Geophys. J. Int. (1999) **136,** 637–650

Optimal experiment design: cross-borehole tomographic examples

Andrew Curtis

Schlumberger Cambridge Research, High Cross, Madingley Road, Cambridge, CB3 0EL, *UK. E-mail: curtis@cambridge.scr.slb.com*
and *Utrecht University, Department of Theoretical Geophysics, Postbus* 80.021, 3508 *TA Utrecht, the Netherlands*

Accepted 1998 October 1. Received 1998 September 30; in original form 1998 June 12

SUMMARY

Experiment design optimization requires that the *quality* of any particular design can be both quantified and then maximized. In this study, experiment quality is defined to measure the constraints on a particular model offered by the anticipated experimental data (that is, it measures anticipated model information post-experiment). Physical and financial constraints define the space of possible experimental designs. The definitions used here require that the relationship between model parameters and data can be linearized without significant loss of information.

Two new measures of model information are introduced and compared to three previously known measures. One of the new measures can be calculated extremely efficiently allowing experiments constraining large model spaces to be designed. This efficiency trades off with a lack of sensitivity to poorly constrained parts of the model. Each measure is used independently to design a cross-borehole tomographic survey including surface sources and receivers (henceforth called nodes) which maximally constrains the interborehole velocity structure. The boreholes are vertical and the background velocity is assumed to be approximately constant. Features common to most or all optimal designs form robust design criteria—'rules of thumb'—which can be applied to design future experiments. These are:

(1) surface nodes significantly improve designs;
(2) node density increases steadily down the length of each well;
(3) surface node density is increased slightly around the central point between the wells;
(4) average node density on the ground surface is lower than that down each well.

Three of these criteria are shown to be intuitively reasonable (the fourth is not), but the current method is quantitative and hence may be applied in situations where intuition breaks down (for example, non-vertical wells with multilateral splays; combining different data types; inversion for anisotropic model parameters). In such cases the optimal design is usually not obvious, but can be found using the quantitative methods introduced and discussed herein.

Key words: cross-borehole tomography, experiment, information, optimal design, survey.

1 INTRODUCTION

Many geophysical experiments or surveys are expensive to perform. For this reason, planning an experiment in terms of cost, logistics and the goals which an experiment will achieve becomes of crucial importance. This study addresses the following question: given the goals of an experiment, how can the experiment be designed such that the highest number of goals are achieved within the constraints imposed by cost and logistics?

To perform such a design process it is necessary to define the set of possible experiment designs, and the information that any particular experiment design would provide. This may then be compared with the goals, and the match or mismatch quantified. In many experiments the set of possible experiment designs is constrained mainly by physical or financial constraints. The experimental aim is often to constrain values of a given model (for example, a velocity model of the Earth) using data measurements of an observed response (traveltimes of body waves travelling through the Earth) to a stimulus

(seismic source), where this response depends on the model through a relationship that is understood (traveltime is the integral of slowness along each ray path). It is often also the case that this relationship itself depends on the experiment design (moving a seismic source causes body waves to traverse a different portion of the Earth).

Assuming that the relationship between parameters and responses can be linearized without significant loss of information, linear inverse theory can be employed to analyse the quality of any experimental design. Thus, this study will be restricted to cases where the relationship is either truly linear, or where a sufficiently good prior estimate of the model parameters exists such that the non-linear problem can be linearized locally around that estimate without significant loss of accuracy.

Fig. 1 illustrates a simple experiment design process: the aim is to design an experiment consisting of four source (circle)–receiver (square) ray paths along which traveltime measurements of waves will be made that allows most information about the wave velocity in each of four square cells to be obtained, assuming the velocity is constant within each cell. If we use the simple parallel-ray experiment on the left then only two independent pieces of information can be retrieved [the average velocities $(V_1+V_2)/2$ and $(V_3+V_4)/2$]. The two differences in velocity (V_1-V_2) and (V_3-V_4) cannot be constrained and hence not all of the velocities $V_1, \ldots, V_4$ can be constrained by this experiment. This is reflected by the eigenvalue spectrum of the linearized inverse problem shown below the experiment design (a full description of what this means is given later): two positive eigenvalues imply that only two pieces of information are resolvable, the two zero eigenvalues imply that two pieces of information cannot be resolved by this experiment. However, using the experiment design on the right-hand side of the figure each datum constrains the velocity in exactly one cell and we obtain four independent pieces of information (V_1, V_2, V_3 and V_4). In the eigenvalue spectrum we obtain four positive eigenvalues, one corresponding to each piece of information.

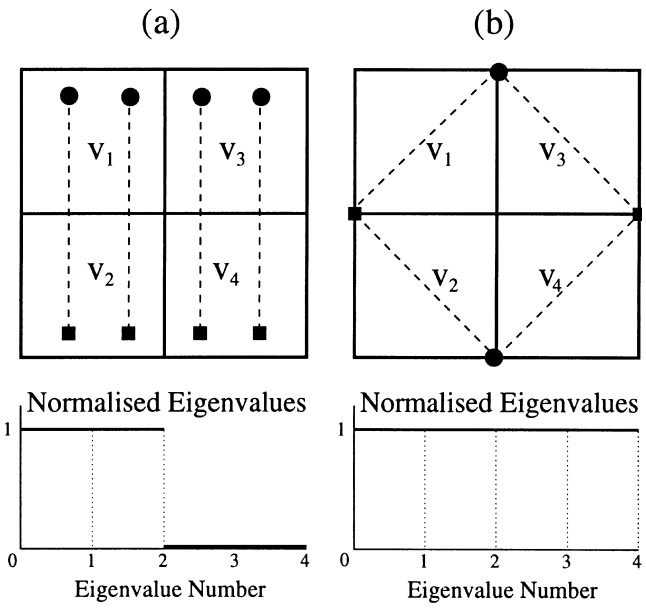

Figure 1. Comparison of eigenvalue spectra from the inverse problem of obtaining (constant) velocities $V_1 \ldots V_4$ within each cell given traveltime data along each of the dashed ray paths. Twice as much independent information is available using the ray path geometry in experiment design (b) than in design (a), and this is reflected in the corresponding eigenvalue spectra (shown beneath each geometry).

Notice that this extra information is obtained using half the number of sources and receivers as in the first experiment. This illustrates that intelligent experiment designs can often return more information at lower cost than designs based on poor design criteria.

All cells in both experiment designs have equal areas, and equal fractional ray-path coverage (cumulative [ray length within each cell] divided by [total ray length]). Therefore, ray-path coverage is not a good measure of the information about a given model which may be retrieved. However, the eigenvalues (with corresponding eigenvectors) contain *all* information relevant to the problem of translating data information into the model space, and as illustrated above the positivity of the eigenvalue spectrum gives an excellent measure of how much information can be retrieved, if the inverse problem is approximately linear.

In the past, many measures of eigenvalue positivity have been proposed in order to quantify the quality of any given experiment design. This paper begins by introducing a new measure which requires significantly less computation than most of those used previously. This measure was proposed by Curtis & Snieder (1997) to design model parametrizations with which the most possible information could be retrieved from a given experiment geometry. In the current paper the experiment geometry is varied to obtain the experiment design that gives maximum information about a model with pre-defined parametrization, as in Fig. 1. Curtis & Snieder show how the measure can be estimated efficiently using the 'reconditioning' algorithm. An important departure from their algorithm that increases the accuracy of the measure estimate by several orders of magnitude is described herein. A *quality vector* is then defined which consists of this, plus a selection of other measures which have been used in previous studies, together with one new variation. In practice one would probably use a weighted sum of these measures to design an optimal experiment. An understanding of the primary sensitivities of each measure, and the expected information about the model which would be retrieved from optimal experiments based on each measure individually, is necessary before any particular weighting system can be defined. These sensitivities will be discussed, illustrated and summarized in a table for ease of reference.

In the examples presented here, a cross-borehole traveltime experiment is designed to resolve best the interborehole velocity structure. Sources and receivers are constrained to lie within two wells or on the ground surface, and the background velocity model is assumed to be approximately constant between the wells. Designing an optimal source–receiver geometry in such experiments is non-trivial in practice because wells always deviate from the vertical (for example, due to borehole drilling drift, or when multilateral splays exist). Here the experiment design process is applied to vertical boreholes with limited numbers of sources and receivers because in this case rules of thumb for future survey design problems can be constructed (especially when cross-borehole surveys with and without the addition of sources and receivers on the surface are compared). These rules of thumb agree with human intuition,

validating both this intuition and the design techniques. However, the techniques presented here are applicable without modification to any irregular experiment geometry, provided it does not contravene the key assumption of approximate linearity of the model–data relationship.

2 METHODOLOGY

In the following sections various measures of experiment design quality will be used and compared. Hence, in this section a vector of five quality measures $\mathbf{\Theta}$ is introduced, rather than any single measure. Although any such vector can never be comprehensive, the vector presented here includes some commonly used measures based on inverse problem conditioning, and the five particular measures were selected to span a range of sensitivities to the inverse problem eigenspectrum (defined below).

For any given model vector $\mathbf{m}$ within model space M, let the forward problem of estimating the corresponding data vector $\mathbf{d}$ in data space D be accomplished by the matrix operator $\mathbf{A}$, i.e.

$$\mathbf{d} = \mathbf{A}\mathbf{m}. \tag{1}$$

Then, for a given data vector $\mathbf{d}_0$, we often wish to find the model vector $\mathbf{m}_0 \in M$ such that $|\mathbf{d}_0 - \mathbf{A}\mathbf{m}_0|^2$ is minimized. Theoretically this is accomplished by pre-multiplying eq. (1) by $\mathbf{A}^{\mathrm{T}}$ and taking a matrix inverse:

$$\mathbf{m} = (\mathbf{A}^{\mathrm{T}}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{d}. \tag{2}$$

Instability in the solution arises because the $N \times N$ square matrix,

$$\mathbf{L} = \mathbf{A}^{\mathrm{T}}\mathbf{A}, \tag{3}$$

is often near-singular; that is, some of its eigenvectors, $\{\mathbf{e}_i : i = 1, \ldots, N\}$ say, have extremely small eigenvalues $\{\lambda_i : i = 1, \ldots, N\}$. Measurement errors in the data space D propagate into solution $\mathbf{m}$ parallel to each eigenvector $\mathbf{e}_i$ with an amplification $1/\lambda_i$; hence, when small eigenvalues exist the solution becomes unstable and unreliable, and the inverse problem is said to be ill-conditioned (Menke 1992).

