Survey design strategies for linearized nonlinear inversion

Andrew Curtis and Carl Spencer, Schlumberger

SUMMARY

Standard nonlinear survey or experimental design criteria generally contain an underlying assumption of local linearity and global unimodality of the data misfit function. This is a severe limitation since many nonlinear problems are multimodal. We present more fundamental quality criteria that must be fulfilled before measures of the type described above are meaningful. This allows us to define a more robust, nonlinear design strategy.

INTRODUCTION

Many factors affect the design of a survey or experiment (henceforth, survey): physical or logistic constraints usually define the types of designs that are possible in practice, minimising cost is often paramount and this usually trades off with maximising the amount of information which we expect to obtain from the survey. This work shows how survey designs which maximize the expected model information can be constructed in cases when the relationship between the survey data and model to be constrained may be nonlinear, but where the model information will be obtained through iterated linearized inversion of the recorded data.

Typically, surveys are designed such that the predicted model uncertainties post-survey are as small as possible, that is, most possible information is obtained. Inevitably then, survey design requires an understanding of how data and model uncertainties are related. Figure 1 (a-d) illustrates this for 1-dimensional problems with varying degrees of nonlinearity. In each case the top plot shows the model–data relationship (the forward function G), the lower plot is a data misfit function E (the misfit between data predicted by G and some measured data d^t). Given data d^t, the aim of inversion is to find all models that are consistent with d^t to within the data noise, i.e., which have sufficiently low misfit.

The shaded areas in (a) and (b) show how data uncertainties of $\pm\sigma$ around d^t project into model uncertainties around the true model m^t when the forward problem is linear. The steeper the forward function the smaller will be the predicted model uncertainty for any measured data. Crucially, changing the survey geometry changes the forward and hence misfit functions. Hence, for linear or linearized problems, optimal survey design techniques usually adjust the design to create the forward function that has the largest absolute gradient (1-dimensional case) or largest absolute measure of the derivatives of data with respect to model parameters (multi-dimensional case, e.g., Box & Lucas 1959, Kijko 1977, Barth & Wunsch 1990, Maurer & Boerner 1998a, Curtis 1998, Curtis 1999). The most impor-
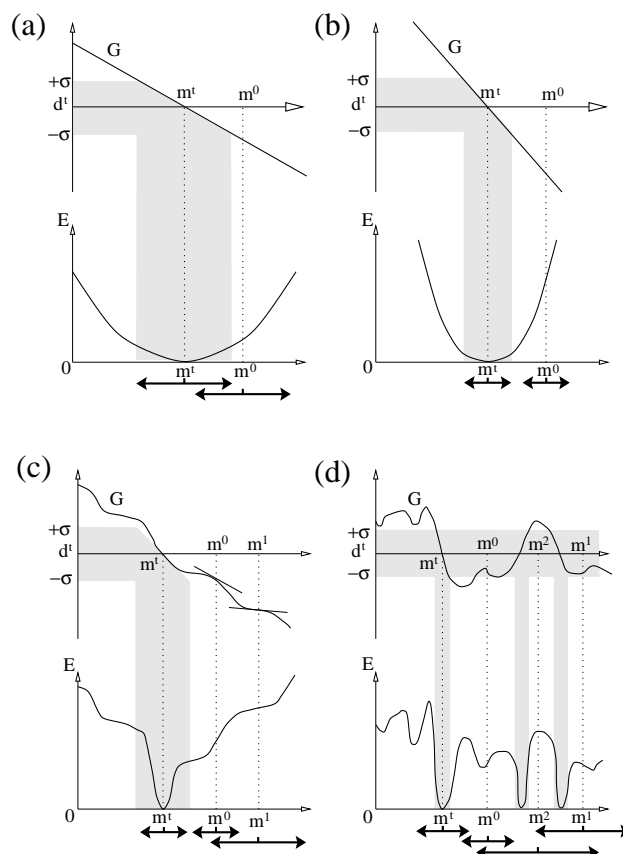

