Geophys. J. Int. (2009) **178,** 145–158 doi: 10.1111/j.1365-246X.2009.04156.x

A method of fast, sequential experimental design for linearized geophysical inverse problems

Darrell A. Coles and Frank Dale Morgan

Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, 77 *Massachusetts Avenue, Cambridge* MA 02139, *USA*
E-mail: darrell.coles@gmail.com

Accepted 2009 February 12. Received 2009 February 11; in original form 2007 June 21

SUMMARY

An algorithm for linear(ized) experimental design is developed for a determinant-based design objective function. This objective function is common in design theory and is used to design experiments that minimize the model entropy, a measure of posterior model uncertainty. Of primary significance in design problems is computational expediency. Several earlier papers have focused attention on posing design objective functions and opted to use global search methods for finding the critical points of these functions, but these algorithms are too slow to be practical. The proposed technique is distinguished primarily for its computational efficiency, which derives partly from a greedy optimization approach, termed sequential design. Computational efficiency is further enhanced through formulae for updating determinants and matrix inverses without need for direct calculation. The design approach is orders of magnitude faster than a genetic algorithm applied to the same design problem. However, greedy optimization often trades global optimality for increased computational speed; the ramifications of this tradeoff are discussed. The design methodology is demonstrated on a simple, single-borehole DC electrical resistivity problem. Designed surveys are compared with random and standard surveys, both with and without prior information. All surveys were compared with respect to a 'relative quality' measure, the post-inversion model per cent rms error. The issue of design for inherently ill-posed inverse problems is considered and an approach for circumventing such problems is proposed. The design algorithm is also applied in an adaptive manner, with excellent results suggesting that smart, compact experiments can be designed in real time.

Key words: Image processing; Numerical solutions; Inverse theory; Downhole methods; Electromagnetic theory.

1 INTRODUCTION

In exploration geophysics, standardized or *ad hoc* surveys cannot guarantee optimally informative data sets; that is, data sets whose inversions assure high model accuracy or minimal model uncertainty. For an arbitrary heterogeneous target, it is reasonable to conjecture that there exists an ideal survey that optimizes the information content of the data and that standard survey geometries must almost certainly be suboptimal in comparison.

The deliberate, computational manipulation of geophysical surveys in order to create data with superior information characteristics is termed optimal experimental design (OED). Tangentially, the terms *survey* and *experiment* will be used interchangeably. OED is distinguished principally by the fact that it treats design as a computational problem, and its prime objective is to create compact sets of observations that produce superior data quality. Thus, experimental design seeks to maximize the connection between the data collected and the model from which they derive.

Geophysical constitutive equations are (typically) non-linear functions that can be cast in terms of a forward operator g that maps a model m to data d by means of an observation at ω, viz.,

$$d = g(m, \omega). \tag{1}$$

A set of observations,

$$\Omega = \{\omega_1, \omega_2, \ldots, \omega_N\}, \tag{2}$$

constitutes an *experiment*, which governs collection of the data set; the set of all experiments is *experiment space*; the set of all data corresponding to individual observations is *data space*; and the set of all earth models is *model space*. Each experiment can be quantitatively ascribed a 'quality metric' indicating how strongly it *informs* the connection between data and model space, and the object of the designer is to search experiment space for the highest quality mapping between data and model space.

This article examines linear or linearized experimental design, which proceeds from the linearization of eq. (1):

$$\Delta \mathbf{d} = \mathbf{G}\Delta \mathbf{m}, \tag{3}$$

where $\mathbf{G}$ is the sensitivity (Jacobian) matrix, comprising the $M \times D$ partials of $g(\mathbf{m}, \omega_i)$ with respect to the discretized model parameters m_j:

$$G_{ij} \equiv \frac{\partial g\,(\mathbf{m}, \omega_i)}{\partial m_j} = \begin{bmatrix} \mathbf{g}\,(\mathbf{m}, \omega_1)^T \\ \vdots \\ \mathbf{g}\,(\mathbf{m}, \omega_i)^T \\ \vdots \end{bmatrix}. \tag{4}$$

The Jacobian and its inverse, $\mathbf{G}$ and $\mathbf{G}^*$, respectively, are the objects upon which linearized experimental design typically operates (e.g. Atkinson *et al.* 2007). Curtis & Spencer (1999) and others have pointed out that because eq. (3) is a linear approximation, it is valid only in the neighbourhood of the Taylor expansion point in model space, implying that $\mathbf{G}$ is only technically valid in this neighbourhood. If $\mathbf{G}$ depends on $\mathbf{m}$, and the design objective function operates on $\mathbf{G}$, then the experimental design depends on $\mathbf{m}$. For non-linear inverse problems, an experimental design should ideally be optimal with respect to the true solution and with respect to all points in model space through which the inversion passes on its trajectory to the true solution. Approaches to the fully non-linear design problem can be found in the literature on Bayesian experimental design methods (e.g. Sebastiani & Wynn 2000; van den Berg *et al.* 2003) and in novel methods for reducing non-uniqueness in non-linear problems (Winterfors & Curtis 2008), but this article does not delve into these. Instead, linearized design theory is adopted because its methodologies are currently more practicable, though less general, than non-linear experimental design. However, we introduce an adaptive design procedure that benefits from the expedience of linearized design, potentially circumventing the computational complexity of non-linear design.