In such situations the *generalized* inverse $\mathbf{L}^{\dagger}$ is often used in place of $\mathbf{L}^{-1}$ in eq. (2). Here $\mathbf{L}^{\dagger} = \underline{\mathbf{L}}^{-1}$, where $\underline{\mathbf{L}}$ is the matrix that remains once all of the near-zero eigenvalues and corresponding eigenvectors have been removed from $\mathbf{L}$. Usually a damping threshold δ specifies the level below which eigenvalues (and eigenvectors) must be removed.

The solution to the inverse problem $\mathbf{m}$ in eq. (2) may be expressed in terms of the basis of eigenvectors of $\mathbf{L}$. Hence, the result of removing small eigenvalues and corresponding eigenvectors from $\mathbf{L}$ is that uncertainty in the model is reduced (since low eigenvalues are removed so data noise will be amplified less in the model space), but that resolution (the degree to which the remaining eigenvectors span the model space and hence can be used as a basis for the true model solution) is poorer.

Optimal experimental design based on matrix inverse methods is almost always concerned with maximizing the eigenvalues of matrix $\mathbf{L}$. This either minimizes the propagation of data errors into the model solution and hence minimizes model uncertainty, or minimizes the number of eigenvalue/eigenvector pairs that must be removed to form $\mathbf{L}^{\dagger}$ thus improving resolution. It can be shown that if $\mathbf{A}$ is real then the eigenvalues of $\mathbf{L}$ are all non-negative. Hence, optimal designs are obtained by maximizing measures sensitive to the magnitude of eigenvalues of $\mathbf{L}$. The particular strategy followed (either minimizing uncertainty, improving resolution or a combination of both) is determined by the sensitivity to the eigenspectrum of the particular measure chosen. Thus, the sensitivity of any measure must be examined and understood in detail before it can be used.

2.1 Measures of eigenvalue positivity

In this study the following measures of eigenvalue positivity are considered. Details of how each measure may be evaluated efficiently, the computational operation count required for their evaluation and their respective sensitivities to the eigenspectrum are described in Appendix A. In the notation here it is assumed that eigenvalues are listed in order of decreasing magnitude (λ_1 largest, λ_N smallest). Each measure corresponds to an element in the quality vector $\mathbf{\Theta}$:

$$\Theta_1 = \sum_{i=1}^{N} \frac{\lambda_i}{\lambda_1} \quad \text{(it can be shown that all } \lambda_i > 0\text{)};$$

$$\Theta_2 = \log \lambda_k \quad \text{for pre-defined fixed index } k;$$

$$\Theta_3 = k \quad \text{such that } \lambda_k > \delta \text{ for some pre-defined tolerance } \delta;$$

$$\Theta_4 = \sum_{i=1}^{N} \frac{-1}{\lambda_i + \delta};$$

$$\Theta_5 = \log |\mathbf{L}|_{\delta} = \sum_{i=1}^{N} \gamma_i, \quad \text{where} \quad \gamma_i = \begin{cases} \log \lambda_i & \text{if} \quad \lambda_i \geq \delta \\ \text{Penalty} & \text{if} \quad \lambda_i < \delta \end{cases}.$$

References are as follows. Θ_1: Curtis & Snieder (1997) and Appendix of this study; Θ_2: Barth & Wunsch (1990) and Smith *et al.* (1992); Θ_3: this study; Θ_4: Maurer & Boerner (1998); Θ_5: Wald (1943), Box & Lucas (1959), John & Draper (1975), Kijko (1977), Rabinowitz & Steinberg (1990) and Steinberg *et al.* (1995); Penalty term introduced in this study. In all examples that follow the Penalty term is set equal to 10.

Apart from the measure of Curtis & Snieder (1997), all measures listed here require the complete eigenvalue spectrum (or, for an equivalent amount of computation, the determinant) to be calculated. The measure of Curtis & Snieder can be obtained in $\sim N^2$ operations (multiplications or additions; see Appendix A), whereas the others require at least $\sim N^3$. Hence, where computation is a limiting factor it may be that only Θ_1 can be used. Note, though, that this relative efficiency trades off with a lack of sensitivity to the smallest eigenvalues.

Measures $\Theta_3, \ldots, \Theta_5$ contain a *damping* parameter δ. The purpose of this parameter is to reduce sensitivity to eigenvalues with magnitudes approximately less than δ. This is often useful if structure associated with eigenvalues below a certain magnitude is likely to be removed, or altered by regularization during inversions of the experimental data: if such structure will not be included in final inversion results then the corresponding eigenvalues should not influence our choice of experiment design.

When an experiment is designed by maximizing any particular measure, the sensitivity of that measure to the eigenspectrum dictates the general form of the *a posteriori* model uncertainty distribution which would be expected when

the experimental data are used to constrain the model. For instance, if measure Θ_2 with $k=1$ is maximized, the resulting experiment design may have a very large value for λ_1 at the expense of very small values for $\lambda_2, \ldots, \lambda_N$; in this case there will only be one direction in model space in which the model estimate is resolved and well-constrained. This is less likely to be the case if Θ_1 is maximized since this measure is sensitive to the magnitude of all eigenvalues.

Similar considerations for each measure in the quality vector allow us to suggest general guidelines about the shape of the model uncertainty distributions that might be expected post-inversion if the experiment conducted was designed by maximizing any particular quality measure. Details are given in Appendix A. These guidelines are most easily summarized by visualizing each case when the model vector has only two elements, m_1 and m_2, and when we assume that the inverse problem is overdetermined for some, but not all, experiment designs (Fig. 2). In this case a typical 2-D, randomly designed experiment may result in an over or underdetermined inverse problem (cases 1 and 2 respectively, left-hand column). Error ellipse axes have length $l_i \propto 1/\lambda_i$, $i=1,2$. Features to notice in optimized experiments are as follows.

(1) Only Θ_2 with $k=1$ guarantees that l_1 will be decreased with respect to a typical random experiment.

(2) Θ_1 is only sensitive to eigenvalue magnitude relative to the largest eigenvalue, hence the ellipse may be large but will definitely be closed, giving an overdetermined inverse problem.

(3) Trade-offs between l_1 and l_2 exist to some extent in all measures other than Θ_2—see Appendix for discussion.

(4) If small eigenvalues (and corresponding eigenvectors) are damped (removed from matrix $\mathbf{L}$ in eq. 3) then the major axis direction is completely ignored in the inversion. Thus the formal model uncertainty is decreased (it becomes equal to the length of the minor axis), but there is no resolution in the major axis direction.

(5) The perceived success of design optimization depends to a large extent on the typical error distribution in the random case—most results in the right-hand column are better than case 2, but few are necessarily better than case 1.

Many of these features will be observed in the experimental results in later sections.

2.2 Experiment design using the genetic algorithm

The quality vector $\boldsymbol{\Theta}$ may be calculated for any matrix $\mathbf{L}=\mathbf{A}^T\mathbf{A}$ and depends on experiment design in a highly non-linear manner. Designs of higher quality are those for which one or more elements Θ_i of vector $\boldsymbol{\Theta}$ are increased. In practice, one would probably decide to maximize a weighted sum,

$$\Phi = \sum_{i=1}^{5} \beta_i \Theta_i , \tag{4}$$

where weights β_i are chosen such that the sum gives some desired sensitivity to the eigenspectrum. In the current study this maximization is achieved by minimizing $-\Phi$ with respect to experiment design. Thus, any non-linearized minimizing algorithm may be used to find an experiment design of increased quality. In general we may not find the global maximum value of Φ, but *any* increase in Φ adds quality to the experiment.

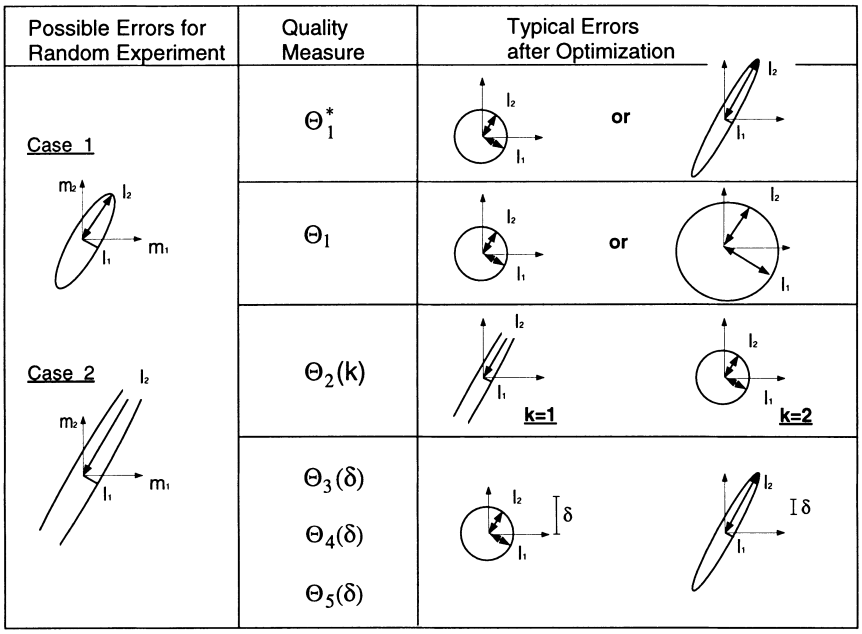

Figure 2. Schematic overview of expected characteristics of model error distributions resulting from either randomly designed experiments (left-hand column) or those designed by maximizing different measures of design quality (middle and right-hand columns—see text and Appendix A for definitions of quality measures $\Theta_1, \ldots, \Theta_5$ and Θ_1^*). Options marked 'or' in the right-hand column are those over which there is little or no control. It is assumed that the inverse problem is overdetermined for some but not all physically realizable experiment designs. Error ellipse axes have length $l_i \propto 1/\lambda_i$, $i=1, 2$. For underdetermined problems the *a posteriori* error ellipses shown are truncated (since they are of infinite extent). In the right-hand column axis labels m_1 and m_2 are omitted for clarity, and note that in general ellipses may be rotated compared to those in the left-hand column. See text for a discussion of this figure.

