Theory of model-based geophysical survey and experimental design

Part 2—nonlinear problems

ANDREW CURTIS, *Schlumberger Cambridge Research, U.K.*

This is the second of a two-part tutorial (the first part was in October's *TLE*). It provides a theoretical framework from the field of statistical experimental design (SED, a field of statistics) within which model-based survey and experimental design problems and methods can be understood. Specifically, these tutorials describe methods pertinent to the detection and inference of models of physical properties of rocks in the laboratory, or of the earth. Most of this part can be understood without reading Part 1, but if any discussion is unclear, it is likely that reading Part 1 will help.

The choice of method to use to design experiments depends greatly on how one can measure information about a model that is expected to be gleaned from the data acquisition. This in turn depends principally on whether the relationship between those data and the model parameters of interest is approximately linear, or significantly nonlinear. Consequently, Part 1 of this tutorial dealt with cases where this relationship is approximately linear; Part 2 deals with theory for nonlinear problems.

Since Part 1 and Part 2 are sequential, the equation numbers and figure numbers continue from Part 1. Part 1 had equations 1-8 and, hence, Part 2 starts at equation 9; Part 1 had Figures 1-9 and Part 2 starts at Figure 10.

This part begins with a brief recap of key concepts in linear design theory that were covered in Part 1, including two equations without which Part 2 is incomprehensible. A short section then reviews aspects of probability theory; these are used to quantify information in nonlinear design problems. The section thereafter creates a general, theoretical framework of design for nonlinear model-data relationships, followed by a section that covers an approximation and an alternative to the full theory that may be employed in some special cases. Finally, a discussion of profitable areas of future research and development in model-based design theory completes the tutorial.

Recap of Part 1. Models of interest were presented in vector notation— $\mathbf{m}=[m_1,...,m_P]$, i.e., it was assumed that there are P elements of a model that need to be estimated, using data from the experiment or survey to be designed. It was also assumed that any survey design can be described by a vector S (S might describe equipment location and type, for example). The set of surveys that can reasonably be carried out given logistical and financial constraints was represented by Σ. Then survey S can be carried out if S is a member of Σ.

The relationship between the earth model $\mathbf{m}$ and the data $\mathbf{d}$ that would be recorded if a model $\mathbf{m}$ was true is $\mathbf{F_S}$:

$$\mathbf{d} = \mathbf{F_S(m)} \qquad (2)$$

As the notation $\mathbf{F_S}$ implies, the form of the model-data relationship (often referred to as the forward function) depends on the survey design $\mathbf{S}$. Part 1 showed that most
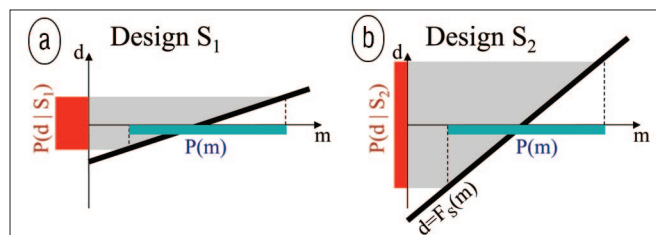

Figure 10. *Similar to top plots in Figure 1 (in Part 1) except that in this case the prior model pdf is fixed (blue-green) and projected through the forward function to produce the prior data pdf (red). (a) and (b) show different linear forward functions.*

survey design theory can be thought of as a set of methods that attempt to optimize the relationship $\mathbf{F_S}$, such that expected information about the model parameters $\mathbf{m}$ provided by data $\mathbf{d}$ is maximized.

If the forward function is linear, then equation 2 can be represented by a linearized system of equations,

$$\mathbf{d} = A_S\mathbf{m} \qquad (4)$$

where A_S is a matrix. Then, Part 1 of this tutorial showed that experiments could be designed using the criterion that the most informative experiment would be the one that maximized some measure of positivity of the eigenvalue spectrum of matrix $A_S^T A_S$ (T stands for transpose). This is because the eigenvalue spectrum (together with corresponding eigenvectors) describes exactly what information will be transferred from the expected data measurements into information about the model parameters of interest. Equations 7 in Part 1 provide example measures of eigenvalue positivity, and equations 8 gave measures adapted so that resulting survey designs would focus information on a subspace of the model space (for example, on a subset of the model parameters).

When the relationship $\mathbf{F_S(m)}$ in equation 2 is nonlinear, it can not be represented by constant-matrix equation 4. It is therefore necessary to use different measures of how information is transferred from data to model space (Ford et al., 1989). In order to construct such measures, it is useful at this point to move to a probabilistic framework to describe data and model constraints. A brief revision of probability density functions is given next. This leads to the construction of an entropy-based measure of information that can be used in linear, and in nonlinear design problems.

Probability distributions. The following notation will be used: A capital variable name (e.g., X) will represent a random variable (that is, a variable whose value varies randomly according to some probability distribution). Associated with each random variable X is a probability density function (pdf; defined below) describing how X varies, represented by $P(x)$. The probability of an event is repre-

sented by Pr(...).

