JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 114, B04307, doi:10.1029/2008JB005948, 2009

Iteratively constructive sequential design of experiments and surveys with nonlinear parameter-data relationships

T. Guest[1,2] and A. Curtis[1,2]

Received 23 July 2008; revised 26 January 2009; accepted 26 February 2009; published 28 April 2009.

[1] In experimental design, the main aim is to minimize postexperimental uncertainty on parameters by maximizing relevant information collected in a data set. Using an entropy-based method constructed on a Bayesian framework, it is possible to design experiments for highly nonlinear problems. However, the method is computationally infeasible for design spaces with even a few dimensions. We introduce an iteratively constructive method that reduces the computational demand by introducing one new datum at a time for the design. The method reduces the multidimensional design space to a single-dimensional space at each iteration by fixing the experimental setup of the previous iteration. Both a synthetic experiment using a highly nonlinear parameter-data relationship and a seismic amplitude versus offset (AVO) experiment are used to illustrate that the results produced by the iteratively constructive method closely match the results of a global design method at a fraction of the computational cost. This work thus extends the class of iterative design methods to nonlinear problems and makes fully nonlinear design methods applicable to higher dimensional real-world problems.

Citation: Guest, T., and A. Curtis (2009), Iteratively constructive sequential design of experiments and surveys with nonlinear parameter-data relationships, *J. Geophys. Res.*, *114*, B04307, doi:10.1029/2008JB005948.

1. Introduction

[2] Large sums of money are invested every year in scientific surveys and experiments by both industry and scientific funding agencies. For each experiment a design process must first take place. Physical and logistical constraints define bounds on the types of experiments that are feasible, while minimizing cost often trades off with maximizing the amount of information we expect to record from the experiment. For this reason, optimizing designs of experiments in terms of cost, logistics, and the information which the experiment is expected to provide becomes of critical importance to maximizing return on investment [*Curtis and Maurer*, 2000; *Maurer and Boerner*, 1998].

[3] Statistical experimental design (SED) is a field of statistics devoted to developing methods to design experiments so as to optimize the expected information obtained. However, in many practical fields the design process is carried out using heuristics (rules of thumb) rather than physical relationships and mathematical modeling. Within the Geosciences for example, where enormous sums are spent on data collection, formal SED theory has only been applied in a handful of cases to design tomographic surveys [*Barth and Wunsch*, 1990; *Curtis*, 1999a, 1999b; *Curtis et al.*, 2004], earthquake monitoring surveys [*Kijko*, 1977a, 1977b; *Rabinowitz and Steinberg*, 2000; *Steinberg et al.*, 1995; *Winterfors and Curtis*, 2008], microseismic monitoring surveys [*Curtis et al.*, 2004], resistivity surveys [*Furman et al.*, 2004, 2007; *Maurer et al.*, 2000; *Stummer et al.*, 2004; *Wilkinson et al.*, 2006], geological expert elicitation processes [*Curtis and Wood*, 2004], and amplitude versus offset experiments [*van den Berg et al.*, 2003]. See *Curtis* [2004a, 2004b] for a review of SED related to the Geosciences.

[4] In this paper we focus on designing experiments such that the predicted postexperimental uncertainties on parameters are minimized while satisfying other constraints (for example, logistical, physical or financial). Optimal experimental design therefore requires an understanding of how the recorded data and postexperimental parameter uncertainties are related [*Atkinson and Donev*, 1992; *Box and Lucas*, 1959; *Silvey*, 1980].

[5] Let function $\mathbf{F}_\xi$ represent the relationship between parameters $\mathbf{m}$ and data $\mathbf{d}$, such that if we ignore measurement noise for now,

$$\mathbf{d} = \mathbf{F}_\xi(\mathbf{m}) \tag{1}$$

would be recorded if parameters $\mathbf{m}$ were true. The subscript ξ in the forward function $\mathbf{F}_\xi$ indicates that the parameter-data relationship is dependant on the experimental design ξ where ξ is a vector representing, for example, source and receiver types and locations, or any other pertinent aspects of the design.

[6] One of the main reasons that SED theory has not gained general acceptance in the Geosciences is that the relationship $\mathbf{F}_\xi$ is commonly highly nonlinear. The results of applying linear SED theory are therefore not necessarily robust, while

[1]Grant Institute, School of GeoSciences, University of Edinburgh, Edinburgh, UK.
[2]Also at Edinburgh Collaborative of Subsurface Science and Engineering, Edinburgh, UK.

0148-0227/09/2008JB005948$09.00

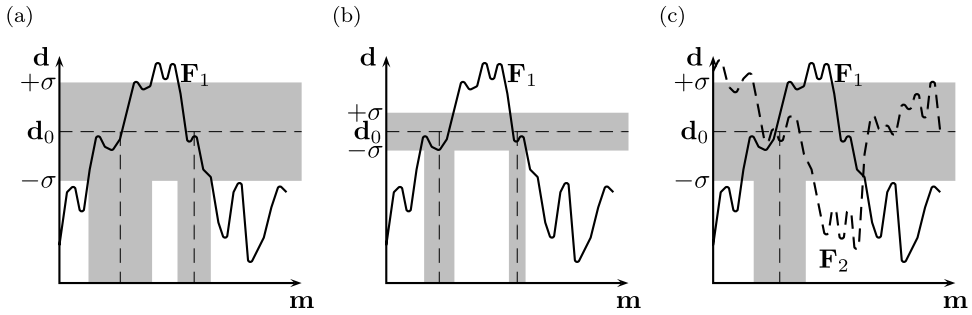

Figure 1. Parameter-data relationships for three experimental designs. Designs in Figures 1a and 1b use the same experimental setup so the forward function $\mathbf{F}_\xi$ is identical. (a) The first experiment only uses a single measurement to predict the parameter and hence has a large uncertainty ($\pm\sigma$). (b) The second experiment represents repeated measurements resulting in a reduced data uncertainty and hence a smaller parameter uncertainty. In Figure 1c, two data are recorded, both with value $\mathbf{d}_0$, but each with a different forward function (solid and dashed lines).

applying fully nonlinear design theory is too computationally costly. Consequently, in the above list of cited work in this field, only the papers of *van den Berg et al.* [2003] and *Winterfors and Curtis* [2008] apply nonlinearized design theory, and these only addressed one-dimensional and two-dimensional parameter vectors respectively. Since naturally occurring design problems in many other fields are also nonlinear, we suspect that the uptake of SED techniques will be similarly hampered. There is thus a great need for new methods of SED that are applicable to multidimensional, nonlinear problems without linearization.

[7] Experimental error and data noise always result in uncertainty on the measurement of data $\mathbf{d}$. Figures 1a and 1b show schematically the results of two experiments characterized by the same nonlinear parameter-data forward function. Experiment (a) collects a single data measurement $\mathbf{d}_0$ with associated uncertainty σ, while experiment (b) collects multiple identical measurements that reduce the data uncertainty. The aim is to invert equation (1) to infer the postexperimental (a posteriori) parameter range. Figures 1a and 1b show that both inversions result in a bimodal parameter estimate, however experiment (b) is better constrained. Figure 1c shows an estimate of the parameters that could be inferred from an experiment using two different parameter-data relationships. Since both experiments are used to constrain the same parameters the results must be consistent. The resultant parameter range is given by the intersection of the two inversion solutions and is shown in Figure 1c to be unimodal.

[8] Consider designing an experiment that allows $n = 10$ data measurements to be recorded so as to minimize postexperimental uncertainties. Say 20 unique designs are available for each datum, all of which have the same cost. This results in 20^{10} possible designs, and assuming that the optimality of the overall experimental design can be quantified in some way, finding the globally optimal design out of so many possibilities is generally computationally infeasible. *Ryan* [2003] proposed a method to calculate the optimal design using Markov chain Monte Carlo techniques without any parameter-space or data-space discretization (which is in contrast to our method), while *Muller and Parmigiani* [1996] used the assumption that the quantifiable design space is a smooth surface to reduce the computational cost. As noted by *Hamada et al.* [2001], when the dimension of the data-space becomes greater than 3 the numerical integration required to compute the optimality measure becomes computationally infeasible.

[9] *Curtis et al.* [2004] introduced a deterministic, iterative design algorithm that was applied to linearized design of seismic tomography and microseismic monitoring surveys. Their method starts with the design containing the maximum number of receivers conceivable, then in turn, the receiver that provides the least postexperimental information is iteratively removed until a final design with an affordable number of data measurements is obtained. We refer to this kind of algorithm as ''iteratively destructive''. *Curtis et al.* [2004] also described an iteratively constructive version of their algorithm where at each iteration the datum that provides the most additional information to some minimal design is added until a cost constraint is reached. Such an algorithm was applied by D. Coles and F. Morgan (A method of fast, sequential experimental design for linearized inverse problems with applications for borehole dc resistivity, submitted manuscript, 2009), *Stummer et al.* [2002, 2004], and *Wilkinson et al.* [2006] to design linearized geoelectrical tomography surveys.

[10] Since these methods are iterative they do not necessarily find globally optimal designs. Furthermore, these methods have only been developed for linearized design problems using linear algebraic results to compute information increases and losses cheaply.

