcal{E}_K; g) \equiv \inf_{\hat{g}\in\mathcal{E}_K} \rho_\Theta(\hat{g}; g). \tag{6}$$

Because $\mathcal{T}$ is a separable Banach space, it is measurable with respect to its Borel σ-algebra. Suppose that Θ is a measurable subset of $\mathcal{T}$, and let π be a probability measure on Θ. The expectation operator over the (prior) distribution π on Θ is denoted as $\mathbb{E}_\pi$. Thus, for example,

$$\mathbb{E}_\pi \hat{g}(X) = \int_\Theta \int_{\mathcal{X}_K} \hat{g}(x)\mathbb{P}_{\eta,K}(\mathrm{d}x)\pi(\mathrm{d}\eta). \tag{7}$$

The *(Bayes) ρ-risk of $\hat{g}$ (for estimating g) for prior π* is

$$\rho_\pi(\hat{g}; g) \equiv \int_\Theta \rho_\eta(\hat{g}; g[\eta])\pi(\mathrm{d}\eta). \tag{8}$$

The *(Bayes) ρ-risk (for estimating g) for prior π (over estimators in $\mathcal{E}_K$)* is

$$\rho_\pi(\mathcal{E}_K; g) \equiv \inf_{\hat{g}\in\mathcal{E}_K} \rho_\pi(\hat{g}; g). \tag{9}$$

The notation is summarized in table 1.

3. Resolution

3.1. Backus–Gilbert resolution

Consider a linear forward problem $X = K[\theta] + \epsilon$ with $\mathbb{E}\epsilon = 0$ and $\mathbf{Cov}(\epsilon) = \Sigma$, where Σ is a positive definite matrix. For the moment, suppose that $\Theta = \mathcal{T}$; i.e., there are no extra constraints on θ. In geophysical inverse problems, $\mathcal{T}$ is typically a space of functions of position $\theta = \theta(r)$, with r limited to some domain $\mathcal{D} \subset \mathbb{R}^k$. Restrict attention to real-valued linear parameters $g[\theta]$. Backus and Gilbert [4, 7] showed that the only linear parameters that can be estimated with finite bias whatever be $\theta \in \mathcal{T}$ are of the form

$$g[\theta] = \beta \cdot K[\theta] \equiv \sum_{j=1}^{n} \beta_j \kappa_j[\theta] \tag{10}$$

Table 1. Notation.

$\mathcal{T}$	A separable Banach space
Θ	The possible states of the world, a subset of $\mathcal{T}$
θ	An element of Θ; the particular state the world is in
η	A generic element of Θ
K	An abstract index relating to the forward mapping
$\mathcal{X}_K$	A separable Banach space of possible observations corresponding to the forward mapping with index K
x	A generic element of $\mathcal{X}_K$
X	Data: a random variable that takes values in $\mathcal{X}_K$; $X \sim \mathbb{P}_{\theta,K}$
$\lambda(K)$	The cost of observing X for the forward mapping with index K
L	Maximum tolerable cost $\lambda(K)$
$\mathcal{G}$	A set of mappings from $\Theta \to \mathcal{B}$, where $\mathcal{B}$ is a separable Banach space
g	An element of $\mathcal{G}$, a parameter
$T(\cdot)$	A figure of merit for parameters g. $T : \mathcal{G} \to \mathbb{R}^+$
$\mathcal{A}$	A measurable space of *actions*
$\mathcal{E}_K$	A set of estimators, measurable mappings from $\mathcal{X}_K$ into $\mathcal{A}$
$\hat{g}$	An element of $\mathcal{E}_K$, an estimator
$\mathbb{E}_\eta$	Expectation operator when the true state of the world is $\eta \in \Theta$
$\rho_\eta(\hat{g}; g[\eta])$	Risk of estimating g using $\hat{g}$ when the world is in state η. For fixed g and K, $\rho : \mathcal{E}_K \times \Theta \to \mathbb{R}^+$.
$\rho_\Theta(\hat{g}; g)$	$\sup_{\eta\in\Theta} \rho_\eta(\hat{g}; g)$: maximum risk of estimating g using $\hat{g}$
$\rho_\Theta(\mathcal{E}_K; g)$	$\inf_{\hat{g}\in\mathcal{E}_K} \rho_\Theta(\hat{g}; g)$: minimax risk for estimating g using an estimator in $\mathcal{E}_K$
π	A prior probability distribution on Θ
$\rho_\pi(\hat{g}; g)$	$\int_\Theta \rho_\eta(\hat{g}; g[\eta])\pi(\mathrm{d}\eta)$: average risk for prior π for estimating g using $\hat{g}$
$\rho_\pi(\mathcal{E}_K; g)$	$\inf_{\hat{g}\in\mathcal{E}_K} \rho_\pi(\hat{g}; g)$: Bayes risk for prior π for estimating g using an estimator in $\mathcal{E}_K$
M	Maximum tolerable risk

with $\beta \in \mathbb{R}^n$. Moreover, the estimator

$$\hat{g}(X) = \sum_{j=1}^{n} \beta_j X_j = \beta \cdot X \tag{11}$$

is unbiased for $(\beta \cdot K)[\theta]$ and has variance

$$\beta^T \cdot \Sigma \cdot \beta. \tag{12}$$

Since $\beta \cdot X$ is unbiased for $(\beta \cdot K)[\theta]$,

$$\mathrm{MSE}_\eta(\beta \cdot K[\eta], \beta \cdot X)) = \mathbf{Var}_\eta(\beta \cdot X) = \beta^T \cdot \Sigma \cdot \beta, \qquad \eta \in \mathcal{T}. \tag{13}$$

Suppose we are unwilling to tolerate mean squared error greater than M. Within the set of linear functionals that are estimable with $\mathrm{MSE}_\eta \leqslant M$ for all $\eta \in \mathcal{T}$, we can choose to estimate one that optimizes some criterion. Because $\mathcal{T}$ is a Hilbert space of functions of position $r \in \mathcal{D}$, the linear functionals $\kappa_j[\cdot]$ have 'representers' $\kappa_j(r)$ that are themselves elements of $\mathcal{T}$. Let $K(r)$ denote the vector of functions $(\kappa_j(r))_{j=1}^n$, and for $\gamma \in \mathbb{R}^n$, define

$$\gamma \cdot K(r) \equiv \sum_{j=1}^{n} \gamma_j \kappa_j(r). \tag{14}$$

In Backus–Gilbert theory, the vector of coefficients β is chosen so that the function $\beta \cdot K(r)$ satisfies $\int \beta \cdot K(r)\,\mathrm{d}r = 1$. That is, the average of the model given by $\beta \cdot K[\theta]$ is the integral of θ against an 'averaging kernel' $\beta \cdot K(r)$ that has unit area or volume. Backus–Gilbert resolution theory shows how to find β to minimize the spread of $\beta \cdot K(r)$ around some target

point r_0, for several definitions of 'spread'. The heuristic is that we can almost estimate $\theta(r_0)$ if we can estimate an average of θ that is spatially concentrated near r_0. A definition of the spread of $\beta \cdot K(r)$ about r_0 for models with one spatial dimension is

$$S_{r_0}(\beta \cdot K) \equiv 12 \int (r - r_0)^2 (\beta \cdot K(r))^2 \, \mathrm{d}r. \tag{15}$$

The factor of 12 makes the spread of a unit-area boxcar function centred at r_0 equal to its width.