In this study a genetic algorithm (GA) was used to minimize $-\Phi$ with respect to source and receiver locations in cross-borehole tomographic examples, keeping the velocity model parametrization fixed (see Holland 1975 or Sambridge & Drijkoningen 1992 for a description of GAs). Two sets of experiments were performed. First, since the only quality measure which is relatively new to geophysical survey design is the 'reconditioning' measure Θ_1 of Curtis & Snieder (1997), experiments were performed using this measure alone to demonstrate its strengths and weaknesses. (Measure Θ_3 is also new to survey design, but as explained in the Appendix and illustrated below, it represents an improvement only in efficiency and ease of use over Θ_2, not in the results obtained). Second, similar experiments are performed using measures $\Theta_2, \ldots, \Theta_5$ in turn and the results are compared. Hence, throughout this study only one β_i in the sum of eq. (4) will be non-zero at a time. This illustrates the sensitivities of each measure Θ_i individually. As noted above, an understanding of these sensitivities is a prerequisite to designing a more complex weighting system.

Each experiment design was described by a vector of Cartesian source and receiver locations, and these are coded digitally within the GA (in contrast to the usual binary encoding). This has the advantage that the internal operations of the GA become more intuitive: the cross-over operation involves swapping locations between different source–receiver sets, while the mutation operation requires that a completely new source–receiver vector is chosen at random from a macro set of all possible locations. This macro set is pre-defined as all points within boreholes or on the ground surface, with the single limitation that only sources can be placed in the left borehole and only receivers in the right borehole (or vice versa). Cross-over and mutation steps are performed between successive GA generations.

A GA requires tuning. The particular values used for the probability of cross-over P_c and mutation P_m, and the size of the population of source–receiver sets which are stored at each generation can have a large effect on the performance of the GA. Throughout this study the values $P_c = 0.2$ and $P_m = 0.02$ with population size 64 over 2000 generations were used. The low mutation rate causes the GA to converge around a minimum value of $-\Phi$ fairly quickly whilst still allowing some freedom to test and reproduce source–receiver sets which are completely different from those around the (possibly local) minimum. It is likely that further experimentation would have allowed improved values of P_c and P_m to be chosen, but these will always be problem-dependent so there was little value in searching for them in this study. Each minimization was performed five times and the run which gave the best design overall is presented below.

Currently the most commonly used cross-borehole survey design assigns regular and densely spaced sources and receivers in respective boreholes with no sources or receivers on the ground surface. Therefore, in each experiment this 'standard' survey geometry is compared with the best survey design found with the additional freedom of surface sources and receivers while keeping the total number of sources and receivers fixed and equal to that in the regular geometry. Hence, comparison of information provided by these two designs implicitly provides a measure of the added value of including surface measurements in cross-borehole surveys.

3 RESULTS

3.1 Experiment set 1: introducing Θ_1

To illustrate the concepts of optimal survey design and to introduce the new quality measure Θ_1, survey design is performed first for a problem similar to that which was illustrated in Fig. 1. The space between two vertical boreholes is divided into four equidimensional rectangular cells, and only two sources and two receivers are used. The only difference from the situation in Fig. 1 is that sources and receivers (henceforth called *nodes*) can only be placed on three sides of the model (the ground surface and two boreholes, one on either side). In this case it is less obvious how to construct the optimal geometry than it was in Fig. 1(b), and this example illustrates many interesting features of the method in a simple and accessible way.

Fig. 3(a) shows ray paths in the standard experiment with regularly spaced sources and receivers down each borehole for this simple case. Notice that we can see immediately that one zero eigenvalue will exist for this and indeed every symmetric cross-borehole tomography experiment with vertical boreholes and without surface nodes: the difference between average velocities on the left- and right-hand sides of the velocity model cannot be resolved for a similar reason to that discussed for Fig. 1(a).

The experiment was redesigned by maximizing Θ_1 using the genetic algorithm, and the best source–receiver geometry found is shown in Fig. 3(b). Although this design may not represent the absolute global maximum of measure Θ_1, it should be close since all five independent runs of the GA produced almost identical designs, apart from a reflection about the line $x = 50$ m in two cases (due to symmetry of the entire design problem about this line). Such an irregular geometry is by no means obvious given the regular shape of the velocity cells. At each stage during the GA search, the running-best design was stored (that is, any design with Θ_1 higher than the previous best design was stored). There were nine such designs during the 2000 generations and Fig. 3(c) shows the true normalized eigenvalue spectra of these with the regular and the best design spectra highlighted. The zero eigenvalue of the regular design is removed and the second eigenvalue greatly increased by using the optimal design, showing that in this case all four cell velocities may be constrained.

Notice that these advantages are gained at the slight expense of eigenvalue 3, which decreases in magnitude. This demonstrates that Θ_1 contains no explicit penalty for eigenvalue reduction, nor even for zero eigenvalues, as indicated in Fig. 2.

The quality measure Θ_1 for each of the nine running-best models is shown in Fig. 3(d), which illustrates the increase in Θ_1 over the period of optimization using the GA. Note that on average the actual work done by the GA to find each consecutive design in this plot increases with each consecutive design. That is, it takes far more work to improve design 6 by finding design 7 than it does to improve design 1 by finding design 2. Since *any* increase in the quality measure indicates an increase in the amount of information which may be retrieved from the data, we may choose to stop the GA search at any point and use the current best design.

From Fig. 3(c) it is clear that even in this simple example one needs to calculate Θ_1 with high accuracy to obtain improvements in the smallest eigenvalue of each spectrum.

