Geophys. J. Int. (1999) **139,** 205–215

Optimal design of focused experiments and surveys

Andrew Curtis

Schlumberger Cambridge Research, High Cross, Madingley Road, Cambridge, CB3 0EL, *UK. E-mail: curtis@cambridge.scr.slb.com*

Accepted 1999 June 17. Received 1999 June 17; in original form 1998 September 16

SUMMARY

Experiments and surveys are often performed to obtain data that constrain some previously underconstrained model. Often, constraints are most desired in a particular subspace of model space. Experiment design optimization requires that the *quality* of any particular design can be both quantified and then maximized. This study shows how the quality can be defined such that it depends on the amount of information that is focused in the particular subspace of interest. In addition, algorithms are presented which allow one particular focused quality measure (from the class of focused measures) to be evaluated efficiently. A subclass of focused quality measures is also related to the standard variance and resolution measures from linearized inverse theory. The theory presented here requires that the relationship between model parameters and data can be linearized around a reference model without significant loss of information. Physical and financial constraints define the space of possible experiment designs. Cross-well tomographic examples are presented, plus a strategy for survey design to maximize information about linear combinations of parameters such as bulk modulus, $\kappa = \lambda + 2\mu/3$.

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

1 INTRODUCTION

Many geophysical experiments or surveys are expensive to perform. For this reason, designing an optimal experiment in terms of its cost and the goals which it will achieve is of crucial importance. The *quality* of an experiment design can be defined as the degree to which the experiment will satisfy some desired criteria (for example, physical, logistic or financial constraints, resolution of desired information). If the quality is measured by some function Φ, then typically Φ will increase as the number of experiment goals which will be fulfilled increases, and as the cost of the experiment decreases. Both the cost and the degree to which an experiment fulfils its goals will usually depend on the details of the particular experiment performed, for example, the exact experimental geometry, the equipment used, etc. Once Φ has been defined rigorously, the optimal experiment to conduct will be the one for which Φ is maximized.

Often the primary goal of an experiment is to obtain maximum information about a specific portion of a model. For example, one might wish a geophysical survey to provide information that is focused on a portion of an earth model containing an oil or gas reservoir, or perhaps on specific properties (for example, density, particular isotropic or anisotropic elastic moduli, conductivity, porosity) of either the whole subsurface volume or just a portion of it. However in many cases the data obtained from the experiment will also depend on the rest of the model (for example, the propagation of seismic energy between the ground surface and the reservoir depends on the model of the overburden). Hence, in general we wish to obtain maximum information about a particular subspace of the model parameter space, but to do so we must also obtain some information about the rest of the model in order to relate the model subspace to the data.

In many previous geophysical applications, model 'information' has been defined in terms of data coverage [for example, the number of seismic rays passing through a point in the earth model, or the fold of seismic reflections at a point, e.g. Özdenvar *et al.* (1996)]. Such methods suffer from the deficiency that they attempt to evaluate the amount of information about the model within the data domain, rather than within the model domain itself. A few geophysical studies have measured model information in the model domain using eigenvalues of the linearized inverse problem (explained below; Kijko 1977; Barth & Wunsch 1990; Rabinowitz & Steinberg 1990; Steinberg *et al.* 1995; Curtis & Snieder 1997; Maurer & Boerner 1998; see Curtis 1999 for a review and comparison of these). All of these studies designed experiments by maximizing the global model information, that is, the total amount of information about all model parameters which the experiment would be expected to deliver.

This paper presents a class of measures that are sensitive to the *distribution* of information in the model domain. By maximizing any of these measures, experiments or surveys

can be designed such that information retrieved is focused on certain pre-specified model properties or parameters, or more generally on a subspace of the earth model parameter space. Algorithms are described which allow certain particular focusing measures to be calculated efficiently for any experimental design, and example optimal survey designs are illustrated for a cross-borehole tomographic survey focused on the interborehole section of a hydrocarbon reservoir.

2 METHODOLOGY

The quality measure Φ will be decomposed into two (not necessarily distinct) parts:

$$\Phi = \mu\Phi^{\mathrm{G}} + (1-\mu)\Phi^{\mathrm{F}}, \tag{1}$$

where Φ, Φ^{G} and Φ^{F} are all measured in the earth model domain and will be defined rigorously later. Here Φ^{G} is a measure of global information. Φ^{F}, on the other hand, is the amount of information which will be focused in the specific model subspace in which we are interested. In practice, one would usually want to balance the distribution of information between the whole model and the model subset, so a relative weighting factor μ is introduced. If $\mu = 1$ then maximizing Φ will maximize the total amount of information across the whole model; if $\mu = 0$ then maximizing Φ will maximize the amount of information which is distributed across the specified model subset with no regard for information distributed elsewhere in the model.

Notice that if cost can be estimated for each experimental design then Φ above may easily be augmented to include the cost of the experiment. Alternatively, the financial and physical constraints may be used to define the set of possible experiment designs as illustrated in the examples below. Also, instead of adding the weighted focusing term Φ^{F}, it is possible to add a sequence of weighted terms $\Phi^{\mathrm{F1}}, \Phi^{\mathrm{F2}}, \ldots$, where each term measures the degree of focusing in distinct model subspaces $F1, F2, \ldots$, allowing different weights to be assigned to each.

The main advance presented in this methodology is to show how focusing of information can be measured. Crucially, whereas Φ^{G} is usually defined in terms of the eigenvalues of the linearized inverse problem of translating data into model constraints (defined below), the focusing term Φ^{F} will also take directions of associated eigenvectors into account.

2.1 Preparatory steps

Throughout the rest of this section it is assumed that the inverse problem of constraining the model parameters given the experimental design can be linearized to give a linear inverse problem (e.g. Aki & Richards 1980). This ensures that for any experimental design a matrix $\mathbf{A}$ can be constructed which describes the linearized relationship between the earth model parameters and the data that would be recorded during the experiment. Crucially, $\mathbf{A}$ depends on the particular experimental design. Hence, if a particular experimental design is described by $\mathbf{S} = \{\mathbf{x}_1, \ldots, \mathbf{x}_{N_d}, q_1, \ldots, q_{N_d}\}$, where $\mathbf{x}_1, \ldots, \mathbf{x}_{N_d}$ are the spatial locations of equipment of type $q_1, \ldots, q_{N_d}$ respectively, then the matrix $\mathbf{A}$ is a function of $\mathbf{S}$. Hence, any measures of experimental quality (Φ, Φ^{G} or Φ^{F}) which depend on $\mathbf{A}$ are in turn also functions of the experimental design $\mathbf{S}$. To simplify notation I continue to write $\mathbf{A}$ rather than $\mathbf{A}(\mathbf{S})$; the $\mathbf{S}$ dependence of $\mathbf{A}$ will be understood.

