Geophys. J. Int. (2003) **155,** 411–421

Optimal nonlinear Bayesian experimental design: an application to amplitude versus offset experiments

Jojanneke van den Berg,[1,*] Andrew Curtis[2,3] and Jeannot Trampert[1]

[1] *Faculty of Earth Sciences, Utrecht University,* PO Box 80021, 3508 TA, *Utrecht, the Netherlands. E-mail: J.vandenberg@phys.uu.nl*
[2] *Schlumberger Cambridge Research, High Cross, Madingley Road,* CB3 0EL *Cambridge, United Kingdom*
[3] *Department of Geology and Geophysics, Edinburgh University, Grant Institute, West Mains Road, Edinburgh* EH9 3JW, *United Kingdom*

Accepted 2003 May 5. Received 2003 May 6; in original form 2002 October 22

SUMMARY

When designing an experiment, the aim is usually to find the design which minimizes expected post-experimental uncertainties on the model parameters. Classical methods for experimental design are shown to fail in nonlinear problems because they incorporate linearized design criteria. A more fundamental criterion is introduced which, in principle, can be used to design any nonlinear problem. The criterion is entropy-based and depends on the calculation of marginal probability distributions. In turn, this requires the numerical calculation of integrals for which we use Monte Carlo sampling. The choice of discretization in the parameter/data space strongly influences the number of samples required. Thus, the only practical limitation for this technique appears to be computational power. A synthetic experiment with an oscillatory, highly nonlinear parameter–data relationship and a simple seismic amplitude versus offset (AVO) experiment are used to demonstrate the method. Interestingly, in our AVO example, although overly coarse discretizations lead to incorrect evaluation of the entropy, the optimal design remains unchanged.

Key words: Bayesian, design, experimental, inversion, nonlinear, survey.

1 INTRODUCTION

Finding an optimal geometry, or design, of a practical experiment often means finding the design which maximizes the expected post-experimental information of particular model parameters of interest. This is equivalent to minimizing the expected post-experimental uncertainties in those model parameters. Thus, experimental design requires an understanding of the relationship between data and post-experimental model parameter uncertainties (Box & Lucas 1959; Atkinson & Donev 1992; Curtis 1999a,b; Curtis & Spencer 1999; Curtis & Maurer 2000).

Consider a linearized problem in which the forward problem of estimating data **d** given a model parameter vector **m**, and the associated inverse problem solution are, respectively,

$$\mathbf{d} = \mathbf{G_{m_0}} \mathbf{m} \tag{1}$$

$$\mathbf{m} = \left(\mathbf{G_{m_0}}^T \mathbf{G_{m_0}}\right)^{-1} \mathbf{G_{m_0}}^T \mathbf{d} = \mathbf{L} \mathbf{d}, \tag{2}$$

where $\mathbf{G_{m_0}}$ is a matrix of derivatives of **d** with respect to **m** calculated at a reference model $\mathbf{m_0}$, matrix **L** is defined in eq. (2), and the matrix inversion in eq. (2) represents the classical least-squares solution and must be replaced by $\mathbf{m} = \mathbf{G_{m_0}}^T(\mathbf{G_{m_0}} \mathbf{G_{m_0}}^T)^{-1}$ **d** if the problem is under-determined. In the presence of null spaces, the inverse operator **L** needs to be replaced by a more general expression including regularization (Tarantola 1987). To be consistent with geophysical literature we refer to **m** as the model and call $\mathbf{G_{m_0}}$ the forward function (which is fixed for any particular experimental design). Note that this differs from terminology in statistical experimental design literature (thus in many references cited in this paper) where the 'model' usually includes both $\mathbf{G_{m_0}}$ and **m**.

Uncertainties in the data are projected into the model parameter space as $\mathbf{L}\mathbf{C_d}\mathbf{L}^T$, where $\mathbf{C_d}$ is the covariance matrix, so any method for performing experimental design must alter either $\mathbf{G_{m_0}}$ or the uncertainties in **d**. For a linear problem (where $\mathbf{G_{m_0}}$ is constant with respect to $\mathbf{m_0}$, hence we can write simply **G**), most methods for experimental design are based on optimizing the eigenvalue spectrum of $(\mathbf{G}^T\mathbf{G})$. That is, the spectrum should have as large eigenvalues and be as flat as possible (Curtis 1999a,b). This is illustrated in Fig. 1, which shows the geometry for a simple 1-D experiment with one source and one receiver. The aim of the experiment is to choose the best offset *x* for the estimation of the single slowness *m* of the half-space below the surface from the traveltime *d* of a direct wave between source and receiver. For this example a slowness of $1/1500$ s m^{-1} was used as the true slowness of the medium.

Fig. 2 shows model-data relationships between *m* and *d* for 50 m and 100 m offsets. The plots show the projection of an uncertainty in a measured datum d_0 of ±0.01 seconds into the model parameter

[*] Now at: Institute for Marine and Atmospheric Research, Utrecht University, Princetonplein 5, 3584 CC, Utrecht, the Netherlands.

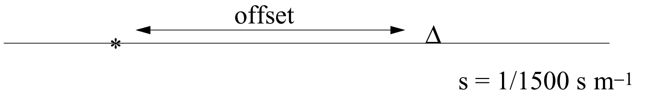

Figure 1. Geometry of a simple example of an experimental design. The theory is given by $d = xm$, where d is the time [s], m is the slowness [s/m] and x is the offset [m], or the distance between the source * and the receiver Δ. The experimental aim is to find the optimal offset for retrieving the true slowness of $1/1500$ s m^{-1}.

space. These figures show that the function relating model parameters m and data d is steeper for the larger offset. For a constant data uncertainty this results in a smaller uncertainty region around the true model parameter value. So, for this simple experiment, the larger offset is recommended for more accurate model parameter estimates, and generally, as this problem is linear, the longer the path through the medium the more accurate the results are likely to be (Johnson & Leone 1977; Squires 1985; Atkinson & Donev 1992).

The example above illustrates that the aim of linear experimental design can usually be thought of as increasing gradients **G** such that post-experimental uncertainties are minimized. Therefore, **G** must be estimated and maximized appropriately prior to conducting the experiment. In the above 1-D example, the single eigenvalue of $\mathbf{G}^T$ **G** is the gradient squared. Hence, maximizing the gradient in Fig. 2 is equivalent to maximizing the eigenvalue of $\mathbf{G}^T$ **G**.

In a linear problem **G** is constant over all reference model parameters $\mathbf{m_0}$ for any particular experimental design $\boldsymbol{\xi}$. Hence, it does not matter at which model parameter values the expected post-experimental uncertainty is estimated, since the same estimates for post-experimental uncertainty will be obtained whatever $\mathbf{m_0}$ is chosen.

This is not true in nonlinear or even pseudo-linear problems. In such situations $\mathbf{G_{m_0}}$ varies as a function of $\mathbf{m_0}$ and the true model parameter values are unknown. This leads to errors in the estimated post-experimental uncertainties if a single erroneous reference model $\mathbf{m_0}$ is used (Curtis & Spencer 1999). This problem is usually dealt with using classical nonlinear estimates for the quality of an expected design $\boldsymbol{\xi}$ (Box & Lucas 1959; Johnson & Leone 1977; Ford et al. 1989; Atkinson & Donev 1992; Chaloner & Verdinelli 1995; Curtis & Spencer 1999):

$$\Phi(\boldsymbol{\xi}) = \int_M \phi(\boldsymbol{\xi}, \mathbf{m})\rho(\mathbf{m})\, d\mathbf{m}. \tag{3}$$

In this equation $\phi(\boldsymbol{\xi}, \mathbf{m})$ is called a quality measure and usually consists of some measure of elements or eigenvalues of the forward operator $\mathbf{G}_\mathrm{m}$ estimated at model parameter value **m**, where this measure reflects the expected quality of a specific design $\boldsymbol{\xi}$ (the most commonly used measure is $\phi(\boldsymbol{\xi}, \mathbf{m}) = \det\,(\mathbf{G_m}^T\mathbf{G_m})$ for design $\boldsymbol{\xi}$). That is, $\phi(\boldsymbol{\xi}, \mathbf{m})$ reflects the expected post-experimental uncertainty of model parameter estimates if **m** contains the true model parameter values and the model-data relationship is approximately linear around **m** (within the data uncertainties). So, instead of using one reference model parameter value $\mathbf{m_0}$ to estimate the post-experimental uncertainties, in eq. (3) a distribution of reference models $\rho(\mathbf{m})$ is used where $\rho(\mathbf{m})$ embodies the prior knowledge about the likelihood of possible model parameter values being the true values. In this way the quality measure $\Phi(\boldsymbol{\xi})$ is an average measure of expected information over the entire feasible portion of the model parameter space.