Generally, there is a tradeoff between spread and variance: picking β so that $g_\beta[\theta]$ has small spread—i.e., so that $\beta \cdot K[\theta]$ is an average of θ concentrated near r_0—typically requires $\mathbf{Var}_\theta(\beta \cdot K) = \beta^T \cdot \Sigma \cdot \beta$ to be large. See [28] for examples. The smallest value of the spread among the linear functionals that can be estimated with zero bias and variance $\leqslant M$ is the *Backus–Gilbert resolution at variance M*.

The coefficients β can be chosen to optimize a penalty other than 'spread'. For example, one might try to make $\beta \cdot K$ approximate integration against a boxcar function centered at r_0, a step function at r_0, or some other target function, depending on the scientific goal. See, e.g., [14, 24, 29].

Backus–Gilbert resolution is intrinsic to the inverse problem: it quantifies the spread of the most concentrated linear estimator whose MSE is not too big. A related notion, *the resolution of a linear estimate*, is a property of both the inverse problem and the linear estimate. In a linear inverse problem, any linear estimator $\hat{g}(X) = \beta \cdot X$ is an unbiased estimator of an average of the model, namely, $\int \beta \cdot K(r)\theta(r) \, \mathrm{d}r$. One can calculate the spread of the averaging kernel $\beta \cdot K(r)$ around any given point r_0. For example, when $\mathcal{T}$ is a Hilbert space, a norm-regularized least squares estimate with a fixed tradeoff parameter χ is

$$\hat{\theta}_{RLS} = \arg\min_{\eta \in \Theta} \{(K[\eta] - X)^T \cdot \Sigma^{-1} \cdot (K[\eta] - X) + \chi \|\eta\|^2\}. \tag{16}$$

At each r, $\hat{\theta}_{RLS}(r)$ is of the form $\sum_{j=1}^n \beta_j(r) X_j$, so the expected value of the estimate at r_0 is $\beta(r_0) \cdot K[\theta]$, the variance is $\beta(r_0)^T \cdot \Sigma \cdot \beta(r_0)$, and the spread about r_0 is $S_{r_0}(\beta(r_0) \cdot K)$. That spread is *the resolution of the estimate* $\beta(r_0) \cdot X$. The MSE of $\beta(r_0) \cdot X$ as an estimate of $\beta(r_0) \cdot K[\theta]$ could exceed M—the spread of this estimator is not quite the same as the intrinsic resolution at a given MSE.

3.2. Minimax resolution

Backus–Gilbert resolution finds the functional with smallest spread among those linear functionals that can be estimated by a linear estimator with MSE $\leqslant M$ for all $\eta \in \mathcal{T}$.

The intrinsic minimax resolution just replaces some of the ingredients of Backus–Gilbert resolution with more abstract ingredients. Let $\mathcal{G}$ be a set of parameters on Θ. In Backus–Gilbert resolution, $\mathcal{G}$ are linear functionals corresponding to averages with weights that integrate to unity. Let $\mathcal{E}_K$ be a class of estimators; in Backus–Gilbert resolution, $\mathcal{E}_K$ is the class of linear estimators. More generally, we might consider arbitrary measurable estimators. Or we might restrict $\mathcal{E}_K$ to estimators whose computational cost does not exceed a given constant. Let $T : \mathcal{G} \to \mathbb{R}^+$ be a cost function on $\mathcal{G}$; T plays the role of 'spread' in Backus–Gilbert resolution. Let $\rho_\eta(\hat{g}, g)$ be the risk of using the estimator $\hat{g}$ to estimate $g[\theta]$ when in fact $\theta = \eta$; ρ plays the role of MSE.

Definition 1. *The (intrinsic minimax) T-resolution for parameters in $\mathcal{G}$ and estimators in $\mathcal{E}_K$ at ρ-risk M is*

$$R_\Theta(M; T, K, \rho, \mathcal{G}, \mathcal{E}_K) \equiv \inf_{g \in \mathcal{G}: \rho_\Theta(\mathcal{E}_K; g) \leqslant M} T(g). \tag{17}$$

That is, $R_\Theta(M; T, K, \rho, \mathcal{G}, \mathcal{E}_K)$ is the smallest value of $T(g)$ among parameters $g \in \mathcal{G}$ that can be estimated with maximum ρ-risk not exceeding M over models in Θ by some estimator $\hat{g}$ in $\mathcal{E}_K$. It is implicit that the estimator $\hat{g}$ uses $\mathcal{X}_K$-valued data X that are a sample from some $\mathbb{P}_{\eta,K} \in \mathcal{P}_K$ and that $\mathcal{G}$ are $\mathcal{B}$-valued functions; R also depends on $\mathcal{X}_K$ and $\mathcal{P}_K$.

Whether the optimization problem (17) can be solved—and whether the solution is scientifically useful—depends on T, K, Θ, $\mathcal{E}_K$, $\mathcal{B}$, $\mathcal{G}$ and $\mathcal{P}_K = \{\mathbb{P}_{\eta,K} : \eta \in \Theta\}$, but the abstract problem makes sense even if

- the model space $\mathcal{T}$ is not a Hilbert space;
- there are constraints, so $\Theta \neq \mathcal{T}$;
- the set of possible data values $\mathcal{X}$ is not $\mathbb{R}^n$;
- the forward problem is nonlinear;
- the set $\mathcal{G}$ of parameters includes nonlinear parameters;
- the possible parameter values $\mathcal{B}$ are not real numbers;
- the elements of $\mathcal{G}$ do not have unbiased estimators;
- the set $\mathcal{E}_K$ of estimators includes nonlinear estimators and biased estimators;
- the penalty T is not a measure of spread;
- the measure of risk is not mean squared error.

Definition (17) coincides with the Backus–Gilbert resolution at r_0 if $\mathcal{T}$ is a Hilbert space of functions of position, $\Theta = \mathcal{T}$, $\mathcal{X} = \mathbb{R}^n$, $\mathbb{P}_{\eta,K} \sim K[\eta] + \epsilon$ where $K = (\kappa_j)_{j=1}^n \subset \mathcal{T}$, $E\epsilon = 0$, $E(\epsilon\epsilon') = \Sigma$, $\mathcal{B} = \mathbb{R}$, $\mathcal{G} = \mathcal{T}$, $\mathcal{E}_K$ is the class of linear functionals on $\mathcal{X}_K$ ($\beta \cdot X$, $\beta \in \mathbb{R}^n$), $\rho(\hat{g}; g) = \mathrm{MSE}(\hat{g}; g)$, and $T(g)$ is one of the Backus–Gilbert measures of spread around r_0. The definition also reproduces some existing extensions of Backus–Gilbert resolution, including those in [14, 24, 29, 30], which amount to taking $T(g) \equiv \|g - g_0\|^2$, for a fixed function g_0.