2.2 The forward and linearized inverse problems

The complete set of model parameters to be estimated during the experiment and all possible combinations of their values span the model space, M, say. For any given model vector $\mathbf{m}'$ of dimension N within the complete model space M, let the forward problem of calculating the corresponding data vector $\mathbf{d}'$ in data space D be accomplished by the function $\mathbf{G}$, i.e.

$$\mathbf{d}' = \mathbf{G}(\mathbf{m}'). \tag{2}$$

Then, for a given measured data vector $\mathbf{d}_0$, we often wish to find the model vector $\mathbf{m}_0 \in M$ such that

$$O(\mathbf{d}_0, \mathbf{m}_0) = [\mathbf{d}_0 - \mathbf{G}(\mathbf{m}_0)]^2 \tag{3}$$

is minimized, where $O(\mathbf{d}_0, \mathbf{m}_0)$ is called the misfit or objective function, and measures the compatibility between model $\mathbf{m}_0$ and data $\mathbf{d}_0$. Let $\overline{\mathbf{m}}$ be the best estimate of the correct model prior to inversion, and $\overline{\mathbf{d}} = \mathbf{G}(\overline{\mathbf{m}})$ be the corresponding data. Then $\mathbf{d} = \mathbf{d}_0 - \overline{\mathbf{d}}$ is the data perturbation required to fit the recorded data $\mathbf{d}_0$. This is achieved by perturbing the prior model $\overline{\mathbf{m}}$ by a small amount, $\mathbf{m}$, say. Then from eq. (3),

$$O(\mathbf{d}_0, \mathbf{m}_0) = [\mathbf{d} + \overline{\mathbf{d}} - \mathbf{G}(\overline{\mathbf{m}} + \mathbf{m})]^2 \tag{4}$$

$$\simeq [\mathbf{d} + \overline{\mathbf{d}} - \mathbf{G}(\overline{\mathbf{m}}) - \mathbf{A}\mathbf{m}]^2 \quad (-\text{linearization step}) \tag{5}$$

where $\mathbf{A}$ is a matrix of derivatives evaluated at $\overline{\mathbf{m}}$,

$$\mathbf{A}_{ij} = \left.\frac{\partial \mathbf{G}(\mathbf{m})_i}{\partial \mathbf{m}_j}\right|_{\overline{\mathbf{m}}} \tag{6}$$

$$= [\mathbf{d} - \mathbf{A}\mathbf{m}]^2. \tag{7}$$

Minimizing expression (7) is accomplished by setting

$$\mathbf{m} = (\mathbf{A}^{\mathrm{T}}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{d}. \tag{8}$$

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

$$\mathbf{L} = \mathbf{A}^{\mathrm{T}}\mathbf{A}, \tag{9}$$

is often near-singular; that is, some of its (unit length) 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 the model solution $\mathbf{m}_0$ parallel to each eigenvector $\mathbf{e}_i$ with an amplification $1/\lambda_i$ (eq. B5); hence, when small eigenvalues exist the solution becomes unstable and unreliable, and the inverse problem is said to be ill-conditioned (Menke 1989).

In such situations an approximation to the *generalized* inverse $\mathbf{L}^{\dagger}$ is often used in place of $\mathbf{L}^{-1}$ in eq. (8). 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.

2.3 Global information Φ^{G}

Since the set of eigenvectors is orthogonal, each can be considered to be an independent piece of information which the experiment would return about the model space. The magnitude of the corresponding eigenvalue defines how certain that piece of information will be. Matrix $\mathbf{L}$ is positive indefinite;

that is, the eigenvalues are all positive or zero. Hence, to design an experiment which provides the maximum total constraint on the model space (the maximum global information), the experiment design must be varied such that the eigenvalues of matrix $\mathbf{L}$ are maximized, thereby minimizing the propagation of data errors into the model solution (minimizing model uncertainty, eq. B5) and/or minimizing the number of eigenvalues below the damping threshold δ (improving resolution, eq. B4) (Kijko 1977; Barth & Wunsch 1990; Rabinowitz & Steinberg 1990; Steinberg *et al.* 1995; Curtis & Snieder 1997; Maurer & Boerner 1998; Curtis 1998, 1999). In order to achieve this it is necessary to construct some measure of eigenvalue positivity.

Curtis (1998, 1999) suggested using a vector $\Phi^{\mathrm{G}} = [\Phi_1^{\mathrm{G}}, \ldots, \Phi_5^{\mathrm{G}}]^{\mathrm{T}}$ of global positivity measures, most of which were selected or adapted from previous literature:

$$\left.\begin{aligned} \Phi_1^{\mathrm{G}} &= \sum_{i=1}^{N} \lambda_i^n, \\ \Phi_2^{\mathrm{G}} &= \log \lambda_k, \quad \text{for pre-defined fixed index } k, \\ \Phi_3^{\mathrm{G}} &= k \quad \text{such that } \lambda_k > \delta \text{ for some pre-defined threshold } \delta, \\ \Phi_4^{\mathrm{G}} &= \sum_{i=1}^{N} \frac{-1}{\lambda_i + \delta}, \\ \Phi_5^{\mathrm{G}} &= \log |\mathbf{L}|_\delta = \sum_{i=1}^{N} \gamma_i, \quad \text{where} \quad \gamma_i = \begin{cases} \log \lambda_i & \text{if} \quad \lambda_i \geq \delta \\ \text{Penalty} & \text{if} \quad \lambda_i < \delta \end{cases}, \end{aligned}\right\} \tag{10}$$

where we regard Φ_i^{G}, $i = 1, \ldots, 5$ to be functions of the eigenvalues $\Phi_i^{\mathrm{G}}(\lambda_1, \ldots, \lambda_N)$. The references are as follows. Φ_1^{G}: Curtis & Snieder (1997), Curtis (1998, 1999) ($n = 1$); Φ_2^{G}: Barth & Wunsch (1990), Smith *et al.* (1992); Φ_3^{G}: Curtis (1998, 1999); Φ_4^{G}: Maurer & Boerner (1998); Φ_5^{G}: Wald (1943), Box & Lucas (1959), John & Draper (1975), Kijko (1977), Rabinowitz & Steinberg (1990), Steinberg *et al.* (1995), penalty term introduced in Curtis (1998, 1999). Here the N eigenvalues must be ordered according to magnitude, with λ_1 the largest, and δ is the damping threshold. In each expression, the eigenvalues λ_i may also be replaced by the *relative* eigenvalues λ_i / λ_1. This ensures that the measures are sensitive to the 'flatness' of the eigenspectrum, or the degree of information homogeneity across the eigenvectors (Curtis & Snieder 1997). Generally all of these measures increase if some (possibly relative) eigenvalues increase, but each has a different sensitivity to the eigenvalue spectrum.

2.4 General definition of the term Φ^{F}

In maximizing measures of the eigenvalue magnitudes without using information in the eigenvectors, all previous geophysical experiment design work based on the eigenspectrum has maximized some measure of the global information on the model space, since there is no consideration of *where* in model space the experiment will place constraints. Hence, previous work defines and maximizes the term Φ^{G} in eq. (1).

The term Φ^{F}, on the other hand, must account for where in model space the information will be focused. Let F be the model subspace where we desire most experimental information to be focused. Most information from the experiment (that is, smallest uncertainties in the model estimate) will be placed in the subspace spanned by eigenvectors for which the corresponding eigenvalues are large (Fig. 1a). Hence, to measure whether this information is in the desired model subspace F, it is necessary to assess how well these eigenvectors span F. If such eigenvectors lay entirely within F, maximum resolution of F would be achieved, and the lowest uncertainty is then obtained when the associated eigenvalues are large (Fig. 1b). Hence, the degree of focusing is quantified by measuring the component of each eigenvector which lies within F, and by weighting this component by the magnitude of the associated eigenvalue.