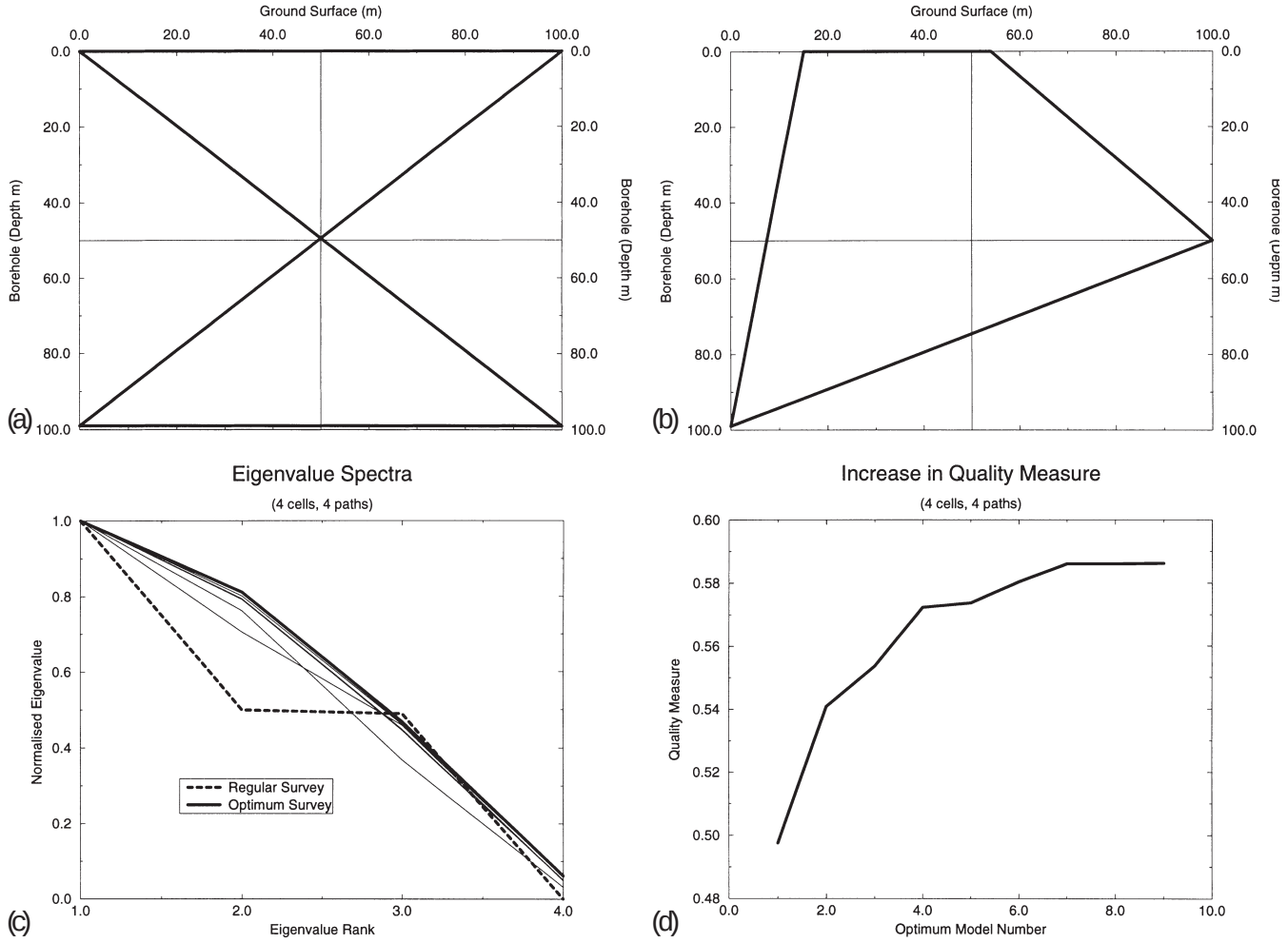

Figure 3. (a) Ray paths in the common experiment design using regularly spaced sources and receivers in each well, and no surface sources. In this particular example only four constant-velocity model cells (solid thin outlines), two sources and two receivers, and hence four ray paths (bold lines), are used. In this and all following experiments (unless otherwise stated), the ground surface bounds the top of the plot and two vertical boreholes bound the sides of the box. Sources may be placed in the left borehole, receivers in the right borehole, and both sources and receivers may be placed on the surface (although there are no surface nodes in this particular figure). (b) Optimal experiment design found by maximizing Θ_1 allowing surface sources. Figure key and experiment description as in (a). (c) Normalized eigenvalue spectra of the regular survey geometry in (a) (dashed line), the optimal geometry in (b) (bold solid line) and the running-best models before the optimal model was found (thin solid lines). Corresponding quality measures Θ_1 are given in (d). (d) Increase in Θ_1 with each successive running-best survey design found by the GA. Corresponding normalized eigenvalue spectra are shown in (c).

This is true to a much greater extent in all of the following examples in this paper since it is usually not possible to increase all eigenvalues above zero, so the smallest positive eigenvalues usually approach machine precision. In Appendix A it is shown how Θ_1 can be estimated in $\sim N^2$ operations. The main inaccuracy in estimating Θ_1 derives from the estimate $\overline{\lambda}_1$ in eq. (A3) (Appendix A). Curtis & Snieder (1997) took $\overline{\lambda}_1$ to be the average of several estimates of λ_1, whereas in the Appendix it is shown that in fact one should achieve far greater accuracy using the maximum of these estimates.

The simple experiment described above illustrates most of the concepts introduced in the current study in an accessible way. Figs 4(a–e) illustrate the results when the method is applied to a more complex example involving 20 sources and 20 receivers in a more finely parametrized model space (100 cells), still with a homogeneous background velocity model. The geometry of the resulting best design shown in Fig. 4(b) should not be taken as exactly optimal for use in practice (since the GA may not find the exact maximum of Θ_1, and since the model cells still have an appreciable size which affects the exact location of sources and receivers). However, certain general patterns are common to all of the designs which gave high-quality measures, and hence these may be used as rules of thumb in designing future surveys.

Two such common patterns exist for the designs found in this experiment: First, in each borehole the distribution of nodes in the best designs found (e.g. Fig. 4b) has a gradually increasing density throughout the length of each well until the base of the well, where nodes are densely packed. Second, surface nodes are required, and their density is lower than the average density down the boreholes. Each borehole usually had 16 nodes along its length, while the surface only had eight nodes (note that each borehole and the interborehole surface all have equal lengths, and are spanned by equal numbers of model cells).

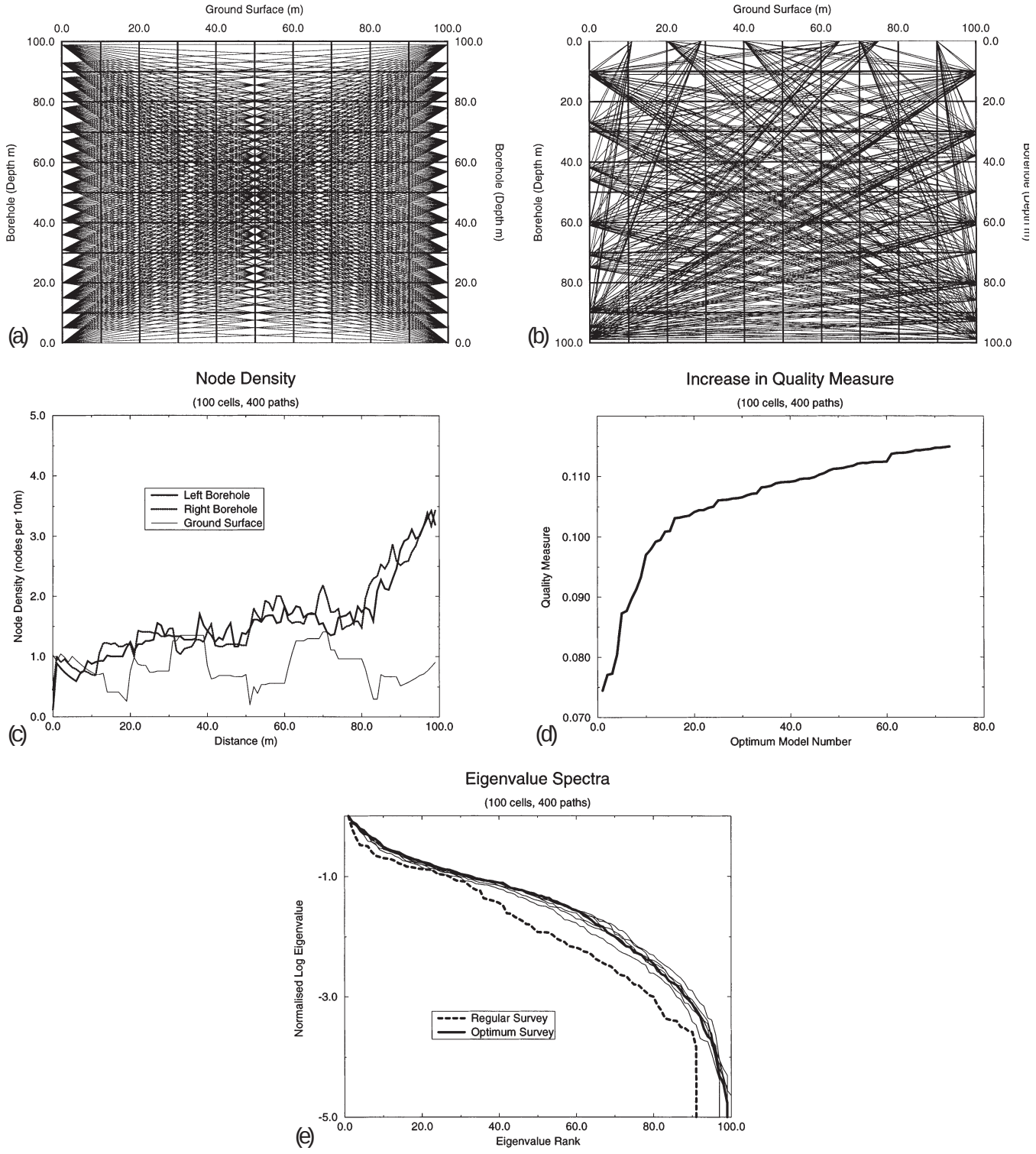

Figure 4. (a) Common experiment design using 20 regularly spaced sources and receivers in each well and no surface sources. In this example 100 cells on a regular grid (thick outlines) and 400 ray paths (thin lines) are used. For figure key and experiment description see caption to Fig. 3(a). (b) Optimal experiment design found by maximizing Θ_1 and allowing surface sources. For figure key and experiment description see caption to Fig. 3(a). (c) Node density down the boreholes and along the ground surface as a function of distance, averaged over 10 m bins (corresponding to the width of model cells in a and b), and over all running-best models with $\Theta_1 > 0.1$ (60 in total—see d). (d) Increase in Θ_1 with each successive running-best survey design found by the GA. Some corresponding eigenvalue spectra are shown in (e). (e) Log-normalized eigenvalue spectra of the regular survey geometry in (a) (regular survey—design 1), the optimal geometry in (b) (design 73) and some running-best models before the optimal model was found (designs 10, 20, . . . , 60, thin lines). Corresponding quality measures Θ_1 are given in (d).

The above two patterns were common to all of the best designs found by the GA (that is, all designs with quality measure close to the maximum found), and this is illustrated in Fig. 4(c), in which the average node density in each borehole and along the ground surface is plotted as a function of distance. In this plot the node density (number of nodes in each interval) has been smoothed with a 10 m averaging kernel (corresponding to the width of 1 model cell), and averaged over all running-best designs with condition number greater than 0.1 (60 designs in total, see Fig. 4d). The optimal borehole densities decrease to zero at the top of each well, increase gradually over the length of each well, and increase more sharply at the base of each well. The roughness of the average ground surface density curve is due partly to the low number of nodes used in this experiment (for example, in Fig. 4(b) only eight nodes are distributed across the 10 model cells which touch the surface) and partly to strong interdependency amongst the 60 models which were averaged (the GA tends to duplicate running-best designs or swap parts of them with other designs in each population). However, it is clear that the average surface density is lower than the average in the boreholes, and from Fig. 4(b) that sources and receivers should be spatially alternating (since ray paths emanate from surface nodes alternately to the left and right boreholes, which contain only sources and only receivers respectively).

The increase in Θ_1 with each successive reduction found is shown in Fig. 4(d). It is clear from this plot that an approximately optimal experiment design could have been found with very much less computation than was actually performed, and also that further small increases in Θ_1 are almost certainly possible by continuing the search further. Thus one must always decide to stop the genetic algorithm when a survey design has been found that is 'good enough'.