[11] In this paper we introduce an iteratively constructive method that works for linear or nonlinear problems, and which reduces the number of designs considered to nx where x is the number of possible single-datum designs, and n is the total number of data to be recorded. In the above example the total number of designs that need to be considered would be reduced to 200. In brief, the algorithm works by first calculating the optimal design for a single-datum measurement (x possible designs to consider). It then designs an experiment that records two data measurements, but instead of considering all x^2 possible two-data designs it fixes the first datum to that found in the previous step. There are therefore only x new designs to consider corresponding to possible designs for the second datum. This iterative method is repeated for subsequent data one at a time, fixing previously found designs, until the required final number of data measurements has been attained or a cost threshold exceeded.

[12] By not considering every one of the x^n possible design we find a locally optimal design rather than the globally optimal design. However, note first that considering the entire design space is generally infeasible, second we will show that no linearization of physics is necessary in iterative methods, and third the final design is found in a more scientifically rigorous manner than when using rules of thumb or other heuristic methods. Additionally, as shown below the designs found can be close to globally optimal, even in nonlinear problems.

[13] It is important to distinguish "iterative design" methods discussed here, from so-called "sequential design" methods [*Fedorov*, 1972; *Ford et al.*, 1989; *Hu*, 1998]. In sequential methods the data collected from the previous experiment are used to estimate the parameter probability density function (p.d.f.) which is then used to design the next experiment. Sequential design methods thus build on previous experiments. Iterative design methods on the other hand use a constant prior parameter p.d.f. when adding successive data measurements, and are methods to find a (sub)optimal design for a single experiment when full global optimization is infeasible or too costly to find.

[14] The iteratively constructive design method introduced herein requires a quantitative measure of the design quality that is valid for any parameter-data relationship, however nonlinear. Both linearized design methods [*Curtis*, 2004a, 2004b; *Silvey*, 1980; *Steinberg et al.*, 1995] and classical nonlinear design methods [*Atkinson and Donev*, 1992; *Box and Lucas*, 1959; *Chaloner and Verdinelli*, 1995; *Ford et al.*, 1989] employ a quality measure based on linearized physics and so are not robust in highly nonlinear situations as explained by *van den Berg et al.* [2003]. Instead, we adopt a Bayesian approach for parameter inference in which p.d.f.s represent states of information about parameters, and expected postexperimental information can be estimated to assess design quality without linearization. Our quality measure and example application (designing amplitude versus offset (AVO) experiments) is similar to the study of *van den Berg et al.* [2003]. However, that previous paper found globally optimal solutions and as a consequence was limited to designing a single-datum experiment due to the huge computational demand. In this paper we show how our new, iteratively constructive algorithm reduces the computation required for nonlinear design problems such that we can extend the range of application to many more data points. More generally, we believe that iteratively constructive or destructive methods may represent one of the best available options to extend fully nonlinear design methods to real-world, high-dimensional scientific problems.

2. Method

2.1. Nonlinear Design Theory

[15] According to Bayes' theorem the posterior p.d.f. of parameters $\mathbf{m}$ given recorded data $\mathbf{d}$ and design ξ is given by

$$\sigma(\mathbf{m}|\mathbf{d},\xi) = \frac{\theta(\mathbf{d}|\mathbf{m},\xi)\rho(\mathbf{m})}{\sigma(\mathbf{d}|\xi)} \tag{2}$$

where $\theta(\mathbf{d}|\mathbf{m},\xi)$ represents the p.d.f. of data $\mathbf{d}$ being observed given parameters $\mathbf{m}$, and design ξ, $\rho(\mathbf{m})$ is the p.d.f. representing the prior information on parameters $\mathbf{m}$ and $\sigma(\mathbf{d}|\xi)$ is the marginal distribution over observed data and contains all information about which data are likely to be recorded during the experiment:

$$\sigma(\mathbf{d}|\xi) = \int \theta(\mathbf{d}|\mathbf{m},\xi)\rho(\mathbf{m})d\mathbf{m}. \tag{3}$$

[16] In designing an experiment we wish to maximize the information on the parameters $\mathbf{m}$ expected to be contained in the posterior p.d.f. $\sigma(\mathbf{m}|\mathbf{d},\xi)$. We therefore need to quantify the information I represented by a p.d.f. The entropy of any probability distribution $f(\mathbf{x})$ is related to Shannon's measure of information [*Shannon*, 1948] as

$$Ent(\mathbf{X}) = -I\{f(\mathbf{x})\} + c = -\int_{\mathbf{x}} f(\mathbf{x})\log\{f(\mathbf{x})\}d\mathbf{x} \tag{4}$$

where *Ent* is the entropy function, $f(\mathbf{x})$ is the p.d.f. of the random variable $\mathbf{X}$, I is the information measure as defined by *Shannon* [1948], and c is a constant. The optimal experimental design can be determined by maximizing the information expected to be contained, postexperiment, in the posterior parameter p.d.f. (equation (2)). A fully nonlinear quality measure can therefore be defined [after *Lindley*, 1956] as

$$\Phi(\xi) = -\int_D Ent\{\sigma(\mathbf{m}|\mathbf{d},\xi)\}\sigma(\mathbf{d}|\xi)d\mathbf{d} \tag{5}$$

where $-Ent\{\sigma(\mathbf{m}|\mathbf{d},\xi)\}$ represents the amount of information contained in the posterior p.d.f. about the parameters $\mathbf{m}$ given a particular data measurement $\mathbf{d}$ recorded using the design ξ. The quality measure is calculated by taking the expectation of the entropy function with respect to the marginal distribution (equation (3)) over all data measurements.

[17] To evaluate $Ent\{\sigma(\mathbf{m}|\mathbf{d},\xi)\}$ requires that the distribution in equation (2) can be evaluated. In many cases this represents finding the solution to an inverse problem (finding constraints on $\mathbf{m}$ given a value for $\mathbf{d}$), which in nonlinear problems can be extremely demanding computationally. Fortunately under some common assumptions an alternative, cheaper computation is equivalent. *Shewry and Wynn* [1987] showed that

$$-\Phi(\xi) + Ent\{\sigma(\mathbf{d}|\xi)\} = Ent\{\sigma(\mathbf{d},\mathbf{m}|\xi)\} = b \tag{6}$$

where b is a constant, and the second equality is true if both the prior parameter distribution and the measurement noise is design independent. The parameter-data relationship is often assumed to have the form

$$\mathbf{d} = f_\xi(\mathbf{m}) + \epsilon, \tag{7}$$

where ϵ is a vector of independent random errors which do not depend on either the parameters or on the design. It can be shown that equation (6) holds for relationships of this type [*Shewry and Wynn*, 1987]. Instead of maximizing $\Phi(\xi)$ equation (6) shows that the optimal design can be determined by maximizing $Ent\{\sigma(\mathbf{d}|\xi)\}$. For this

latter measure, only information about $\sigma(\mathbf{d}|\xi)$ is required, and in most cases $\sigma(\mathbf{d}|\xi)$ consists of the prior information on the parameters projected through the physical relationship $\theta(\mathbf{d}|\mathbf{m}, \xi)$, that is, it only involves solving the forward problem.

[18] In general, the p.d.f. required to calculate the entropy is not known analytically and must therefore be deduced numerically. The Monte Carlo method we adopt here varies from that used by *van den Berg et al.* [2003]. Samples of $\rho(\mathbf{m})$ are generated and projected through the physical relationship in equation (7) into a discretized data space. The physical relationship used incorporates the random associated measurement error ϵ (equation (7)). The resulting data space histogram of all projected samples is normalized to have unit volume, wherefore it represents a numerical approximation to the posterior p.d.f. $\sigma(\mathbf{d}|\xi)$. This approximation of $\sigma(\mathbf{d}|\xi)$ is then used to assess the quality of the experimental design by calculating the entropy,

$$Ent\{\sigma(\mathbf{d}|\xi)\} \approx \sum_i \hat{\sigma}(\mathbf{d}_i) \log\{\hat{\sigma}(\mathbf{d}_i)\} \tag{8}$$

where $\mathbf{d}_i$ represents the center of the ith discretized bin in the data space. For any given experimental design, $\hat{\sigma}(\mathbf{d}|\boldsymbol{\xi})$ is affected by the number of parameter samples, the data space discretization bin size, and the measurement error.

[19] In linearized problems the parameter-data relationship is in the form

$$\mathbf{d} = \mathbf{G}_{\mathbf{m}_0}\mathbf{m} + \epsilon \tag{9}$$

where $\mathbf{G}_{\mathbf{m}_0}$ is a matrix of derivatives of $\mathbf{d}$ with respect to $\mathbf{m}$ calculated at a reference model $\mathbf{m}_0$. In linear problems $\mathbf{G}_{\mathbf{m}_0}$ is constant with respect to the reference model and the optimal design is commonly determined by maximizing measures sensitive to the magnitudes of eigenvalues of $(\mathbf{G}^T\mathbf{G})$ [*Curtis*, 1999a, 1999b]. A common measure used in linear problems with Gaussian uncertainties is

$$\Phi_{\text{linear}}(\xi) = k \det\left(\mathbf{G}^T\mathbf{G}\right) \tag{10}$$

where k is a constant. For linearized (but not truly linear) problems, if all posterior distributions are Gaussian then the classical nonlinear estimate for the quality of an experimental design can be used [*Atkinson and Donev*, 1992; *Box and Lucas*, 1959; *Chaloner and Verdinelli*, 1995; *Ford et al.*, 1989]:

$$\Phi_{\text{linearized}}(\xi) = \int_M k \det\left(\mathbf{G_m}^T\mathbf{G_m}\right)\rho(\mathbf{m})d\mathbf{m}. \tag{11}$$

Rather than using a single set of parameter values, the expected design quality measure is calculated over the prior distribution of parameters values. Therefore equation (11) represents an average measure of linearized design quality over the prior parameter p.d.f. While equations (11) and (5) have similar forms equation (5) requires no linearization of the forward function or assumptions regarding the posterior distributions. Thus equation (5) represents a measure of design quality applicable to truly nonlinear problems.