Let $\{\mathbf{f}_i : i = 1, \ldots, N_{\mathrm{F}}\}$ be any set of unit length basis vectors for subspace F, where each $\mathbf{f}_i$ has dimension N and $N_{\mathrm{F}} < N$. Then the component of any eigenvector $\mathbf{e}_j$ which lies in

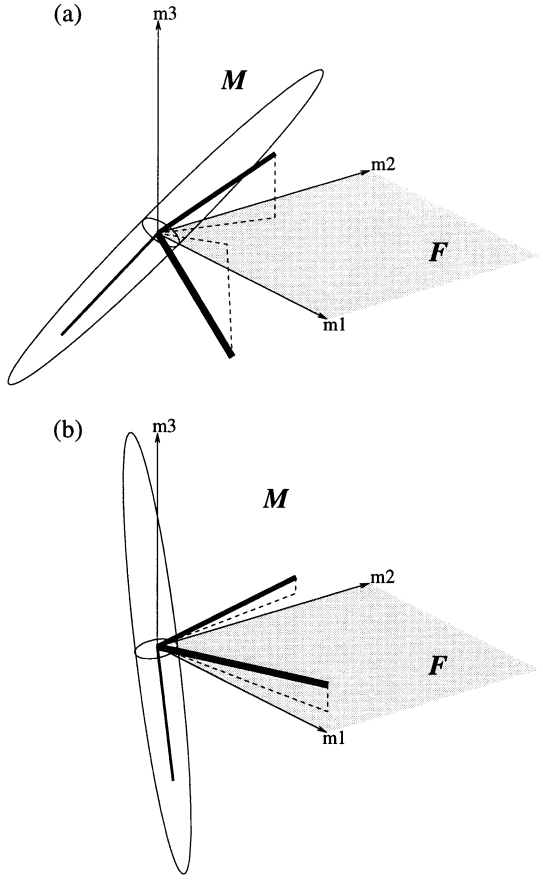

Figure 1. Focusing of information for two experiment designs (a and b). The Cartesian axes represent parameters m_1, m_2 and m_3 of a three-parameter model space $M = \{[m_1, m_2, m_3]^{\mathrm{T}}\}$. Define subspace F to be spanned by parameters m_1 and m_2 (shaded area). Eigenvectors of matrix $\mathbf{L}$ in eq. (9) are shown as solid bold unit vectors; the thickness of each represents the magnitudes of corresponding eigenvalues (thickest ~ largest eigenvalue). Since data uncertainties are magnified by 1/eigenvalue in the direction of corresponding eigenvectors, typical model uncertainty ellipses might look like those shown (thin solid lines). Clearly the projection of these uncertainties onto subspace F would be larger for design (a) than (b). Projections onto F of eigenvectors with largest eigenvalues are shown as dashed thin lines. Hence, experimental uncertainties in subspace F are minimized when eigenvectors with large eigenvalues are parallel to F; this occurs when projections of these eigenvectors onto F are maximized.

subspace F is the projection of $\mathbf{e}_j$ onto F:

$$\sqrt{\sum_{i=1}^{N_F} (\mathbf{e}_j \cdot \mathbf{f}_i)^2}\,. \tag{11}$$

Define the *focused component* $\overline{\lambda}_j$ of eigenvalue λ_j by

$$\overline{\lambda}_j^2 = \lambda_j^2 \sum_{i=1}^{N_F} (\mathbf{e}_j \cdot \mathbf{f}_i)^2\,; \tag{12}$$

that is,the original eigenvalue times the component of $\mathbf{e}_j$ that lies in F. From (12), $\overline{\lambda}_j$ is large when both λ_j and the projection (11) of $\mathbf{e}_j$ onto F is large. Hence, we can construct measures of how much information is focused in subspace F by defining

$$\Phi_i^F(\lambda_1, \ldots, \lambda_N) = \Phi_i^G(\overline{\lambda}_{O_1}, \ldots, \overline{\lambda}_{O_N})\,, \tag{13}$$

or, if relative eigenvalue components in F are of interest,

$$\Phi_i^F(\lambda_1, \ldots, \lambda_N) = \Phi_i^G\left(\frac{\overline{\lambda}_{O_1}}{\overline{\lambda}_{O_1}}, \ldots, \frac{\overline{\lambda}_{O_N}}{\overline{\lambda}_{O_1}}\right), \tag{14}$$

where the focused eigenvalue components $\{\overline{\lambda}_{O_1}, \ldots, \overline{\lambda}_{O_N}\}$ have been reordered using ordering $\{O_1, \ldots, O_N\}$ so that $\overline{\lambda}_{O_1}$ is the largest, $\overline{\lambda}_{O_N}$ the smallest. An example of such a measure using eq. (13) ($i=1$, $n=2$) is

$$\Phi_1^F = \sum_{j=1}^{N} \overline{\lambda}_{O_j}^2 = \sum_{i=1}^{N} \sum_{j=1}^{N_F} \lambda_j^2 (\mathbf{e}_j \cdot \mathbf{f}_i)^2\,. \tag{15}$$

Here the squared component of each eigenvector $\mathbf{e}_j$ inside F (given by expression 11) is weighted by λ_j^2 and summed. The weights ensure that eigenvector components in F for which the corresponding eigenvalues are large contribute most to the sums in eq. (15). A slight generalization of this example that will be useful below is

$$\Phi_1^F = \sum_{i=1}^{N} \sum_{j=1}^{N_F} \lambda_j^n (\mathbf{e}_j \cdot \mathbf{f}_i)^2\,, \tag{16}$$

in which the eigenvalue weights may be raised to any power.

The most significant terms in expressions represented by (13) (or 14) are those for which both the component of $\mathbf{e}_{O_j}$ in F and the corresponding λ_{O_j} (or $\lambda_{O_j}/\lambda_{O_1}$) are large. From here onwards, expression (15) is used in examples, but this may be replaced by any similar expression without loss of generality.

Maximizing Φ_i^F in eq. (13) for some i with respect to experimental design will provide the design for which model eigenvectors with large eigenvalues lie within subspace F. This is equivalent to designing experiments in which the most possible information is focused in F. Hence Φ_i^F defined as in eq. (13) can be used as the focusing term Φ^F in eq. (1).

It appears that to calculate Φ^F using eq. (13) one must first calculate the complete eigenvalue and eigenvector spectrum of the matrix $\mathbf{L}=\mathbf{A}^T\mathbf{A}$. However, for the example in eq. (15) when $N_F < N$, Appendix A shows how Φ^F can be calculated exactly using more efficient methods.

2.5 Normalization of Φ^G and Φ^F

To ensure that the absolute values of μ are meaningful (so that for instance $\mu=0.5$ gives an equal balance between global and focused information), it is necessary to normalize both measure Φ^G and measure Φ^F. In most cases this is not easy to achieve analytically, but may be achieved approximately by taking a number of evaluations of each of Φ^G and Φ^F over a set of random experimental designs, and dividing each measure by the mean value obtained, i.e.

$$\Phi_{\text{new}}^G = \frac{\Phi_{\text{old}}^G}{\phi_G}\,, \tag{17}$$

$$\Phi_{\text{new}}^F = \frac{\Phi_{\text{old}}^F}{\phi_F}\,, \tag{18}$$

where ϕ^G and ϕ^F are the average values of the global and focused information measures over the set of random experiments. When the means are obtained using a relatively small number of experiment designs, this option is likely to be most successful when Φ^G and Φ^F are of the same form, i.e. $\Phi=\mu\Phi_i^G+(1-\mu)\Phi_j^F$ with $i=j$. However, a pragmatic approach of checking the results of any design process by inspecting standard measures of expected resolution and uncertainty (see the example below) allows any combination of global and focused measures to be used.

2.6 Information, resolution and uncertainty

The quality of inverse problem solutions is often assessed using the standard linearized estimates of resolution and uncertainty. I now show how these are related to the class of focused measures defined by eq. (16). Both resolution and uncertainty are calculated using the matrix $\mathbf{L}=\mathbf{A}^T\mathbf{A}$ after the eigenvalues below the damping threshold (and corresponding eigenvectors) have been removed from $\mathbf{L}$, that is, as though the inverse problem solution had been constructed explicitly. Here it is assumed that this was achieved by replacing the exact matrix inverse $\mathbf{L}^{-1}$ by the generalized inverse $\mathbf{L}^\dagger$ in eq. (8).