Figure 1: Forward function G and misfit function E for four different experiment designs: (a) and (b) linear, (c) pseudolinear and (d) strongly nonlinear. Horizontal axis represents the model space, vertical axis represents the data space (labels omitted for clarity). $\mathbf{m}^t$ denotes the true model corresponding to measured datum $\mathbf{d}^t$, $\mathbf{m}^0, \ldots, \mathbf{m}^2$ represent possible best model estimates prior to inversion. Shaded region in (a) and (b) shows how $\pm\sigma$ uncertainties on the measured datum translate into uncertainties in the model domain — represented schematically by double-headed arrows beneath each model estimate. In case (c), the projection is inaccurate because the linearization of G at m^t (the diagonal boundary of the shaded region) deviates from the true function — the *local linearization error*. In addition, in case (d) the linearized uncertainty only approximates the left-hand shaded region; there is no provision for finding the two other regions of good-fit models — the *multimode error*.

tant point about linear problems is that it does not matter at which model we evaluate the gradient $\partial G/\partial m$. Hence, if our best estimate of the true model m^t prior to the survey was m^0, we may still evaluate $\partial G/\partial m|_{m_0}$ to estimate the post-survey model uncertainties (double-headed arrows).

In nonlinear situations (c) and (d), this is not the case. In the pseudo-linear case (c), if we evaluated the derivative at m^t we would predict that the post-survey uncertainties will be fairly small. However, m^t is not known prior to the survey; if our best prior model estimate was m^0 or m^1 for example, $\partial G/\partial m$ may vary greatly (tangents drawn on plot) and we might obtain very different estimates for the post-survey uncertainty. Minimizing the latter predicted uncertainties could lead to poor survey designs. This situation is even more extreme in the strongly nonlinear case (d).

Classical Bayesian nonlinear experiment design theory tackles this problem by calculating the *expected* model uncertainties given all possible prior models (e.g., Ford et al., 1989, Atkinson & Donev 1992, Maurer and Boerner 1998b). Hence for example, if prior information constrains the model to lie with a uniform distribution on the section of the horizontal axis shown, the *average* predicted (logarithm of) model uncertainty across that set of models is calculated and minimized. Typically this is done by calculating the gradient of G at a set of model samples distributed according to the prior model probability distribution and averaging a measure of the associated predicted linearized model uncertainties.

PROBLEMS

Notice that standard Bayesian methods assume that for each model sample the corresponding linearized (gradient-based) predicted uncertainty is correct. Figure 1(c) shows the linearized uncertainty estimate for measured data $d^t \pm \sigma$ in a weakly nonlinear case. The gradient, evaluated at the true model m^t in this case, is used to estimate where function G will intercept the $d^t \pm \sigma$ levels. In this particular case the estimated uncertainty is too large because the gradient at m^t is less than the gradient slightly below m^t (causing the shaded region to deviate from the true function G). In other cases the estimate may be too small. This is the *local linearization error*.

Case (d) illustrates an additional problem. More than one model explains the measured data perfectly (three models have zero misfit). Since these models are separated by poorly fitting models, the misfit function is *multimodal*. Given the measured data $d^t \pm \sigma$, evaluating the gradient of G at m^t provides the linearized uncertainty estimate spanned by the shaded region around that model. However, the true model uncertainty is the union of all three shaded regions since given the data there is no way to distinguish between any of the models contained therein. Linearized uncertainty estimates do not even approximate multimodal uncertainty; this is the *multimode error*.

When surveys are designed for subsequent nonlinear inversion the multimode error is potentially extremely serious. For computational efficiency, linearized inverse theory is commonly used to solve high-dimensional inverse problems. Such methods require an initial model estimate from which successive steps are taken in model space that move in the locally most 'downhill' direction on the misfit function. The process stops at the model with minimum misfit that can be found with only downhill steps. Clearly, if we use different initial models in case (d) we will end up at different local minimum-misfit solutions, which may (initial model m^2) or may not (initial model m^0) give a good fit to the measured data. Using such methods there is no indication of whether a better solution exists; hence multimodal misfit functions may result in completely wrong model estimates.