Pdfs are always positive or zero, and have the following relationships to probabilities (presented here for a vector random variable $\mathbf{X}$):

$$\Pr(\mathbf{a} \leq \mathbf{X} \leq \mathbf{b}) = \int_{\mathbf{a}}^{\mathbf{b}} P(\mathbf{x})\, d\mathbf{x} \tag{9}$$

$$\Pr(-\infty \leq \mathbf{X} \leq \infty) = \int_{-\infty}^{\infty} P(\mathbf{x})\, d\mathbf{x} = 1 \tag{10}$$

for real vectors $\mathbf{a}$ and $\mathbf{b}$, where elements $b_i > a_i$ for all i up to the dimensionality of $\mathbf{X}$, and where the integrals run over each dimension of the vector $\mathbf{x}$. Equation 9 shows that the pdf contains all information necessary to calculate the probability of any particular event occurring (e.g., the event that $\mathbf{X}$ lies between $\mathbf{a}$ and $\mathbf{b}$). Indeed, this equation constitutes the definition of a pdf. Equation 10 shows that pdfs are normalized so that the total probability that $\mathbf{X}$ takes any value at all is 1.

If $\mathbf{X}$ and $\mathbf{Y}$ are two random variables, then the joint random variable that describes the variation of both together is denoted $(\mathbf{X},\mathbf{Y})$ with pdf $P(\mathbf{x},\mathbf{y})$. Given knowledge of the joint distribution function, the marginal distribution of $\mathbf{X}$ and the conditional distribution of $\mathbf{X}$ given knowledge of the value of $\mathbf{Y}$ [denoted $(\mathbf{X}|\mathbf{Y})$], are respectively defined as

$$P(\mathbf{x}) = \int_{-\infty}^{\infty} P(\mathbf{x}, \mathbf{y})\, d\mathbf{y} \tag{11}$$

$$P(\mathbf{x}|\mathbf{y}) = \frac{P(\mathbf{x}, \mathbf{y})}{P(\mathbf{y})} \tag{12}$$

The marginal distribution of $\mathbf{X}$ in equation 11 is obtained by integrating the joint pdf over all possible values of $\mathbf{Y}$. Thus we obtain the pdf of $\mathbf{X}$ given no knowledge at all of the value of $\mathbf{Y}$. At the other extreme, the conditional distribution in equation 12 provides the pdf of $\mathbf{X}$ with complete knowledge of the value of $\mathbf{Y}$. This is given by the joint distribution $(\mathbf{X},\mathbf{Y})$ divided by the marginal distribution of $\mathbf{Y}$.

The expectation of a function $\mathbf{f}(\mathbf{X})$ of a random variable $\mathbf{X}$ is defined by

$$E[\mathbf{f}(\mathbf{X})] = \int_{-\infty}^{\infty} P(\mathbf{x})\mathbf{f}(\mathbf{x})\, d\mathbf{x} \tag{13}$$

The expectation of a function of a random variable is the function's average or mean value.

Entropy, information, and nonlinear design theory. Shannon (1948) showed that there is a unique measure of the amount of information about $\mathbf{X}$ within the associated pdf $P(\mathbf{x})$ which satisfies various desirable properties of such a measure. This measure $I(\mathbf{X})$ is usually defined in terms of the entropy $Ent(\mathbf{X})$ of a continuous random variable $\mathbf{X}$ as

$$I(\mathbf{X}) = -Ent(\mathbf{X}) = \int_{-\infty}^{\infty} P(\mathbf{x}) \log(P(\mathbf{x}))\, d\mathbf{x} = E\big[\log(P(\mathbf{x}))\big] \tag{14}$$

where the integral exists. The information and entropy are negatives of each other: While $I(\mathbf{X})$ measures information about $\mathbf{X}$, $Ent(\mathbf{X})$ measures uncertainty in $\mathbf{X}$.

Say we partition the random vector $\mathbf{X}$ into two parts, $\mathbf{X} = [\mathbf{X_H}, \mathbf{X_{H'}}]$ where $\mathbf{H}$ represents the set of indices that are in the first part and $\mathbf{H'}$ those in the second part. Then using equation 12 to expand $P(\mathbf{X_H}, \mathbf{X_{H'}})$ and inserting this into equation 14 we obtain the identity

$$I(\mathbf{X}) = I(\mathbf{X_H}) + E_{X_H}\big[\, I(\mathbf{X_{H'}}|\mathbf{X_H})\,\big] \tag{15}$$

where the expectation in this equation is taken by integrating only over $\mathbf{X_H}$ using equation 13, and $I(\mathbf{X_{H'}}|\mathbf{X_H})$ is the information in the conditional distribution of $\mathbf{X_{H'}}$ given $\mathbf{X_H}$ in equation 12.