There is a problem with this approach arising from the definition of $\phi(\boldsymbol{\xi}, \mathbf{m})$, which is generally a gradient-based measure. Curtis & Spencer (1999) showed that in the case of a truly nonlinear situation, an error is committed whenever models within disconnected regions in the model parameter space might fit the measured data to within their given uncertainties. We illustrate this with the following example, extended from Curtis & Spencer (1999).

Fig. 3 shows two 1-D sawtooth functions with different periods of oscillation. Let G_{m_0} be the derivative matrix (in this case a single value) $[df/dm|_{m_0}]$, where f is the sawtooth function and we ignore points where G_{m_0} is not defined, let $\phi(\xi, m_0)$ be the determinant of $G_{m_0}^T G_{m_0}$, the classical experimental design measure (Box & Lucas 1959; Ford et al. 1989; Atkinson & Donev 1992; Chaloner & Verdinelli 1995). In linear problems, this determinant is constant with respect to m_0 and represents a measure of the extent to which measurement uncertainties propagate into expected post-experimental model parameter uncertainties, similar to the slowness example above. Let us now examine the use of this measure in the nonlinear sawtooth example.

Data uncertainties of plus or minus 0.2 around a datum $d_0 = 2$ are projected into the model parameter space using the linearized local gradient method around a single model parameter value m_0 in each plot in Fig. 3. This is visualized by the error bar projections flanked by solid lines in the figures. As the derivative of the right function is larger than the derivative of the left function, the corresponding projected uncertainty is smaller in the right plot. In each plot individually, identical values for $\phi(\xi, m_0)$ are found for any model parameter value m_0, since the magnitude of $G_{m_0}^T G_{m_0}$ is constant at all points on the sawtooth functions (ignoring the extrema). However, the quality measure $\Phi(\xi)$ calculated using eq. (3) gives $\Phi(\xi) = 1$ for the left figure and $\Phi(\xi) = 4$ for the right figure. This is because, for fixed data uncertainties, the linearized model

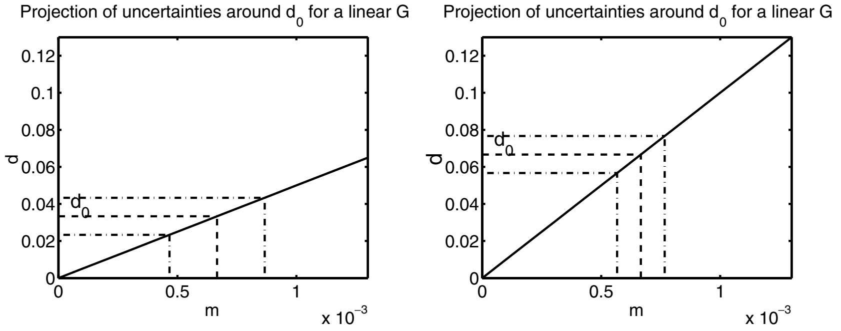

Figure 2. Model-data relationships for the geometry shown in Fig. 1. In the left figure an offset of $x = 50$ m has been used, for the right figure $x = 100$ m. The figures show the change in gradient G as a function of the offset. The dashed line and the dash-dotted lines represent the projection of the observed datum d_0 with an observational uncertainty of ±0.01 into the model parameter space.

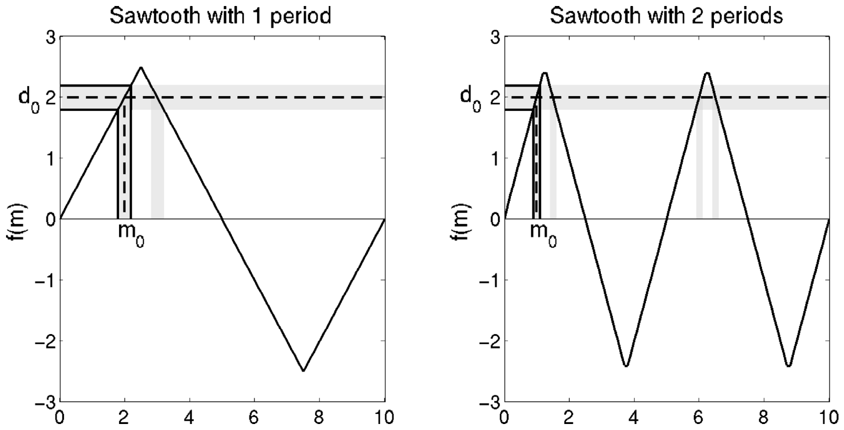

Figure 3. Sawtooth functions with 1 period (left) and 2 periods (right) between $m = 0$ and $m = 10$. In both figures a datum $d_0 = 2 \pm 0.2$ is projected into the model parameter space using the linearized local gradient method, shown by the error bar projections between the solid lines. The union of all of the grey error bar projections onto the model parameter space represents the true model parameter uncertainty in each case.

parameter uncertainties bounded by the solid lines are smaller in the right figure than in the left figure. So, according to the classical estimate for design quality, eq. (3), the experimental design producing the right figure would be a better one than the design producing the left figure.

However, the uncertainties described above are local, linearized approximations and account only for one of several possible regions in model parameter space that fit the measured data. In fact, the true uncertainty in model parameter space is given by the union of all of those possible regions, represented by the set of all vertical grey regions in Fig. 3. As a result, the true post-inversion uncertainty for both figures is exactly equal (the sum of the uncertainties in the model parameter space in each case is 0.8). So in contrast to the result from classical design theory above, the design producing the left figure is probably a better design: the solution to the inverse problem of estimating model parameters given any measured data d_0 is easier to calculate and represent, since it is less fragmented than in the right figure (Curtis & Spencer 1999).

When designing a nonlinear experiment $\Phi(\xi)$ is usually maximized. In the above case this would mean that the final design would produce neither of the two functions in Fig. 3, but instead would produce a sawtooth with as many extrema as possible. Similarly to above, this would fragment the region of model parameter space fitting any observed data, and in reality would provide no reduction at all in post-experimental model parameter uncertainties. Maximizing $\Phi(\xi)$ according to classical measures for the quality of a design would therefore result in a more difficult inverse problem to solve with no expected gain in information. Hence, classical nonlinear (linearization-based) experimental design measures are not robust in nonlinear situations.

2 THEORY AND METHOD

To be able to construct a measure for experimental design that will work for any nonlinear experiment, a framework without linearization is now introduced. We use a Bayesian approach for model parameter inference in which probability density functions (p.d.f.s) represent a given state of information.

According to Tarantola & Valette (1982), the solution to an inverse problem is given by the posterior, or post-experimental p.d.f.,

$$\sigma(\mathbf{d}, \mathbf{m}) = \frac{\rho(\mathbf{d}, \mathbf{m})\theta(\mathbf{d}, \mathbf{m})}{\mu(\mathbf{d}, \mathbf{m})}, \qquad (4)$$

where $\rho(\mathbf{d}, \mathbf{m})$ represents the prior knowledge on data **d** and model parameters **m**, $\theta(\mathbf{d}, \mathbf{m})$ represents the information about the physics relating data and model parameters, and $\mu(\mathbf{d}, \mathbf{m})$ is called null information and represents an objective reference state of minimum information (Tarantola & Valette 1982; Tarantola 1987). We will adopt the following convention (which differs from that of Tarantola & Valette 1982). We include within $\theta(\mathbf{d}, \mathbf{m})$ the entire (uncertain) relationship between the actual data measurements recorded and the model parameters. Thus, $\theta(\mathbf{d}, \mathbf{m})$ includes both (i) the relationship between **m** and idealized, noise-free data, and (ii) the relationship between these idealized data and the data values actually recorded in the experiment. Then, $\rho(\mathbf{d})$ includes only information about the actual data values recorded (and not on their assumed uncertainties). Thus, $\theta(\mathbf{d}, \mathbf{m})$ represents the physical relationship between data and model parameters including all uncertainties over which we have some influence through the experimental design, and $\rho(\mathbf{d}, \mathbf{m})$ contains only *a priori* information over which we have no control, other then through $\theta(\mathbf{d}, \mathbf{m})$. In contrast, Tarantola & Valette (1982), include relationship (i) within $\theta(\mathbf{d}, \mathbf{m})$ and relationship (ii) within $\rho(\mathbf{d})$.

In designing an experiment we aim to maximize the information about model parameters **m** that are expected to be contained within $\sigma(\mathbf{d}, \mathbf{m})$. Therefore, it is necessary to be able to quantify the information content of a p.d.f. The entropy of any random vector **X** may be defined in relation to Shannon's measure for information (Shannon 1948); see also Tarantola & Valette (1982), Shewry & Wynn (1987) and Sebastiani & Wynn (2000),

$$Ent(\mathbf{X}) = -I(f(\mathbf{x})) + c = -\int_X f(\mathbf{x}) \log(f(\mathbf{x}))\, d\mathbf{x}, \qquad (5)$$

where $f(\mathbf{x})$ is the p.d.f. of **X**, and I is the information content of a p.d.f. as defined by Shannon (1948). The measure of information I is equal to minus the entropy, except for a constant c assuming a uniform null distribution.