Linearized methods are sometimes also used to provide computationally efficient linearized uncertainty estimates around the final solution found. However, as illustrated above, both the local linearization and multimode errors may cause these estimates to be very poor. Again, however, these methods provide no indication that this is the case, and the linearized uncertainty estimates are often accepted.

Since previous Bayesian methods do not detect either the local linearization of multimode errors where these errors may greatly affect the value of models infered from the data, such methods are clearly inadequate for nonlinear design and may lead to poorly designed surveys. New criteria to assess quality that are sensitive to such errors are required.

SOLUTIONS

The choice of survey design directly influences the form of the forward function G, and hence the possible misfit functions. If (goal 1) a survey could be designed such that the misfit function was unimodal there would be no multimodal error. If (goal 2) the survey could be designed such that the forward problem was as linear as possible, the local linearization error would be minimized. Only after these two fundamental survey design goals have been achieved (especially goal 1) should the Bayesian techniques described above be used. If left unfulfilled, assumptions on which the Bayesian uncertainty measures are based are invalid and post-survey linearized inversion techniques may give incorrect results.

Two measures of survey design quality that assess the degree to which goals (1) and (2) are achieved are Φ_1 and Φ_2 respectively:

Φ_1: Number of minima in the possible misfit functions

Φ_2: $\frac{1}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} \mathrm{var}\left(\frac{\partial G_i}{\partial m_j}\right)$
where the data space and hence the forward function $\mathbf{G}$ have dimension N, model vector $\mathbf{m}$ has dimension M, and variance is calculated over the prior model dis-

Nonlinear survey design

Φ_3: A standard Bayesian quality measure

Measure Φ_3 is any standard Bayesian measure as described earlier. Varying the survey design such that Φ_1 is minimized clearly reduces the degree of multimodality of the misfit function and hence the multimode error (goal 1). Similarly, minimizing Φ_2 reduces the average variance of the gradient ($N = M = 1$) or derivative matrix (N or $M > 1$) over the prior distribution of models. Since variation of such derivatives is the definition of nonlinearity, minimizing Φ_2 explicitly reduces the degree of nonlinearity and hence the linearization error (goal 2). Only if these fundamental survey design criteria have been achieved does measure Φ_3 become a useful estimate of expected post-survey model uncertainty/information and should be minimized or maximized as appropriate.

Estimating quality measures Φ_1 and Φ_2
Measure Φ_2 can be estimated for any survey design by calculating derivatives $\partial G / \partial m_j$ at each of a set of models sampled from the prior model distribution. The variances and relevant sums can then be estimated.

Measure Φ_1 is more difficult to estimate since it requires knowledge of the global structure of the misfit function. Unless the function is analytic, this requires that samples of the misfit function over the prior model distribution (possibly linearly inverted to find the local best-fit model from each sample) are analyzed for signs of multimodality. An additional problem is that in principal a misfit function can only be constructed once data are available (post-survey). However, to evaluate misfits G must be evaluated for each of the model samples selected, giving a set of possible data samples that are consistent with our prior information. Each data sample defines a different possible misfit function, so that if L samples are taken, L misfit functions must be analyzed.

Usual methods to assess multimodality involve performing cluster analysis on the set of model samples that provide low misfits (e.g., Vasco et al., 1996). However, this only makes use of the misfit function values for each model. To evaluate measure Φ_2, functional derivatives must be evaluated at each model sample. These provide extra information that can also be used to estimate the number of misfit minima for Φ_1.

Consider the case where E is the n-norm (n even) of the data misfit. Then,

$$E = \left(\mathbf{d}^t - \mathbf{G(m)} \right)^n, \qquad \text{and} \qquad (1)$$

$$\frac{\partial E}{\partial m_i} = nE^{\frac{n-1}{n}} \cdot \frac{\partial \mathbf{G(m)}}{\partial m_i} \qquad (2)$$

where m_i is the ith model parameter in vector $\mathbf{m}$. Hence, misfit values $E(\mathbf{m})$ and derivatives of G with respect to each model parameter (calculated for Φ_2) are sufficient information to calculate $-\nabla E(\mathbf{m}) = [\partial E(\mathbf{m}) / \partial m_1, \ldots, \partial E(\mathbf{m}) / \partial m_M]^T$.