Equation 15 is a general identity for partitioned random vectors. Hence, it will be true for any such vector that we create. Define $\mathbf{D}$ and $\mathbf{M}$ to be random vectors of the data measurements and model parameters respectively. That is, they are vectors that vary randomly according to uncertainties in the data and model parameter values, respectively. Clearly these random vectors depend on each other as we have already stated in equation 2: knowledge of $\mathbf{M}$ would change the data that we might expect to record, while knowledge about actual data recordings would place constraints on possible values of $\mathbf{M}$. Let $\mathbf{X} = (\mathbf{D},\mathbf{M})$ be the joint random vector. Then equation 15 implies that

$$I(\mathbf{D},\mathbf{M}) = I(\mathbf{D}) + E_{\mathbf{D}}\big[\, I(\mathbf{M}|\mathbf{D})\,\big] \tag{16}$$

This equation states that the total information in the model and data spaces together is given by the information in the data space plus the expected information in the model space given the data. In fact, it is possible that all of the terms in equation 16 depend on the experimental design $\mathbf{S}$ since the data distribution does. Since we wish to optimize information about the model by changing the design equation 16 can be rewritten so the dependence on $\mathbf{S}$ is made explicit:

$$I(\mathbf{D},\mathbf{M}|\mathbf{S}) = I(\mathbf{D}|\mathbf{S}) + E_{\mathbf{D}}\big[\, I(\mathbf{M}|\mathbf{D},\mathbf{S})\,\big] \tag{17}$$

Notice that the term $E_{\mathbf{D}}[I(\mathbf{M}|\mathbf{D},\mathbf{S})]$ represents exactly the information that we wish to maximize when designing an experiment. It represents the total amount of information that we expect in the model space given experimental data $\mathbf{D}$ acquired using design $\mathbf{S}$. When we design experiment $\mathbf{S}$ we have no data, so the expectation is over our pre-experimental (usually called *prior*) best estimate of the distribution of $\mathbf{D}$. We usually have little direct prior information about the data that will be recorded, but some can be derived indirectly from our prior information about the model.

The bad news about the term $E_{\mathbf{D}}[I(\mathbf{M}|\mathbf{D},\mathbf{S})]$ is that whereas the known forward function $\mathbf{F_S(m)}$ takes a particular model $\mathbf{m}$ and predicts the corresponding data $\mathbf{d}$ that would be recorded if that model was true, the distribution $\mathbf{M}|\mathbf{D},\mathbf{S}$ requires that information about the data $\mathbf{D}$ and design $\mathbf{S}$ is imposed on the model $\mathbf{M}$. This requires that the (unknown) inverse of function $\mathbf{F_S}$ is calculated, and hence this is called an inverse problem. Theory to solve such problems is called inverse theory, and in practical, geophysical, nonlinear problems it is a significant expense in terms of effort and computing power (indeed, in many nonlinear problems it is infeasible to calculate $\mathbf{M}|\mathbf{D},\mathbf{S}$). Hence, it would be better if we did not have to maximize the term directly.

Notice that if it was the case that the total information on $\mathbf{D}$ and $\mathbf{M}$ on the left side of equation 17 was independent of $\mathbf{S}$, then we could maximize term $E_{\mathbf{D}}[I(\mathbf{M}|\mathbf{D},\mathbf{S})]$ with respect to design $\mathbf{S}$ by minimizing term $I(\mathbf{D}|\mathbf{S})$—the calculation of which does not require an inversion. Below it is shown that the former condition holds for an important class of problems.

Since our initial definition of random vector $\mathbf{X} = [\mathbf{D},\mathbf{M}]$ to form equations 16 and 17 was arbitrary, let us now reverse the order of $\mathbf{M}$ and $\mathbf{D}$. Since $I(\mathbf{D},\mathbf{M}) = I(\mathbf{M},\mathbf{D})$, we obtain

$$I(\mathbf{D},\mathbf{M}|\mathbf{S}) = I(\mathbf{M}) + E_{\mathbf{M}}\big[\, I(\mathbf{D}|\mathbf{M},\mathbf{S})\,\big] \tag{18}$$

(e.g., Lindley, 1956) where we have dropped the dependence on **S** from the first term on the right since the prior pdf of **M** is independent of design **S**. Notice that this implies that the left side is also independent of design **S** if $I(\mathbf{D}|\mathbf{M},\mathbf{S})$ is. Sebastiani and Wynn (2000) showed that the latter is indeed the case for the general regression problem

$$\mathbf{d} = \mathbf{F_S}(\mathbf{m}) + \varepsilon \qquad (19)$$

for any $\mathbf{F_S}(\mathbf{m})$, if the errors in the data ε are also independent of the design **S**. This is because, for any **m**, the function $\mathbf{F_S}(\mathbf{m})$ is fixed so the entropy of **D**|**M,S** is equal to the entropy of ε. In such cases equation 17 can be written

$$I(\mathbf{D},\mathbf{M}) = I(\mathbf{D}|\mathbf{S}) + E_{\mathbf{D}}\left[I(\mathbf{M}|\mathbf{D},\mathbf{S})\right] \qquad (20)$$

While equation 19 and the condition that ε does not depend on design **S** may not be representative of all practical nonlinear design problems, wherever this can be used as an approximation, the implications are profound for nonlinear design: No inverse problems need be solved in order to design experiments. The constant left side of equation 20 with respect to design **S** implies that maximizing the expected information in the rightmost term of equation 20 (our experimental design objective) is exactly equivalent to minimizing the information in the first term on the right side, $I(\mathbf{D}|\mathbf{S})$.