When designing an experiment a data set from that experiment is not available, so we set $\rho(\mathbf{d}) = \mu(\mathbf{d})$, assuming that the prior distribution can be decomposed as $\rho(\mathbf{d}, \mathbf{m}) = \rho(\mathbf{d})\rho(\mathbf{m})$ and that the null distribution can be decomposed similarly. The post acquisition information on model parameters **m** is described by the marginal posterior distribution,

$$\sigma(\mathbf{m}) = \int_D \sigma(\mathbf{d}, \mathbf{m})\, d\mathbf{d}. \qquad (6)$$

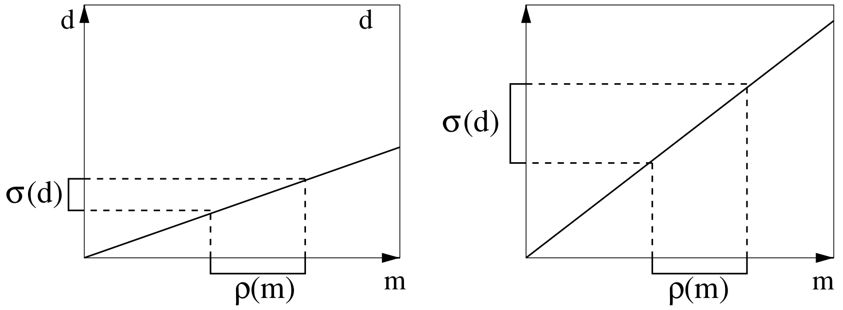

Figure 4. Uncertainty distributions in the data space corresponding to a fixed uniform uncertainty distribution $\rho(m)$ in the model parameter space for two linear forward functions with different gradients. The left figure has a small data space uncertainty and a low value of $Ent(\mathbf{d}\,|\,\boldsymbol{\xi})$, the right figure has a large data space uncertainty and a high value of $Ent(\mathbf{d}\,|\,\boldsymbol{\xi})$.

One might expect to be able to design an experiment such that information expected to be in $\sigma(\mathbf{m})$ is maximized. However, $\theta(\mathbf{d}, \mathbf{m})$ can often be decomposed as $\theta(\mathbf{d}, \mathbf{m}) = \theta(\mathbf{d}\,|\,\mathbf{m})\,\mu(\mathbf{m})$, i.e. $\theta(\mathbf{d}, \mathbf{m})$ incorporates no additional information on $\mathbf{m}$. Therefore, using $\rho(\mathbf{d}) = \mu(\mathbf{d})$, we often obtain from eq. (4) that $\sigma(\mathbf{d}, \mathbf{m}) = \rho(\mathbf{m})\;\theta(\mathbf{d}\,|\,\mathbf{m})$ and hence $\sigma(\mathbf{m}) = \rho(\mathbf{m})$. This does not vary with the experimental design $\boldsymbol{\xi}$ and hence cannot be used to determine $\boldsymbol{\xi}$.

Instead, experiments can be designed by maximizing the information expected to be contained in the conditional posterior p.d.f.,

$$\sigma(\mathbf{m}\,|\,\mathbf{d}) = \frac{\sigma(\mathbf{d}, \mathbf{m})}{\sigma(\mathbf{d})} \tag{7}$$

where $\sigma(\mathbf{m}\,|\,\mathbf{d})$ represents the probability of $\mathbf{m}$ being the true value for the model parameter given any data measurement $\mathbf{d}$. In eq. (7), $\sigma(\mathbf{d})$ is the marginal posterior distribution on the data $\mathbf{d}$:

$$\sigma(\mathbf{d}) = \int_M \sigma(\mathbf{d}, \mathbf{m})\, d\mathbf{m}. \tag{8}$$

Prior to conducting an experiment, $\sigma(\mathbf{d})$ embodies all information about what data are likely to be recorded during the experiment. In those frequent cases when $\sigma(\mathbf{d}, \mathbf{m}) = \rho(\mathbf{m})\;\theta(\mathbf{d}\,|\,\mathbf{m})$, we see that $\sigma(\mathbf{d})$ simply contains prior information on the model parameters projected into the data space through the physical relationship $\theta(\mathbf{d}\,|\,\mathbf{m})$.

A quality measure for nonlinear (*nl*) experimental design can then be defined as,

$$\Phi_{nl}(\boldsymbol{\xi}) = -\int_D Ent(\mathbf{m}\,|\,\mathbf{d}, \boldsymbol{\xi})\sigma(\mathbf{d})\, d\mathbf{d}, \tag{9}$$

where *Ent* is the entropy function and $-Ent(\mathbf{m}\,|\,\mathbf{d}, \boldsymbol{\xi})$ represents the amount of information contained in the conditional p.d.f. $\sigma(\mathbf{m}\,|\,\mathbf{d})$ about the model parameters given a particular data measurement $\mathbf{d}$ recorded using experimental design $\boldsymbol{\xi}$. This measure of information is weighted by the likelihood that data measurement $\mathbf{d}$ will be obtained when performing the experiment, $\sigma(\mathbf{d})$. Integration over all possible data measurements $\mathbf{d}$ results in $\Phi_{nl}(\boldsymbol{\xi})$.

$\Phi_{nl}(\boldsymbol{\xi})$ above and $\Phi(\boldsymbol{\xi})$ from classical nonlinear methods (equation 3) have similar form. The most important difference between $\Phi_{nl}(\boldsymbol{\xi})$ and $\Phi(\boldsymbol{\xi})$ is that the former requires no linearization of the model-data relationship. The concept of maximizing a gradient has not been used.

According to Shewry & Wynn (1987),

$$-\Phi_{nl}(\boldsymbol{\xi}) + Ent(\mathbf{d}\,|\,\boldsymbol{\xi}) = Ent(\mathbf{d}, \mathbf{m}\,|\,\boldsymbol{\xi}) = b, \tag{10}$$

where b is a constant, if $Ent(\mathbf{d}, \mathbf{m}\,|\,\boldsymbol{\xi})$ is design-independent. They demonstrate that this is the case for many geophysical problems. For example, the data-model parameter relationship $\theta(\mathbf{d}, \mathbf{m})$ used for most geophysical problems is,

$$\mathbf{d} = f(\mathbf{m}) + \boldsymbol{\epsilon} \tag{11}$$

where $\boldsymbol{\epsilon}$ is a vector of independent, random errors, which do not depend on either the model parameter space or the design. It can be demonstrated that eq. (10) holds for relationships of this type. Therefore, instead of maximizing $\Phi_{nl}(\boldsymbol{\xi})$, the optimal design can also be found by maximizing $Ent(\mathbf{d}\,|\,\boldsymbol{\xi})$. For this measure only information about $\sigma(\mathbf{d})$ is required, hence the calculation is simplified.

Eq. (10) can be explained intuitively using Fig. 4. For linear examples we showed earlier that the optimal design can be found by maximizing the gradient of the forward relationship between data and model parameters. Consider the case where $\sigma(\mathbf{d}, \mathbf{m}) = \rho(\mathbf{m})$ $\theta(\mathbf{d}\,|\,\mathbf{m})$ and $\theta(\mathbf{d}\,|\,\mathbf{m}) = D(f(\mathbf{m}) - \mathbf{d})$ where D is the Dirac delta function and $f(\mathbf{m})$ is a noise-free model-data relationship. Suppose we have two designs producing different 1-D, linear forward relationships $f(m)$, and that we have a fixed uniform prior model parameter distribution $\rho(m)$. Distribution $\sigma(d, m)$ projects this distribution through each forward relationship. We find that the corresponding posterior measurement uncertainties $\sigma(d)$ are uniform and different for the two designs. For the left figure (small gradient of $f(m)$, least informative design), the corresponding measurement uncertainty is small, hence we obtain a low value for $Ent(d\,|\,\xi)$. For the right figure (high gradient of $f(m)$, most informative design), the corresponding measurement uncertainty is large, hence we obtain a high value for $Ent(d\,|\,\xi)$. Therefore, the optimal design is the one which gives the highest value for $Ent(d\,|\,\xi)$. The theory above generalizes this for nonlinear and uncertain forward relationships in $\theta(\mathbf{d}, \mathbf{m})$.

For most nonlinear problems the p.d.f.s required to calculate $\Phi_{nl}(\boldsymbol{\xi})$ are not known analytically. Hence, for a generally applicable method all integrations must be defined numerically. After testing several methods we opted to use Monte Carlo integrations, (Lepage 1978),

$$\int_X g(\mathbf{x})\, d\mathbf{x} \approx \frac{1}{N} \sum \frac{g(\mathbf{x})}{s(\mathbf{x})}, \tag{12}$$

where N is the number of samples taken of a function $g(\mathbf{x})$, and $s(\mathbf{x})$ is the sampling distribution-the probability that a sample is drawn at position $\mathbf{x}$.