2.2. Iterative Design Theory

[20] In the method of *van den Berg et al.* [2003], the entropy is calculated for every possible design. The design that corresponds to the maximum entropy value is deemed the optimal design. In iteratively constructive methods on the other hand,

$$\xi_j = \arg\max\left[Ent\left\{\sigma\left(\mathbf{d}|\xi_j\right)\right\}\right], \text{ given such that } \xi_{j-1} \text{ fixed,} \tag{12}$$

meaning that the new optimal design ξ_j combines the data measurements of the previous iteration ξ_{j-1} augmented by the single datum measurement that maximizes the entropy of the posterior data p.d.f. given that ξ_{j-1} remains fixed.

[21] The same method can be used in an iteratively destructive manner were all possible data are recorded and one at a time the datum that provides least information is removed until the final experimental design is found. Although this method is plausible in principle, as shown below it would prove to be computationally infeasible.

2.3. Design of Numerical Implementation

[22] Our ability to evaluate the right side of equation (12) constitutes the computational limiting factor in the new method. While computation is greatly reduced compared to that required to find the global optimum of arg max$[Ent\{\sigma(\mathbf{d}|\xi)\}]$ used in previous studies, it nevertheless increases significantly for increasing data space dimensions, as shown in examples below. We show in section 5 that the introduction of a second design subproblem, to design dynamically the number of data space samples required to design each new ξ_j accurately, greatly enhances the method's performance.

3. Synthetic Test: Sawtooth Functions

[23] We now evaluate the performance of the new design method by applying it to a highly nonlinear, synthetic problem for which there is an analytic solution for the projected data space marginal distribution $\sigma(\mathbf{d})$. We imagine that the experimental situation is such that by choosing different designs we can alter the forward function $f_\xi(m)$ in equation (7) to be a sawtooth function with an integer number ξ of periods in the range $m \in [0, 10]$ (Figure 2). Prior information on m is assumed to be uniform in this range. *van den Berg et al.* [2003] illustrated that classical Bayesian (linearised) nonlinear measures (equation (11)) fail when applied to a one-dimensional sawtooth forward function. Such measures deliver as optimal design a sawtooth with as many periods as possible, corresponding to a forward function with the highest possible average gradient. This results in a postexperimental parameter-space posterior p.d.f. that is highly multimodal, but which in one dimension provides no Shannon information gain over a single-period sawtooth function.

[24] Using the entropy value as our optimality measure we assess the accuracy of our iteratively constructive method against the globally optimal results for multidimensional sawtooth functions

$$\mathbf{f}_\xi(\mathbf{m}) = \left\{f_{\xi_1}(m), \ldots, f_{\xi_n}(m)\right\}. \tag{13}$$

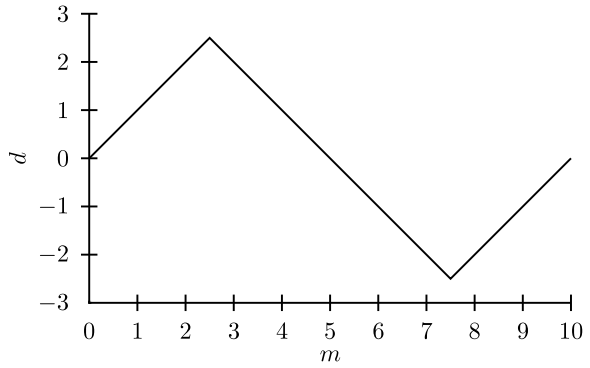

Figure 2. Single-period sawtooth function with maximum amplitude 2.5 in the parameter range $m \in [0, 10]$.

The distribution $\theta(\mathbf{d}, \mathbf{m})$ is defined to have Gaussian error with standard deviation 0.1 around mean $\mathbf{f}_\xi(\mathbf{m})$, truncated at ±3 standard deviations from the mean. The analytical value of $\sigma(\mathbf{d})$ for any of the sawtooth functions (without truncation of the Gaussian) is then given by

$$\sigma(\mathbf{d}) = 0.1\left[-erf\left\{5\sqrt{2}(-2.5+d)\right\} + erf\left\{5\sqrt{2}(2.5+d)\right\}\right] \tag{14}$$

where *erf* is the error function. We therefore use equation (14) as a very close approximation to the truncated Gaussian distribution for any single sawtooth forward function.

[25] Consider first a sawtooth function with a single period over the parameter range (Figure 2). The analytical value for the entropy of the single sawtooth function is equal to 1.645. $\sigma(\mathbf{d})$ is approximated numerically using different data space discretization lengths and total number of parameter space samples. Figure 3 shows the numerical approximation for the entropy as a function of data space discretization length and number of parameter space samples for the single period sawtooth function. The plot shows that as the data space discretization becomes small the total number of samples needed to accurately sample the data space becomes large. When the data space discretization length becomes large then even several thousand parameter space sample points are insufficient to approximate equation (14) accurately. For the remaining sawtooth function examples we use a data space discretization length of 0.1 with 500,000 random parameter space samples.

[26] It is instructive to consider the simple problem of designing a two-datum experiment where each datum could result in sawtooth functions with either $\xi_i = 1$ or 2 periods between $m \in [0, 10]$. If the second function had $\xi_2 = 1$ period, say, this leaves two possible designs: either a repeat experiment $\xi_1 = 1$, or a combination experiment $\xi_1 = 2$ using a single instance of 1 and 2 periods. The same Uniform prior parameter range and Gaussian errors on each datum are used as in the single-sawtooth function experiment, and Figure 4 shows the resulting $\sigma(\mathbf{d})$ functions. The repeat experiment (b) shows a linear trend since any given parameter value results in approximately the same datum for both sawtooth functions. The design using both sawtooth functions in (a) shows lower $\sigma(\mathbf{d})$ values than (b) because a larger proportion of the data space has been sampled using the two different period functions.

[27] The entropy $Ent\{\sigma(\mathbf{d}|\xi)\}$ in equation (8) for the repeat experiment in (b) is 1.126 while the entropy for the experiment in (a) using both one-period and two-period functions is 2.184. The data expected to be recorded from repeat experiments will therefore always provide less post-experimental information on parameters than an experiment using different sawtooth functions for each datum.

[28] We now assess the performance of the iteratively constructive method using this challenging forward function for each f_i between 1 and 10 periods in the range $m \in [0, 10]$. Since *van den Berg et al.* [2003] showed that the entropy value is constant for any period of sawtooth function in the single-datum problem, we begin the iteratively constructive method at the globally optimal design for a two-data scenario. Figure 5 shows the entropy of $\sigma(\mathbf{d})$ for all possible two-data experiments. The optimal design uses $\xi_1 = 9$ and $\xi_2 = 10$ periods. As was observed in the previous example, performing a repeat experiment (represented by cells on the diagonal) provides minimal postexperimental information compared with any other design.

[29] According to our iterative method the optimal design for a three-sawtooth function setup is calculated by first fixing the periods of the two-sawtooth design at $\xi = [9, 10]$. The only sawtooth period to be designed is that of the third function, ξ_3. The entropy is calculated for the 10 possible resulting three-sawtooth experiments, and the maximum was found to correspond to period $\xi_3 = 8$. To compare this to the globally optimal result, the entropy of all possible 10^3 designs using 3 sawtooth functions was calculated. In this case, the locally optimal design found using the iterative method exactly matches the global optimum.

[30] The iterative method is used again to calculate the optimal four-sawtooth design, and predicts that the functions with periods 8, 9, 10 and 10 are optimal. In this case the optimal design uses a repeated data measurement rather than using another function with a different period. The globally optimum design was also located from the possible 10^4 designs. Again, the local optimum of the iterative

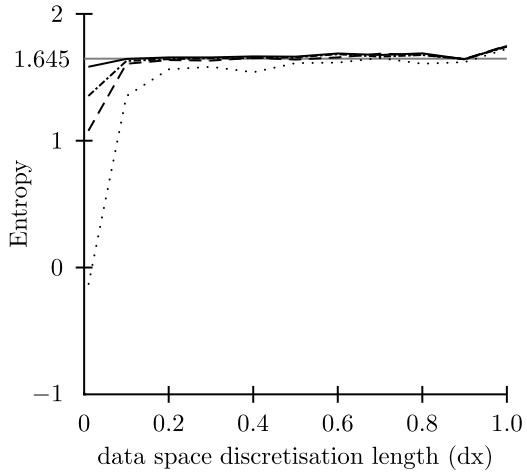

Figure 3. Entropy values for the single-period sawtooth function as a function of data space discretization length (dx) for four experiments with different numbers of parameter space samples (dotted, 100; dashed, 500; dot-dashed, 1000; solid, 5000). The gray line represents the analytical value of 1.645.