Definition (17) also leads to an alternative definition of the resolution of an estimator:

Definition 2. *The T-resolution of the estimator $\hat{g}$ for parameters in $\mathcal{G}$ at ρ-risk M is*

$$R_\Theta(\hat{g}, M; T, K, \rho, \mathcal{G}) \equiv \inf_{g \in \mathcal{G}: \rho_\Theta(\hat{g}; g) \leqslant M} T(g). \tag{18}$$

The intrinsic minimax resolution $R_\Theta(M; T, K, \rho, \mathcal{G}, \mathcal{E}_K)$ is related to the resolution of $\hat{g}$, $R_\Theta(\hat{g}, M; T, K, \rho, \mathcal{G})$, through

$$R_\Theta(M; T, K, \rho, \mathcal{G}, \mathcal{E}_K) = \inf_{\hat{g} \in \mathcal{E}_K} R_\Theta(\hat{g}, M; T, K, \rho, \mathcal{G}). \tag{19}$$

3.3. Bayes resolution

We can also define measures of resolution using Bayes risk rather than maximum (or minimax) risk.

Definition 3. *The intrinsic Bayes T-resolution for parameters in $\mathcal{G}$ and estimators in $\mathcal{E}_K$ at ρ-risk M for prior π is*

$$R_\pi(M; T, K, \rho, \mathcal{G}, \mathcal{E}_K) \equiv \inf_{g \in \mathcal{G}: \rho_\pi(\mathcal{E}_K; g) \leqslant M} T(g). \tag{20}$$

Definition 4. *The Bayes T-resolution of the estimator $\hat{g}$ for parameters in $\mathcal{G}$ at ρ-risk M for prior π is*

$$R_\pi(\hat{g}, M; T, K, \rho, \mathcal{G}) \equiv \inf_{g \in \mathcal{G}: \rho_\pi(\hat{g}; g) \leqslant M} T(g). \tag{21}$$

The intrinsic Bayes resolution and the Bayes resolution of $\hat{g}$ for prior π are related through

$$R_\pi(M; T, K, \rho, \mathcal{G}, \mathcal{E}_K) = \inf_{\hat{g} \in \mathcal{E}_K} R_\pi(\hat{g}, M; T, K, \rho, \mathcal{G}). \tag{22}$$

3.4. Confidence-based resolution

The new definitions of resolution do not require $\mathcal{E}_K$ to be a collection of point estimators. For example, $\mathcal{E}_K$ could be a collection of confidence interval estimators. The new definitions thus encompass new measures of the resolution for confidence sets, by picking $\rho_\eta(\hat{g}; g)$ suitably when $\hat{g}$ is a confidence set.

Let $\mathcal{E}_K$ be a collection of randomized set-valued estimators and for fixed $\alpha \in (0, 1)$ define

$$\rho_\eta(\hat{g}; g[\eta]) \equiv \begin{cases} \infty, & \mathbb{P}_{\eta,K}\{\hat{g}(X) \ni g[\eta]\} < 1 - \alpha \\ \mathbb{E}_\eta \mu_\eta(\hat{g}(X)), & \mathbb{P}_{\eta,K}\{\hat{g}(X) \ni g[\eta]\} \geqslant 1 - \alpha. \end{cases} \tag{23}$$

Here $\mu_\eta(S)$ is the measure of the set S, and the expectations and probabilities are with respect to the distribution of the data X and the distribution of $\hat{g}(x)$ for each fixed x when the model is η. It is possible—and sometimes scientifically desirable—to allow the measure μ_η to depend on η [32]. Then the intrinsic minimax resolution $R_\Theta(M; T, K, \rho, \mathcal{G}, \mathcal{E}_K)$ is the smallest value of $T(g)$ among parameters $g \in \mathcal{G}$ that can be estimated by a randomized confidence set in the class $\mathcal{E}_K$ that has coverage probability at least $1 - \alpha$ for all $\eta \in \Theta$ and has maximum expected μ_η-measure not larger than M. There is an analogous definition of intrinsic resolution based on Bayesian credible regions, and a definition based on a Bayesian average size of frequentist confidence regions with respect to a prior.

Techniques for finding confidence intervals for one or more parameters of the model, such as 'strict bounds' [12, 21, 35, 37] and 'funnel functions' [25], fit into this framework. Moreover, the framework suggests how those techniques might be generalized and optimized to provide sharper inferences.

Strict bounds find simultaneous confidence intervals for a collection of pre-specified parameters of the model, $\{g_j[\theta]\}_{j \in \mathcal{J}}$. The confidence intervals are projections of a confidence region for the whole model θ. Suppose we observe $X = K[\theta] + \epsilon$ where the forward mapping $K : \Theta \to \mathbb{R}^n$ can be nonlinear, $\theta \in \Theta$, and the observational errors $\{\epsilon_j\}_{j=1}^n$ are modeled as independent zero-mean Gaussian random variables with unit variance. Let $\chi^2_{\alpha,n}$ denote the $1 - \alpha$ critical value of the χ-square distribution with n degrees of freedom. Then the ball $\left\{x \in \mathbb{R}^n : \|X - x\|_2^2 \leqslant \chi^2_{\alpha,n}\right\}$ is a $1 - \alpha$ confidence set for $K[\theta]$. Its pre-image, $\mathcal{D} \equiv \left\{\eta \in \Theta : \|X - K[\eta]\|_2^2 \leqslant \chi^2_{\alpha,n}\right\}$ is a confidence set for θ. The intervals

$$\left\{\left[\inf_{\eta \in \mathcal{D}} g_j[\eta], \sup_{\eta \in \mathcal{D}} g_j[\eta]\right]\right\}_{j \in \mathcal{J}} \tag{24}$$

are simultaneous $1 - \alpha$ confidence intervals for $\{g_j[\theta]\}$:

$$\mathbb{P}_{\nu,K}\left\{\cap_{j \in \mathcal{J}} \left\{\left[\inf_{\eta \in \mathcal{D}} g_j[\eta], \sup_{\eta \in \mathcal{D}} g_j[\eta]\right] \ni g_j[\nu]\right\}\right\} \geqslant 1 - \alpha, \qquad \forall \nu \in \Theta. \tag{25}$$

However, they are not typically the smallest (in expectation) $1 - \alpha$ confidence sets, because $\mathcal{D}$ is not tailored to the set $\mathcal{G}$ and the information $\theta \in \Theta$ is used in a crude way.