In continuous inverse problems resolution can be described by the resolving kernels of Backus & Gilbert (1970). In discrete problems, however, the resolution matrix is defined by

$$\hat{\mathbf{m}} = \mathbf{R}\mathbf{m}_{\text{true}}\,, \tag{19}$$

where $\hat{\mathbf{m}}$ is the estimate of the model that would be retrieved from the inversion if $\mathbf{m}_{\text{true}}$ is the representation of the true model in the current parametrization (data uncertainties are not considered explicitly in the definition of resolution). Thus, if $\mathbf{R}$ is an identity matrix, the model estimate will be equal to the true model representation. As $\mathbf{R}$ becomes less like the identity, the model estimate will be an averaged version of the true model representation, the rows of $\mathbf{R}$ acting as averaging kernels. Since each row of $\mathbf{R}$ is a vector of unit length, the resolution (or smoothing) of a particular model parameter can be estimated by examining the corresponding diagonal element of $\mathbf{R}$: if this element is equal to one the parameter is perfectly resolved; if the element is less than one the parameter estimate will be averaged over more than one of the true parameters.

Uncertainty in the estimate $\hat{\mathbf{m}}$ is usually quantified by the model covariance matrix $\mathbf{C_m}$ (which takes data uncertainties into account—eq. B5). The diagonal elements of the model covariance matrix are the variances of each respective model parameter estimate.

Hence, the resolution and covariance over any subspace F spanned by a subset of the model parameters can be assessed by examining the diagonal elements of $\mathbf{R}$ and $\mathbf{C_m}$ corresponding to those model parameters. Possible measures of subspace resolution and variance are thus given by the sum of these diagonal elements.

Order the model parameters such that those which span subspace F are the first N_F parameters. Define the notation $\mathbf{e}_{j,i}$ to mean the ith component of eigenvector $\mathbf{e}_j$. Then, assuming that (uncorrelated) data variances have been normalized to unity, Appendix B shows that we can write the resolution and variance measures on subspace F as:

$$|\text{Resolution over } F| = \sum_{\substack{j=1 \\ \dagger}}^{N} \sum_{i=1}^{N_F} (\mathbf{e}_{j,i})^2 , \tag{20}$$

$$|\text{Variance over } F| = \sum_{\substack{j=1 \\ \dagger}}^{N} \sum_{i=1}^{N_F} \frac{1}{\lambda_j} (\mathbf{e}_{j,i})^2 , \tag{21}$$

$$\text{Focusing over } F = \Phi^F = \sum_{j=1}^{N} \sum_{i=1}^{N_F} \lambda_j^2 (\mathbf{e}_{j,i})^2 \tag{22}$$

(with Φ_1^F from eq. 16 with $n=2$ expressed in the current notation using orthogonal expansion A2). Here the † symbol denotes that the first sum in eqs (20) and (21) is performed only over those eigenvalues and eigenvectors that are not damped out of the inversion.

Eqs (20)–(22) illustrate that the fundamental difference between this focusing measure and the resolution and variance measures for the focusing subspace F is the power of eigenvalues λ_j that weight the eigenvector components. Thus, since the specific subset of N_F independent parameters span F and hence form a basis of F, the resolution and variance measures are just particular *damped* members of the generalized class of focused quality measures given by eq. (16), with n taking values of 0 and -1.

3 EXAMPLE: CROSS-WELL TOMOGRAPHIC SURVEY DESIGN

There now follows an example application of focused experimental design in which optimal cross-well tomographic arrays are designed. The global information measure used here is $\Phi^G = \Phi_4^G$ (eq. 10), which is always negative and increases as the combined set of eigenvalues λ_i increases (Maurer & Boerner 1998). The damping factor δ ensures that the measure is not sensitive to eigenvalues much smaller than some pre-specified δ, and in the following example $\delta = 0.0001 \times \lambda_1$, where λ_1 is the largest eigenvalue. Measure Φ^F will be defined as in eqs (15) and computed using the efficient form (A9). Optimal designs are found by maximizing the total quality measure Φ in eq. (1) with respect to experiment design (in these examples, with respect to source/receiver geometries), after normalization of Φ^G and Φ^F using eqs (17) and (18), and with various values of the weighting parameter μ in eq. (1).

3.1 Problem formulation

In this example, it is assumed that six seismic sources and six geophones (receivers) are available and may be located anywhere within either of two neighbouring vertical boreholes or on the ground surface between the boreholes (Fig. 2). The inter-borehole earth model contains a subvolume (a hypothetical hydrocarbon *reservoir*, say) within which we are particularly interested in focusing information (spanned by crosses in Fig. 2). Three cases will be considered: $\mu = 0.9$, which requires a low degree of focusing, $\mu = 0.1$, which requires a high degree of focusing on the subvolume, and $\mu = 0.5$, a compromise between these two extremes.

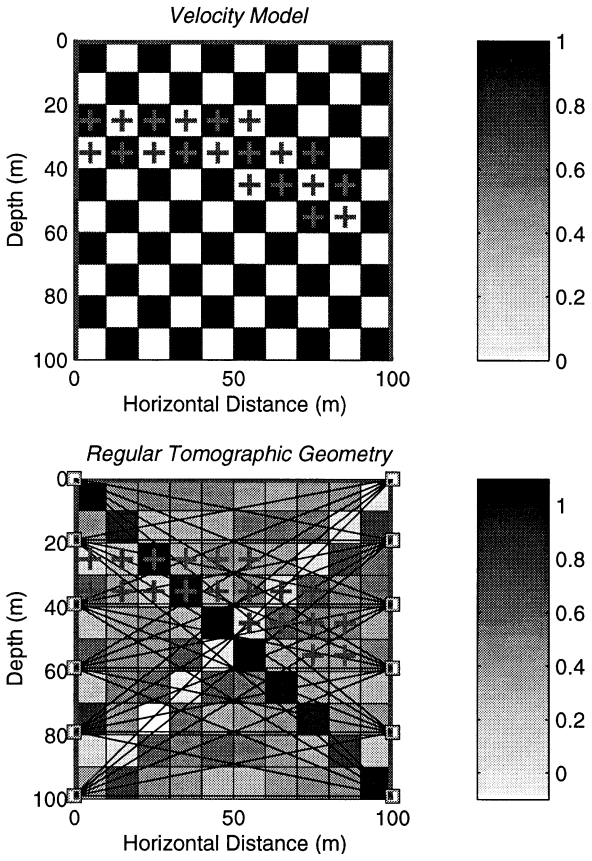

Figure 2. Experiment geometry, velocity model and performance of regular experimental design. The upper plot shows the experiment geometry: grey boundaries—two vertical boreholes and ground surface; large squares—cells of slowness model parametrization; crosses—cells forming model subspace of particular interest (e.g. a hydrocarbon reservoir); greyscale—velocity perturbations of 'true' synthetic model (dimensionless). Lower plot shows the regular cross-well array design: small squares in boreholes—sources/receivers; black lines—ray paths between sources/receivers; greyscale—velocity perturbations of model recovered from inversion of synthetic data (created for the given ray paths propagated through the chequerboard model in the top plot, and damped with threshold $\delta = 10^{-4} \times \lambda_1$, where λ_1 is the largest eigenvalue).

3.2 Linearized physical relationships

In traveltime tomography the traveltimes of waves travelling in rays through the Earth are measured and inverted for the Earth's slowness structure (where slowness is 1/velocity). An experimental design $\mathbf{S} = \{\mathbf{x}_1, \ldots, \mathbf{x}_{N_d}, q_1, \ldots, q_{N_d}\}$ consists of locations $\mathbf{x}_i$ of seismic sources or geophones, with either source or geophone specified by the corresponding q_i. The experimental design $\mathbf{S}$ determines which portions of the earth model will be traversed by rays, and hence which parts may be constrained by the traveltime data. The data vector $\mathbf{d}$ will consist of deviations of these traveltimes from those ($\bar{\mathbf{d}}$ say) which would be expected if the waves had travelled through some reference earth slowness model $\bar{\mathbf{m}}$ (usually $\bar{\mathbf{m}}$ is the best estimate of the slowness structure prior to the inversion).