In our method, samples are drawn only from the region where $\sigma(\mathbf{d}, \mathbf{m})$ is non-zero. The samples are generated using $\rho(\mathbf{m})$ in the model parameter space and are distributed uniformly around the forward function $f(\mathbf{m})$ in the data space, whichever the shape of the distribution of expected measurement uncertainties. In the 1-D examples used in this paper, the true data uncertainty region around

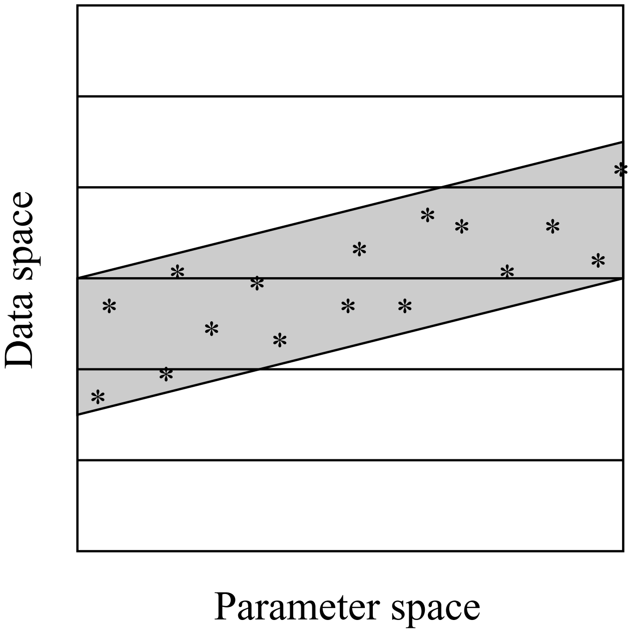

Figure 5. Parameter-data space geometry; this figure shows the discretization of the data space (horizontal lines). Samples are drawn only in the grey region where $\sigma(\mathbf{d}, \mathbf{m})$ is non-zero. For each sample, $\sigma(d, m)$ is calculated using eq. (4).

$f(m)$ for any m is Gaussian, but samples are drawn uniformly from a region $f(m) \pm 3\delta$, where $f(m)$ represents the forward function relating model parameters and data in eq. (11), see Fig. 5. δ is the standard deviation of the Gaussian uncertainty.

To calculate $\sigma(\mathbf{d})$, the data space is discretized into regular intervals (Fig. 5). Applying eq. (12) to approximate eq. (8), the marginal $\sigma(\mathbf{d})$ is approximated by,

$$\sigma(\mathbf{d}) \approx \frac{1}{N} \sum \frac{\sigma(\mathbf{d}, \mathbf{m})}{s(\mathbf{m})}, \tag{13}$$

where N now is the number of samples inside one discretization interval of the data space and $s(\mathbf{m})$ is the distribution as a function of $\mathbf{m}$ of only those samples. Because $s(\mathbf{m})$ is not always known analytically within each data interval (as $f(\mathbf{m})$ may not be analytic and may be nonlinear), a numerical approximation to $s(\mathbf{m})$ is made by binning all the locations of the samples in the model parameter space for each data interval in data space, and normalizing the histogram to have unit volume. The obtained histogram is used as an approximation to $s(\mathbf{m})$ where for each individual sample a linear interpolation method gives the final value of $s(\mathbf{m})$. The procedure is illustrated for one dimension in Fig. 6.

Finally, using the result for $\sigma(\mathbf{d})$ from eq. (13), the complete quality measure for experimental design is given by

$$Ent(\mathbf{d} \mid \boldsymbol{\xi}) \approx \frac{1}{M} \sum \frac{\sigma(\mathbf{d})}{s(\mathbf{d})}, \tag{14}$$

where $s(\mathbf{d})$ is a uniform distribution in the data space, as $\sigma(\mathbf{d})$ has been approximated using a regular discretization, Fig. 5. M is the total number of intervals in the data space and hence the total number of values of $\sigma(\mathbf{d})$ available for the calculation of the entropy. M is directly related to the size dx of the discretization interval in the data space. This means that the total number of samples T will approximate NM.

3 THE ENTROPY OF A SAWTOOTH FUNCTION

In the introduction, an example with sawtooth functions was used to show that classical nonlinear design methods fail in multi valued problems. The entropy method should be capable of dealing with any problem, including those that are strongly nonlinear. This is demonstrated by calculating $Ent(d \mid \xi)$ for sawtooth forward functions with periods 1, 2, 5 and 10 similar to those in Fig. 3. All sawtooth functions have an amplitude of 2.5 and for each model parameter value m, the distribution $\theta(d \mid m)$ is given by a Gaussian shaped uncertainty in the data space with standard deviation 0.1 around $f(m)$ where $f(m)$ is the sawtooth forward function and the Gaussian is truncated at three standard deviations from the mean. The model parameter space runs from 0 to 10 and $\rho(m)$ is uniform and equal to 1/10.

The analytical solutions for the entropy $Ent(d \mid \xi)$ for each of these sawtooth functions are identical, and are equal to 1.645. This shows that the entropy method correctly evaluates the designs. It does not necessarily favour designs with steeper gradients in contrast to the classical gradient based methodology which does.

In practice, entropies are calculated numerically. There are three sources of errors in the numerical calculations, assuming perfectly known physics:

(1) the discretization interval size in the data space dx, Fig. 5,

(2) the discretization interval size in the model parameter space dm, Fig. 6,

(3) the total number of samples T.

In the following sections, these effects are investigated individually.

3.1 Discretization of the data space

As a sawtooth function is essentially a sequence of linear functions, $\sigma(d)$ can be calculated analytically and is equal to

$$\sigma(d) = 0.1(-erf(5\sqrt{2}(-2.5 + d)) + erf(5\sqrt{2}(2.5 + d))) \tag{15}$$

where *erf* is the errorfunction. $\sigma(d)$ is evaluated at different values corresponding to midpoints of the discretization intervals, renormalized, and the corresponding entropy $Ent(d \mid \xi)$ is calculated numerically from those analytical values as in eq. (14). Thus, the dependencies on the number of samples N and the discretization interval size of the model parameter space dm are removed from the problem and it can be seen easily which discretization interval size in the data space dx is needed for a good approximation of the entropy.

In Fig. 7 (left) the entropy is plotted as a function of the discretization interval in the data space dx. Fig. 7 (right) shows the sensitivity of the entropy to uncertainties in the value of the distribution $\sigma(d)$ (since $\sigma(d)$ is approximated numerically). Clearly, even large uncertainties do not have a significant influence on the calculation of the entropy. Also, the effects are not particularly sensitive to the size of the discretization interval below $dx = 0.25$. For discretizations larger than $dx = 0.3$ it is not expected that any number of samples T is sufficient, since even with an analytically known $\sigma(d)$ the entropy can not be evaluated correctly.

3.2 Discretization of the model parameter space

The influence of the number of samples N and the discretization interval dm was checked. Assuming that the sampling distribution

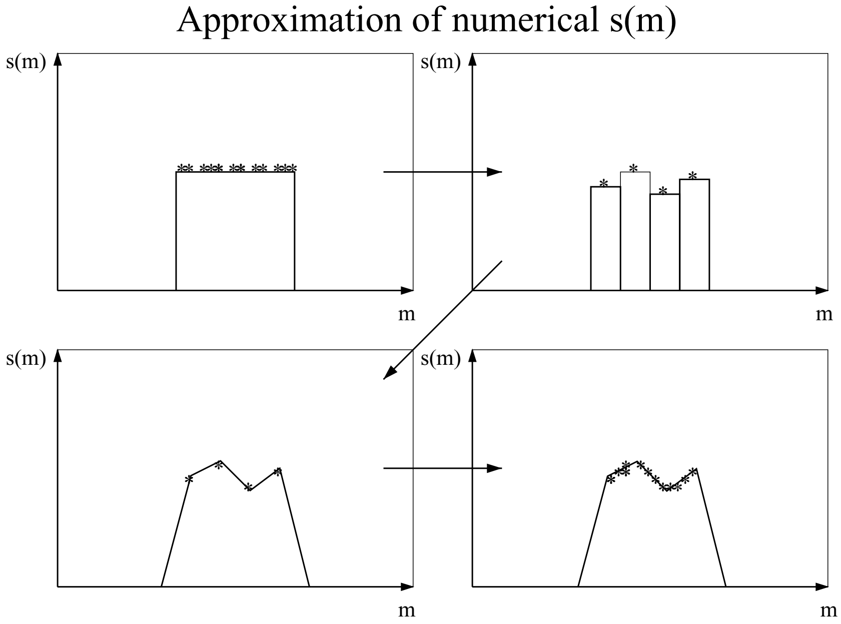

Figure 6. The sequential steps in estimating the sampling distribution s(m). First the samples are drawn from the analytical sampling distribution (in this figure a boxcar function (top left)). Then a histogram is made in which the locations of the samples in the model parameter space are binned (top right). Then this histogram is normalized and with a linear interpolation (bottom left), the value of the numerical s(m) for each individual sample is deduced (bottom right).

s(*m*) will be close to a Gaussian or a boxcar depending on whether the prior distributions are Gaussian or uniform, the tests were performed for both a boxcar and a Gaussian distribution *s*(*m*), where the Gaussian was truncated at three standard deviations from the mean. Samples were drawn from these analytical functions. Then, the numerical approximation for *s*(*m*) was estimated as illustrated in Fig. 6.