The resolution of strict bounds can be defined as follows. Let $\mathcal{G}$ be a collection of parameters on Θ; let $T : \mathcal{G} \to \mathbb{R}^+$. For example, $\mathcal{G}$ could be normalized boxcar averages centered at r_0 and $T(g)$ could be the width of the boxcar. Fix $\alpha \in (0, 1)$. Let $\ell_g(X)$ be the lower endpoint of the strict bounds confidence interval for $g[\theta]$, and let $u_g(X)$ be the upper endpoint. Define

$$\rho_\eta([\ell_g, u_g], g[\eta]) \equiv \mathbb{E}_\eta(u_g(X) - \ell_g(X)), \tag{26}$$

the expected Lebesgue measure of the strict bounds confidence interval for $g[\theta]$ when the true model is $\eta \in \Theta$. Then the T-resolution of strict bounds for parameters in $\mathcal{G}$ at maximum expected length M (using confidence level $1-\alpha$) is

$$R_\Theta(\text{strict bounds}, M; T, K, \rho, \mathcal{G}) \equiv \inf_{g\in\mathcal{G}:\rho_\Theta(\hat{g};g)\leqslant M} T(g). \tag{27}$$

In this cartoon, R is the width of the narrowest boxcar average centered at r_0 for which the expected length of a strict bounds $1-\alpha$ confidence interval is at most M, whatever be $\theta \in \Theta$. We could also define a 'conditional resolution' for strict bounds using attained length of the confidence intervals after the data are collected, rather than using the minimax expected length before the data are collected.

3.5. Design resolution

Another layer of optimization can be added to the definitions of intrinsic resolution: optimizing the set of observations to make. (See, e.g., [9, 31, 38].) Suppose we can control some aspect of the forward problem, for example, where on Earth's surface to place instruments or sources, the spectrum of a source wavelet for exploration seismology, the shape of an antenna for observing the cosmic microwave background, the shape of the 'galactic cut' for estimating the spectrum of the cosmic microwave background, the signal-to-noise ratio of a transducer, the integration time of a CCD collector, or the number of observations to make. Each such 'experimental design' K has a cost $\lambda(K)$ in time, money and other valuables. Changing the design K can change the space $\mathcal{X}_K$ of possible observations and the probability distribution of the data, $\mathbb{P}_{\theta,K}$.

Let $\mathcal{K}$ denote the set of experimental designs K we are willing to consider. For each experimental design K, there is a set $\mathcal{E}_K$ of estimators we are willing to use (for example, those with computational cost not greater than a given constant, or those with a specific kind of functional dependence on the data); we assume that all take values in the same action space $\mathcal{A}$. The risk ρ depends implicitly on K, but I make no notational distinction. Let $\mathcal{E} \equiv \{\mathcal{E}_K\}_{K\in\mathcal{K}}$.

Definition 5. *The minimax intrinsic design T-resolution for parameters in $\mathcal{G}$, estimators in $\mathcal{E}$ and designs in $\mathcal{K}$ at ρ-risk M and λ-cost L is*

$$R_\Theta(M, L; T, \rho, \mathcal{G}, \mathcal{E}, \mathcal{K}) \equiv \inf_{K\in\mathcal{K}:\lambda(K)\leqslant L} \inf_{g\in\mathcal{G}:\rho_\Theta(\mathcal{E}_K;g)\leqslant M} T(g). \tag{28}$$

Definition 6. *The Bayes intrinsic design T-resolution for parameters in $\mathcal{G}$, estimators in $\mathcal{E}$ and designs in $\mathcal{K}$ at ρ-risk M and λ-cost L for prior π is*

$$R_\pi(M, L; T, \rho, \mathcal{G}, \mathcal{E}, \mathcal{K}) \equiv \inf_{K\in\mathcal{K}:\lambda(K)\leqslant L} \inf_{g\in\mathcal{G}:\rho_\pi(\mathcal{E}_K;g)\leqslant M} T(g). \tag{29}$$

These definitions treat $\lambda(K)$ as deterministic. But the cost of K could also be random, with a distribution that could depend on unknown parameters. Indeed, the cost might depend on θ. If the cost is random, we might work with the Bayes expected cost or the minimax expected cost.

These definitions can also be adapted to find, for example, the lowest (fixed-, expected- or maximum-) cost design that would yield a given resolution at a given risk. The definitions of design resolution are related to quantities in information-based complexity (IBC) [38]. IBC seeks to find the optimal information (essentially, K) and optimal estimator (essentially, the best element of $\mathcal{E}$) to estimate a fixed parameter $g\theta$ within some error (usually a deterministic measure of risk). Design resolution adds an additional layer to the IBC problem: selecting the parameter g optimally. See section 4.3.

4. Examples

Consider a very simple inverse problem with a single datum linearly related to the unknown. The model space is $\mathcal{T} = L_2[-1, 1]$, Lebesgue square-integrable functions on the interval $[-1, 1]$. We observe

$$X = \int_{-1}^{1} \frac{1}{2}\theta(r)\,\mathrm{d}r + \epsilon, \tag{30}$$

with $\epsilon \sim N(0, \sigma^2)$. Because $\mathcal{T} = L_2[-1, 1]$ is its own dual, there is an isometry between bounded linear functionals on $\mathcal{T}$ and elements of $\mathcal{T}$: for any bounded linear functional $g : \mathcal{T} \to \mathbb{R}$, there is an element $g(r) \in \mathcal{T}$ such that for all $\eta \in \mathcal{T}$,

$$g[\eta] = \int_{-1}^{1} g(r)\eta(r)\,\mathrm{d}r. \tag{31}$$

Identify $g[\cdot]$, the linear functional, with $g(r)$, the element of $\mathcal{T}$ that 'represents' $g[\cdot]$. For $g \in \mathcal{T}$, define

$$T_0(g) \equiv 12 \int_{-1}^{1} r^2 g^2(r)\,\mathrm{d}r. \tag{32}$$

The functional $T_0(g)$ measures the spread of $g(r)$, the representer of g, around $r = 0$; it is one of the measures of concentration Backus and Gilbert consider. Let $\mathcal{G}$ be the set of linear functionals for which $T_0(g)$ is finite and $\int_{-1}^{1} g(r)\,\mathrm{d}r = 1$. Use MSE to measure risk, and take $\mathcal{E}_K$ to consist of estimators that are linear in X—that is, estimators of the form βX, $\beta \in \mathbb{R}$.

There are no constraints, so Backus–Gilbert theory says the only linear parameters that can be estimated with finite MSE are linear combinations of the measurement functionals. The only linear combination of the measurement functionals that corresponds to an average with unit area is $g_1(r) = \frac{1}{2}1_{[-1,1]}(r)$. Thus, $\mathcal{G}$ is the singleton set $\{g_1\}$. The spread of g_1 about $r = 0$ is

$$T_0(g_1) = 12 \int_{-1}^{1} r^2 (1/2)^2\,\mathrm{d}r = 2. \tag{33}$$

The corresponding unbiased estimator is $\hat{g}(X) = X$; its MSE as an estimator of $g[\theta]$ is equal to its variance, σ^2. If $\sigma^2 \leqslant M$, the resolution at $r = 0$ for MSE $\leqslant M$ is 2; otherwise, the resolution is infinite.