Minimizing the information in **D**|**S** as a design criterion, or equivalently maximizing its entropy, lead to this method being called maximum entropy sampling (Shewry and Wynn, 1987). The beauty of maximizing the entropy of **D**|**S** is that calculation of the entropy requires only evaluation of the forward function in equation 19. This is because all that is needed is to calculate the prior pdf of the data **D** given design **S,** and calculate its entropy using equation 14.

The prior data pdf may be calculated as illustrated in Figure 10a for a linear problem. Let us represent our prior information about the model as a prior pdf, $P(\mathbf{m})$; in Figure 10 this is represented by a uniform distribution over the blue range of the model axis. If we project this through the forward function in equation 2, we obtain $P(\mathbf{d}|\mathbf{S})$, the prior data distribution (also uniform, represented in red). The design criterion to be maximized is simply the entropy of this distribution.

We can evaluate the prior data distribution using different designs and hence different forward functions $\mathbf{F_S}(\mathbf{m})$. For the design $\mathbf{S_2}$ in Figure 10b, we see that the forward function is steeper and consequently the uniform prior data distribution is wider. Hence, the information in the prior data distribution with design $\mathbf{S_2}$ is lower than the information with design $\mathbf{S_1}$ (the range of possible data d under design $\mathbf{S_1}$ is more restricted). Since minimizing the information $I(\mathbf{D}|\mathbf{S})$ is equivalent to maximizing the expected model information $E_{\mathbf{D}}[I(\mathbf{M}|\mathbf{D})]$, we deduce that $\mathbf{S_2}$ is the better design.

The analysis above made no assumptions about the form of the forward function $\mathbf{F_S}(\mathbf{m})$, and indeed is valid for fully nonlinear problems. Figure 11 illustrates exactly the same procedure applied to two nonlinear forward functions. Again, the information in the prior data distribution with design $\mathbf{S_2}$ is less than the information with design $\mathbf{S_1}$, and we deduce that design $\mathbf{S_2}$ is expected to supply greater information about the model **m**, and hence is the better design.

Notice that in the linear case in Figure 1 in Part 1, we projected expected data measurement uncertainties into the model space in order to evaluate the quality of any survey design—that is, we simulated carrying out an inversion. To
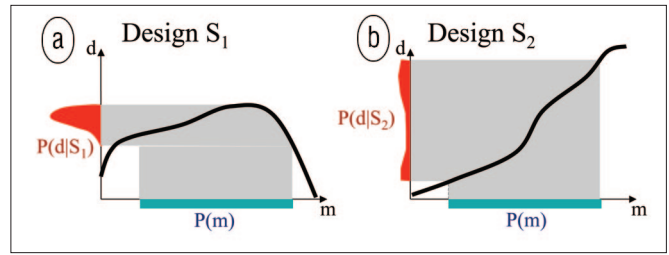

Figure 11. *Similar to Figure 10, but with nonlinear forward functions.*
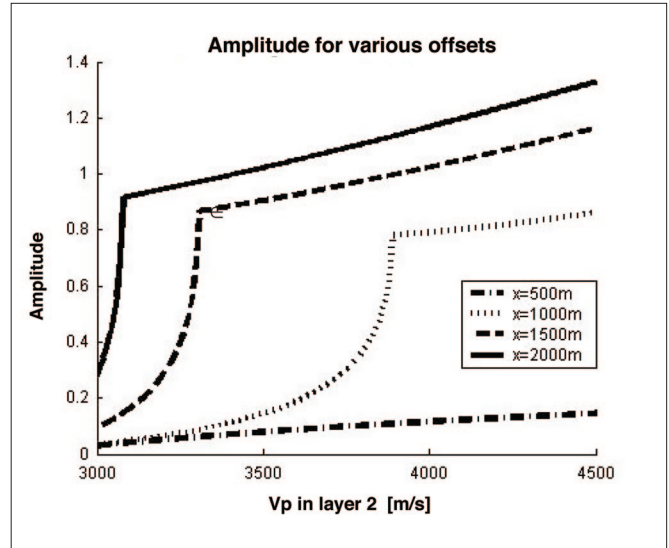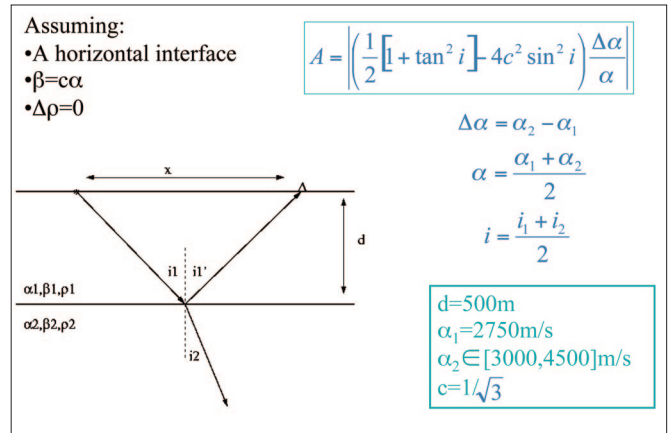