Fig. 8 shows the average percentage difference between analytical (*ana*) and numerically estimated (*num*) distributions for *s*(*m*), defined by

$$\epsilon = \left| \frac{num - ana}{ana} \right| * 100, \qquad (16)$$

over the entire box or truncated Gaussian as a function of the discretization interval *dm* and the number of samples *N*. Instead of the actual size of the discretization interval, the quantity $(m_{\max} - m_{\min})/dm$ is used. So, if $(m_{\max} - m_{\min})/dm$ equals 5, there are five intervals within the boundaries of the box or Gaussian. For highly nonlinear problems note that the samples may occupy multiple disconnected regions in the model parameter space. $(m_{\max} - m_{\min})$ should then be replaced by local values $(m_{\max-\mathrm{local}} - m_{\min-\mathrm{local}})$.

When calculating the entropy for a particular design, the sampling distribution *s*(*m*) is estimated for each discretized data-interval. Hence, the number of samples *N* considered in Fig. 8 is not the same as the total number of samples *T* required to calculate the entropy. It is the number of samples *N* required for the calculation of $\sigma(d)$ using eq. (13) for each specific value of *d*.

Fig. 8 shows two important results. First, both figures look very much alike. However, in the case of a truncated Gaussian the errors are larger. This is probably due to the way the mean percentages were calculated. Eq. (16) shows that small values for the analytical *s*(*m*) (low probability regions in model parameter space) can cause larger values for ϵ due to sporadic errors in the numerical histogram itself. Since in the case of a uniform distribution all analytical values

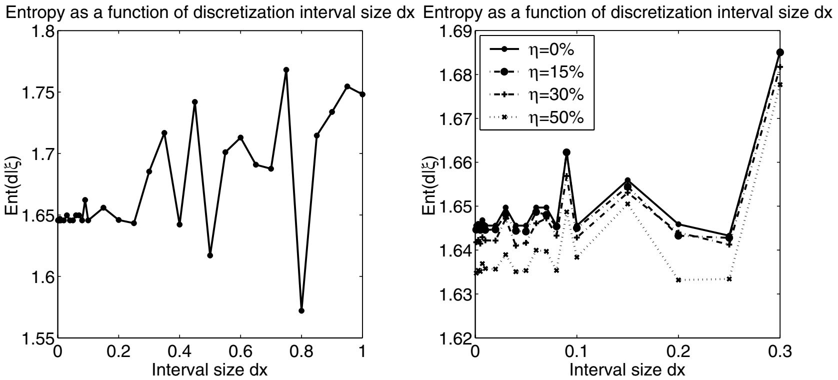

Figure 7. Left: convergence towards the analytical value for *Ent*(*d* | ξ) as a function of the data space discretization interval size *dx*. The analytical value of the entropy is 1.645. Samples from $\sigma(d)$ are taken at the center of each discretization interval and these samples are renormalized to obtain the numerical approximation for $\sigma(d)$. Right: convergence towards the analytical *Ent*(*d* | ξ) as a function of *dx* in the case where a random perturbation is added to the analytical value of $\sigma(d)$. η represents the maximum range of uncertainty and the actual uncertainty is a random number between 0 and η.

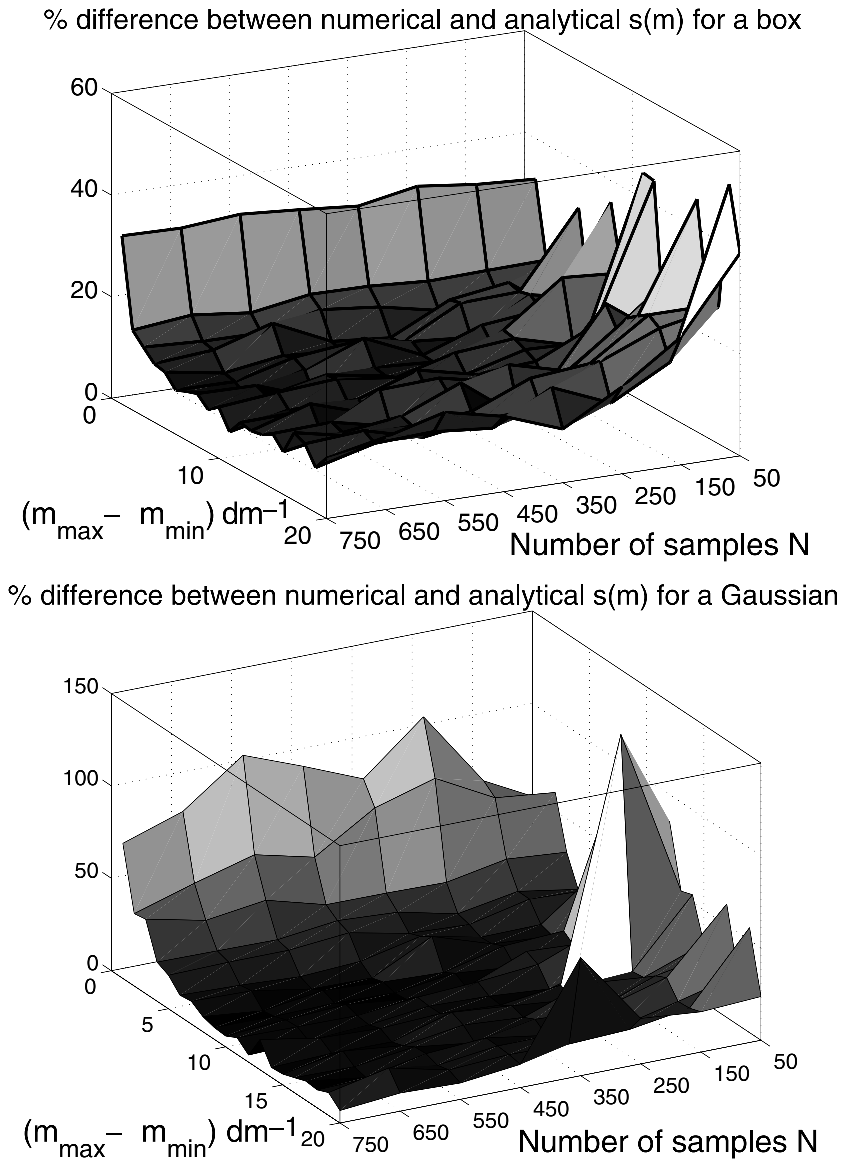

Figure 8. The percentage difference between analytical and numerical $s(m)$ as a function of discretization and number of samples N, when the analytical $s(m)$ is a boxcar function (top figure) or a truncated Gaussian (bottom figure). The samples are drawn according to the analytical $s(m)$, then counted into a histogram. This histogram is normalized and assumed to be an approximation for $s(m)$ as shown in Fig. 6. Then, the numerical value for $s(m)$ for each individual sample is deduced by linearly interpolating the histogram. The percentage difference shown here is the average percentage difference over the entire box or truncated Gaussian for one single run of the algorithm.

are equal, this tendency to emphasize areas with low probability has no effect.

Second, Fig. 8 shows that it is more important to have sufficient samples inside an interval than to have a large number of intervals. The errors due to undersampling can become very large, while errors due to a coarse discretization are relatively limited even if only 1 or 2 intervals are used.

The accurate estimation of $s(m)$ is important for the calculation of the entropy, but since this is calculated as an integration, it is expected that up to a certain limit errors in $s(m)$ may be cancelled if the errors are random.