Define $g_\delta(r) = \frac{1}{2\delta}1_{[-\delta,\delta]}(r)$ for $\delta \in (0, 1]$, and let $g_\delta[\theta] \equiv \int_{-1}^{1} g_\delta(r)\theta(r)\,\mathrm{d}r$. The spread of g_δ about $r = 0$ is

$$T_0(g_\delta) = 12 \int_{-\delta}^{\delta} r^2 \left(\frac{1}{2\delta}\right)^2 \mathrm{d}r = 2\delta \leqslant 2. \tag{34}$$

For $\delta < 1$, $g_\delta[\theta]$ is not identifiable and there is no estimator of $g_\delta[\eta]$ that has finite MSE for all $\eta \in \Theta$.

4.1. Constraints

Define $\|\eta\|_\infty \equiv \operatorname{ess\,sup}|\eta(r)|$, $\eta \in \mathcal{T}$, where ess sup denotes the essential supremum. Suppose we have the additional constraint $\theta \in \Theta \equiv \{\eta \in \mathcal{T} : \|\eta\|_\infty \leqslant \tau\}$. Then $|g_\delta[\theta]| \leqslant \tau$, so the bias of estimators of $g_\delta[\theta]$ can be controlled. Let $\mathcal{G}$ be the set $\{g_\delta : \delta \in (0, 1]\}$ and let $\mathcal{E}_K$ be the set of affine estimators, $\mathcal{E}_K = \{\beta X + \gamma : \beta, \gamma \in \mathbb{R}\}$. Let $T(g_\delta) = \delta$, the half-width of the boxcar function g_δ. (Note that for boxcar functions, $T(g_\delta) = S(g_\delta)/2$.) We seek

$$R_\Theta(M; T, K, \mathrm{MSE}, \mathcal{G}, \mathcal{E}_K) = \inf_{\beta,\gamma\in\mathbb{R}\ \delta\geqslant 0:\sup_{\|\eta\|_\infty\leqslant\tau}\ \inf_{\mathrm{MSE}(g_\delta[\eta],\beta X+\gamma)\leqslant M}} \delta. \tag{35}$$

Note that this resolution problem has a constraint on the model, a restriction on the class of estimators $\mathcal{E}_K$, a restriction on the class of target parameters $\mathcal{G}$, and a different penalty measure (the half-width δ).

For $\delta \in (0, 1]$,

$$g_1[\eta] = \delta g_\delta[\eta] + \left(\int_{-1}^{-\delta} + \int_{\delta}^{1} \right) \frac{1}{2} \eta(t)\, \mathrm{d}t. \tag{36}$$

Thus

$$\sup_{\eta \in \Theta} |g_1[\eta] - \delta g_\delta[\eta]| = \tau(1 - \delta). \tag{37}$$

Estimating $g_\delta[\theta]$ from X is thus equivalent to estimating ν from the observation $X \sim N(\delta\nu + \varepsilon, 1)$ where $|\nu| \leqslant \tau$ and $|\varepsilon| \leqslant \tau(1 - \delta)$. (Identify $\nu = g_\delta[\theta]$ and $\varepsilon = g_1[\theta] - \delta g_\delta[\theta]$.) We can find the minimax MSE affine estimator by calculus.

By symmetry, the optimal additive constant γ in the affine estimator is zero, so the optimal affine estimator is linear. Let $Z \sim N(0, 1)$. The MSE of the linear estimator $\hat{g}_\delta(X) = \beta X$ of ν is

$$E_\eta[\beta\delta\nu + \beta\delta + \beta Z - \nu]^2 = ((\beta\delta - 1)\nu + \beta\delta)^2 + \beta^2. \tag{38}$$

For fixed β, this is largest when $(\beta\delta - 1)\nu$ and $\beta\delta$ have the same sign. When $\beta < 0$ or $\beta > 1/\delta$, the maximum is attained at $\nu = \tau$, $\delta = \tau(1 - \beta)$; for $0 \leqslant \beta \leqslant 1/\delta$, the maximum is attained at $\nu = -\tau$, $\delta = \tau(1 - \beta)$. Thus

$$\sup_{\eta \in \Theta} \mathrm{MSE}_\eta(g_\gamma[\eta], \beta X) = \begin{cases} \tau^2(\beta - 1)^2 + \beta^2, & \beta < 0 \text{ or } \beta > 1/\delta \\ \tau^2[1 + \beta(1 - 2\delta)]^2 + \beta^2, & 0 \leqslant \beta \leqslant 1/\delta. \end{cases} \tag{39}$$

We need to minimize this maximum MSE (which is continuous in β) by choosing β well. The top expression on the right-hand side of (39) is monotone decreasing for $\beta < 0$ and monotone increasing for $\beta > 1/\delta$. The second expression is monotone increasing on $[0, 1/\delta]$ if $\gamma < 1/2$, and has a minimum in $(0, 1/\delta)$ if $\delta > 1/2$. The optimal β, β^*, is thus 0 if $\delta \leqslant 1/2$, or the unconstrained minimum of $\tau^2[1 + \beta(1 - 2\delta)]^2 + \beta^2$ if $\delta > 1/2$:

$$\beta^* = \begin{cases} 0, & \delta \leqslant \dfrac{1}{2} \\ \dfrac{\tau^2(2\delta - 1)}{\tau^2(2\delta - 1)^2 + 1}, & \delta > \dfrac{1}{2}. \end{cases} \tag{40}$$

The corresponding affine minimax risk is

$$\mathrm{MSE}_\Theta(g_\beta, \beta X) = \begin{cases} \tau^2, & \delta \leqslant \dfrac{1}{2} \\ \dfrac{\tau^2}{1 + \tau^2(2\delta - 1)^2}, & \delta > \dfrac{1}{2}. \end{cases} \tag{41}$$

We can find the resolution by solving (41) for the smallest δ for which the minimax affine MSE is at most M:

$$R_\Theta(M; T, K, \mathrm{MSE}, \mathcal{G}, \mathcal{E}_K) = \begin{cases} \infty, & M < \min\left(\tau^2, \dfrac{\tau^2}{1 + \tau^2}\right) \\ 0, & M \geqslant \tau^2 \\ \dfrac{1}{2}\left(1 + \sqrt{\dfrac{1}{M} - \dfrac{1}{\tau^2}}\right), & \dfrac{\tau^2}{1 + \tau^2} \leqslant M < \tau^2. \end{cases} \tag{42}$$

Take $M = 1$, corresponding to the MSE of the linear estimator of g_1, In the units we have been using in this example, the Backus–Gilbert resolution is 1. But because of the constraint, the resolution of affine estimators is 0 if $\tau^2 \leqslant 1$, or $\left(1 + \sqrt{1 - \frac{1}{\tau^2}}\right)/2 < 1$ if $\tau^2 > 1$.

4.2. Resolution of 'strict bounds'

We shall calculate the resolution of the strict bounds $1 - \alpha$ confidence procedure—a nonlinear set estimator—for this inverse problem.