This is a vector that points in the direction of maximum decrease of E at model $\mathbf{m}$.

Figure 2 shows a (shaded) misfit surface with six minima (crosses) over a portion of a 2-dimensional model space over which we assume that the prior model distribution is uniform. The misfit function E and its gradient ∇E are calculated at 100 model samples (left) and 50 samples (right). In M-dimensional model spaces the gradient vector is tangent to a plane of dimension $M - 1$. For any point $\mathbf{m}^i$, define a second point $\mathbf{m}^j$ to be *downhill* of the first if and only if $\mathbf{m}^j$ lies on the side of the tangent plane at $\mathbf{m}^i$ to which the gradient vector $\nabla E(\mathbf{m}^i)$ points. The following algorithm traces paths between samples approximately downhill on the misfit function E (*descent paths*); this information is used to estimate the number of minima:

1. Select an initial sample $\mathbf{m}^i$, $i = 0$, that has never before been visited at random from the sample set.

2. Find the closest sample downhill of sample $\mathbf{m}^i$. We use the distance metric
   $$||\mathbf{m}^{i+1} - \mathbf{m}^i|| = [\mathbf{m}^{i+1} - \mathbf{m}^i] \cdot \nabla E / |\nabla E|$$
   where $| \cdot |$ is the standard Euclidean norm. Thus we preferentially select models $\mathbf{m}^j$ for which $\mathbf{m}^j - \mathbf{m}^i$ lies close to the gradient direction $\nabla E(\mathbf{m}^i)$. If it exists, denote the new model $\mathbf{m}^{i+1}$. If it does not exist this path propagates outside of the sample set without a well-defined minimum being found. Warning! — sample size may be insufficient. Go to step 5.

3. If $\mathbf{m}^{i+1}$ is already in a previous path, return to step 1.

4. If $\mathbf{m}^{i+1} = \mathbf{m}^j$ for some $j < i$ then this path contains a loop — minimum found!

5. If previously unvisited samples exist, return to step 1.

All descent paths are shown by thin lines in figure 2. All path branches either end at another path (detected in step 3), or continue downhill towards a minimum. The main new concept in this algorithm is in step 4. A path must either ultimately end in one of two ways: (1) outside the range our current sample set (detected in step 2) — in this case either the sample set is inadequate, or the minimum does not lie within our prior model distribution. (2) The path must double back on itself forming a loop through two or more samples (detected step 4). Example loops are shown in figure 2. Loops are formed when the closest downhill sample $\mathbf{m}^{i+1}$ to sample $\mathbf{m}^i$ lies across a minimum. Hence, each loop indicates a misfit minimum lies approximately within, or close to the loop. Figure 2 (left hand plot) shows that this is the case, and that if the misfit function is sampled sufficiently densely then the number of loops equals the number of minima. The algorithm seems to be fairly robust — even with only 50 samples, many of which lie far from the minima, five out of the 6 minima are found (right hand plot).
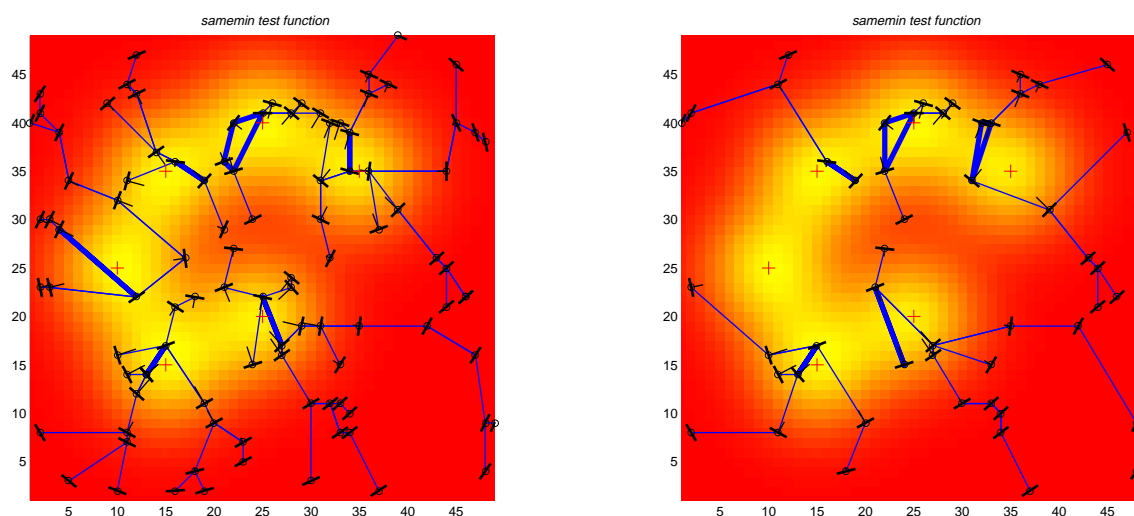
Nonlinear survey design