3.3 The number of samples *T*

We now fix a discretization and estimate the required total number of samples T. The discretization in the data space is set to $dx = 0.01$; Fig. 7 shows that deviations from the analytical value are then negligible. Taking into account the results in Fig. 8, the discretization of the model space is fixed to $(m_{\max} - m_{\min})/dm = 5$, for each discretization interval in the data space. From hereafter a sawtooth function with a period of 10 is used. Fig. 9 (top) shows the percentage difference between the analytical and the numerical values for entropy as a function of the number of samples T. The solid line is the average percentage difference over 50 runs, the dashed lines are the minimum and maximum percentage differences over those 50 runs. Some artefacts are clearly visible. For very low numbers of samples T the entropy converges rapidly towards the analytical value. But after approximately 500 samples the curve starts to diverge, and beyond roughly 3000 samples the entropy converges properly towards the analytical value. Also, the error around the average percentage difference increases after 3000 samples before decreasing again after approximately 4000 samples. Since the data

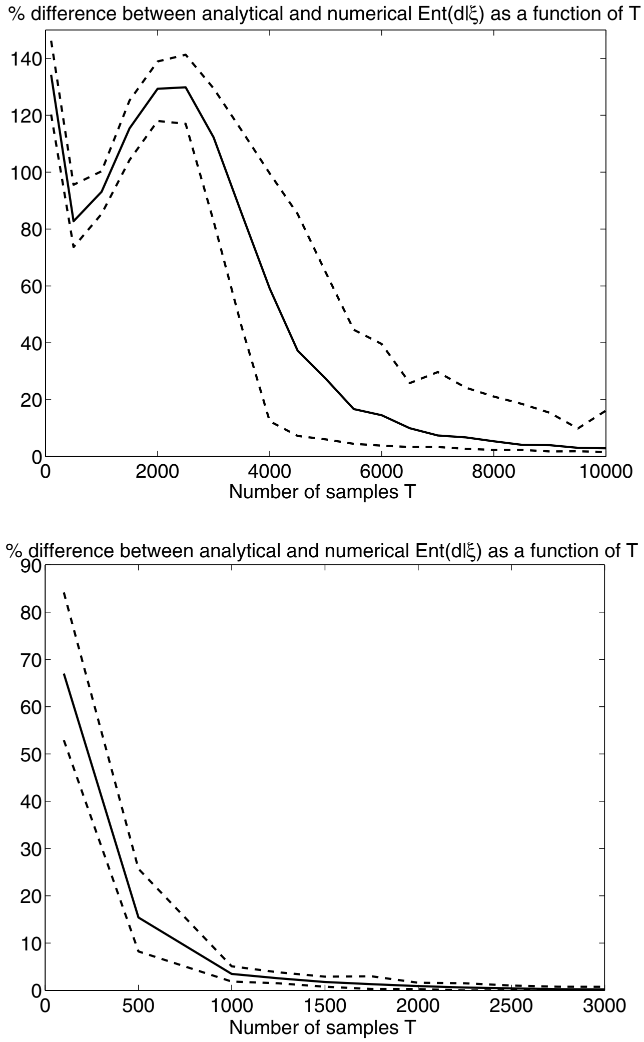

Figure 9. Percentage difference between the analytical and numerical solution for $Ent(d \mid \xi)$ for the sawtooth example with period 10, as a function of the number of samples T. A Gaussian uncertainty with standard deviation 0.1 was assumed as data uncertainty around the sawtooth forward function $f(m)$. The solid line is the average percentage difference over 50 runs, the dashed lines are the minimum and maximum percentage differences over 50 runs. The analytical solution is 1.645. The entropies were calculated with $dx = 0.01$ in the data space and $dm = 0.002$ in the model parameter space (top), and $dx = 0.05$ and $dm = 0.05$ (bottom), equivalent to $(m_{\max} - m_{\min})/dm = 1$.

space discretization interval $dx = 0.01$ and the data are running from approximately -2.9 to 2.9 (which means roughly 550–600 discrete intervals), it is likely that the former artefacts are due to undersampling. Using 600 samples T, for instance, means that for the calculation of each value for $\sigma(d)$ only 1 or 2 samples N are available, which is clearly too few (Fig. 8). The local minimum lies roughly at the point where each interval on average has 1 sample N. The local maximum is roughly at the point where each interval has on average 4–5 samples N.

The test was therefore repeated with $dx = 0.05$ and $(m_{\max} - m_{\min})/dm = 1$. Figs 7 (left plot) and 8 show that in general this will degrade accuracy, but Fig. 7 (right plot) shows that large uncertainties in the histograms only have a minor effect on the calculation of the integral. The advantage of more coarse discretizations is that fewer samples are required.

Fig. 9 (bottom) shows the results for a sawtooth with period 10. The figure shows that with this discretization, only about 1000 samples T are required to get within 5 per cent of the analytical solution.

Fig. 10 shows that with certain discretizations only 200 samples are required for the sawtooth to approximate the entropy to within 5 per cent of the analytical value – far fewer than the number of samples suggested by Fig. 9. This shows the importance of choosing a sensible discretization.

3.4 Results

We calculated the entropy of $\sigma(d)$ with $dx = 0.05$ in the data space, $dm = 0.005$ in the model parameter space and 5000 samples for 4 sawtooth functions with periods 1,2,5 and 10. Since these sawtooth functions have different gradients, this means that $(m_{\max} - m_{\min})/dm$ varied for each sawtooth. However, the discretization dm and the number of samples T should be sufficient to estimate the value of $Ent(d \mid \xi)$ correctly for each sawtooth function. Fig. 11 shows that indeed all sawtooth functions have the same entropy and are very close to the analytical value of 1.645.

We can thus use the entropy criterion in a fully numerical implementation to obtain correct results in situations where the standard Bayesian design theory fails. The most important factor in this calculation is the choice of discretization, since this strongly influences the number of samples, and hence the calculation time required.

4 AN AVO EXAMPLE

4.1 Theory and geometry

We present a detailed example using amplitude versus offset (AVO) data. Using the amplitudes of waves generated by a surface source reflected at a specific depth d and recorded at the surface again (Fig. 12), it is possible to estimate the velocity $\alpha 2$ of the layer below the reflector, for given assumptions on the other model parameters. The design problem is to choose the optimal source-receiver distance which is expected to give the most accurate post-inversion estimate for $\alpha 2$.

Aki & Richards (1980) approximate the reflection coefficient for P waves at a single interface by

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

where α is the average P-wave velocity and equal to $(\alpha 1 + \alpha 2)/2$, β is the average S-wave velocity and equal to $(\beta 1 + \beta 2)/2$, ρ is the average density and equal to $(\rho 1 + \rho 2)/2$, i equals $(i_1 + i_2)/2$, $\Delta\alpha$ means $(\alpha 2 - \alpha 1)$, and $\Delta\beta$ and $\Delta\rho$ are defined similarly. p is the slowness given by Snell's law:

$$p = \frac{\sin i_1}{\alpha 1} = \frac{\sin i_2}{\alpha 2}. \qquad (18)$$

Rewriting eq. (17) yields (Yilmaz 2001),

$$R_P = \left[\frac{1}{2}(1 + \tan^2 i)\right]\frac{\Delta\alpha}{\alpha} - \left[4\frac{\beta^2}{\alpha^2}\sin^2 i\right]\frac{\Delta\beta}{\beta} + \left[\frac{1}{2}\left(1 - 4\frac{\beta^2}{\alpha^2}\sin^2 i\right)\right]\frac{\Delta\rho}{\rho}. \qquad (19)$$

If we assume $\beta = c\alpha$ with some constant c ($c = 1/\sqrt{3}$ for a Poisson medium) and $\Delta\rho = 0$, this equation simplifies to

$$R_P = \left(\frac{1}{2}\left[1 + \tan^2 i\right] - 4c^2 \sin^2 i\right)\frac{\Delta\alpha}{\alpha}. \qquad (20)$$

This is a nonlinear equation as a function of one model parameter $\alpha 2$, given $\alpha 1$, c and i. Instead of using angle i, one usually works

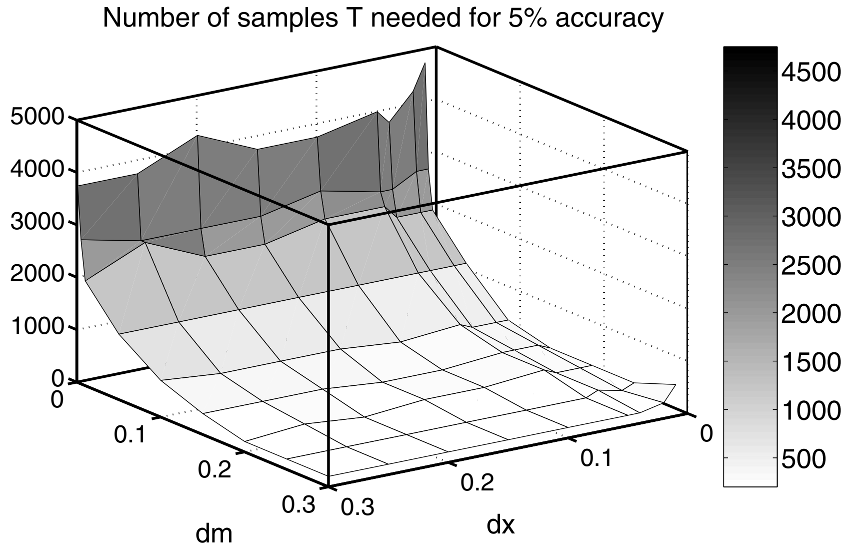

Figure 10. Number of samples T required to reduce the maximum percentage difference between the numerical and analytical value for the entropy over 50 runs to within 5 per cent of the analytical value for the entropy (1.645). This for a sawtooth with period 10 with a Gaussian uncertainty with standard deviation 0.1 in the data space. For discretizations dx larger than 0.3 it is not expected that any number of samples is sufficient, Fig. 7. The corresponding maximum dm is then also 0.3 ($(m_{\max} - m_{\min})/dm = 1$).