We first find the strict bounds confidence intervals as a function of the observation x and the half-width δ of the boxcar. Let z be the $1 - \alpha/2$ quantile of the normal distribution, so that if $Z \sim N(0, 1)$, $\mathbb{P}(|Z| > z) = \alpha$. Under the assumptions, the confidence set for θ is

$$\mathcal{D} = \{\eta \in \Theta : |g_1[\eta] - X| \leqslant z\}. \tag{43}$$

The strict bounds confidence interval for $g_\delta[\theta]$ is

$$\left[\inf_{\eta\in\mathcal{D}} g_\delta[\eta], \sup_{\eta\in\mathcal{D}} g_\delta[\eta]\right]. \tag{44}$$

This interval is empty if $|X| > \tau + z$; to simplify things, we shall take the confidence interval to consist of the point $-\tau$ if $X < -\tau - z$ and to consist of the point τ if $X > \tau + z$. This does not change the expected length of the interval, and—if $g_\delta[\theta]$ happens to be $\pm\tau$—increases the coverage probability.

As noted before, estimating $g_\delta[\theta]$ is equivalent to estimating ν from the observation $X \sim N(\delta\nu + \varepsilon, 1)$ where $|\nu| \leqslant \tau$ and $|\varepsilon| \leqslant \tau(1 - \delta)$. When $X = x$, the lower endpoint of the strict bounds confidence interval for $g_\delta[\theta]$ is

$$\ell_\delta(x) = \begin{cases} -\tau, & x \leqslant \tau(1 - 2\delta) + z \\ (x - z - (1 + \delta)\tau)/\delta, & \tau(1 - 2\delta) + z < x \leqslant \tau + z \\ \tau, & x > \tau + z. \end{cases} \tag{45}$$

The upper endpoint is

$$u_\delta(x) = \begin{cases} -\tau, & x \leqslant -\tau - z \\ (x + z + (1 + \delta)\tau)/\delta, & -\tau - z < x \leqslant -\tau(1 - 2\delta) - z \\ \tau, & x > -\tau(1 - 2\delta) - z. \end{cases} \tag{46}$$

The length of the confidence interval is

$$w_\delta(x) = u_\delta(x) - \ell_\delta(x) \tag{47}$$

$$= \begin{cases} 0, & x \leqslant -\tau - z \\ (x + z + (1 + 2\delta)\tau)/\delta, & -\tau - z < x \leqslant \tau(1 - 2\delta) + z \\ 2(z + (1 + \delta)\tau)/\delta, & \tau(1 - 2\delta) + z < x \leqslant -\tau(1 - 2\delta) - z \\ (-x + z + (1 + 2\delta)\tau)/\delta, & -\tau(1 - 2\delta) < x \leqslant \tau + z \\ 0, & x > \tau + z. \end{cases} \tag{47}$$

(The third condition can be met only if $\delta > (z + \tau)/2\tau > 1/2$.) The length decreases monotonically with $|x|$, so the expected length is maximized when $\delta\nu + \varepsilon = 0$. In that case,

$$\begin{aligned} \mathbb{E}_0 w_\delta(X) = 2[&\phi(\tau + z) - \phi(\tau(1 - 2\delta) + z(1 + 2\delta)\tau/\delta(\Phi(\tau + z) - \Phi(-\tau(1 - 2\delta))) \\ &+ 1_{\delta>(z+\tau)/2\tau}(z + (1 + \delta)\tau)/\delta(2\Phi(-\tau(1 - 2\delta) - z) - 1)]. \end{aligned} \tag{48}$$

The resolution is the smallest value of δ such that $\mathbb{E}_0 w_\delta(X) \leqslant M$. (That problem needs to be solved numerically, but the expected length is monotonic in δ, so the search is straightforward.) For illustration, the resolution is 0.697 for $\tau = 1, \alpha = 0.05$, and $M = 2 \times 1.96$—the length of a naive fixed-length confidence interval for $g_1[\theta]$. That is, for

$\tau = 1$ we can estimate $g_{0.697}[\theta]$ with a 95% confidence interval whose expected length does not exceed 2×1.96, whatever be $\theta \in \Theta$. Without the prior information $\theta \in \Theta$, we would only be able to estimate $g_1[\theta]$, and the length of the 95% confidence interval would be 2×1.96.

4.3. Design resolution: quadrature of Lipschitz functions

This section finds the minimax design resolution in a deterministic problem: numerical integration of a Lipschitz function over the shortest possible interval for which the spacing between samples is not too small and the maximum possible error is not too large.

Let Θ be the set of functions $\eta = \eta(r)$ on the interval $[0, 1]$ that satisfy the Lipschitz condition $|\eta(r) - \eta(r')| \leqslant C|r - r'|$ for a known constant C. For $\eta \in \Theta$ and $\delta \in (0, 1]$, define $g_\delta[\eta] \equiv \delta^{-1} \int_0^\delta \eta(r)\,\mathrm{d}r$. Let $\mathcal{G} \equiv \{g_\delta : \delta \in (0, 1]\}$, and define $T(g_\delta) = \delta$. Let $\mathbf{r}$ be any finite set $\{r_j\} \subset [0, 1]$; let $n \equiv \#\mathbf{r}$; and assume that the points r_j are ordered so that $r_1 < r_2 < \cdots < r_n$. Define $K_\mathbf{r}[\eta] = (\eta(r_j))_{j=1}^n$, that is, $K_\mathbf{r}$ samples η at the points in $\mathbf{r}$. Let $\mathcal{K}$ denote the set of all such forward mappings $K_\mathbf{r}$. We can observe $X = K[\theta]$ for any $K \in \mathcal{K}$—at a price. The cost of observing $K_\mathbf{r}[\theta]$ is

$$\lambda(K_\mathbf{r}) \equiv \max_{j=2}^{n} (r_j - r_{j-1})^{-1}. \tag{49}$$

This is a qualitative model for a situation where technology for sampling faster or on a finer scale costs more, but for any given technology, each observation has negligible cost.

Since there is no observational error in this model, $\mathbb{P}_{\eta,K}$ is a point mass at $(\eta(r_j))_{j=1}^n$. Define

$$\rho_\eta(\hat{g}; g[\eta]) \equiv \mathbb{E}_\eta|\hat{g} - g[\eta]| = |\hat{g} - g[\eta]|. \tag{50}$$

We consider the set $\mathcal{E}$ of estimators with arbitrary functional dependence on the data. From the development in [26, 38], it is clear that the worst-case data are $X = 0$ and that the optimal estimator is linear, a modified trapezoid rule with the points $\{r_j\}$ chosen non-adaptively.