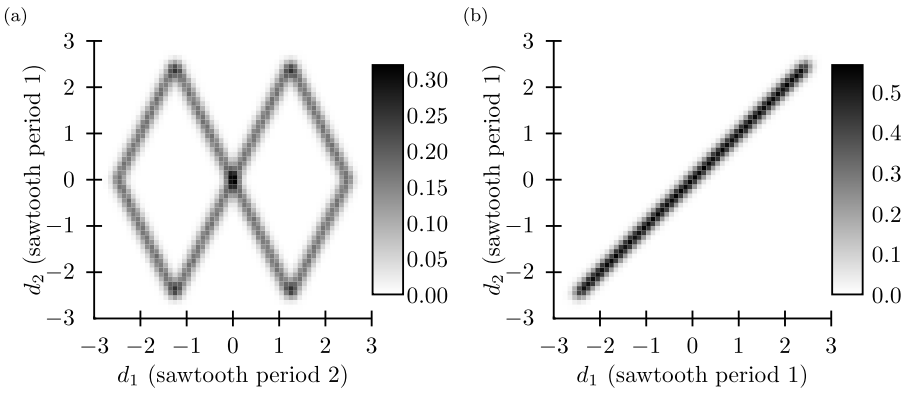

Figure 4. Plots of $\sigma(\mathbf{d})$ for (a) a design that uses both period sawtooth functions and (b) a repeat experimental design, calculated using 500,000 parameter-space samples and a data-space discretization length of 0.1. Scale bar represents the normalized $\sigma(\mathbf{d})$ value for each data-space discretization.

method exactly matches the global optimum while only searching a 10-element design space, a saving of three orders of magnitude in computing power.

4. Geophysical Application: Multiple Amplitude-Versus-Offset (AVO) Receivers

[31] Having gained confidence in the new method, we now apply it to the practical geophysical problem addressed by *van den Berg et al.* [2003]. The computational savings offered by our method allows us to extend their results to far more complex designs than they found computationally tractable.

[32] Our data will be the amplitudes of planar seismic waves generated with unit amplitude at the ground surface, reflected from a subsurface boundary between two geological layers at depth d, and recorded when they arrive again at the ground surface (Figure 6). Each is a function of the incident angle of the wave at the boundary, density ρ_i and the elastic media properties (summarized by the P wave velocity α_i, and S wave velocity β_i, for an isotropic medium) of both layers $i = 1, 2$. The recorded amplitudes A_1, A_2 are given in full by the solution to the Zoeppritz equations,

$$\begin{aligned}
\cos i_1 A_1 + \frac{\alpha_1}{\beta_1}\sin j_1 B_1 + \frac{\alpha_1}{\alpha_2}\cos i_2 A_2 - \frac{\alpha_1}{\beta_2}\sin j_2 B_2 &= \cos i_1 \\
-\sin i_1 A_1 + \frac{\alpha_1}{\beta_1}\cos j_1 B_1 + \frac{\alpha_1}{\alpha_2}\sin i_2 A_2 + \frac{\alpha_1}{\beta_2}\cos j_2 B_2 &= \sin i_1 \\
-\cos 2j_1 A_1 - \sin 2j_1 B_1 + \frac{\rho_2}{\rho_1}\cos 2j_2 A_2 - \frac{\rho_2}{\rho_1}\sin 2j_2 B_2 &= \cos 2j_1 \\
\sin 2i_1 A_1 - \frac{\alpha_1^2}{\beta_1^2}\cos 2j_1 B_1 + \frac{\rho_2}{\rho_1}\frac{\beta_2^2}{\beta_1^2}\frac{\alpha_1^2}{\alpha_2^2}\sin 2i_2 A_2 + \frac{\rho_2}{\rho_1}\frac{\alpha_1^2}{\beta_1^2} & \\
\cdot \cos 2j_2 B_2 &= \sin 2i_1
\end{aligned} \tag{15}$$

where for a horizontal boundary i_1 is the P wave angle of incidence and reflection, i_2 the P wave angle of refraction, j_1

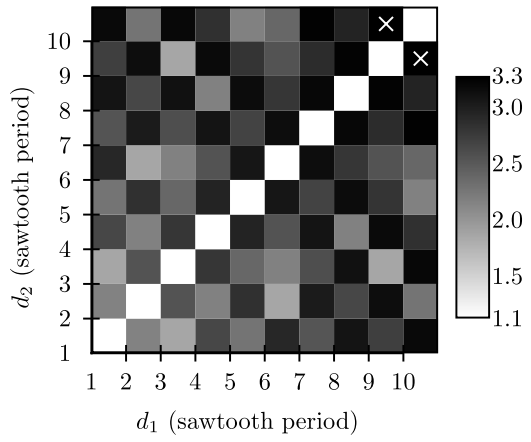

Figure 5. Two-data entropy values of entropy of $\sigma(\mathbf{d})$ for the complete design space for pairs of sawtooth functions with integer periods ranging from 1 to 10 over the parameter range of 0–10. For each experiment, the data space discretization length is 0.1 and 500,000 samples are drawn at random from a Uniform prior distribution over the parameter range. The white crosses indicate the optimal two-data experimental design.

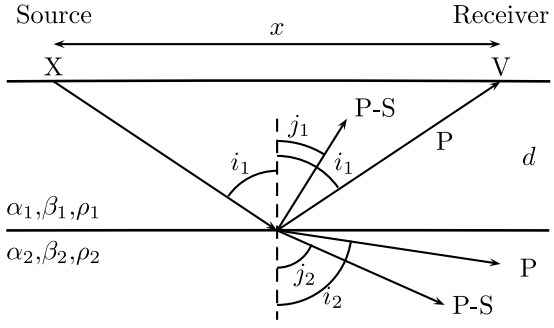

Figure 6. Geometry of an AVO experiment with a single interface. The distance between the source (X) and the receiver (V) is called the offset (x). The depth to the interface is d. At the interface, the incident P wave energy is split into a reflected P wave and P-S wave conversion and is also transmitted into the second layer as a P wave and P-S wave conversion. The amplitudes of each wave are given by equation (15). The properties of the subsurface are given by density (ρ), P wave velocity (α), and S wave velocity (β) in each layer.

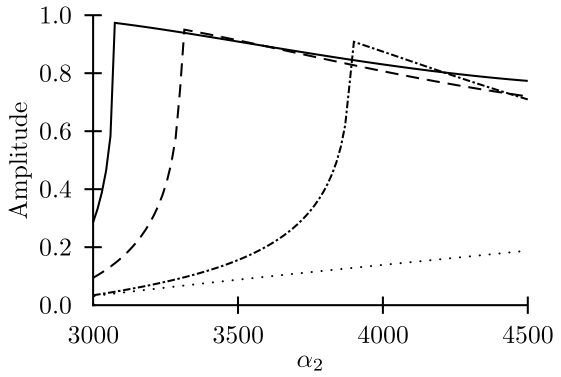

Figure 7. Reflected P wave amplitude data as a function of parameter α_2 for four different offsets: x = 500 m (dotted line), x = 1000 m (dot-dash line), x = 1500 m (dashed line), and x = 2000 m (solid line). The velocity of the top layer is 2750 m s^{-1}, and the depth to the reflecting boundary is 500 m.

the converted P-S wave angle of reflection and j_2 the converted P-S wave angle of refraction [*Yilmaz*, 2001]. The amplitudes of the waves are represented by A_1 for the reflected P wave, A_2 for the refracted P wave, B_1 for the reflected S wave and B_2 for the refracted S wave. From the reflected P wave amplitude it is possible, given values for the other elastic properties, to estimate the P wave velocity of the lower layer (α_2)

$$A_1 = f_x(\alpha_2). \tag{16}$$

Our initial design problem is to find the single offset (x) between one source and one receiver that best constrains the value of α_2, given the P wave reflection coefficient A_1 measured at that offset.

[33] We assume that both layers are represented by a Poisson medium in which $\beta = c\alpha$ where $c = 1/\sqrt{3}$, and that there is no significant density contrast between the layers. Figure 7 shows the P wave reflection amplitudes as a function of the parameter α_2 for 4 different offsets using a depth of 500 m for the reflecting boundary and a P wave velocity of 2750 m s^{-1} for the top layer. The discontinuities in the curves occur when the critical angle of incidence is reached for the given velocity structure (the angle at which the refracted P wave becomes horizontal).

[34] For the design problem the prior parameter information about α_2 is specified by a Uniform distribution ranging from 3000 m s^{-1} to 4500 m s^{-1}, and $\theta(d, m)$ is given by the numerical solution of equations (15) with added Gaussian uncertainty of standard deviation 0.01 to simulate measurement uncertainties.

[35] Entropies are calculated for offsets ranging from 0 m to 5000 m at 25 m intervals. For each offset the data discretization length is 0.01 and 500,000 samples are drawn at random from the uniform parameter space. *van den Berg et al.* [2003] used the same prior information and geometrical setup to find the optimal receiver offset. Figure 8 shows the entropy values as a function of offset for our results and the most densely sampled results of *van den Berg et al.* [2003].