Figure 12. *The amplitude response of a reflection from an interface at various source-receiver offsets x (see Box 1). Offsets shown are 500 m, 1000 m, 1500 m, and 2000 m.*

Box 1: *Definition of an AVO-type problem. Lower-left, a layer-over-a-halfspace earth model is shown with source and receiver separated by offset x. The reflection ray paths to and from the interface at depth d have angles of incidence and reflection i_1. P velocity, S velocity and density in each layer are α_i, β_i and ρ_i, for i=1,2. The formula top-right gives the amplitude A of the reflection coefficient at the interface under assumptions shown top-left (where c is a positive constant, and $\Delta\rho=\rho_2-\rho_1$). Parameter values used for numerical examples of Figures 12 and 13 are shown in the lower right.*

evaluate the entropy criterion we do the reverse: We project prior model information into the data space (Figure 10). Since both linear and nonlinear surveys can be evaluated using the latter technique, this generalizes to nonlinear problems the linear concepts outlined in Part 1.

AVO example. Box 1 defines a simple AVO design problem in an elastic, two-layer earth model, that was considered by

van den Berg et al. (2003). The single source and single receiver have offset x. Assuming that the P velocity α of the top layer is known from other measurements, that density ρ is constant, and that shear velocity β is proportional to P velocity, the following is a one-parameter design problem: what source-receiver offset x provides most information about the velocities in the lower layer?

The forward function $d=F_S(m)$ between the model parameter m of interest (P-velocity in layer 2, α_2) and the amplitude data d observed at the receiver for an assumed unit impulse source are shown in Box 1 (top right). The survey design **S** in this case is defined by a single source-receiver offset (x). Figure 12 shows realizations of the forward function for four different offsets, for a depth to interface of 500 m and other parameters specified in Box 1 (lower right). This function is clearly nonlinear. Assuming that the prior model pdf is uniform along the entire length of the m axis in Figure 12, then additionally it is not easy to decide which would provide the best survey using intuition from linear design theory (steeper gradients are to be preferred) since the gradients change significantly within the range of m.

Van den Berg et al. projected this prior model uncertainty through the forward function onto the data axis for a large range of source-receiver offsets. For each prior data distribution that they obtained, they calculated the entropy, and Figure 13 shows the entropy as a function of offset.

The maximum entropy point occurs for a source-receiver offset of around 1450 m. Notice that this is roughly three times the depth to the interface. A rule of thumb sometimes used for AVO design is that the maximum source-receiver offset should be roughly 2-3 times the depth of the interface of interest, so this "rule" only just spans the optimal offset in this case. Figure 13 also shows that the minimum possible information about the model is given for an offset roughly equal to the interface depth (around 550 m). It would be very difficult or impossible to obtain such results intuitively: In order to design surveys effectively in nonlinear situations, it is necessary to use formal design approaches such as the one used here.

Linearized, nonlinear problems and other approximations. Two other approaches to design surveys in nonlinear situations are described in this section. The first is termed "Bayesian design" in the statistical literature. The second is an alternative referred to here as "multimode reduction." These are described briefly as they may be useful in cases where the entropy criterion is simply too computationally expensive to use.

Bayesian design is an extension of the linear techniques presented earlier (Chaloner and Verdinelli, 1995). It essentially involves averaging a linearized measure of design quality over the prior model probability distribution. Let **P(m)** represent the probability distribution representing prior information about the model, and let **Φ(m)** be a linearized quality measure such as one of those in equations 7 or 8 in Part 1. Φ is a function of **m** in this case because the problem is nonlinear (gradients of **F_S(m)** and hence the eigenspectrum of A_S both depend on **m**). Bayesian design uses a quality measure

$$\bar{\Phi} = E_M\left[\Phi(\mathbf{m})\right] = \int_{\mathbf{m}} \Phi(\mathbf{m})\, P(\mathbf{m})\, d\mathbf{m} \qquad (21)$$

which is the average expected linearized measure of quality, **Φ(m)**. The measure $\bar{\Phi}$ developed from techniques suggested by Box and Lucas (1959) who noted that in nonlinear problems before an experiment has taken place, the most likely value of **Φ(m)** is the one calculated at that model, **m̂** say, for which the prior distribution P(**m**) takes a maximum value (i.e.,
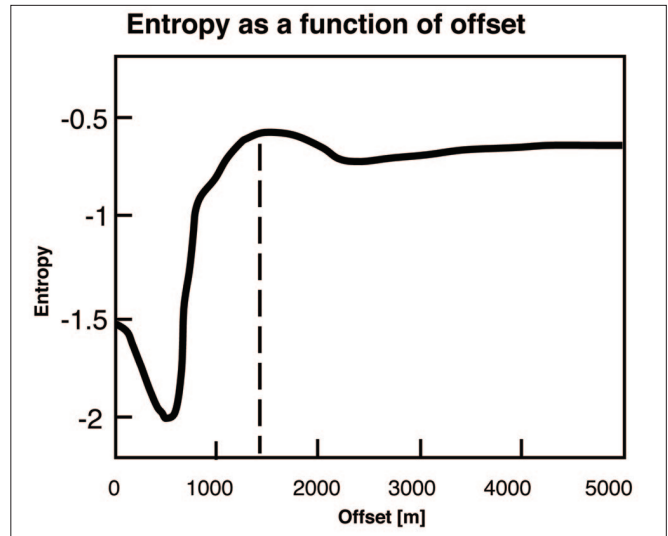

Figure 13. Entropy of the prior data distribution as a function of source-receiver offset for the forward function defined in Box 1, examples of which are shown in Figure 12. The dashed line marks the maximum entropy design of approximately 1450-m offset.

the model at which the mode of P(**m**) occurs). They suggested that $\bar{\Phi}(\mathbf{\hat{m}})$ could be maximized as a design criterion. Equation 21 extends this by taking the average such measure over the whole prior distribution P(**m**).