For a given set of points $\{r_j\}$, let $n_\delta \equiv \max\{j : r_j < \delta\}$ and $\bar{r}_\delta = (r_{n_\delta} + r_{n_\delta+1})/2$. The maximum risk of the modified trapezoid rule for data $X = 0$ is

$$\sup_{\eta\in\Theta:\eta(r_j)=0} \delta^{-1} \int_0^\delta \eta(r)\,\mathrm{d}r = \frac{C}{\delta}\left[\frac{r_1^2}{2} + \sum_{j=2}^{n_r} \frac{(r_j - r_{j-1})^2}{4} + \left\{\begin{array}{ll} \dfrac{(\delta - r_{n_\delta})^2}{2}, & \delta \leqslant \bar{r}_\delta \\ \dfrac{(r_{n_\delta+1} - r_{n_\delta})^2}{4} - \dfrac{(r_{n_\delta+1} - \delta)^2}{2}, & \delta > \bar{r}_\delta \end{array}\right\}\right]. \tag{51}$$

The maximum risk decreases as the spacing between observations shrinks. There is no cost for taking more samples at the minimum spacing L, so: (1) the observations might as well be equally spaced by L, (2) the first sample is at $r_1 < L$, and (3) unless there is a sample at δ, there might as well be a sample at a point $r_j > \delta$ if $\delta - r_{n_\delta} > L/2$ (but taking more than one does not help). Thus the optimal $\mathbf{r}$ is of the form $r_j = r_1 + (j - 1)L$, $j = 1, \ldots, n$. Moreover, one can show that for the optimal $\mathbf{r}$, $n = \lfloor \frac{\delta}{L} \rfloor$ and $r_1 = (\delta - nL)/2$. The maximum risk of the modified trapezoid rule for the optimal $\mathbf{r}$ is thus

$$\sup_{\eta\in\Theta} |\hat{g}_\delta - g_\delta(\eta)| = \frac{C}{4\delta}\left[4r_1^2 + (n - 1)L^2\right]. \tag{52}$$

The minimax design resolution in this problem is therefore

$$R_\Theta(M, L; T, \rho, \mathcal{G}, \mathcal{E}, \mathcal{K}) = \min \left\{ \delta \in (0, 1] : \frac{CL^2}{4\delta} \left[\left(\frac{\delta}{L} - \left\lfloor \frac{\delta}{L} \right\rfloor \right)^2 + \left\lfloor \frac{\delta}{L} \right\rfloor - 1 \right] \leqslant M \right\}. \tag{53}$$

5. Discussion

The definitions of intrinsic resolution wrap minimax or Bayes estimation of a parameter g inside an additional optimization problem to pick g: find the parameter g with smallest $T(g)$ in the class $\mathcal{G}$ that can be estimated with minimax or Bayes ρ-risk not exceeding M using an estimator in $\mathcal{E}_K$. So, results on minimax or Bayes estimation can help find the intrinsic minimax or Bayes resolution. (For example, see [10] on minimax estimation of linear functionals in inverse problems in Hilbert spaces with convex constraints. Reference [36] gives an example in geomagnetism.) When $\mathcal{G}$ is a singleton set, calculating the resolution reduces to a classical minimax or Bayes estimation problem.

The definitions of intrinsic design resolution wrap minimax and Bayes estimation in a third layer: picking the set of observations to make, subject to a bound on their cost. Equivalently, they wrap information-based complexity problems [38] in an extra layer of optimization to select the parameter to estimate. When $\mathcal{G}$ is a singleton set, calculating the design resolution reduces to an information-based complexity problem.

For fixed g, subject to some technical conditions on ρ, Θ and $\mathcal{E}_K$, minimax risk and Bayes risk are related: the minimax risk is the Bayes risk for the 'least favorable prior'. That is,

$$\rho_\Theta(\mathcal{E}_K; g) = \sup_{\pi \in \Pi} \rho_\pi(\mathcal{E}_K; g). \tag{54}$$

See, e.g., [11, 12, 19, 20]. Reference [32] give a numerical algorithm for building minimax estimators by constructing a least-favorable prior iteratively using numerical optimization and Markov Chain Monte Carlo (MCMC).

In the definitions developed here, 'resolution' might have nothing to do with estimating a local property of θ, depending on the class $\mathcal{G}$, the possible parameter values $\mathcal{B}$ and the penalty T. For example, estimating a difference in averages in two parts of the model [14] fits the definition.

5.1. Systematic errors

The new definitions can incorporate systematic errors, including uncertainties in the forward problem, by introducing nuisance parameters into the stochastic model. (Nuisance parameters are parameters that affect the probability distribution of the data, but are not themselves of interest.) This can be helpful when the forward problem involves a numerical simulation, for example. Here is a cartoon that incorporates errors in the forward model. Suppose the observations are

$$X = K[\theta] + \epsilon \in \mathbb{R}^n, \tag{55}$$

where K might be nonlinear. We know *a priori* that $\theta \in \Theta$ and that $K \in \mathcal{K}$, but we do not know which element of $\mathcal{K}$ was used to take the measurements. The n-vector ϵ of additive errors has a known distribution that does not depend on the value of θ or K. Let K_0 be a fixed element of $\mathcal{K}$. For $\eta \in \Theta$ and $K_0 \in \mathcal{K}$, define

$$\sigma(K_0, \mathcal{K}, \eta) \equiv \{K'[\eta] - K_0[\eta] : K' \in \mathcal{K}\}. \tag{56}$$

This lets us embed the problem into the current formalism by extending the model η to include an n-vector of nuisance parameters, the differences between $K[\eta]$ and $K_0[\eta]$. The set Θ expands to be the set $\Theta^*_{K_0} \equiv \{(\eta, \eta') : \eta \in \Theta, \eta' \in \sigma(K_0, \mathcal{K}, \eta\}$. The new forward model is

$$X = K^*[\theta^*] + \epsilon, \tag{57}$$

where $\theta^* \in \Theta^*$ and for $\eta^* = (\eta, \eta') \in \Theta^*$, $K^*[\eta^*] = K_0[\eta] + \eta'$. This treats the set $\mathcal{K}$ as deterministic; it is straightforward to treat it as stochastic instead. Other sources of systematic error can be treated similarly.

5.2. Application notes

To use the new definitions in practice, one must specify a number of sets and functions. For example, the intrinsic minimax resolution depends on the class $\mathcal{G}$ of target parameters, the penalty T, the set $\mathcal{E}_K$ of estimators, the risk function ρ, and the tolerable risk M. Science and taste are involved in choosing all of these. Both $\mathcal{G}$ and T relate most closely to the science: $\mathcal{G}$ should contain parameters that are scientifically interesting, and T imposes a preference order on those parameters. Choosing the risk function ρ determines the sense in which one wants to estimate those parameters well, and M specifies just how well. The choice of the set $\mathcal{E}_K$ of estimators might be driven by the theoretical and computational tractability of the resulting optimization problem, or by the cost of calculating the estimators.

Finding the minimax risk and Bayes risk for estimating each fixed $g \in \mathcal{G}$ are standard problems in statistical decision theory. Similarly, for fixed g, finding optimal information $K \in \mathcal{K}$ and an optimal estimator $\hat{g} \in \mathcal{E}_K$ are standard problems in information-based complexity. So, there is a body of theory that can be drawn upon to solve some of the optimization problems in the definitions of resolution and design resolution. However, specific results depend on properties of the sets and functions involved: linearity, convexity, boundedness, and so on. For many combinations of T, $\mathcal{G}$, $\mathcal{K}$, $\mathcal{E}_K$, ρ and $\mathcal{P}_K$, the optimization problems will be intractable.