[36] The discrepancy between the results is due to different forward functions being used to calculate the reflection coefficient. *van den Berg et al.* [2003] used the *Aki and Richards* [2002] approximation to the solution to the Zoeppritz equations:

$$R_P = \frac{1}{2\cos^2 i}\frac{\Delta\alpha}{\alpha} - 4\beta^2 p^2 \frac{\Delta\beta}{\beta} + \frac{1}{2}\left(1 - 4\beta^2 p^2\right)\frac{\Delta\rho}{\rho} \tag{17}$$

where α is the average P wave velocity, $\Delta\alpha$ is the difference in the P wave velocities in the upper and lower layers, and β, $\Delta\beta$, ρ, and $\Delta\rho$ are similarly defined. i is the average of the P wave reflection and refraction angle, and p is the horizontal slowness. Assuming a Poisson medium ($\beta = c\alpha$, where $c = 1/\sqrt{3}$) and no density contrast between the layer ($\Delta\rho = 0$) simplifies the equation to that used by *van den Berg et al.* [2003]:

$$R_P = \left\{\frac{1}{2}\left[1 + \tan^2 i\right] - 4c^2 \sin^2 i\right\}\frac{\Delta\alpha}{\alpha} \tag{18}$$

[37] The approximation of *Aki and Richards* [2002] assumes that the angles i_1, i_2, j_1 and j_2 (Figure 6) are real and close to 90°. This assumption holds for small offsets (less than 1000 m) and low layer-2 P wave velocities (less than 3500 m s^{-1}). At values greater than these the approximation fails and only the Zoeppritz equations correctly predict the reflection amplitude coefficient. We therefore use the full solution to the Zoeeppritz equations for our results (equation (15)).

[38] The new results show that the optimal location for a single receiver is at an offset of 1050 m. This is in contrast to the *van den Berg et al.* [2003] optimal offset of approximately 1500 m. Both methods produce a local minimum at an offset of 500 m, a distance equal to the interface depth. A standard rule of thumb approximation for an AVO experiment is to use offsets between 1 and 3 times the depth to the interface under study. For the single receiver case this rule is confirmed.

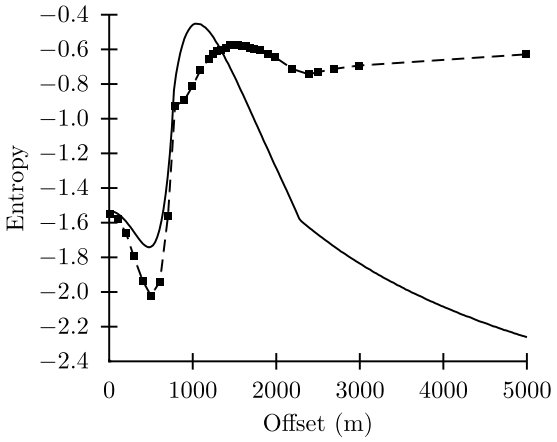

Figure 8. Entropy values as a function of offset for a single receiver. The most densely sampled results of *van den Berg et al.* [2003] are represented by the dashed line (data space discretization length of 0.001 using 300,000 samples). The solid line represents our results using a data discretization length of 0.01 using 500,000 samples.

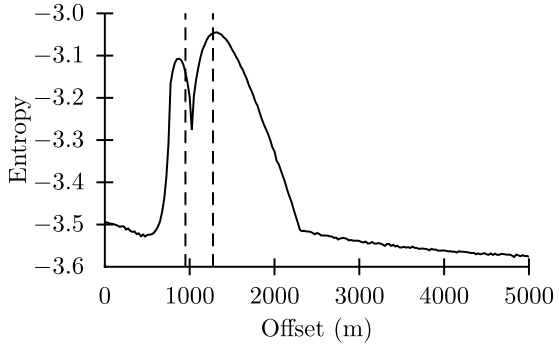

Figure 9. Entropy values as a function of offset for a two-receiver experimental design using the iteratively constructive method with the first receiver fixed at 1050 m. The entropy at each offset is calculated using 500,000 samples and a data space discretization of 0.01. Entropies are calculated at offsets ranging from 0 m to 5000 m at 25-m intervals. The dashed lines represent the optimal offsets found by searching the complete design space (Figure 10).

[39] Using the Zoeppritz equations we now consider a two-receiver experiment. Using both the iteratively constructive method and calculating the entropy values across the complete design space we can see how closely the two resulting experimental designs match. For the iterative method the first receiver is fixed at an offset of 1050 m. The second receiver is varied between offsets ranging from 0 m to 5000 m at 25 m intervals. The same data space discretization and number of parameter space samples are used as in the single receiver experiment.

[40] Figure 9 shows the entropy values as a function of offset for the two-receiver design. The optimal design would therefore place the second receiver at an offset of 1350 m, resulting in a two-receiver design with offsets of 1050 m and 1350 m. Similarly to the single receiver experiment there is a local entropy minimum at an offset of approximately 500 m and a decreasing entropy trend with increasing offsets greater than 2500 m. The entropy notch at 1050 m represents a repeat experiment; although this would provide extra information about velocity α_2, the extra information obtained by placing the second receiver at a slightly larger or smaller offset significantly outweighs this advantage.

[41] Figure 10 shows the entropy values obtained for every possible two-receiver configuration. The globally optimal receiver locations are at offsets of 1275 m and 950 m. These do not coincide with the locations found using the iterative method as the offsets lie either side of the optimal single-receiver offset. However, the entropy map shows several global features that are also seen in Figure 9. First, placing either or both receivers at offsets less than 500 m (equal to the depth of the reflector) or at offsets greater than 2500 m results in an experiment that is expected to record relatively little information that can be used to constrain the parameter. Also, a repeat experimental design in the offset range 750 m−2250 m provides less postexperimental information than having a small offset separation.The information gain expected by performing the experiment is given by

$$Ent\{\sigma(\mathbf{d}|\xi)\} - Ent\{\sigma(\mathbf{d}|\mathbf{m},\xi)\} \tag{19}$$

where $Ent\{\sigma(\mathbf{d}|\xi)\}$ is the calculated optimal entropy value, and $Ent\{\sigma(\mathbf{d}|\mathbf{m},\xi)\}$ represents the entropy of the measurement noise and is described by a Gaussian. The difference

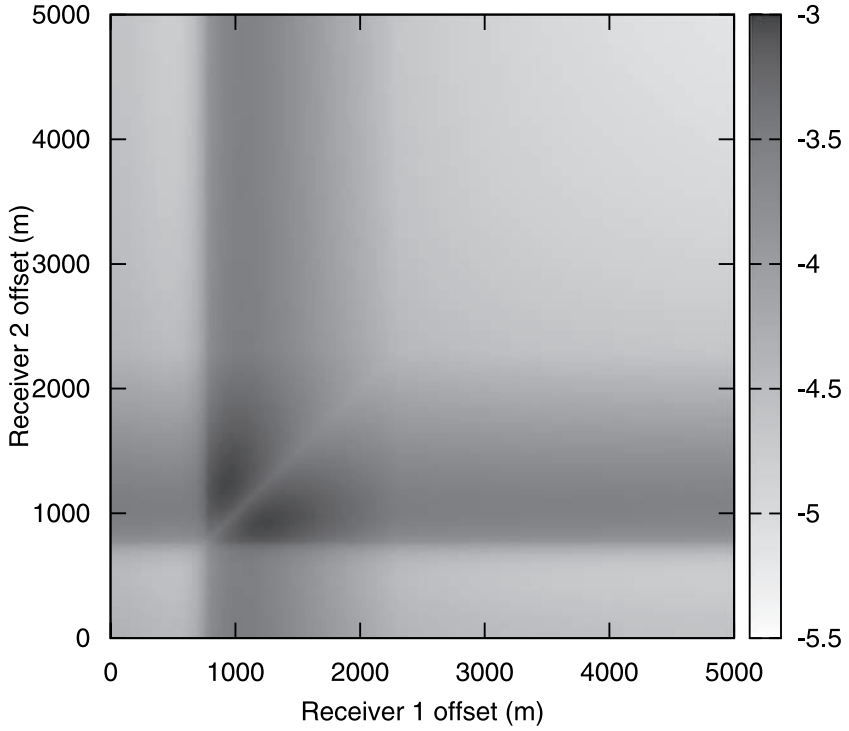

Figure 10. Entropy values for every possible two-receiver experimental design. The plot consists of 40,401 entropy values. The globally optimal receiver locations are at offsets of 1275 and 950 m.

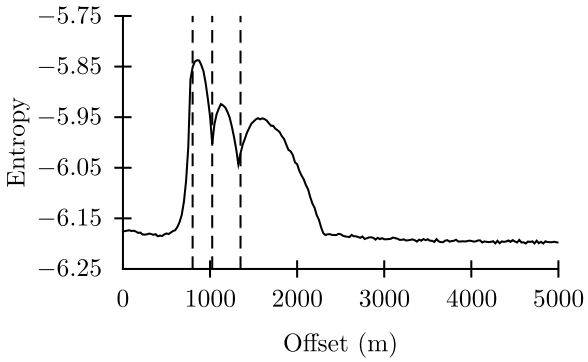

Figure 11. Entropy values as a function of offset for a three-receiver experimental design using the iteratively constructive method with the first two receivers fixed at offsets of 1050 and 1350 m. The dashed lines represent the globally optimal three-receiver design (Figure 12).

in the information gain found using the iteratively constructive method compared to that found by performing a complete design space search is 1.7%, while the global search takes 201 times longer to perform.