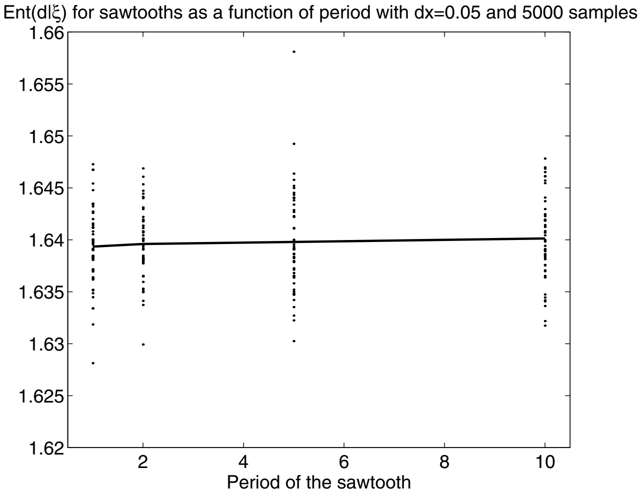

Figure 11. $Ent(d\mid\xi)$ with $dx = 0.05$ in the data space, $dm = 0.005$ in the model parameter space and 5000 samples T for 4 sawtooth functions with periods 1, 2, 5 and 10, with an amplitude of 2.5 and a Gaussian uncertainty with standard deviation 0.1. Each dot represents one run and the solid line connects the averages of 50 such points.

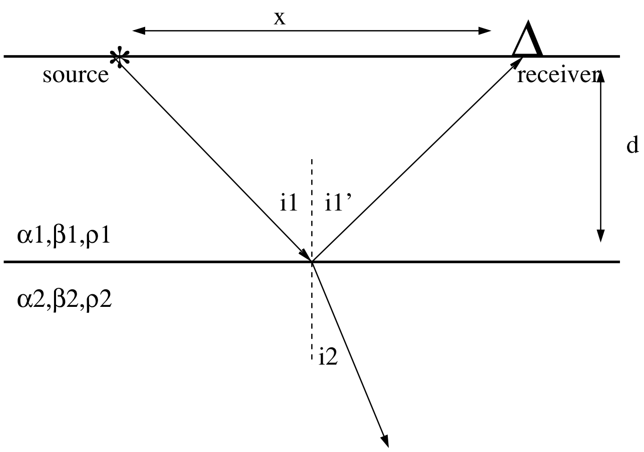

Figure 12. Geometry of a single interface with a P wave source and a P wave reflection and transmission. The distance or offset between source ($*$) and receiver (Δ) is given by x, the depth of the reflector is d, the incident angle is i_1 and assuming a horizontal reflector this is equal to the angle i_1', the transmission angle is given by i_2. The characteristics of the medium are the velocities of the P ($\alpha_{1,2}$) and S ($\beta_{1,2}$) wave velocities and the densities ($\rho_{1,2}$).

with offsets x, Fig. 12. Assuming a horizontal interface and a known depth d, the offset translates into i as

$$i_1 = \arctan\left(\frac{x}{2d}\right) \tag{21}$$

$$i_2 = \arcsin\left(\frac{\alpha 2}{\alpha 1}\sin i_1\right). \tag{22}$$

The amplitude data are given by $A_1 = A_0|R_P|$, where A_0 is the amplitude at the source and A_1 is the measured amplitude at the receiver and where we assume that the data have been adjusted to remove the effect of geometrical spreading, so that A_0 is 1, and $A_1 = |R_P|$. The amplitude-data can be calculated given specific values for $\alpha 1$, $\alpha 2$, the constant c and the depth of the reflector d.

4.2 Results

We investigated two experimental design problems, specified by

$$\alpha 1 = 2750 \text{ m s}^{-1} \tag{23}$$

$$\alpha 2 = [3200, 3300] \text{ m s}^{-1} \text{ (first)}, \quad \alpha 2 = [3000, 4500] \text{ m s}^{-1} \text{ (second)} \tag{24}$$

$$c = \frac{1}{\sqrt{3}} \tag{25}$$

$$d = 500 \text{ m}, \tag{26}$$

where $[a, b]$ m s^{-1} means that the prior information on $\alpha 2$ is given by a uniform distribution between $\alpha 2 = a$ m s^{-1} and $\alpha 2 = b$ m s^{-1},

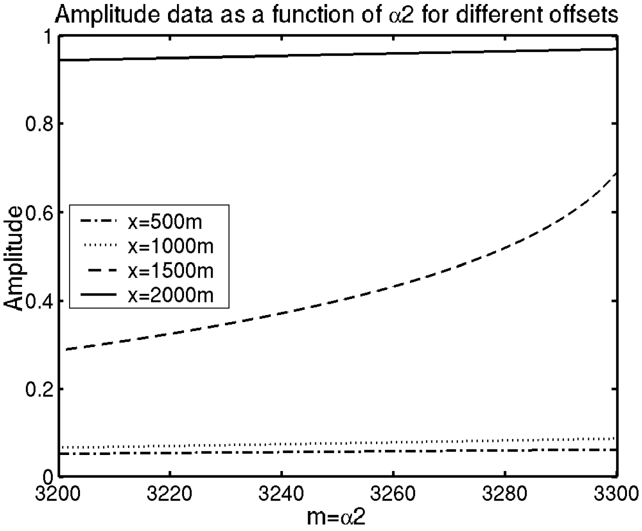

Figure 13. Amplitude data as a function of the model parameter $\alpha 2$ for 4 different offsets, $x = 500$ m, $x = 1000$ m, $x = 1500$ m and $x = 2000$ m. The velocity of the top layer is 2750 m s^{-1} and the depth of the reflector is 500 m. The functions are calculated using $dx = 0.001$, $dm = 100dx$ and 100 000 samples T.

and $\theta(d, m) = U[f(m), \delta]$, where $f(m)$ is the absolute value of R_P, eq. (20), and $\delta = 0.01$ represents the standard deviation of a Gaussian expected measurement uncertainty.

Fig. 13 shows the forward function $f(m)$, as a function of the model parameter $\alpha 2$, for the offsets $x = 500$ m, $x = 1000$ m, $x = 1500$ m and $x = 2000$ m. The maximum gradient is of the order of 0.01. The results of the previous section indicate that for the calculation of $\sigma(d)$ in each data-interval, $(m_{\max} - m_{\min})/dm = 1$ does not cause large errors. After checking this was also true for the AVO example, all tests in this section were performed using $dm = 100dx$, where dm is the discretization interval size in model parameter space and dx is the discretization interval size in the data space.

It is impossible to compare the results of the tests to an analytical solution, as this is a nonlinear example and the required integrals cannot be calculated analytically. Hence, the experimental design problem was calculated for different discretizations and several numbers of samples. Fig. 14 shows the entropy as a function of offset for $dx = 0.2$ and 500 samples, $dx = 0.1$ and 500 samples, $dx = 0.01$ and 1000 samples, and $dx = 0.001$ and 100 000 samples. Two things are clear from this figure. First, comparing the curves for $dx = 0.01$ and $dx = 0.001$ we see that the entropy shows no significant changes at small discretization interval sizes. Second, while it is clear that for large discretization interval sizes, the values for the entropy do not converge to the correct value, this does not change the results of the experimental design problem: the overall shape of the entropy-curves remains roughly the same and the desired maximum entropy design always occurs at the same offset. This is an indication that it is safe to use a coarser discretization since the errors due to undersampling are much larger and less predictable than the errors due to too coarse a discretization.

The optimal offset is found to be at approximately 1500 m from the source. Fig. 13 shows that this optimal offset is found where the data are most sensitive to changes in the model parameters. This suggests that we could have used a gradient based method. This is probably due to the fact that this was not a strongly nonlinear example (Fig. 13). The design problem is therefore repeated with a different $\rho(m)$. In the first example $\rho(m)$ was uniform between $\alpha 2 =$ 3200s m s^{-1} and $\alpha 2 = 3300$ m s^{-1}. For the second example, $\rho(m)$ is set to be uniform between $\alpha 2 = 3000$ m s^{-1} and $\alpha 2 = 4500$ m s^{-1}, a much larger interval. Fig. 15 shows the relationship between data and model parameters for four different offsets, $x = 500$ m, $x = 1000$ m, $x = 1500$ m and $x = 2000$ m. Clearly, this problem is more strongly nonlinear. The discontinuity in the curves is at the velocity where the angle of incidence reaches the value for the critical angle of incidence.