Figure 2: Misfit function (shading) with six minima (crosses) and descent paths (thin lines) between samples (T-shapes). Left plot has 100 samples, right plot 50 samples. At each sample both the gradient downhill direction (T stem) and its normal vector (T top) are shown. Bold sections of descent paths are loops — 6 loops in left hand plot, 5 in right hand plot.

CONCLUSIONS: A STRATEGY FOR NONLINEAR SURVEY DESIGN

In the case where the survey data is nonlinearly related to model parameters and will be used to constrain those parameters using linearized inverse methods, the survey design strategy is now simple to define.

1. Minimize Φ_1 to minimize multimode errors in subsequent survey design stages and in linearized inversion.

2. Minimize Φ_2 to minimize local linearization errors in stage 3, and to validate linearized post-survey uncertainty estimates.

3. Minimize/Maximize Φ_3, a standard Bayesian measure of survey quality as defined previously.

These stages can be performed sequentially or simultaneously using a sensible relative weighting system for Φ_1, Φ_2 and Φ_3. Stages 1 and 2 are more fundamental than stage 3 since they validate commonly used Bayesian measures Φ_3. Local linearization errors can be severe when curvature of the misfit function is large. On the other hand, if multiple modes in the misfit function exist then the local linearized approximation to the forward function may approximate only a small section of the function that provides good fit to the data. In this sense the multimode error is the most fundamental, and since it can also lead to completely wrong model estimates from linearized inversion of survey data, probably priority should usually be given to minimizing Φ_1. Hence, for potentially multimodal nonlinear problems the stages listed above are in decreasing order of importance. This is a completely new perspective on standard nonlinear survey design criteria where Bayesian measures Φ_3 are the only measures considered.

REFERENCES

Atkinson, A. C., and Donev, A. N., 1992, Optimum experimental design: Clarendon Press, Oxford.

Barth, N., and Wunsch, C., 1990, Oceanographic experiment design by simulated annealing: J. Phys. Oceanog., **20**, 1249–1263.

Box, G. E. P., and Lucas, H. L., 1959, Design of experiments in non–linear situations: Biometrika, **46**, 77–90.

Curtis, A., 1998, Optimal cross–borehole tomographic geometries: 1998 SEG Extended Abstracts, New Orleans.

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

Ford, I., Titterington, D. M., and Kitsos, C. P., 1989, Recent advances in nonlinear experimental design: Technometrics, **31**, 1, 49–60.

Kijko, A., 1977, An algorithm for the optimum distribution of a regional seismic network — I: Pageoph., **115**, 999–1009.

Maurer, H., and Boerner, D. E., 1998a, Optimized and robust experimental design: A non–linear application to EM sounding: Geophys. J. Int., **132**, 458–468.

Maurer, H., and Boerner, D. E., 1998b, Optimized design of geophysical experiments: Leading Edge, **17**, 8, 1119–1125.

Vasco, D. W., Peterson, J. E. Jr. and Majer, E. L., 1996, Nonuniqueness in traveltime tomography: Ensemble inference and cluster analysis: Geophysics, **61**, 4, 1209–1227.