6. Conclusions

The Backus–Gilbert notion of resolution can be generalized to apply to nonlinear inverse problems, inverse problems with constraints, measures of optimality other than spatial concentration, parameters other than local averages, nonlinear and biased estimators, set estimators (and others) in addition to point estimators, systematic as well as stochastic errors, and measures of risk other than variance of unbiased point estimators. There are both minimax and Bayesian versions of the resolution. The minimax version reproduces Backus–Gilbert resolution, SOLA and some other measures of resolution as special cases, but also allows some interesting new measures, such as resolution based on confidence sets for restricted parameters. The generalization of resolution wraps classical optimal estimation problems—minimax or Bayes estimation of a parameter—within another optimization problem of selecting the parameter to estimate. The generalized resolution problem is to select the optimal parameter to estimate, among parameters that can be estimated with minimax or Bayes risk bounded by M.

One can also define minimax and Bayesian 'design resolution' problems. These embed the generalized resolution problem within another optimization problem: selecting the best observations to make, subject to a constraint on the cost of making the observations. That is, the design resolution problem is to select the optimal parameter to estimate, among parameters that can be estimated with minimax or Bayes risk bounded by M, using observations whose total

cost is bounded by L. This amounts to wrapping a problem in information-based complexity within the problem of selecting an optimal parameter to estimate.

Acknowledgments

I am grateful to Andrew Curtis and David Freedman for helpful conversations, and to Janne Huttunen, Luis Tenorio and two anonymous referees for comments on a previous draft. This work was supported by the AX Division at the Lawrence Livermore National Laboratory through the Department of Energy under contract W-7405-Eng-48.

References

[1] Backus G 1970 Inference from inadequate and inaccurate data: I. *Proc. Natl Acad. Sci.* **65** 1–7
[2] Backus G 1970 Inference from inadequate and inaccurate data: II. *Proc. Natl Acad. Sci.* **65** 281–7
[3] Backus G 1970 Inference from inadequate and inaccurate data: III. *Proc. Natl Acad. Sci.* **67** 282–9
[4] Backus G and Gilbert F 1968 The resolving power of gross Earth data *Geophys. J. R. Astron. Soc.* **16** 169–205
[5] Backus G E 1988 Comparing hard and soft prior bounds in geophysical inverse problems *Geophys. J.* **94** 249–61
[6] Backus G E 1989 Confidence set inference with a prior quadratic bound *Geophys. J.* **97** 119–50
[7] Backus G E and Gilbert F 1970 Uniqueness in the inversion of inaccurate gross Earth data *Phil. Trans. R. Soc. Lond. A.* **266** 123–92
[8] Casella G, Hwang J T G and Robert C 1993 A paradox in decision-theoretic interval estimation *Statistica Sin.* **3** 141–55
[9] Chaloner K and Verdinelli I 1995 Bayesian experimental design: a review *Stat. Sci.* **10** 273–304
[10] Donoho D L 1994 Statistical estimation and optimal recovery *Ann. Stat.* **22** 238–70
[11] Evans S N, Hansen B and Stark P B 2005 Minimax expected measure confidence sets for restricted location parameters *Bernoulli* **11** 571–90
[12] Evans S N and Stark P B 2002 Inverse problems as statistics *Inverse Problems* **18** R1–R43 (invited)
[13] Gough D O 1985 Inverting helioseismic data *Sol. Phys.* **100** 65–99
[14] Gough D O, Sekii T and Stark P B 1996 Inferring spatial variations in solar properties from helioseismic data *Astrophys. J.* **459** 779–91
[15] Green W R 1975 Inversion of gravity profiles by use of a Backus–Gilbert approach *Geophysics* **40** 763–72
[16] Hon Y C and Wei T 2001 Backus–Gilbert algorithm for the Cauchy problem of the Laplace equation *Inverse Problems* **17** 261–71
[17] Huestis S P 1980 The inversion of heat flow data: a Backus–Gilbert formulation *Geophys. J. Int.* **62** 649–60
[18] Jeffrey W 1988 Inversion of helioseismology data *Astrophys. J.* **327** 987–92
[19] Le Cam L 1986 *Asymptotic Methods in Statistical Decision Theory* (New York: Springer)
[20] Lehmann E L and Casella G 1998 *Theory of Point Estimation 2nd edn* (New York: Springer)
[21] Medin A E, Parker R L and Constable S 2007 Making sound inferences from geomagnetic sounding *Phys. Earth Planet. Inter.* **160** 51–9
[22] Oldenburg D W 1976 Calculation of Fourier transforms by the Backus–Gilbert method *Geophys. J. Int.* **42** 413–31
[23] Oldenburg D W 1979 One-dimensional inversion of natural source magnetotelluric observations *Geophysics* **44** 1218–44
[24] Oldenburg D W 1981 A comprehensive solution to the linear deconvolution problem *Geophys. J. R. Astron. Soc.* **65** 331–57
[25] Oldenburg D W 1983 Funnel functions in linear and nonlinear appraisal *J. Geophys. Res.* **88** 7387–98
[26] Packel E W 1988 Do linear problems have linear optimal algorithms? *SIAM Rev.* **30** 388–403
[27] Parker R L 1971 The inverse problem of electrical conductivity in the mantle *Geophys. J. R. Astron. Soc.* **22** 121–38
[28] Parker R L 1994 *Geophysical Inverse Theory* (Princeton, NJ: Princeton University Press)
[29] Pijpers F P and Thompson M J 1992 Faster formulations of the optimally localized averages method for helioseismic inversions *Astron. Astrophys.* **262** L33–6
[30] Pijpers F P and Thompson M J 1994 The SOLA method for helioseismic inversion *Astron. Astrophys.* **281** 231–40
[31] Pukelsheim F 1993 *Optimal Design of Experiments* (New York: Wiley)

[32] Schafer C M and Stark P B 2006 Constructing Confidence Sets of Optimal Expected Size, *Tech. Report* 836, Department of Statistics, Carnegie Mellon University
[33] Schomburg B and Berendt G 1987 On the convergence of the Backus-Gilbert algorithm *Inverse Problems* **3** 341–6
[34] Snieder R 1991 An extension of Backus–Gilbert theory to nonlinear inverse problems *Inverse Problems* **7** 409–33
[35] Stark P B 1992 Inference in infinite-dimensional inverse problems: discretization and duality *J. Geophys. Res.* **97** 14,055–14,082
[36] Stark P B 1992 Minimax confidence intervals in geomagnetism *Geophys. J. Int.* **108** 329–38
[37] Stark P B, Parker R L, Masters G and Orcutt J A 1986 Strict bounds on seismic velocity in the spherical Earth *J. Geophys. Res.* **91** 13,892–13,902
[38] Traub J F and Werschulz A G 1998 *Complexity and Information* (Cambridge: Cambridge University Press)
[39] Westwater E R and Cohen A 1973 Application of Backus-Gilbert inversion technique to determination of aerosol size distribution from optical scattering measurements *Appl. Opt.* **12** 1340–8