[42] Figure 11 show the entropy values obtained for the three-receiver optimal experiment calculated using the iteratively constructive method. In the two-receiver plot (Figure 9) there were 2 entropy peaks of similar magnitude. The three-receiver experiment clearly locates the third receiver at an offset of 850 m. The local entropy minima seen at approximately 500 m offset in both the single and two-receiver setup is now less prominent. The entropy decline with offsets greater than 2500 m is also less steep than in the previous designs, indicating that any offset greater than 2500 m will provide the same, minimal amount of postexperimental information. The entropy signal at large offsets also shows small amounts of numerical noise, indicating poor sampling in the data space. In the single-receiver case 500,000 parameter samples populated a maximum of 100 data space discretizations, whereas in the three-receiver case the same number of samples are used to populate a data space with 1,000,000 (100^3) discretizations. The entropy signal-to-noise ratio within the offset range 600 m to 2300 m is nevertheless still high as indicated by the smooth entropy function. The two entropy notches at offsets of 1050 m and 1350 m represent offsets where a receiver is already located from previous iterations.

[43] We also calculated the globally optimal three-receiver experiment by searching the entire experimental design space. Calculating every possible experimental design at 25 m offset spacing was too expensive computationally, so Figure 12 shows an entropy isosurface for the three-receiver design produced using 100 m offset spacings. Since the maximum seems to occur in the offset range 500 m to 2000 m a more detailed search at 25 m spacings was performed over that interval. The globally optimal experimental design uses offsets of 800 m, 1025 m and 1350 m. This is almost an exact replication of the design produced by our iterative method.

[44] A total of 359,632 entropy values were calculated to locate the three globally optimal receiver offsets. Only 603 entropies were calculated using the iteratively construc-

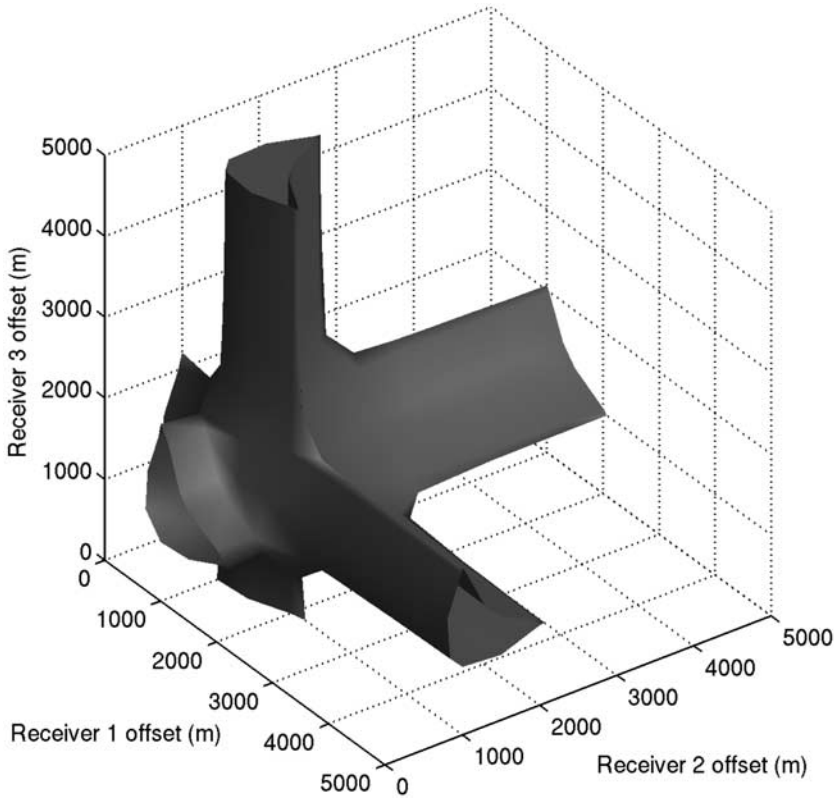

Figure 12. Entropy isosurface for a three-receiver experimental design using global sampling. The isosurface represents entropy values equal to 75% of the maximum.

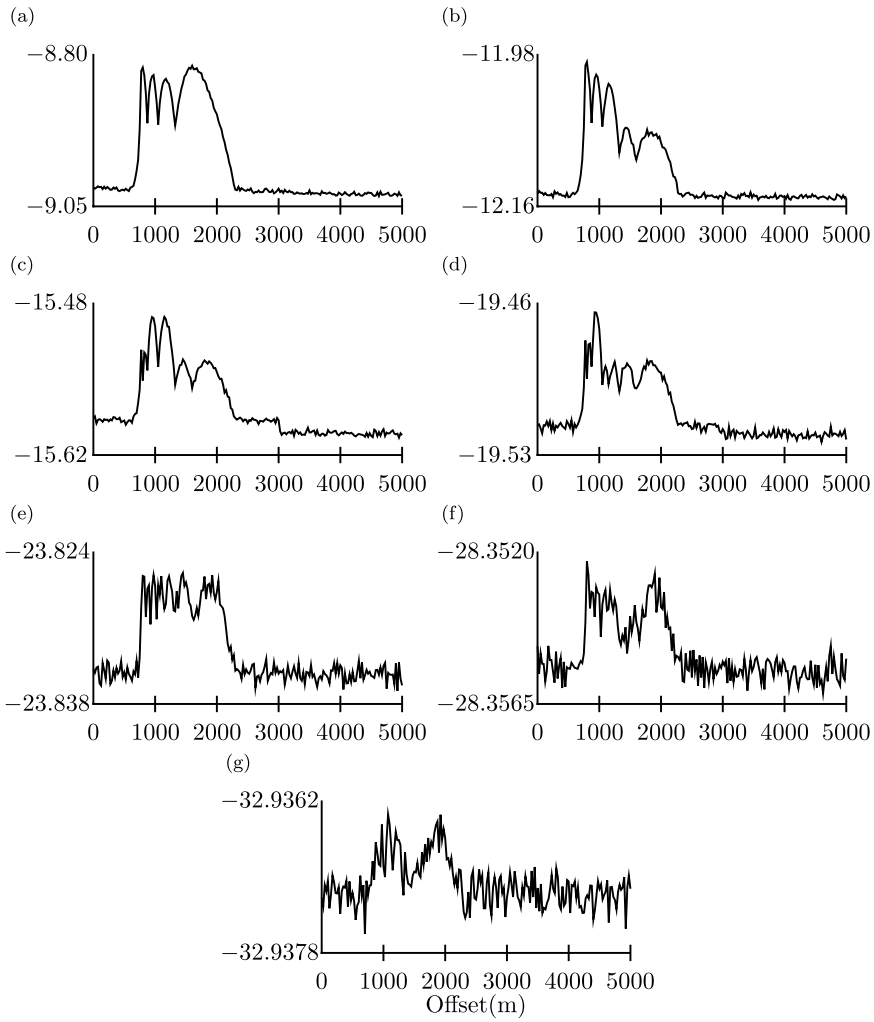

Figure 13. Entropy plots for successive receiver locations using our iteratively constructive method. (a) The entropy curve used to place the fourth receiver, and (g) the curve used to place the tenth receiver. For each plot, 500,000 parameter samples are used and the discretization length in each data-space dimension is 0.01.

tive method, a computational saving of over 99%. The same general features of the two-receiver design (Figure 10) are seen in the 3 receiver case (Figure 12). Offsets greater than 2500 m result in experiments with a low associated entropy value, as do experiments with offsets smaller than 500 m.

[45] The iteratively constructive method was then used to place 10 receivers in total. For each iterative design the same number of parameter space samples and data space discretization lengths have been used. Figure 13 shows the entropy results for successive experimental designs. Clearly the last few plots in Figure 13 are contaminated by numerical noise, a point to which we return below.

[46] The final 10-receiver experimental design is illustrated in Figure 14. Note that 9 of the 10 receivers are placed within the range 500 m to 1500 m (between 1 and 3 times the depth to the reflector), the range often used in AVO rule-of-thumb design methods. However, 1 of the receivers lies outside of this range, corresponding to the fourth receiver. The fourth receiver offset (Figure 13a) is located with high numerical accuracy indicating that the classical heuristic rules fail to match the mathematics-based design criterion in this simple problem.

[47] As the number of receivers increases so does the noise level in the entropy signal. We can locate the seventh receiver (Figure 13d) offset with some certainty but for the eighth, ninth, and tenth receivers the poor data space sampling causes the noise level to be too large to locate each receiver offset accurately. However, there is a clear

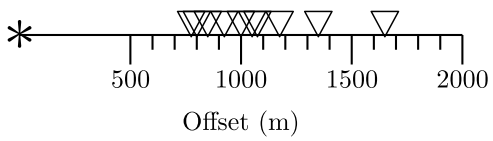

Figure 14. Experimental design using 10 receivers placed using the iteratively constructive method and 500,000 parameter-space samples. The source is denoted by asterisk and each receiver by an inverted triangle.