However, the value of Θ_1 is not the only criterion that might be used to terminate the search. Fig. 4(e) shows the log-normalized eigenvalue spectra of several of the running-best models throughout the search. What is immediately clear from this plot is that the best designed model with respect to Θ_1 that was found (number 73 in Fig. 4d) does not have the maximum possible number of non-zero eigenvalues. This is again due to the fact that Θ_1 is primarily sensitive to the largest eigenvalues and contains no explicit penalty for zero or near-zero eigenvalues. If the data uncertainties in this experiment were expected to be large enough that eigenvalues less than approximately $10^{-1.7}\lambda_1 \simeq \lambda_1/50$ would have to be damped (that is, those with normalized log value less than 1.7 in Fig. 4e would have to be removed) to obtain a reasonable solution then this does not matter since all eigenvalues which would be maintained have indeed been increased during the optimization. However, if the data are very accurate then this characteristic of Θ_1 does pose a significant problem since we have no control over the level at which eigenvalues may be reduced rather than increased.

In conclusion, despite the relative efficiency with which Θ_1 can be calculated, it should be used only (i) for problems where the data are expected to have reasonably large uncertainties, (ii) in overdetermined problems (with any data accuracy) where all eigenvalues contribute significantly to the sum in eq. (A3), (iii) if we only care about structure spanned by eigenvectors corresponding to the largest few eigenvalues for any other reason or (iv) if the computation involved in calculating the other measures in quality vector $\boldsymbol{\Theta}$ is too great for the computing power available (that is, if the number of model parameters N is too large).

3.2 Experiment set 2: comparison of $\Theta_1, \ldots, \Theta_5$

Experiment designs optimized with respect to each of the five different quality measures $\Theta_1, \ldots, \Theta_5$ in the quality vector $\boldsymbol{\Theta}$ are now compared. Experiments were designed for the same tomography problem as that shown in Figs 4(a) and (b) using each of the other four measures in turn. Figs 5(a) and (b)

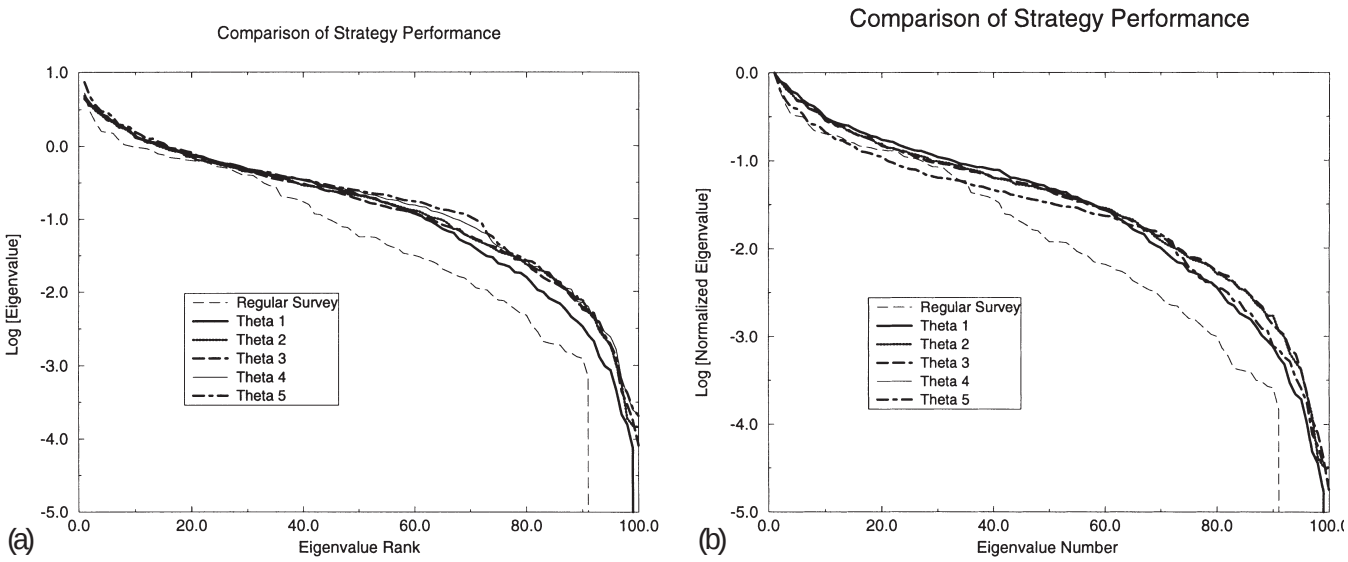

Figure 5. (a) Log eigenvalue spectra of the regular survey geometry in Fig. 4(a) and the optimal geometries found using each of the quality measures in vector $\boldsymbol{\Theta}$, including that using Θ_1 illustrated in Figs 4(b) and (e). In each measure involving a δ term, $\delta = 0.001$ was used; for Θ_2, $k = 90$ was used. (b) Log-normalized eigenvalue spectra (normalized by the largest eigenvalue) of the regular survey geometry in Fig. 4(a) and the optimal geometries found using each of the quality measures in vector $\boldsymbol{\Theta}$, including that using Θ_1 illustrated in Figs 4(b) and (e). In each measure involving a δ term, $\delta = 0.001$ was used; for Θ_2, $k = 90$ was used.

show a comparison of the absolute and normalized eigenvalue spectra respectively of the optimum model found in each case, and Fig. 6 shows the corresponding node density distributions.

The most obvious feature of the optimized eigenvalue spectra is that they all represent far better conditioned inverse problems than the regular geometry with evenly spaced nodes down each borehole and no surface nodes, despite the fact that the number of nodes used in all cases was the same. Hence, the redistribution of nodes to include surface sources and receivers can provide significantly more information.

Within the scatter of the optimum spectra some interesting patterns have emerged. From Fig. 5(a) it seems that measure Θ_5 has performed best, giving largest eigenvalues for ranks 1–3 and 40–70, with no significantly low values across the other ranks. As expected from the previous discussion, the reconditioning measure Θ_1 performed worst for small eigenvalues of ranks greater than 50. The damped reciprocal measure Θ_4 also performed noticeably better than Θ_2 or Θ_3 throughout most of the eigenvalue range; this is also expected from the discussion in the Appendix because it is sensitive to all eigenvalues to some extent, not only to the one closest to the damping level δ.

The relative log eigenvalues plotted in Fig. 5(b) (normalized by λ_1) allow us to compare the flatness of the spectra (how closely the standard error ellipses approximate hyper-spheres). A perfectly flat spectrum would give homogeneous resolution and uncertainty across the whole model space. The reconditioning measure Θ_1 optimizes spectrum flatness explicitly, and indeed performs this function best for the largest eigenvalues but again worst for the smallest eigenvalues. The lowest values for ranks 2–65 are obtained using the determinant measure Θ_5, clearly due to the large value obtained for λ_1

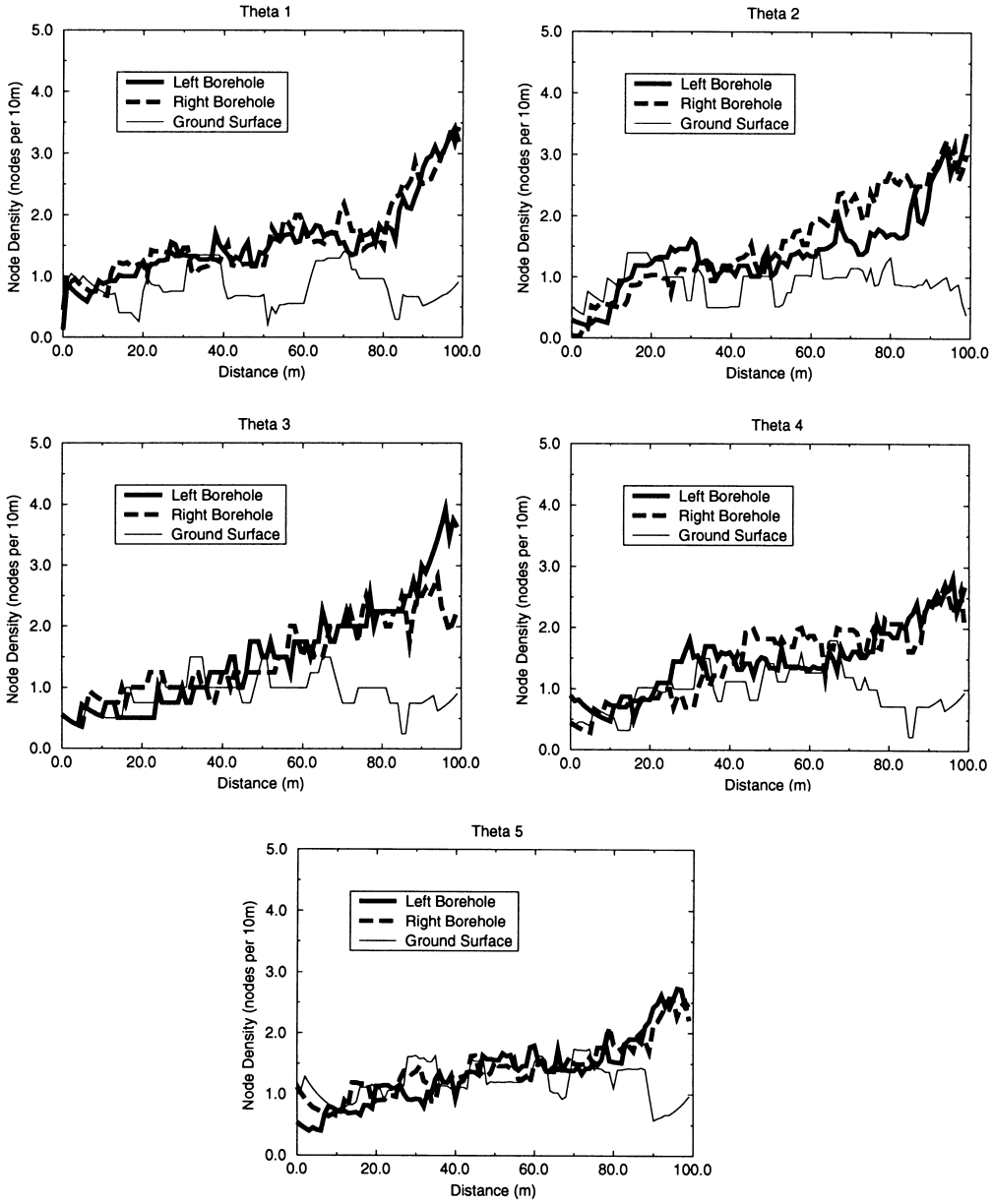

Figure 6. Node density across the surface and within each borehole for the optimum survey geometries found using each measure in the quality vector $\boldsymbol{\Theta}$. Each corresponds to an eigenvalue spectrum shown in Figs 5(a) and (b).

using Θ_5 (see Fig. 5a). Again, all measures provide more homogeneous resolution than the regular cross-borehole tomography experiment.