The slowness model $\mathbf{m}$ for which we invert is a vector of slowness deviations from $\overline{\mathbf{m}}$ which provide the best fit to data deviations $\mathbf{d}$.

In the current example, the reference velocity model $\overline{\mathbf{m}}$ is taken to be constant (so that ray paths are straight), and only the traveltime of the direct wave is measured as data. Similarly to the study of Curtis (1998, 1999), the earth model is subdivided into a set of N mutually exclusive cells. The standard linearized forward problem is then

$$\mathbf{d}_i = \sum_{j=1}^{N} l_{ij}\mathbf{m}_j \tag{23}$$

or

$$\mathbf{d} = \mathbf{A}\mathbf{m}, \tag{24}$$

where $A_{ij} = l_{ij}(\mathbf{S}, \overline{\mathbf{m}})$ is the length of ray path i in cell j. This defines the matrix A in eqs (7) and (8), and consequently $L = \mathbf{A}^{\mathrm{T}}\mathbf{A}$ from eq. (9).

3.3 Optimization

Basis vectors $\{\mathbf{f}_i : i = 1, \ldots, N_{\mathrm{F}}\}$ for the reservoir subspace F are defined to be unit vectors pointing at each subspace cell in turn, similar to eqs (A7) and (A8). Thus, Φ^{F} can be calculated using the highly efficient formulation in eq. (A9).

Notice that although eq. (23) is linear with respect to model parameters, $l_{ij}(\mathbf{S}, \overline{\mathbf{m}})$ may be highly non-linear with respect to the experimental design $\mathbf{S}$; hence so may $\mathbf{A}_{ij}$ by eq. (24), $\mathbf{L}$ by eq. (9) and Φ^{G} and Φ^{F} by eqs (10) and (13) respectively. Hence, maximization of quality measure Φ with respect to design $\mathbf{S}$ is potentially a highly non-linear task. In this example a genetic algorithm (GA—Holland 1975; Sambridge & Drijkoningen 1992) was used to perform this maximization by minimizing $-\Phi$ with similar parameters to Curtis (1999). Although it is possible (even probable) that the GA does not find the globally optimum value of Φ, *any* increase in Φ indicates an improved experimental design. Hence, any non-linear minimization algorithm will do, including manual repositioning of sources and receivers by trial and error!

3.4 Results

In this synthetic example, the velocity anomalies which will represent the 'true' (slowness) perturbations to the reference model $\overline{\mathbf{m}}$ are shown at the top of Fig. 2. The chequerboard pattern of anomalies is one of the most difficult models to recover by ray-based tomography; every cell boundary has a contrast across it which must be resolved using synthetic traveltimes which are weighted averages across neighbouring cells traversed by the ray paths (eq. 23). Crosses mark the model cells covering the reservoir volume in which we are particularly interested.

The lower plot shows the usual, regularly spaced cross-well geometry of sources and receivers (small squares down each borehole) and source–receiver ray paths (black lines). Since only the ray paths are important to the physics above, sources and receivers are not graphically discriminated and both are henceforth called *nodes*.

Synthetic traveltime data are created by integrating the traveltime along each ray path in the 'true' chequerboard model in the upper plot using eq. (23). These data are then inverted using eq. (8) and the generalized inverse approach described earlier. No noise was added to the data. The resulting velocity model (1/slowness) is represented by the grey shades in the lower figure. Clearly the chequerboard model has not been recovered. Indeed, the only recognizable feature of the original model is the cross-like structure down the diagonals.

Such poor performance was expected. The slowness model $\mathbf{m}$ consists of 100 independent parameters (cells) to be constrained, whereas six sources and six receivers provide only 36 independent pieces of traveltime information. However, if the experiment was designed appropriately it might be possible to resolve the slowness structure across the 20 cells of particular interest, although with so few data this is still a difficult problem.

The design optimization algorithm was run using a weighting factor $\mu = 0.9$ in eq. (1), that is, with focusing contributing only 10 per cent to the quality measure Φ. Fig. 3 shows the final results. Top left is the best experiment design found by the genetic algorithm. Two aspects are immediately clear: first, ray paths are concentrated around the focusing subvolume, and second, within that subvolume the chequerboard is recovered better than was possible using the regular cross-well experimental geometry.

The diagonal of the resolution matrix $\mathbf{R}$ is shown top right (each diagonal element corresponds to one model parameter). As any of these values approaches 1 the velocity in that model cell will be better resolved by the inversion. The best resolution is concentrated within or in the neighbourhood of the reservoir. Similarly, the (square root of) the diagonal of the covariance matrix is plotted lower right. Note that these are calculated using unit data variances and hence have large absolute values; in the current context only their relative magnitudes are relevant. The general pattern of cells with large uncertainty approximately follows that of cells with good resolution, illustrating the trade-off that always exists between these two measures.

Log eigenvalue spectra of matrix $\mathbf{L}$ for the regular and optimized experiment designs are shown bottom left. The optimized design contains relatively more information in the largest 12 eigenvalues at the expense of information in the smallest eigenvalues. This is because the largest eigenvalues for the optimized design correspond to eigenvectors that approximately span the reservoir subspace. Hence, Φ can be maximized by concentrating more information (larger eigenvalues) in those eigenvectors at the expense of information across the rest of the model.

Similar plots are shown for the cases $\mu = 0.5$ and $\mu = 0.1$ in Figs 4 and 5 respectively. The case $\mu = 0.1$ (90 per cent focusing in the quality measure Φ) is the opposite extreme to that just discussed. Notice that in this case the area of high resolution is extremely focused within the reservoir subspace as desired, but unfortunately does not span the whole of this subspace. In addition, even though the chequerboard is well-defined within the reservoir in this noiseless situation, peaks in the uncertainty attached to this recovered structure are very much larger than in the previous case. This illustrates two points: first, it is not sufficient to design experiments based on ray path coverage alone, since in this case the coverage within the reservoir is good but resolution and covariance are not optimal (demonstrated explicitly by Curtis & Snieder 1997 and Curtis 1999), and second, it illustrates the danger of choosing a quality measure based principally on the focusing measure Φ^{F} in eq. (13): in