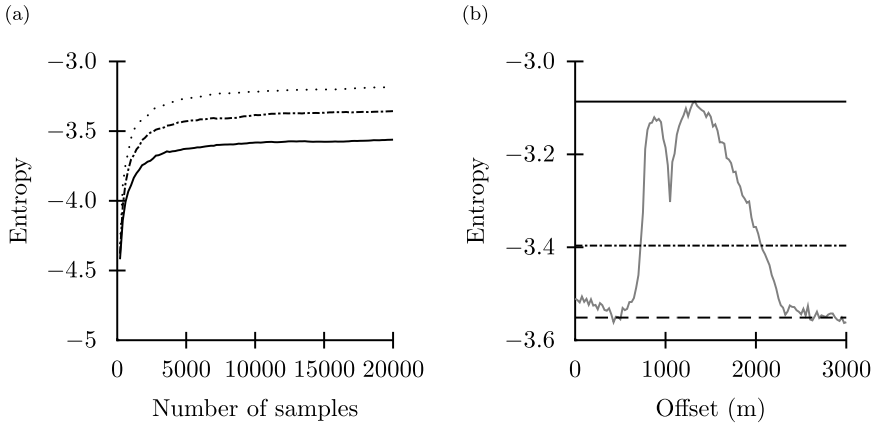

Figure 15. (a) The entropy estimates for offsets of 1000 m (dotted), 2000 m (dot-dash), and 3000 m (solid), as increasing number of samples are drawn from the parameter space. (b) The resulting entropy profile (gray line) produced from 20,000 samples. The solid line indicates the maximum entropy value, the dot-dashed line indicates two standard deviations, and the dashed line indicated three standard deviations from the maximum.

trend seen in the first 9 receiver plots that places all the receivers at offsets between 500 m and 2000 m. Also, in the numerically accurately placed receivers (1–7) a repeat experiment is never the optimal solution.

5. Design of an Efficient Numerical Sampling Scheme

[48] Accurately locating more than 7 receivers using the new method requires a more densely sampled data space than used above to reduce the noise level in the entropy signal. The number of samples required to locate each successive receiver is therefore a tradeoff between the desired signal-to-noise ratio of the entropy response and the amount of computation used. We have approached this tradeoff as requiring a separate design procedure that is performed as the number of samples is increased.

[49] The minimum required number of parameter space samples in the iteratively constructive method is determined by assessing the standard deviation of the entropy estimates. It should be noted that the use of "standard deviation" is only used as one possible variation measure in order to assess whether sufficient samples have been drawn, and not in the classical sense of characterizing a p.d.f. As more samples are drawn from the parameter space the entropy asymptotes to a constant value (Figure 3). We evaluate this convergence by considering the standard deviation of the entropy signature at offsets of 1000 m, 2000 m, and 3000 m which approximately spans the range of offsets likely to be optimal given the above tests. At each of these offsets the entropy value is calculated after each subsequent 200 samples have been drawn from the parameter space. The standard deviations are calculated over each consecutive set of 100 data space entropy values. The standard deviation measures the variation in 100 entropy values over the addition of 20,000 parameter space samples. Hence the standard deviation should be observed to decrease with every successive set of 20,000 samples as the entropy estimate converges.

[50] Figure 15a shows how the entropy estimate changes as samples are drawn from the parameter space for the AVO experiment that locates the second optimal receiver using the iteratively constructive method. The standard deviation is calculated for each offset profile. The maximum standard deviation from the three profiles is retained as the value used in the algorithm below.

[51] Figure 15b shows the resulting entropy profile (gray line) estimated using 20,000 samples. We determine if sufficient samples have been drawn from the parameter-space by assessing the entropy signature between the global maximum (indicated by the solid black line) and 2 standard deviations from this value (dot-dash line). If no other local maxima are found in this range we conclude that with 95% confidence we have found the true maximum value and no more samples are required. If, as in this example, there are local maxima in the two standard deviation range another 20,000 samples are added at each offset. The standard deviation is recalculated as above and the entropy profile reevaluated. This process is repeated until a single peak is found within the two standard deviation lower entropy limit. To reduce the computational time we discard offsets that have associated entropy estimates that fall below 3 standard deviations (dashed line) from the maximum entropy value when adding extra samples these offsets are not considered.

[52] This process creates a dramatic reduction in the computational time required to place multiple receivers. We recalculated the optimal 10-receiver design using the iteratively constructive method and the variable sample size using the above algorithm. The final design is shown in Figure 16 (top). Comparing Figure 16 to Figure 14, the new results locate 2 receivers beyond 1500 m instead of only 1. Harder to see is that the experiments that required fewer than 500,000 samples all located the optimal receiver at the same offset as the previous experiment, showing that we had needlessly over sampled the data space in the previous results. Conversely, in the experiments that required more than 500,000 samples, the receivers were positioned in different locations indicating that the data space was not

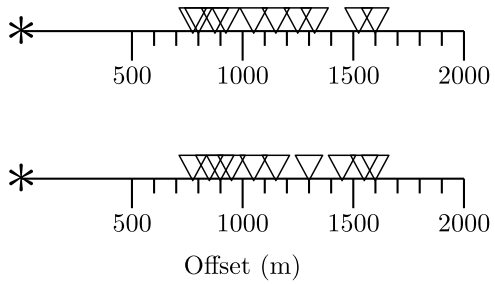

Figure 16. Experimental design using 10 receivers placed using the iteratively constructive method with the variable sample size test. Top design represents optimal 10-receiver design to 95% confidence and lower design to 66% confidence.

previously adequately sampled to reveal the true global maximum due to numerical noise.

[53] Figure 17 shows the actual number of samples required to accurately locate the receivers. In the previous method we had over sampled the first five receiver locations but under sampled the remaining five. In fact, using the variable sample size method also shows that only 40% of the samples were required to place the first 5 receivers when compared to the iteratively constructive method that used a static 500,000 samples for each receiver.

[54] The requirement of 95% confidence in finding global entropy maxima for successive receivers is arbitrary. Figure 16 (bottom) also shows the design obtained if we are satisfied with 66% confidence (one standard deviation), while Figure 17 also shows the number of samples in this case. 9 of the 10 receivers placed with 66% confidence match within ±25 m of the receivers offsets found to 95% confidence. Placing the ninth and tenth receivers with 95% confidence requires approximately 2.5×10^6 more samples than with 66% confidence.

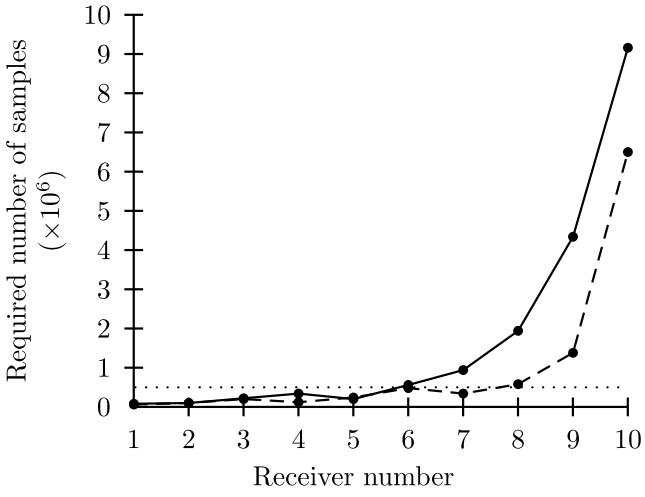

Figure 17. The number of samples required to accurately locate the optimal receiver using the iteratively constructive method to 95% (solid line) and 66% (dashed line) confidence. The dotted line shows the number of samples used in our previous example (Figure 14); we had oversampled the first five experiments but undersampled the final five.

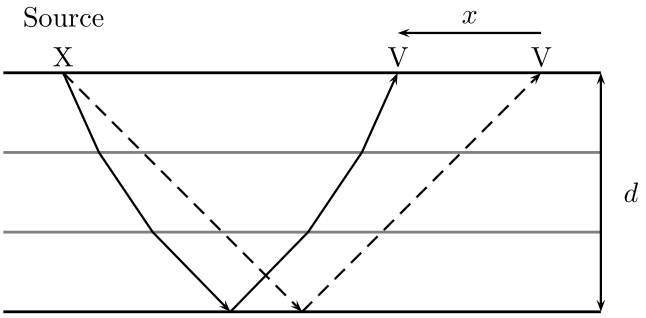

Figure 18. Raypath differences caused by having a different subsurface model. The dashed line represents a model with no refraction before the targeted interface, whereas the solid line refracts at two interfaces (gray lines), both before and after reflection. Both raypaths have the same angle of incidence and reflection at the lower-most interface.

6. Discussion of AVO Results

[55] The final designs for the AVO application appear to contradict accepted heuristic design methods by placing one or more receivers at offsets greater than 3 times the depth to the reflecting layer. However, rather than this being a general, design-independent phenomenon, the final design can be shown to depend strongly on the detailed subsurface Earth structure above the reflecting boundary, on the prior information about parameters, and on the data uncertainty. While the two-layer structure in Figure 6 might be representative of a shallow geophysical experiment, a more representative subsurface structure for experiments incorporating deeper layers would typically contain several layers above the target interface, each with different properties, and velocity would usually increase with depth. At every interface the raypath of the seismic energy would be refracted toward the horizontal at increasing depths in the model. The refraction process results in a reduced offset at the surface for the same angle of incidence and hence reflection coefficient at the subsurface interface (Figure 18). That is, the angular range at the bottom reflecting boundary created using our optimal designs in Figure 16 could also be obtained with receivers placed at smaller offsets.