The average node density plots in Fig. 6 show remarkable similarities. Some anomalies do exist (only measure Θ_1 gives an increased node density around the base of each well as seen previously, and there are some asymmetries in the designs from Θ_2 and Θ_3, although these are not consistent between different runs of the design process so are possibly not significant), but several characteristics are common to most or all designs:

(1) redistributing borehole nodes so that they lie on the surface (or equivalently in a horizontal well above the target model if the entire model is located at greater depth) significantly improves designs;
(2) the node density increases steadily down the length of each well;
(3) the average node density on the ground surface is slightly lower than the average node densities down each well;
(4) measures $\Theta_3, \ldots, \Theta_5$ give optimal designs with surface node density that is increased slightly around the central point between the wells.

Since these features are common to most designs using a variety of quality measures, these can be used as general rules of thumb for designing new surveys.

4 SUMMARY, DISCUSSION AND CONCLUSIONS

The reconditioning measure of experiment quality has been introduced to the field of experiment design and its use was demonstrated by designing a cross-borehole tomographic experiment. Three other measures which have been used previously in geophysics were described and discussed, together with another new variant. This gives a total of five different measures of design quality that comprise the elements of the quality vector $\boldsymbol{\Theta}$. Higher values of Θ_i for any i indicate a better designed experiment with respect to that quality measure.

All of these measures are sensitive only to the amount of information that can, in principal, be translated from measured data into model constraints. None is sensitive to experiment cost, logistics or data leverage [extreme importance of a few principal data in which noise can have a large influence on model errors—see Maurer & Boerner (1998) for a treatment of this problem]. Instead, financial and physical constraints are used loosely to define the space of possible experiments (the number of sources and receivers we can afford, and the locations of the boreholes in which they can be placed).

The principal sensitivities of each measure determine the features of expected model uncertainty and resolution after the optimal experiment in each case has been performed. These error characteristics are summarized in Fig. 2. In practice a weighted sum of these measures may be the most robust global information measure; in this case weights must be assigned such that some desired trade-off between model uncertainty and resolution is attained, using the information in Fig. 2 (and probably some trial-and-error) to guide the choice.

To evaluate the reconditioning measure Θ_1 requires $\sim N^2$ operations, where N is the dimension of the model space, whereas the other measures all require at least $\sim N^3$. This saving is significant: a GA has been used in this study to find the minimum of Θ_i for each i (and hence the optimal design). The GA requires many thousands of evaluations of Θ_i, which quickly becomes prohibitive in terms of computational cost as the number of model parameters N increases. However, the computational saving by using Θ_1 is achieved at the expense of sensitivity to small eigenvalues (roughly for eigenvalues smaller than $\lambda_1/50$ in the examples presented herein, where λ_1 is the largest eigenvalue). Since remaining small eigenvalues and corresponding eigenvectors are likely to be damped out of inversions, optimizing designs with respect to Θ_1 is likely to decrease estimate uncertainty but give poorer resolution than might be attained using other measures, which can be tuned to be sensitive to small eigenvalues. Hence, Θ_1 should only be used when either (i) the data–model relationship is over-determined (no small eigenvalues), (ii) where there is no choice due to the large computational workload or (iii) when we are only interested in the structure corresponding to the largest few eigenvalues for any other reason.

The optimal designs found using any measure may not be unique. For instance, in the symmetric cross-borehole examples presented above a reflection about $x = 50$ m always gives another optimal design. Hence, in practice one may ask the question does another equally good design exist that is easier or cheaper to acquire given currently available tools and technology? One way to answer this would be to find as many optimal designs as possible and compare each of them for ease and cost of acquisition. However, if the solution is very non-unique then finding all optimal design possibilities may be difficult. Usually a better method will be to build acquisition or cost constraints into the design process itself, either by subtracting penalty terms (for increased cost or for designs that are incompatible with current tools) from the quality measure being maximized, or by limiting directly the space of designs from which the GA samples so that only tool-compatible and cheap experiments or surveys will be tested. The resulting optimal survey designs will give maximum information while remaining cheap and easy to acquire so that any remaining non-uniqueness need not concern us.

Despite the possible non-uniqueness, optimal cross-borehole tomographic source–receiver geometries assuming an inter-borehole background medium which is approximately homogeneous show remarkable consistency, no matter which measure is used. The common features may therefore be used as rules of thumb in future experiment designs. These may be observed in Fig. 6 and are summarized as follows:

(1) surface nodes (or equivalently nodes in a horizontal well above the target model if the entire model is located at greater depth) significantly improve designs;
(2) the node density increases steadily down the length of each well;
(3) the average node density on the ground surface is lower than the average node densities down each well;
(4) measures $\Theta_3, \ldots, \Theta_5$ give optimal designs with surface node density, which is increased slightly around the central point between the wells.

The first, second and fourth of these features are intuitive: model cells at depth are least constrained by the surfaces nodes and hence require an increased number of proximal borehole nodes; cells near the centre of the space between the wells are the least constrained by the well nodes and hence require more proximal surface nodes. However, the relative node densities on the surface compared to the boreholes shown in Fig. 6 are

not intuitively obvious, and neither is the *rate* of increase in node density down each well. The reason that this extra information is not intuitive is that it is quantitative: although we might expect that certain characteristic patterns in a survey design might give more information than others, from intuition alone we can never quantify exactly which pattern should be implemented in practice. Thus, the automated design process presented here provides information that confirms intuitive experimental quality criterion, but provides additional quantitative information that cannot be obtained in a straightforward manner by other means.

The techniques presented here are especially valuable when irregular experimental situations are encountered. For instance, the wells may deviate significantly from the vertical with several multilateral splays with different depth ranges; we may want to define the optimal trajectory along which to drill lateral wells within which logging tools can be run for improved tomographic resolution; we may wish to include different data types (e.g. electromagnetic or gravity data); we may wish to resolve anisotropic velocity parameters. In all of these cases conventional rules of thumb are likely to be much less effective than in the simple situation above. Optimal experiments can then only be designed using quantitative methods such as those introduced and discussed in this study.

ACKNOWLEDGMENTS

Sincere thanks to Roel Snieder, Carl Spencer and Alberto Michelini for many fruitful discussions. Part of this work was supported by the Edinburgh Parallel Computing Centre through a European Community Training and Mobility of Researchers grant—very many thanks to Bob Pearce for facilitating this.

REFERENCES

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

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

Curtis, A. & Snieder, R., 1997. Reconditioning inverse problems using the genetic algorithm and revised parameterization, *Geophysics,* **62,** 1524–1532.

Holland, J.H., 1975. *Adaptation in Natural and Artificial Systems,* University of Michigan Press.

John, R.C. & Draper, N.R., 1975. D-optimality for regression designs: a review, *Technometrics,* **17,** 15–23.

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

Matsu'ura, M. & Hirata, N., 1982. Generalized least-squares solutions to quasi-linear inverse problems with a priori information, *J. Phys. Earth,* **30,** 451–468.

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

Menke, W., 1992. *Geophysical Data Analysis: Discrete Inverse Theory,* revised edn), *International Geophysics Series,* **45,** Academic Press, Harcourt Brace, Jovanovich.

Press, F., Flannery, B.P., Teukolsky, S.A. & Vetterling, W.T., 1992. *Numerical Recipes in FORTRAN, the Art of Scientific Computing,* 2nd edn, Cambridge University Press, Cambridge.

Rabinowitz, N. & Steinberg, D.M., 1990. Optimal configuration of a seismographic network: a statistical approach, *Bull. seism. Soc. Am.,* **80,** 187–196.

Sambridge, M. & Drijkoningen, G., 1992. Genetic algorithms in seismic wave-form inversion, *Geophys. J. Int.,* **109,** 323–342

Smith, M.L., Scales, J.A. & Fischer, T.L., 1992. Global search and genetic algorithms, *Leading Edge,* January, 22–26.

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

Wald, A., 1943. On the efficient design of statistical investigations, *Ann. Math. Stat.,* **14,** 134–140.

APPENDIX A: MEASURES OF EXPERIMENT QUALITY

In this Appendix each quality measure in turn is introduced with a description of the algorithms required for their calculation, the computational workload, and distinctive features which might be expected in *a posteriori* model error distributions when each measure is maximized to design optimal experiments. The characteristic features of expected model errors are summarized in Fig. 2, and many of these features are observed in the experimental results in later sections.