What is really happening in equation 21 is illustrated in Figure 14a for a uniform prior model pdf over the blue range marked. We can imagine the integral in equation 21 to be selecting infinitely many model samples m from the uniform prior pdf, and for each evaluating measure $\Phi(m)$. For each sample (e.g., sample m_1 in Figure 14a), the interpretation is the following: if m_1 was the true model then to within measurement uncertainty, we would record datum d_1; however, if we recorded datum d_1, the associated measurement uncertainty (gray) would propagate back into model parameter uncertainty around m_1 (red) approximately through the linear projection defined using the gradient at m_1 (black). This is the same procedure as was illustrated in Figure 1 in Part 1, but here is applied locally around m_1. A similar argument holds around model m_2, and around all other model samples (some are shown in green). Hence, a design procedure that maximizes some measure of the average gradient, such as $\bar{\Phi}$ in equation 21, should "on average" provide lowest postsurvey, linearized uncertainty estimates.

Kijko (1977) introduced measure 21 to seismology with Θ_3 in equation 7 taking the role of $\Phi(\mathbf{m})$. Steinberg et al. (1995) approximated design measure 21 with an approximate prior distribution and $\log(\Theta_3)$ for $\Phi(\mathbf{m})$, since this represents the average amount of apparent information if the forward function is linearized around model **m** (Lindley, 1956). Both these studies focused on configuring an optimal monitoring network for locating seismicity simultaneously on multiple possible source zones.

Although they do not describe it as such, the method of Maurer and Boerner (1998) and Maurer et al. (2000) used to design electromagnetic problems is a special case of a Bayesian design procedure. They designed their experiments by maximizing an average of three values of linearized measure Θ_0 in equation 7, the values calculated at each of three prior model samples. This can be represented within the current framework as maximizing the measure in equation 21 where the integral is approximated using only three samples, each assigned equal prior probability.

This approach may work well in many problems, but it

also contains an approximation that requires some care. The linearized measures $\Phi(\mathbf{m})$ are used assuming that the gradient of $\mathbf{F_S(m)}$ approximately controls the relationship between recorded data and model parameter uncertainties. In nonlinear problems this is not necessarily the case as is illustrated in Figure 14b. In this case the gradient of function $F_S(m)$ has high magnitude everywhere and hence measure $\overline{\Phi}$ would indicate this a far better design than the case in Figure 14a. However, consider the case where model m_3 is the correct model, resulting in datum d_3. Associated uncertainties in this datum project back through function $F_S(m)$ to give model uncertainty spanned by the entire set of intervals shown in red, since models anywhere in the red regions would provide data that are indistinguishable from the measured datum. The same would have been true if model m_3 had been placed almost anywhere along the model axis. The reason that the linearized measure $\overline{\Phi}$ is so large in this case is that it considers only the linearized projected uncertainty—that single red region locally around model m_3. Clearly the apparently high quality of this design is illusory.

The problem arises because the linearized projection of data uncertainties onto the model space is not a sufficiently good approximation to the true, nonlinear projection. In particular, in this case this is because there are multiple, discrete regions of model space that fit measured data. This is called a multimodal inverse problem (a mode in this context can be thought of as a distinct region of model space that gives a locally acceptable fit to data given their uncertainties).

A two-step design strategy to tackle such problems was suggested by Curtis and Spencer (1999). In the first step, the survey is designed in order to reduce the number of modes to one; then, while retaining a single mode, in a second step the design is refined to maximize $\overline{\Phi}$ (since the errors incurred in using this measure are likely to be much reduced after the first step). Such a strategy will only be possible in certain nonlinear problems in which it is possible to reduce the number of modes to one.

Microseismic location example. Figure 15 shows an example of applying the two-step procedure outlined above. Figure 15a shows both a random 1D seismic velocity model, and the contours of traveltime of the first-arriving wave emanating from a microseismic event at depth. Traveltimes at receiver locations on the surface are used as synthetic arrival time data in this example. Assuming that both P and S wave arrival times are recorded at each receiver, the difference in arrival times of these waves can be inverted for the event location (Tarantola and Vallette, 1982). Figure 15b shows contours of the (3D) misfit function—the squared misfit to the traveltime data if the event had occurred at any point in space. Minima in the misfit function correspond to modes in Figure 14b. For the receiver geometry in Figure 15b, there are four modes (although the lowermost one still gives a large misfit to the data and may be able to be ignored in this case).