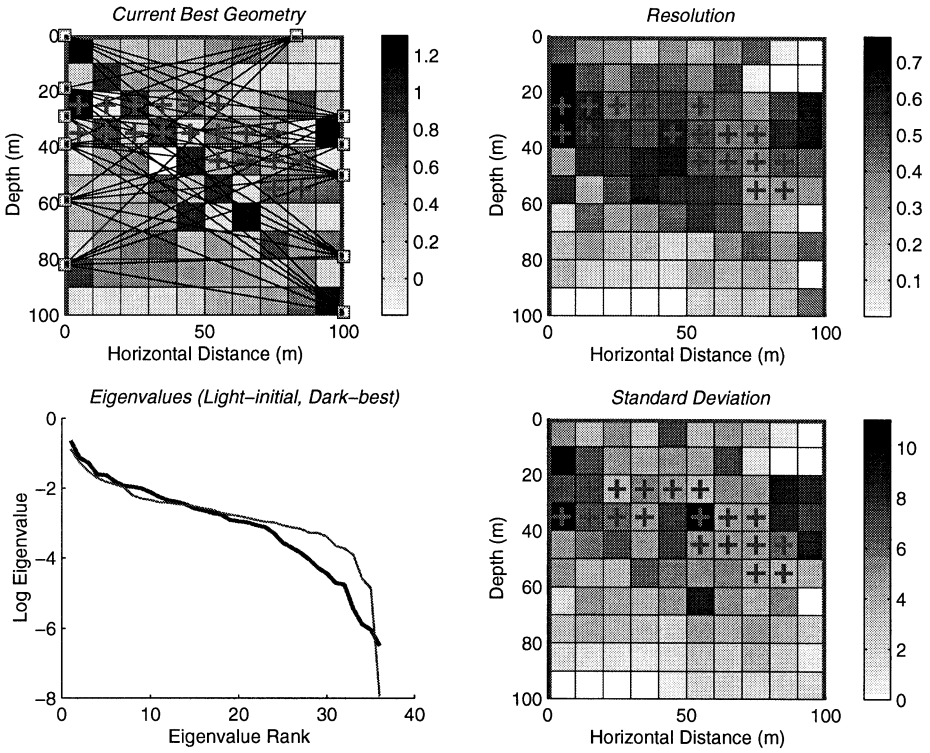

Figure 3. Top left: best experiment geometry found by the genetic algorithm using $\mu = 0.9$ in eq. (1), plus the velocity model recovered from a damped inversion using this geometry (for key see caption to Fig. 2). Right-hand plots: (top) diagonal of the resolution matrix **R**, and (bottom) the square root of diagonal elements of the covariance matrix $\mathbf{C_m}$ (i.e. model standard deviations) assuming unit data variances. Lower left plot: logarithm of eigenvalue spectra of matrix **L** for (light line) the regular experiment design in Fig. 2, and (dark line) the design shown top left in the current figure.

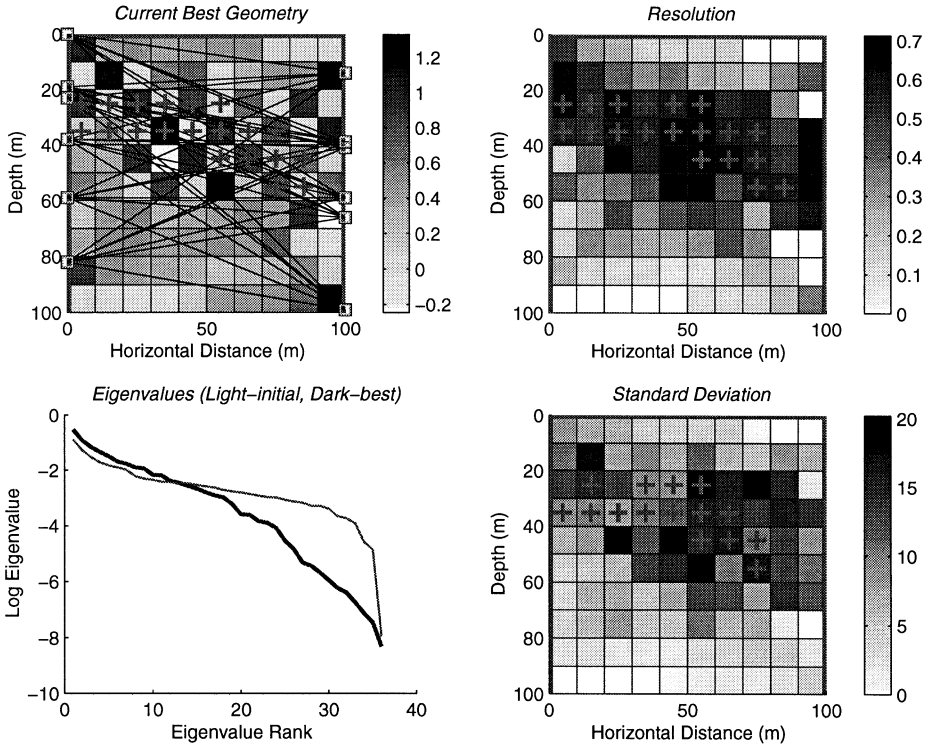

Figure 4. Similar plot to Fig. 3, but here plots relate to the best experiment design found by the genetic algorithm using $\mu = 0.5$ in eq. (1), shown top left.

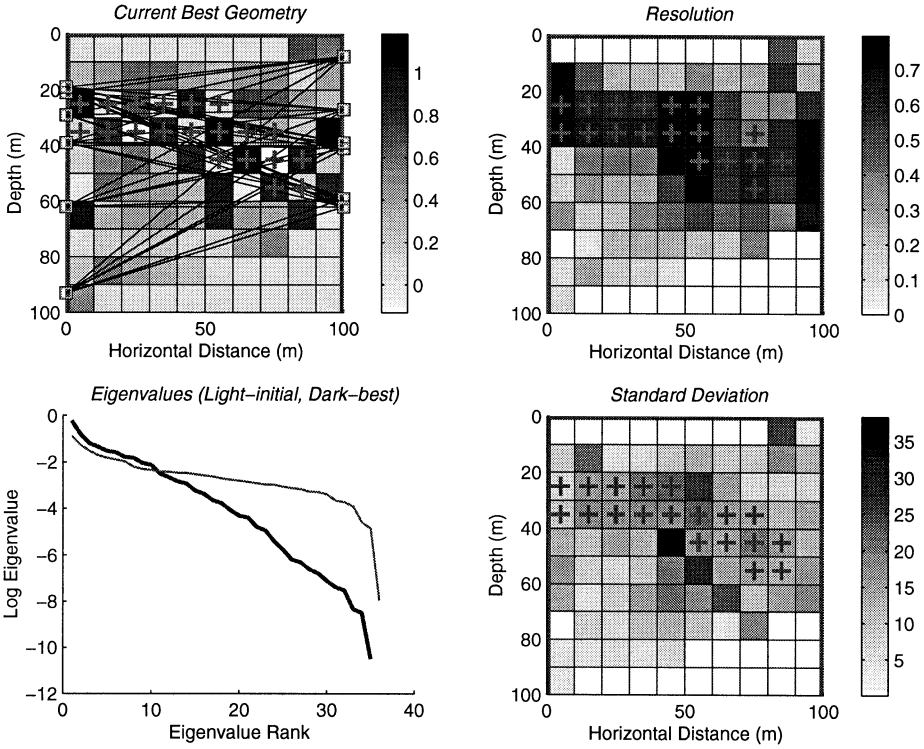

Figure 5. Similar plot to Fig. 3, but here plots relate to the best experiment design found by the genetic algorithm using $\mu = 0.1$ in eq. (1), shown top left.

some circumstances it is possible to maximize Φ^F by locating too few eigenvectors within the desired subspace F by ensuring that the corresponding eigenvalues are large. That is, there is no explicit penalty for not spanning the entire subspace F with well-constrained eigenvectors.

This second point may be circumvented by balancing the focusing measure with a global information measure of eigenvalue positivity Φ^G, as in Fig. 4 for $\mu = 0.5$. Φ^G can be tuned to ensure that many eigenvalues are increased in magnitude rather than only a few; in that case the combined quality measure Φ will be maximized when as many eigenvectors with large eigenvalues as possible lie within F. In Fig. 4 it is clear that the area of high resolution is focused on (and spans) the reservoir, the chequerboard model is recovered almost everywhere within the reservoir (with depleted amplitude in some places but with lower peak uncertainties), and the low eigenvalues have been raised relative to the case $\mu = 0.1$.

The solid line in Fig. 6 shows explicitly the magnitudes of the components of the first few eigenvectors that lie within the reservoir for the geometry in Fig. 4 (only a maximum of 36 eigenvectors have non-zero eigenvalues). This generally lies above the dashed line, which plots the relative magnitudes for the regularly spaced survey geometry—especially true for the first 12 or so eigenvectors for which the corresponding eigenvalues are large. Thus, we confirm that the largest eigenvalues in the redesigned survey correspond to eigenvectors that lie as closely as possible within the target region of the model.