A1 Quality measure Θ_1: $\Sigma_{i=1}^{N} \lambda_i / N\lambda_1$

As illustrated in the Introduction, maximizing the amount of information retrieved from an experiment is achieved by choosing a design that maximizes the eigenvalues of $\mathbf{L}$ in some sense. $\mathbf{A}$ is real so that $\mathbf{L}$ is positive indefinite (that is, all eigenvalues of $\mathbf{L}$ are positive or zero). Hence, let

$$\Theta_1^* = \text{Area under eigenvalue curve} \tag{A1}$$

$$= \sum_{i=1}^{N} \lambda_i . \tag{A2}$$

Then Θ_1^* quantifies the magnitude of the eigenvalues, and maximizing Θ_1^* increases (at least one or more) eigenvalues. This measure may be normalized by the largest eigenvalue λ_1 and by the number of eigenvalues N,

$$\Theta_1 = \frac{\sum_{i=1}^{N} \lambda_i}{N \times \lambda_1} , \tag{A3}$$

as defined (in reciprocal) by Curtis & Snieder (1997). The extra factor N in the numerator ensures that $\Theta_1 \leq 1$, and $\Theta_1 = 1$ *iff* all eigenvalues are equal. The normalization by λ_1 implies that only *relative* magnitudes of the eigenvalues affect Θ_1. This ensures that Θ_1 cannot be maximized by, for instance, maximizing λ_1 alone (as is the case for Θ_1^*).

Curtis & Snieder describe a technique for estimating Θ_1 efficiently. The technique is summarized again here because a departure from their algorithm increases the typical accuracy of the estimate by around three significant figures in the examples in this paper.

The trace of a matrix is invariant under similarity transformation; since a diagonal matrix with elements equal to the eigenvalues of $\mathbf{L}$ can be found by similarity transform of $\mathbf{L}$ alone, we have

$$\sum_{i=1}^{N} \lambda_i = \text{trace}(\mathbf{L}) . \tag{A4}$$

Also, the value of λ_1 may be estimated quickly using the power method as follows. Any random unit vector $\mathbf{U}$ can be expressed in terms of the basis of eigenvectors of $\mathbf{L}$ with coefficients α_i:

$$\mathbf{U} = \sum_{i=1}^{N} \alpha_i \mathbf{e}_i \,. \tag{A5}$$

Then, since $\mathbf{L}\mathbf{e}_i = \lambda_i \mathbf{e}_i$ (by definition of the eigenvalues and eigenvectors),

$$\mathbf{L}^n\mathbf{U} = \sum_{i=1}^{N} \alpha_i \lambda_i^n \mathbf{e}_i \,, \tag{A6}$$

and if $\alpha_1 \neq 0$ then $\mathbf{L}^n\mathbf{U} \rightarrow \alpha_1 \lambda_1^n \mathbf{e}_1$ as $n \rightarrow \infty$ since λ_1 is the largest eigenvalue. Hence

$$\frac{|\mathbf{L}^n\mathbf{U}|^2}{(\mathbf{L}^n\mathbf{U}).\mathbf{U}} \rightarrow \lambda_1^n \text{ as } n \rightarrow \infty \tag{A7}$$

with equality *iff* all eigenvalues are equal. In addition, note that the left-hand side of eq. (A7) tends to λ_1^n from below, as can be seen from

$$\frac{|\mathbf{L}^n\mathbf{U}|^2}{(\mathbf{L}^n\mathbf{U}).\mathbf{U}} = \lambda_1^n \left[\frac{1 + \sum_{i=2}^{N} \left(\frac{\alpha_i}{\alpha_1}\right)^2 \left(\frac{\lambda_i}{\lambda_1}\right)^{2n}}{1 + \sum_{i=2}^{N} \left(\frac{\alpha_i}{\alpha_1}\right)^2 \left(\frac{\lambda_i}{\lambda_1}\right)^{n}} \right] \tag{A8}$$

$$\leq \lambda_1^n \quad \text{since} \quad \frac{\lambda_i}{\lambda_1} \leq 1 \quad \text{and} \quad \lambda_i \geq 0 \quad \forall i = 1, \ldots, N \,. \tag{A9}$$

Hence, problems associated with $\alpha_1 = 0$ can be removed by making several (p) independent estimates of λ_1 using eq. (A7). For each estimate $\mathbf{U}$ is set equal to one of a set of orthonormal vectors in $\Re^N$. As illustrated above, the estimate $\overline{\lambda}_1 \leq \lambda_1$, so the maximum $\overline{\lambda}_1$ obtained over the p trials should be used as an estimate of λ_1. This departure from the technique of Curtis & Snieder (1997) (where the *average* of three independent $\overline{\lambda}_1$ estimates was used) is significant: the accuracy obtained using the current technique is typically increased by three significant figures, and this additional accuracy relates directly to the accuracy of the experiment optimization, as is clear from the examples in the main text. In every case the author has encountered to date, it was not necessary to use $p = N$ trials—the maximum estimate (equal to λ_1 to machine precision) can almost always be obtained within only four or five trials.

Fig. A1 illustrates the improvements offered by the simple adjustment of Curtis & Snieder's algorithm. For three independent runs of the design process illustrated in Figs 3(a)–(d), each using a different value for n (3, 5 and 10) in eqs (A6)–(A9), the average estimate of the largest eigenvalue $\overline{\lambda}_1$ (averaged over a complete orthogonal set of basis vectors $\mathbf{U}$ in eq. A5) is plotted against its true value λ_1 for the running-best models found in the above experiment. In each case, the estimate of λ_1 becomes progressively worse with consecutive designs and, as derived in eq. (A9), all are underestimates of the true value. This behaviour is disturbing since increasingly severe underestimates of λ_1 cause spurious increases in Θ_1 (eq. A3). Hence, Θ_1 can be increased either by increasing the true area under the eigenvalue curve, or by reducing the accuracy of estimate $\overline{\lambda}_1$!

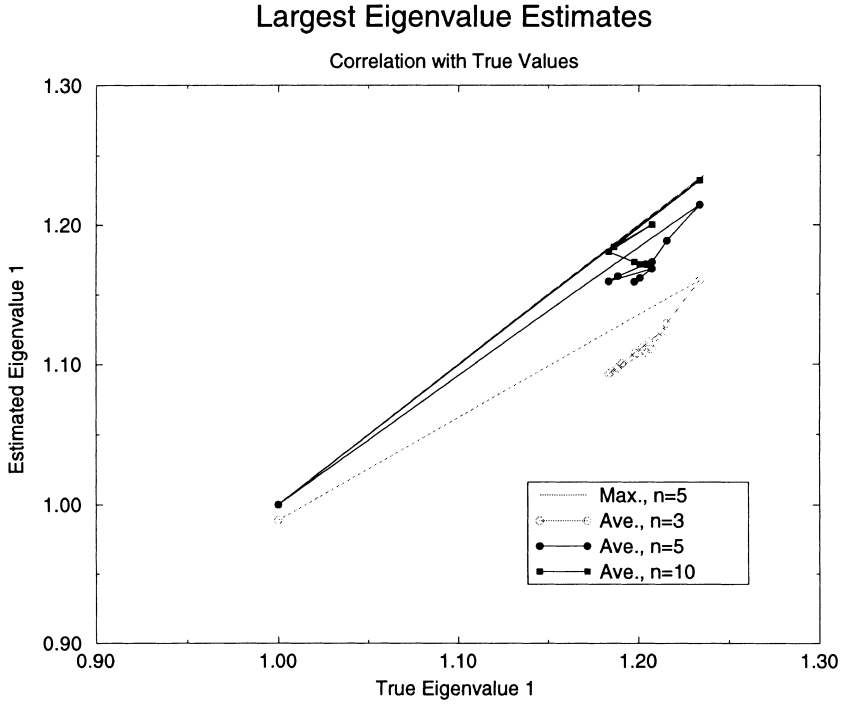

Figure A1. The estimate of the largest eigenvalue $\overline{\lambda}_1$ is plotted against the true value λ_1 for each of the running-best models found in three separate design experiments using Θ_1 with $n = 3$, 5 and 10 in eqs (A6)–(A9), where the estimate is calculated using the averaging method of Curtis & Snieder (1997) over a complete orthogonal set of basis vectors $\mathbf{U}$ (this is more robust than the limited basis used by Curtis & Snieder). In addition, the corresponding plot for $n = 5$ using the current method of taking the maximum estimate (labelled Max., $n = 5$) is shown and exhibits far greater accuracy (it lies exactly along the diagonal).

We can see how an experiment design which reduces the average estimate of $\bar{\lambda}_1$ might be chosen from eq. (A6): if we construct a design such that the eigenvector corresponding to the largest eigenvalue is parallel to one basis vector $\mathbf{U}$, then $\alpha_1 = 0$ and expression (A6) converges to $\alpha_2 \lambda_2^n \mathbf{e}_2$ as n increases, so expression (A7) will tend to λ_2^n. Since $\lambda_2 < \lambda_1$ by definition, we obtain a lower estimate for $\bar{\lambda}_1$. If in addition we choose a design such that $\lambda_2 \ll \lambda_1$ this effect becomes increasingly severe.

The above argument cannot hold for more than one of the p initial vectors $\mathbf{U}$ since they are orthonormal, so the current method of using the maximum of the p estimates of λ_1 should be robust against such problems for $p > 1$. Fig. A1 illustrates this by including the corresponding plot of $\bar{\lambda}_1$ against λ_1 using the current method (using the maximum estimate of λ_1 for $p = 4$ and $n = 5$). No such inaccuracies exist in this case, and indeed in the current example if $n = 10$ is used then $\bar{\lambda}_1 = \lambda_1$ to machine precision.

The resulting operation count (number of additions or multiplications) is N for calculating the trace, and $p \times n \times N^2$ for estimating λ_1. In this study $p = 4$, $n = 5$. Hence, if the velocity model contains significantly more than $N = 20$ parameters, the operation count is of the order of N^2 ($N = 100$ is used in the examples in the text, and $N \sim 1000$ would not be unreasonable under normal experimental situations).