[56] Reducing the size of the prior parameter range both decreases the computational power required to place each successive receiver and changes the final design. Figure 19

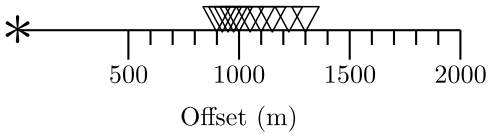

Figure 19. Experimental design using 10 receivers placed using the iteratively constructive method with the variable sample size test and a reduced prior parameter range. The source is denoted by asterisk, and each receiver is denoted by an inverse triangle.

represents the final 10 receiver design after the parameter range has been reduced from [3000 m s^{-1}, 4500 m s^{-1}] to [3375 m s^{-1}, 4125 m s^{-1}]. The design shows that all receivers now conform to the heuristic design by falling within the offsets 500 m to 1500 m, and receiver density decreases with increasing offset.

[57] The final simplification that we used was to have a uniform data uncertainty with increasing offset. In many realistic situations uncertainty might increase with offset. An increased data uncertainty results in a larger volume of the data space being sampled in the entropy calculation. As seen in Figure 4 this results in reduced $\sigma(\mathbf{d})$ values and larger entropy; optimal receiver locations would then tend toward longer offsets than in Figure 19.

[58] From the three factors given above it is clear that the heuristic design statement is oversimplified for AVO design problems. Each situation must be designed independently to optimize information, incorporating as much prior information about subsurface structure and expected data uncertainty as possible.

7. Conclusions

[59] An iteratively constructive design method has been presented which is applicable to nonlinear, multidimensional design problems. The method is based on a Bayesian framework which is suited to nonlinear scenarios where classical design methods, based on forward function gradients, fail.

[60] For both the synthetic sawtooth and practical AVO examples, locally optimum experimental designs produced by the iteratively constructive method closely resemble globally optimal designs, in low-dimensional cases where the later designs can be calculated. However, we have also produced a multireceiver AVO experimental design using the iterative method that would have been computationally intractable using grid-search nonlinear design methods. The parameter space sample size has been treated as both a fixed parameter, and a variable which requires optimization at each iteration. This extra optimization process increases the efficiency of the new method significantly. A simple search of the possible offsets at uniform intervals through the design space has been used to locate the optimal receiver offset. To reduce computation, the only limiting factor to the number of possible data space dimensions, it is highly recommended that more efficient methods are used to search the already-limited design space and to select the optimal discretization length. Given more efficient algorithms, designing experiments for more complex applications which use both the P wave and S wave reflection coefficient for a multilayered subsurface now promises to be computationally feasible.

[61] **Acknowledgments.** We thank Schlumberger for permission to publish this work and Emanuel Winterfors for his insight and stimulating discussions. We acknowledge the support from the Scottish Funding Council for the ECOSSE Joint Research Institute, part of the Edinburgh Research Partnership in Engineering and Mathematics (ERPem).

References

Aki, K., and P. G. Richards (2002), *Quantitative Seismology*, Univ. Science Books, Calif.

Atkinson, A., and A. Donev (1992), *Optimum Experimental Designs*, Oxford Univ. Press, Oxford, U. K.

Barth, N., and C. Wunsch (1990), Oceanographic experiment design by simulated annealing, *J. Phys. Oceanogr.*, *20*(9), 1249–1263.

Box, G., and H. Lucas (1959), Design of experiments in non-linear situations, *Biometrika*, *46*, 77–90.

Chaloner, K., and I. Verdinelli (1995), Bayesian experimental design: A review, *Stat. Sci.*, *10*(3), 273–304.

Curtis, A. (1999a), Optimal experiment design: Cross-borehole tomographic examples, *Geophys. J. Int.*, *136*, 637–650.

Curtis, A. (1999b), Optimal design of focused experiments and surveys, *Geophys. J. Int.*, *139*, 205–215.

Curtis, A. (2004a), Theory of model-based geophysical survey and experimental design: part 1. Linear problems, *Leading Edge*, *23*(10), 997–1004.

Curtis, A. (2004b), Theory of model-based geophysical survey and experimental design: part 2. Nonlinear problems, *Leading Edge*, *23*(10), 1112–1117.

Curtis, A., and H. Maurer (2000), Optimizing the design of geophysical experiments: Is it worthwhile?, *Leading Edge*, *10*, 1059–1062.

Curtis, A.and R. Wood, (2004), Optimal elicitation of probabilistic information from experts, in *Geological Prior Information*, *Geol. Soc. of London Spec. Publ. 239*, edited by A. Curtis and R. Wood, Geol. Soc. of London, London.

Curtis, A., A. Michelini, D. Leslie, and A. Lomax (2004), A deterministic algorithm for experimental design applied to tomographic and microseismic monitoring surveys, *Geophys. J. Int.*, *157*, 595–606.

Fedorov, V. V. (1972), *Theory of Optimal Experiments*, Elsevier, New York.

Ford, I., D. M. Titterington, and C. P. Kitsos (1989), Recent advantages in nonlinear experimental design, *Technometrics*, *31*(1), 49–51.

Furman, A., T. P. A. Ferre, and A. W. Warrick (2004), Optimization of ert surveys for monitoring transient hydrological events using perturbation sensitivity and genetic algorithms, *Vadose Zone J.*, *3*, 1230–1239.

Furman, A., T. P. A. Ferre, and G. L. Heath (2007), Spatial focusing of electrical resistivity surveys considering geologic and hydrologic layering, *Geophysics*, *72*(2), F65–F73.

Hamada, M., H. F. Martz, C. S. Reese, and A. G. Wilson (2001), Finding near-optimal bayesian experimental designs via genetic algorithms, *Am. Stat.*, *55*(3), 175–181.

Hu, I. (1998), On sequential designs in nonlinear problems, *Biometrika*, *85*(2), 496–503.

Kijko, A. (1977a), An algorithm for the optimal distribution of a regional seismic network, I, *Pure Appl. Geophys.*, *115*, 999–1009.

Kijko, A. (1977b), An algorithm for the optimum distribution of a regional seismic network: II. An analysis of the accuracy of location of local earthquakes depending on the number of seismic stations, *Pure Appl. Geophys.*, *115*, 1011–1021.

Lindley, D. (1956), On a measure of the information provided by an experiment, *Ann. Math. Stat.*, *27*(4), 986–1005.

Maurer, H., and D. E. Boerner (1998), Optimized design of geophysical experiments, *Leading Edge*, *17*(8), 1119.

Maurer, H., D. E. Boerner, and A. Curtis (2000), Design strategies for electomagnetic geophysical surveys, *Inverse Probl.*, *16*, 1097–1117.

Muller, P., and G. Parmigiani (1996), Numerical evaluation of information-theoretic measures, in *Bayesian Analysis in Statistics and Economics: Essays in Honor of Arnold Zellner*, edited by D. Berry, K. Chaloner, and J. Geweke, pp. 397–406, Wiley-Interscience, Hoboken, N. J.

Rabinowitz, N., and D. M. Steinberg (2000), A statistical outlook on the problem of seismic network configuration, in *Advances in Seismic Event Location, Modern Approaches in Geophysics*, vol. 13, edited by C. Thurber and N. Rabinowitz, chap. 3, Kluwer Academic Publishers, Amsterdam.

Ryan, K. J. (2003), Estimating expected information gains for experimental designs with application to the random fatigue-limit model, *J. Comput. Graph. Stat.*, *12*(3), 585–603.

Shannon, C. E. (1948), A mathematical theory of communication, *Bell Syst. Tech. J.*, *27*, 623–656.

Shewry, M. C., and H. P. Wynn (1987), Maximum entropy sampling, *J. Appl. Stat.*, *14*, 165–170.

Silvey, S. D. (1980), *Optimal Design - An Introduction to the Theory for Parameter Estimation*, CRC Press, Boca Raton, Fla.

Steinberg, D. M., N. Rabinowitz, Y. Shimshoni, and D. Mizrachi (1995), Configuring a seismographic network for optimal monitoring of fault lines and multiple sources, *Bull. Seismol. Soc. Am.*, *85*(6), 1847–1857.

Stummer, P., H. Maurer, H. Horstmeyer, and A. G. Green (2002), Optimization of dc resistivity data acquisition: Real-time experimental design and a new multielectrode system, *IEEE Trans. Geosci. Remote*, *40*(12), 2727–2735.

Stummer, P., H. Maurer, and A. G. Green (2004), Experimental design: Electrical resistivity data sets that provide optimum subsurface information, *Geophysics*, *69*(1), 120–139.

van den Berg, J., A. Curtis, and J. Trampert (2003), Optimal nonlinear bayesian experimental design: An application to amplitude versus offset experiments, *Geophys. J. Int.*, *155*, 411–421.

Wilkinson, P., J. Chambers, P. Meldrum, R. Ogilvy, and S. Caunt (2006), Optimization of array configurations and panel combinations for the detection and imaging of abadoned mineshafts using 3d cross-hole electrical resistivity tomography, *J. Environ. Eng. Geophys.*, *11*, 213–221.

Winterfors, E., and A. Curtis (2008), Numerical detection and reduction of non-uniqueness in nonlinear inverse problems, *Inverse Probl.*, *24*(2), 025,016, 14 pp.

Yilmaz, Ö. (2001), *Seismic Data Analysis*, Society of Exploration Geophysicists, Tulsa, Okla.

A. Curtis and T. Guest, Grant Institute, School of GeoSciences, University of Edinburgh, The King's Buildings, West Mains Road, Edinburgh EH9 3JW, UK. (andrew.curtis@ed.ac.uk; t.e.guest@sms.ed.ac.uk)