If the resolution and accuracy afforded by the experiment design in Fig. 4 are not sufficient to achieve the goals of the experiment, there are three possibilities: (1) normalization of Φ^F and Φ^G is approximate and may be in error, so other values of μ should be tried and assessed using the resolution and covariance measures; (2) more information must be obtained by adding more sources/receivers to the experimental design $\mathbf{S}$; (3) the experiment must be performed to obtain an updated earth model $\overline{\mathbf{m}} + \mathbf{m}$; using this updated model as the reference model, a second experiment can then be designed to obtain the extra information required.

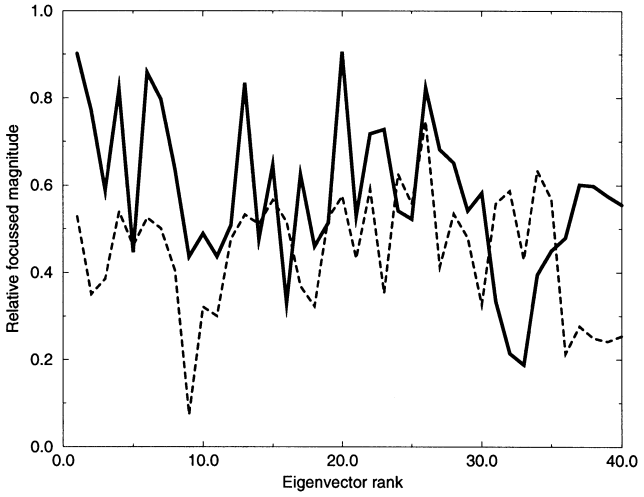

Figure 6. Relative magnitudes of the components of each of the first 40 eigenvectors that lie within the reservoir for the case $\mu = 0.5$ in eq. (1). The dashed line is for the regular geometry, the solid line is for the geometry in Fig. 4.

4 DISCUSSION

The principal contribution presented herein is a method to allow experiments to be designed such that information is focused on any model subspace. A large body of literature on this topic already exists in the field of statistics (see e.g. Atkinson & Donev 1992, Pukelsheim 1993 and references therein). However, these techniques require the *reduced information matrix* P to be calculated as follows: let L be subdivided into blocks

$$L=\begin{bmatrix} L(1,1) & L(1,2) \\ L(1,2)^{\mathrm{T}} & L(2,2) \end{bmatrix}, \tag{25}$$

where $L(1,1)$ is the block corresponding to only the model parameters spanning the subspace of interest F. Then,

$$P=L(1,1)-L(1,2)L(2,2)^{-1}L(1,2)^{\mathrm{T}}, \tag{26}$$

which requires the inverse of $L(2,2)$ to be calculated. Block $L(2,2)$ corresponds to model parameters outside subspace F and may be large if $N_{\mathrm{F}} \ll N$. Hence, this inversion may be very expensive computationally (important since this limits the number of designs that can be analysed). In addition, dependence of these parameters on the data (and hence on elements of L) may be small, so L may be ill-conditioned or singular in many geophysical situations. Since inversion of $L(2,2)$ is not required in the current work, it constitutes an extension of practical significance to these techniques.

Although the examples above include only cross-well tomography, they illustrate by analogy the concept of optimal focused design in several other types of experiment [e.g. vertical seismic profiling and experiments with multiple wells—see the discussions of Curtis & Snieder (1997) and Curtis (1999)]. Of course, the techniques presented in the methodology are not application-specific and need not be applied to seismic data at all. Optimal locations for recording other types of geophysical data such as resistivity, electromagnetic and gravity data may be determined in a similar manner to that above, provided the physics can be linearized robustly in each case. As shown by Curtis (1998, 1999), the current techniques give survey designs that agree with human intuition in simple situations. However, they also provide quantitative information that intuition cannot provide, and are particularly powerful in situations where human intuition fails entirely.

Note that experiments can also be focused on combinations of parameters. As an example of how this might be useful, consider designing an experiment to invert for bulk modulus κ in the Earth, given P- and S-wave traveltimes plus gravity data. Assuming isotropy, seismic velocities depend on density ρ and on Lame's parameters λ and μ (using the usual notation—λ is *not* an eigenvalue!). P velocity V_P, S velocity V_S and the vertical gravity G_z at a point $\mathbf{m}_0$ are given by

$$V_P(\mathbf{m}_0)=\sqrt{\frac{\lambda(\mathbf{m}_0)+2\mu(\mathbf{m}_0)}{\rho(\mathbf{m}_0)}}, \tag{27}$$

$$V_S(\mathbf{m}_0)=\sqrt{\frac{\mu(\mathbf{m}_0)}{\rho(\mathbf{m}_0)}}, \tag{28}$$

$$G_z(\mathbf{m}_0)=\int_x\int_y\int_z G\rho(\mathbf{m})\,\frac{(\mathbf{m}-\mathbf{m}_0)_z}{|\mathbf{m}-\mathbf{m}_0|^3}\,d\mathbf{m}, \tag{29}$$

where $G\simeq 6.672\times10^{-11}$ N m^2 kg^{-2} is the universal gravitational constant, $\mathbf{m}=[x,\,y,\,z]^{\mathrm{T}}$, $(\mathbf{m}-\mathbf{m}_0)_z$ is the z-component of $(\mathbf{m}-\mathbf{m}_0)$, and the integration in eq. (29) is over all volume within which matter exists.

An approximate inverse problem for the spatial distribution of λ, μ and ρ can be constructed and linearized by parametrizing the region of interest using cells, as in the cross-well examples above (corrections are required to remove the effects of matter outside the model volume on the gravity data). In this instance, however, cell i has a vector of three parameters associated with it, $[\lambda,\,\mu,\,\rho]_i^{\mathrm{T}}$. Hence, condition measures Φ^{G} and Φ^{F} can be calculated once the subspace F and its basis vectors have been specified.

If we were interested in focusing the experiment design on a 'reservoir' subvolume, the procedure would be almost identical to the cross-well example above. However, we may be interested in focusing on a specific combination of these parameters, for instance the bulk modulus $\kappa=\lambda+2\mu/3$ in each reservoir cell. This is equally easy: define the unit basis vectors to span the subspace F formed by $\lambda+2\mu/3$ in the reservoir cells:

$$\mathbf{f}_i=[0,\,\ldots,\,0,\,3,\,0,\,\ldots,\,0,\,2,\,0,\,\ldots,\,0]/\sqrt{13}, \tag{30}$$

where 3 appears in the position of parameter λ of a reservoir cell and 2 appears in the position of the corresponding parameter μ. Thus the vectors $\{\mathbf{f}_i\}$ span the required subspace F, and Φ^{F} can be used to measure (and maximize) information focused on the bulk modulus of the reservoir cells.

The only restriction on using these techniques is that the physical relationships between model and data must be linearizable. Indeed, they are most applicable to problems which are truly linear rather than those which must be linearized. Many signal processing algorithms require a linear system $\mathbf{Ax}=\mathbf{b}$ to be solved for vector $\mathbf{x}$ given some vector $\mathbf{b}$. The techniques described herein allow the matrix operator $\mathbf{A}$ to be designed such that the most information possible is transferred from vector $\mathbf{b}$ to vector $\mathbf{x}$, and this can be done in such a way that certain elements of $\mathbf{x}$ are favoured and receive more information than others. Thus, the scope of the current methodology is defined in its most generic sense.

ACKNOWLEDGMENTS

Many thanks are extended to Carl Spencer for his insight in frequent discussions, and to Sarah Ryan and Colin Sayers for their encouragement and support throughout. Excellent suggestions from David Steinberg and an anonymous reviewer are appreciated.

REFERENCES

Aki, K. & Richards, P.G., 1980. *Quantitative Seismology, Theory and Methods,* Vol. 1, W. H. Freeman.