When data errors are approximately Gaussian, expected model errors are Gaussian with confidence ellipse axes of length proportional to $1/\lambda_i$ in the direction of each eigenvector $\mathbf{e}_i$. This is useful as it allows a simple geometric representation of the sensitivity of Θ_1 to expected model errors to be given. Since Θ_1^* may also be maximized to increase the eigenvalues (and since it takes only $\sim N$ operations to calculate) that measure is also discussed here.

The sum of eigenvalues in Θ_1^* is dominantly sensitive to the largest eigenvalues. There is no explicit penalty for zero eigenvalues, and indeed Θ_1^* can be maximized by increasing the largest eigenvalues at the expense of all other eigenvalues. Θ_1, on the other hand, is only sensitive to eigenvalue magnitudes relative to λ_1. Hence, although there is still no explicit penalty for zero eigenvalues, the measure can only be maximized by making the eigenvalues 'more equal' in magnitude (that is, by flattening at least part of the eigenvalue spectrum; *c.f.* Fig. 1).

Fig. 2 illustrates schematically typical error distributions which might exist before and after an experiment has been optimized by maximizing either Θ_1^* or Θ_1. In this example there are only two model dimensions and it is assumed that the inverse problem for m_1 and m_2 can be made overdetermined (all eigenvalues non-zero) within the range of possible experiments. As Θ_1^* is increased, the expected error in at least one direction must be reduced, while that in other directions may be reduced but may also be increased. When Θ_1 is increased, however, λ_1 may decrease but other eigenvalues should increase relative to λ_1. Hence, while maximizing Θ_1^* may make the error ellipse more elongate, maximizing Θ_1 should make it more 'rounded' removing null-space (unconstrained) eigendirections.

In the case when an experiment is inherently underdetermined no matter what design is chosen, many zero eigenvalues may remain after minimizing either of these measures since additive trade-offs exist between eigenvalues in $\sum_{i=1}^{N} \lambda_i$ (large eigenvalues are often increased at the expense of small values). In this case one of the other measures discussed below may be better suited to reduce global posterior model errors.

A2 Quality measure $\Theta_2(k)$: $\log(\lambda_k)$ for pre-defined k

To design ship-track trajectories for oceanographic tomography, Barth & Wunsch (1990) suggested that the eigenvalue spectrum might be increased by maximizing the eigenvalue λ_k at a pre-defined rank k (after ordering the eigenvalues), and this technique was followed by Smith *et al.* (1992). In this study their measure is converted into

$$\Theta_2(k) = \log(\lambda_k) \quad \text{for pre-defined } k\,; \tag{A10}$$

note that, if desired,the measure may easily be normalized by λ_1 using

$$\Theta_2^*(k) = \log(\lambda_k) - \log(\lambda_1) \quad \text{for pre-defined } k\,, \tag{A11}$$

making it sensitive only to relative eigenvalue magnitudes (as in Θ_1).

The operation count for evaluating this measure is approximately equal to that of computing the eigenvalues. This is typically $2N^3/3$ for large N, plus some overhead for eigenvalue ordering ($N \log_2 N$ for a Heapsort algorithm—see Press *et al.* 1992).

In the case of an overdetermined inverse problem we may set $k = N$, and indeed by maximizing Θ_2 the smallest few eigenvalues will be raised (perhaps at the expense of the larger eigenvalues). However, when an experiment is inherently or effectively underdetermined it is not intuitive how one should apply this measure, since when $k < N$ it is inevitable that some eigenvalues λ_i for $k < i \leq N$ will decrease in value during the optimization.

Generally eigenvalues beneath a certain level will be damped or removed when an inverse solution is constructed. The level beneath which eigenvalues should be damped depends on the level of data noise and prior knowledge about the model (Matsu'ura & Hirata 1982). Hence, in practice an experiment should be optimized by increasing the eigenvalues above a certain data-dependent threshold, ignoring eigenvalues beneath this threshold, where the threshold must be estimated based on expected data uncertainties. Using the measure $\Theta_2(k)$, the only way that this could be achieved is to create optimal experiment designs using a range of k values between 1 and N; the final design should be chosen to be the one using the highest k for which the final value of λ_k is greater than our expected threshold. Hence, although this measure seems reasonable, in practice it requires somewhere between $2 \times N^3/3$ and $2 \times N^4/3$ calculations.

Fig. 2 shows the types of errors which might be expected to remain after an experiment has been designed using this measure. As explained above, this depends on the chosen rank k. While λ_k is increased by maximizing Θ_2, eigenvalues at ranks both greater than and less than k may decrease.

A3 Quality measure $\Theta_3(\delta)$: maximum k such that $\lambda_k >$ pre-defined tolerance δ

The practical difficulties involved in applying quality measure Θ_2 can be avoided by using instead

$$\Theta_3(\delta) = \text{maximum } k\,, \quad \text{such that } \lambda_k > \delta\,, \tag{A12}$$

for some pre-defined damping level δ. Remember from the main text that λ_k^{-1} is the error amplification factor in direction $\mathbf{e}_k$. Hence, δ is defined to be the level at which eigenvalues will need to be damped or removed in practice to construct a

constrained inverse problem solution, the damping threshold discussed above. Thus, δ is dependent on expected data uncertainties and is more intuitive to define *a priori* than the fixed-rank k used in $\Theta_2(k)$. In addition, maximizing $\Theta_3(\delta)$ once is equivalent to carrying out the procedure outlined above where a set of trials are performed in which $\Theta_2(k)$ is maximized for a range of values of k.

This measure requires the full eigenvalue spectrum to be calculated and hence requires $\sim 2N^3/3$ operations. No eigenvalue ordering is required, saving $N \log_2 N$ operations compared to Θ_2, and no repeated optimizations are required, giving a further saving of up to $2(N-1)N^3/3$ operations, [although in practice the latter saving is likely to be much lower because fewer than N trials would be performed using $\Theta_2(k)$].

Fig. 2 shows that as long as δ is not specified too low (so that all eigenvalues can be increased above δ within the range of possible experiments), an overdetermined inverse problem will result from maximization of this measure. Note that the measure is insensitive to the largest eigenvalues and hence these may decrease during the optimization in a similar manner to Θ_2. Given the intuitive simplicity of using this measure compared to Θ_2 and the reduced operation count, there seems little reason to use Θ_2 in place of Θ_3. Measure Θ_2 is retained in the quality vector $\mathbf{\Theta}$, however, because a weighted sum of $\Theta_2(k)$ for different k can be used to construct a measure of eigenvalue positivity with arbitrary sensitivity to the eigenspectrum.

A4 Quality measure $\Theta_4(\delta)$: $\Sigma_{i=1}^{N} - 1/(\lambda_i + \delta)$

Maurer & Boerner (1998) minimized the measure

$$\sum_{i=1}^{N} \frac{1}{\lambda_i + \sqrt{\delta}}$$

to design electromagnetic experiments. By analogy, I define

$$\Theta_4(\delta) = \sum_{i=1}^{N} \frac{-1}{\lambda_i + \delta}, \tag{A13}$$

which should be maximized. Again δ should be specified *a priori* based on expected data uncertainties, and the operation count is approximately $2N^3/3$.

Fig. 2 shows the character of expected residual model uncertainties after maximizing $\Theta_4(\delta)$. The measure is most sensitive to the smallest eigenvalues which are approximately greater than δ, and hence its sensitivity is roughly similar to that of Θ_3, although Θ_4 retains sensitivity throughout the spectrum. Its reduced sensitivity to large eigenvalues means that the lengths of the error ellipse major axes may be reduced at the expense of lengthening minor axes.

A5 Quality measure $\Theta_5(\delta)$: threshold determinant $|\mathbf{A}^T\mathbf{A}|_\delta$

The determinant or D criterion,

$$|A^T A| = \prod_{i=1}^{N} \lambda_i, \tag{A14}$$

is the most popular measure of inverse problem conditioning in experiment design and has been used in many studies since the 1940s (see references listed in the main text). The product in eq. (A14) is far more sensitive to small eigenvalues than is the sum in Θ_1. The disadvantage of the D criterion over $\Theta_1, \ldots, \Theta_4$ is that in underdetermined inverse problems the determinant is always zero and the measure cannot be used. Its popularity is mainly due to its intuitive graphical interpretation: if errors in the data are approximately Gaussian, the determinant is inversely proportional to the area of the *a posteriori* error ellipses shown in Fig. 2. Hence, increasing the determinant decreases the model uncertainty hypervolume.

In this study the measure is used in log form with a penalization term for eigenvalues below a certain threshold:

$$\Theta_5(\delta) = \sum_{i=1}^{N} \gamma_i(\delta), \tag{A15}$$

where

$$\gamma_i(\delta) = \begin{cases} \log \lambda_i & \text{if } \lambda_i > \delta \\ 10 & \text{if } \lambda_i \leq \delta \end{cases}. \tag{A16}$$

The penalization term is large enough that $\Theta_5(\delta)$ cannot be minimized unless almost all possible eigenvalues are greater than δ. In a similar manner to Θ_3 and Θ_4 above, δ should be chosen based on uncertainties in the data. Since this information may be estimated *a priori*, the penalization approach in eq. (A16) extends the use of the D criterion to (effectively) underdetermined inverse problems in an intuitively simple manner.

The operation count for this measure is $\sim 2N^3/3$ operations, the same as for Θ_2, Θ_3 and Θ_4, and owing to its relatively homogeneous sensitivity to all eigenvalues in overdetermined problems this is often the preferred measure of design quality. Fig. 2 shows that the typical error characteristics of Θ_3, Θ_4 and Θ_5 might be similar in an overdetermined problem. The differences between using the measures in practice are first due to the details of their differing sensitivities across the eigenspectrum, and second that in the overdetermined case, maximization of Θ_5 guarantees a decrease in standard-error ellipse hypervolume (if δ is sufficiently low), which is not the case using any of $\Theta_1, \ldots, \Theta_4$.