The resulting values for $Ent(d \mid \xi)$ are shown in Fig. 16. Again, the maximum value is at approximately 1500 m, but this is no longer

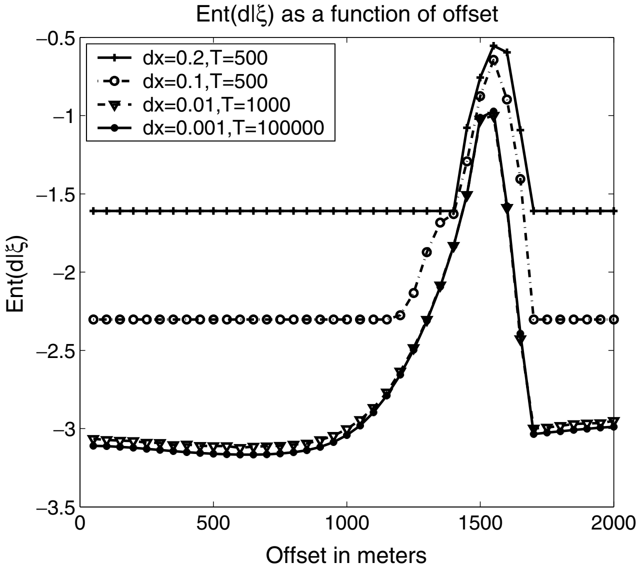

Figure 14. $Ent(d \mid \xi)$ as a function of the offset in metres for the AVO example with a Gaussian uncertainty in the data space with a standard deviation of 0.01. Velocity of the top layer is 2750 m s^{-1}, the depth of the reflector is 500 m and the model parameter space runs from 3200 m s^{-1} to 3300 m s^{-1}. The entropy is calculated using four different discretizations and numbers of samples T. In all cases $dm = 100dx$ has been used.

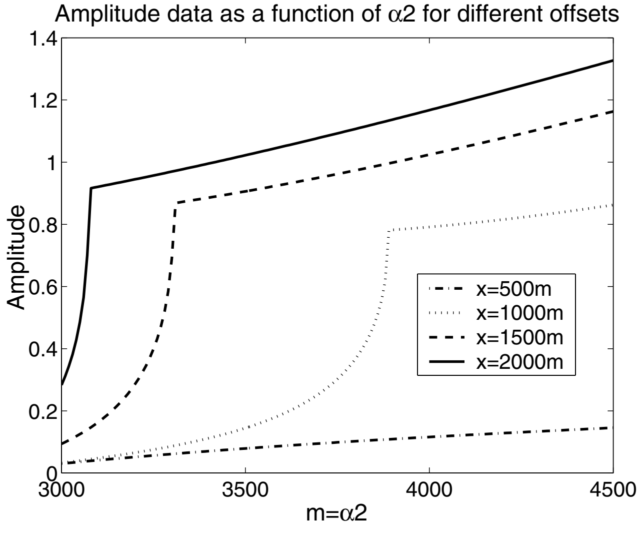

Figure 15. Amplitude data as a function of the model parameter $\alpha 2$ for 4 different offsets, $x = 500$ m, $x = 1000$ m, $x = 1500$ m and $x = 2000$ m. The velocity of the top layer is 2750 m s^{-1} and the depth of the reflector is 500 m. The functions are calculated using $dx = 0.001$, $dm = 100dx$ and 100 000 samples T.

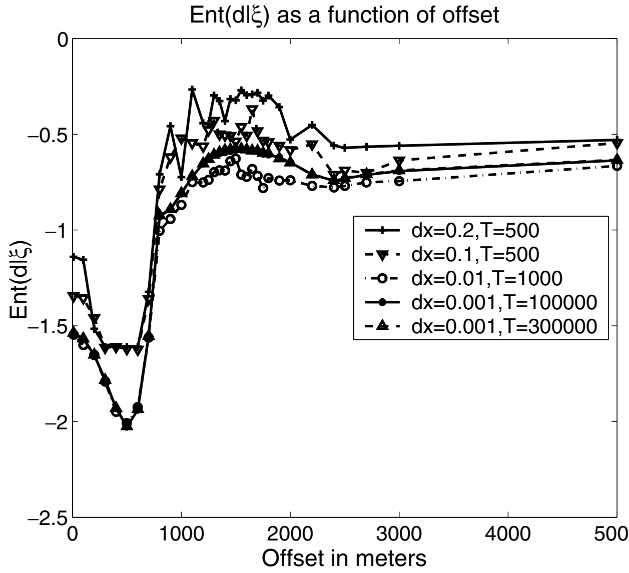

Figure 16. *Ent*$(d \mid \xi)$ as a function of the offset in metres for the AVO example with a Gaussian uncertainty in the data space with a standard deviation of 0.01. Velocity of the top layer is 2750 m s^{-1}, the depth of the reflector is 500 m and the model parameter space runs from 3000 m s^{-1} to 4500 m s^{-1}. The entropy is calculated using five different discretizations and numbers of samples T. In all cases $dm = 100dx$ has been used.

obvious from the gradients in Fig. 15. Otherwise, the results are comparable with the previous example; selection of the maximum entropy design is unaffected by the range of discretizations considered here.

5 CONCLUSIONS

An entropy-based method for nonlinear experimental design has been presented. In principle, this method is applicable to all experimental design problems, but in particular those nonlinear problems where classical nonlinear methods for experimental design fail.

The main difficulty in applying this method lies in choosing a sensible discretization for model parameter and data spaces without knowing the degree of nonlinearity of the problem in advance. This choice strongly influences the number of samples required. Our results suggest that it may be better to choose a coarser discretization to obtain final designs, since the errors associated with undersampling are larger and less predictable than the errors associated with too coarse a discretization interval size. The AVO examples above suggest that the final experimental design remains unaffected by a slightly too coarse discretization. However, further research is necessary to determine to what extent this rule of thumb remains true. For the synthetic sawtooth example, this limit is visible in Fig. 7 (left), since no number of samples is sufficient to estimate the entropy for discretizations dx larger than 0.3.

For the examples shown in this report, a simple search with uniform steps throughout the design-space was used. For larger, high-dimensional problems, it is strongly recommended to use an efficient search-algorithm in order to keep the computational cost as low as possible. The required computational cost appears to be the only practical limitation to the application of this method.

For the practical application of this theory it is important to realize that the examples as discussed in this paper are with one model parameter and one receiver. More receivers automatically increases the dimensionality of the problem. Further research is necessary to apply this theory to realistic experiments with more model parameters and more receivers.

REFERENCES

Aki, K. & Richards, P.G., 1980. *Quantitave Seismology: Theory and Methods,* San Francisco: Freeman.

Atkinson, A.C. & Donev, A.N., 1992. *Optimum Experimental Design,* Oxford Science Publications.

Box, G.E.P. & Lucas, H.L., 1959. Design of Experiments in Non-linear Situations, *Biometrika,* **46,** 77–90.

Chaloner, K. & Verdinelli, I., 1995. Bayesian Experimental Design: A Review, *Statistical Science,* **10,** 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. & Maurer, H., 2000. Optimizing designs of geophysical experiments and surveys: Is it worthwhile?, *EOS, Trans. Am. geophys. Un.,* **81,** 20, 224–225.

Curtis, A. & Spencer, C., 1999. Survey Design Strategies for linearized nonlinear inversion, *69th Annual Int. Meeting, Soc. of Expl. Geophys., Expanded Abstracts,* pp. 1775–1778.

Ford, I., Titterington, D.M. & Kitsos, C.P., 1989. Recent Advances in Nonlinear Experimental Design, *Technometrics,* **31,** 49–60.

Johnson, N.L. & Leone, F.C., 1977. *Statistics and Experimental Design in Engineering and the Physical Sciences,* Wiley series in probabability and mathematical statistics. Wiley, New York.

Lepage, G.P., 1978. A New Algorithm for Adaptive Multidimensional Integration, *Journal of Computational Physics,* **27,** 192–203.

Sebastiani, P. & Wynn, H.P., 2000. Maximum Entropy Sampling and Optimal Bayesian Experimental Design, *J. R. Statist. Soc. B,* **62,** 145–157.

Shannon, C.E., 1948. A Mathematical Theory of Communication, *Bell System Tech. J.,* **27,** 379–423.

Shewry, M.C. & Wynn, H.P., 1987. Maximum Entropy Sampling, *Journal of Applied Statistics,* **14,** 165–170.

Squires, G.L., 1985. *Practical Physics,* 3rd edn, Cambridge Univ. Press, Cambridge.

Tarantola, A., 1987. *Inverse Problem Theory: Methods for Data Fitting and Parameter Estimation,* Elsevier, Amsterdam.

Tarantola, A. & Valette, B., 1982. Inverse Problems = Quest for Information, *J. Geophys.,* **50,** 159–170.

Yilmaz, O., 2001. *Seismic Data Analysis: Processing, Inversion and Interpretation of Seismic Data,* Society of Exploration Geophysics.