Atkinson, A.C. & Donev, A.N., 1992. *Optimum Experimental Designs,* Clarendon Press, Oxford.

Backus, G.E. & Gilbert, F., 1970. Uniqueness in the inversion of inaccurate gross earth data, *Phil. Trans. R. Soc. Lond.,* A**266,** 123–192.

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

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

Curtis, A., 1998. Optimal cross-borehole tomographic geometries, *SEG Extended Abstracts, New Orleans,* 4.

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

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.
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., 1989. *Geophysical Data Analysis: Discrete Inverse Theory,* Revised edn, International Geophysics Series, Vol. 45, Academic Press, Harcourt Brace Jovanovich.
Özdenvar, T., McMechan, G.A. & Chaney, P., 1996. Simulation of complete seismic surveys for evaluation of experiment design and processing, *Geophysics,* **61,** 496–508.
Pukelsheim, F., 1993. *Optimal Design of Experiments,* John Wiley and Sons.
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 waveform 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: COMPUTATIONAL EFFICIENCY

In the first instance it appears that to calculate $\Phi^F = \Phi_1^F$ given by eq. (15) one must first calculate the complete eigenvalue and eigenvector spectrum of the matrix $\mathbf{L} = \mathbf{A}^T\mathbf{A}$. However, for the case when $N_F < N$, a more efficient method exists for calculating Φ^F exactly.

A1 Simplification 1 for increased computational efficiency

An eigenvalue λ_j and corresponding eigenvector $\mathbf{e}_j$ of $\mathbf{L}$ satisfy

$$\mathbf{L}\mathbf{e}_j = \lambda_j \mathbf{e}_j \tag{A1}$$

by definition,and in this study all eigenvectors are taken to be unit vectors. Since $\mathbf{L}$ is real and symmetric the set of eigenvectors $\{\mathbf{e}_j : j = 1, \ldots, N\}$ forms an orthonormal basis of the model space M. Hence, each of the subspace basis vectors $\mathbf{f}_i$ can be expressed as

$$\mathbf{f}_i = \sum_{j=1}^{N} \alpha_j \mathbf{e}_j, \quad \text{where} \quad \alpha_j = \mathbf{e}_j \cdot \mathbf{f}_i . \tag{A2}$$

Hence,

$$\mathbf{L}\mathbf{f}_i = \sum_{j=1}^{N} \alpha_j \mathbf{L}\mathbf{e}_j \tag{A3}$$

$$= \sum_{j=1}^{N} (\mathbf{e}_j \cdot \mathbf{f}_i)\lambda_j \mathbf{e}_j \tag{A4}$$

$$\Rightarrow \quad (\mathbf{L}\mathbf{f}_i)^2 = \sum_{j=1}^{N} \lambda_j^2 (\mathbf{e}_j \cdot \mathbf{f}_i)^2 . \tag{A5}$$

Therefore, Φ^F in eq. (15) for $n = 2$ is given simply by

$$\Phi^F = \sum_{i=1}^{N_F} (\mathbf{L}\mathbf{f}_i)^2 . \tag{A6}$$

Expansion (A6) requires N_F matrix multiplications and does not require the eigenspectrum to be calculated. Typically, calculating the eigenspectrum for expression (13) requires approximately $4N^3/3$ operations (multiplications or additions), which becomes computationally costly as the model space dimension N increases. The N_F matrix multiplications in expression (A6), on the other hand, require only approximately $N_F \times N^2$ operations. Thus expression (A6) gives a saving of approximately $(4N/3 - N_F)N^2$ operations per evaluation of Φ^F, which may be a large reduction if the dimension of the subspace of interest, N_F, is much less than the total number of model parameters, N.

A2 Simplification 2 for a further increase in efficiency

A further simplification is possible when F is the subspace spanned by a subset (not all) of the model parameters. Since $\{\mathbf{f}_i : i = 1, \ldots, N_F\}$ is any pre-specified set of basis vectors which span F, these may be defined however we choose. The simple choice of using unit basis vectors parallel to each of the subset of model parameter axes leads to a simple form for Φ^F.

For example, if the model space has dimension $N = 3$ and the desired subspace F is the space spanned by model parameters 1 and 3, then choose

$$\mathbf{f}_1 = [1, 0, 0]^T , \tag{A7}$$

$$\mathbf{f}_2 = [0, 0, 1]^T . \tag{A8}$$

Then $(\mathbf{L}\mathbf{f}_1)^2$ and $(\mathbf{L}\mathbf{f}_2)^2$ in eq. (A6) are equal to the squared magnitude of the first and third column (or row) vectors of $\mathbf{L}$ respectively. Hence,

$$\Phi^F = \sum_{\substack{i \\ \text{axis } i \in \text{ basis}}} [\text{Column vector } i \text{ of } \mathbf{L}]^2 , \tag{A9}$$

where the sum is over all i such that the ith model parameter axis is a basis vector of F. Expression (A9) requires only $\sim N$ operations, giving a massive computational saving (of the order of N^3 operations) over expression (15) when used directly.

APPENDIX B: LINEARIZED RESOLUTION AND UNCERTAINTY

The discrete resolution matrix is defined by

$$\hat{\mathbf{m}} = \mathbf{R}\mathbf{m}_{\text{true}} \tag{B1}$$

$$= [A^T A]^{\dagger} [A^T A]\, \mathbf{m}_{\text{true}} \tag{B2}$$

using the generalized inverse in eq. (8), where $\hat{\mathbf{m}}$ is the estimate of the model retrieved from the inversion and $\mathbf{m}_{\text{true}}$ is the representation of the true model in the current parametrization. Model uncertainty may be quantified by the model covariance matrix

$$\mathbf{C}_{\mathbf{m}} = \left[A^T \mathbf{C}_{\mathbf{d}}^{-1} A\right]^{-1} \tag{B3}$$

for the estimate $\hat{\mathbf{m}}$, where $\mathbf{C}_{\mathbf{d}}$ is the data covariance matrix.

Let $\mathbf{E} = [\mathbf{e}_1 \quad \mathbf{e}_2 \quad \ldots \quad \mathbf{e}_M]$ be the matrix of eigenvectors of $\mathbf{L}$ (constituting its columns), and $\mathbf{\Lambda} = \mathrm{diag}[\lambda_1, \lambda_2, \ldots, \lambda_M]$ be the diagonal matrix of corresponding eigenvalues. Assume that the measured data have equal variances σ^2 and are uncorrelated so that the data covariance matrix $\mathbf{C_d} = \sigma^2 I$, where I is the appropriate identity matrix. Standard linear inverse theory then gives expressions for the resolution and covariance matrices:

$$\mathbf{R} = \mathbf{E}\mathbf{\Lambda}^{\dagger}\mathbf{\Lambda}\mathbf{E}^{\mathrm{T}}, \tag{B4}$$

$$\mathbf{C_m} = \mathbf{E}\mathbf{\Lambda}^{\dagger}\mathbf{E}^{\mathrm{T}}\mathbf{C_d}. \tag{B5}$$

Thus the resolution matrix is just $\mathbf{E}\mathbf{E}^{\mathrm{T}}$ but with eigenvectors removed for which the corresponding eigenvalues fall below the damping threshold. The expression for model covariance now shows explicitly how the data uncertainties propagate into the model solution proportionally to the inverse of each (non-damped) eigenvalue as stated in the methodology.

In the main text, the measure of resolution and covariance over any subspace F spanned by a subset of the model parameters can be quantified by summing the diagonal elements of $\mathbf{R}$ and $\mathbf{C_m}$ corresponding to those model parameters. Summing the relevant diagonal elements of $\mathbf{R}$ and $\mathbf{C_m}$ in eqs (B4) and (B5) respectively gives expressions (20) and (21) directly (assuming data variances have been normalized to unity).