Curtis and Spencer (1999) provided an algorithm to estimate the expected number of modes for any particular design using a random sampling strategy (i.e., without having to calculate the entire misfit function). Using this algorithm, the expected number of modes can be estimated for any particular design, and hence designs can be altered to reduce the number of modes. Figure 15c shows one design that is modified from that in Figure 15b, but which gives a single mode (at the true event location in Figure 15a). Maintaining this minimum number of modes, in the second step the design is further modified to reduce expected, lin-
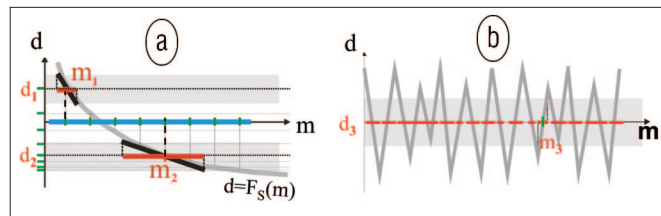

Figure 14. (a) Similar plot to Figure 1a (in Part 1), but for a nonlinear forward function $d=F_S(m)$, and where back-projections of data uncertainties (represented by the heights of gray shaded regions) are carried out locally around each model as illustrated for models m_1 and m_2. (b) Similar to (a), but for a highly nonlinear forward function for which the back-projected model uncertainty consists of many separate intervals (red).

earized uncertainty to a minimum using a design measure $\overline{\Phi}$ in equation 21, and such a design is shown in Figure 15d: the misfit function shows a single, tight minimum that clearly defines the event location. The final configuration is the well-known optimal design for earthquake location in a homogenous medium—one receiver on the expected event hypocenter, the others distributed evenly around a circle centered on the first (e.g., Rabinowitz and Steinberg, 1990). While this example is somewhat artificial, it does illustrate the potential that some problems may indeed be able to be redesigned to be unimodal.

Discussion. Some words of encouragement and caution are probably apposite at this point. While SED methods are often used and are well understood in design problems where the relationship between data and model parameters is (approximately) linear, survey design in situations in which this relationship is nonlinear is more demanding and computationally costly. In the latter case, using the full entropy technique described above would in principal usually solve the design problem. However, evaluating the entropy integral in equation 14 numerically requires many samples of the integrand, resulting in far larger requirements in terms of computer power than for linear problems. Analytic expressions for the entropy exist in special situations (e.g., the entropy of well known analytic pdfs like Gaussians have known analytic expressions) allowing numerical integration of equation 14 to be avoided, but in nonlinear problems it is not clear why the prior data pdfs should fall into one of these special categories. The Bayesian design criterion in equation 21 offers a computationally more tractable alternative, but should be used subject to the caveats about multimodality outlined above.

Areas for further research and development of SED techniques within geophysics include:

- Testing the SED techniques described above and in Part 1 on different practical problems. To date, such techniques have not been widely applied within geophysics).
- Creating versions or implementations of the linearized techniques in Part 1 that can be applied to very large linear problems (millions of data and model parameters). Current techniques may struggle in such situations, unless the data and/or model parameters are subsampled (decimated) first, or unless the matrices involved have special structure or properties (e.g., sparsity).
- Possibly the most important research area is to find alternatives or approximations to the entropy criterion that (1) do not suffer from issues related to multimodality, and (2) can be applied in reasonable time to problems of typical size in any field of application. The compute time to calculate entropy numerically for pdfs involving even only a few data is large by most standards, which reserves this technique for carefully chosen problems.
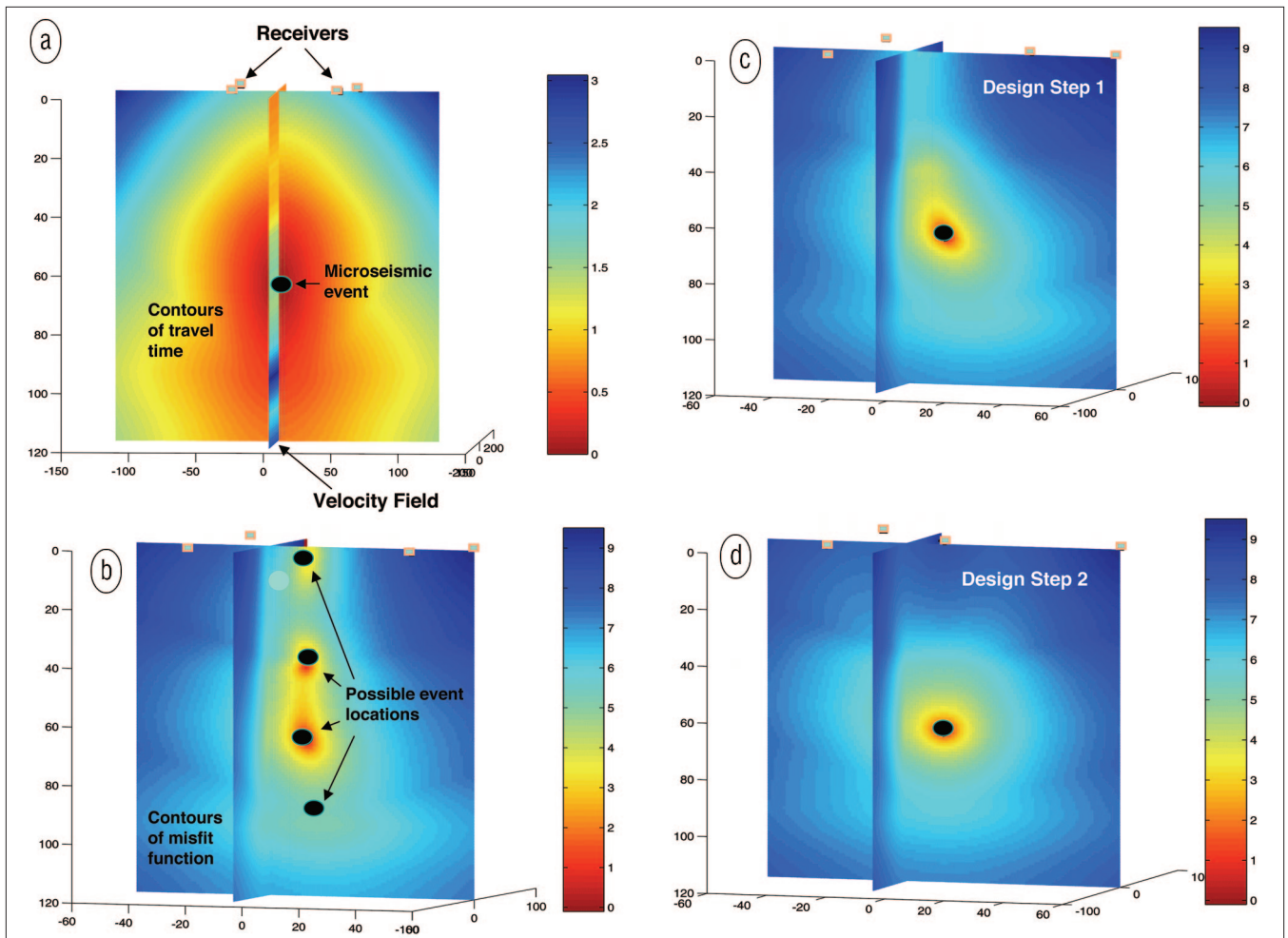
Figure 15. *(a) Panel coming out of the page shows a 1D random velocity model with an underlying trend of increasing velocity with depth. Back panel shows contours of traveltime of the first arriving energy from the microseismic event location shown to any point in the plane. Squares on the surface represent receiver locations. (b) Both panels show slices through the 3D misfit volume for data generated by the event in (a). There are four minima in the misfit surface which represents potential event locations. (c) and (d) are the same as (b), but for different receiver geometries.*

Even in situations where the techniques outlined in this tutorial cannot be applied, surveys must nevertheless be designed! The theoretical framework described herein may provide a basis for understanding issues and difficulties that arise when designing surveys using alternative techniques: heuristics (rules of thumb), data coverage type methods, methods based on a few forward simulations, etc. It is my hope that through improved understanding, perhaps such issues can be mediated in future through research and development of more advanced, and more widely applicable SED techniques. This will surely make endeavors in geophysics more efficient, and ultimately more successful.

Suggested reading. "Design of experiments in nonlinear situations" by Box and Lucas (*Biometrika*, 1959). "Bayesian experimental design: A review" by Chaloner and Verdinelli (*Statistical Science*, 1995). "Survey design strategies for linearized, nonlinear inversion" by Curtis and Spencer (SEG 1999 *Expanded Abstracts*). "Recent advances in nonlinear experimental design" by Ford et al. (*Technometrics*, 1989). "An algorithm for the optimum distribution of a regional seismic network - I" by Kijko (*Pageoph*, 1977). "On a measure of the information provided by an experiment" by Lindley (*Annals of Mathematical Statistics*, 1956). "Optimized and robust experimental design: a nonlinear application to EM sounding" by Maurer and Boerner (*GJI*, 1998). "Design strategies for electromagnetic geophysical surveys" by Maurer et al. (*Inverse Problems*, 2000). "Optimal configuration of a seismographic network: a statistical approach" by Rabinowitz and Steinberg (*SSA Bulletin*, 1990). "Maximum entropy sampling and optimal Bayesian experimental design" by Sebastiani and Wynn (*Journal of the Royal Statistical Society B*, Part 1, 2000). "A mathematical theory of communication" by Shannon (Bell System Technical Journal, 1948). "Maximum entropy sampling" by Shewry and Wynn (*Journal of Applied Statistics*, 1987). "Configuring a seismographic network for optimal monitoring of fault lines and multiple sources" by Steinberg et al. (*SSA Bulletin*, 1995). "Inverse problems = quest for information" by Tarantola and Vallette (*Journal of Geophysics*, 1982). "Optimal nonlinear Bayesian experimental design: An application to amplitude versus offset experiments" by Van den Berg et al. (*GJI*, 2003).

Corresponding author: curtis@cambridge.oilfield.slb.com or andrew.curtis@ed.ac.uk