The chief problem with geoscientific survey design is computational expense. There have been two reasons for this: first, design is a challenging combinatorial optimization problem (Haber *et al.* 2008); second, many of the proposed design objective functions are expensive to calculate. With regard to the former, Curtis (1999a), Maurer *et al.* (2000), Barth & Wunsch (1990) and others have used global search algorithms like the genetic algorithm, simulated annealing, etc. to find critical points of their design objective functions. Such algorithms are ideal for combinatorial optimization problems like experimental design, but they are too slow to be practical for many real problems. There is however an exceptional class of fast algorithms – so-called *greedy algorithms* – that forms the basis for the design methods introduced in this work. More is said about these later.

Nearly all geoscientific design objective functions operate on $\mathbf{G}$ or $\mathbf{G}^*$, as pointed out above (Barth & Wunsch 1990; Rabinowitz & Steinberg 1990; Maurer & Boerner 1998a,b; Curtis 1999a,b; Curtis & Spencer 1999; Curtis & Maurer 2000; Maurer *et al.* 2000; Stummer *et al.* 2002, 2004; Friedel 2003; van den Berg *et al.* 2003; Curtis *et al.* 2004; Furman *et al.* 2004; Narayanan *et al.* 2004; Routh *et al.* 2005; Wilkinson *et al.* 2006a; Furman *et al.* 2007). These objective functions are readily understood in terms of the eigenspectra of $\mathbf{G}$ (Curtis 1999a,b), and indeed many proposed objective functions explicitly operate on the eigenspectrum (Barth & Wunsch 1990; Curtis 1998; Maurer & Boerner 1998a; Curtis 1999b; Maurer *et al.* 2000). However, Maurer *et al.* (2000) point out that eigen-analysis takes $O(N^3)$ operations, making it computationally expensive. Even for objective functions that do not require eigen-analysis but operate indirectly on eigenvalues through the trace or determinant of $(\mathbf{G}^T\mathbf{G})^{-1}$ (Kijko 1977; Rabinowitz & Steinberg 1990; Maurer *et al.* 2000; Narayanan *et al.* 2004), the computational complexity still remains high, owing to the expense of inverting $\mathbf{G}^T\mathbf{G}$ and/or taking its determinant.

Mindful that computational expense has been a hindrance to practical OED, this paper introduces an efficient sequential design technique inspired by the works of Stummer *et al*. (2002, 2004), Wilkinson *et al*. (2006a) and Curtis *et al*. (2004). This approach is an example of greedy optimization (e.g. Papadimitriou & Steiglitz 1998), which trades global optimality for faster computation times. The well-known $\det(\mathbf{G}^T\mathbf{G})^{-1}$ design objective function is used throughout, and, critically, this function is made computationally efficient through the use of special update formulae. Additionally, a simple, adaptive design procedure is introduced that may serve as an alternative to (currently) impractical non-linear design methods and which can be employed in near real time. We demonstrate the methods on a synthetic single-borehole DC resistivity problem. While the chosen example is somewhat artificial, we emphasize that the paper is on survey design and this example does not detract from that focus.

2 METHODS

2.1 The OED objective function and update formulae

A common objective function in experimental design theory, and the one used in this article, is

$$\Theta_{\mathrm{D}} = \left|\left(\mathbf{G}^T\mathbf{G}\right)^{-1}\right|, \tag{5}$$

which is minimized over the space of possible experiments (Box & Lucas 1959; Rabinowitz & Steinberg 1990; Steinberg *et al.* 1995; Sebastiani & Wynn 2000; Narayanan *et al.* 2004). Since $(\det \mathbf{G}^T\mathbf{G})^{-1} = \det\left(\mathbf{G}^T\mathbf{G}\right)^{-1}$, the design problem can also be expressed as the maximization problem,

$$\Omega^* = \arg\max_{\Omega} |\mathbf{G}^T\mathbf{G}|, \tag{6}$$

where Ω is the space of all possible experiments. The stationary point(s) of eq. (5) have been shown to minimize the model entropy (Sebastiani & Wynn 2000; van den Berg *et al.* 2003) or the volume of model uncertainty (Narayanan *et al.* 2004), which are mathematically equivalent measures of global model uncertainty for linearized problems. Below we develop a novel approach for finding quasi-stationary points of eq. (5) without the need to directly calculate expensive determinants or matrix inverses.

2.1.1 Rank-one updates to $\det \mathbf{A}\mathbf{A}^T$ *and* $\det \mathbf{A}^T\mathbf{A}$

This section establishes update formulae for rapidly evaluating $\det \mathbf{A}\mathbf{A}^T$ or $\det \mathbf{A}^T\mathbf{A}$ when rows are added to $\mathbf{A}$. This anticipates the situation where observations are sequentially added to an existing experiment, which corresponds to adding rows to the Jacobian matrix. Proofs for all update formulae here can be found in the Appendices.

Given a block matrix $\mathbf{A}' = [\mathbf{A}^T\ \mathbf{B}^T]^T$, where $\mathbf{A} \in \mathbb{R}^{m\times n}$, $\mathbf{B} \in \mathbb{R}^{p\times n}$ and $m + p \leq n$, the determinant of $\mathbf{A}'\mathbf{A}'^T$ is given by

$$\det(\mathbf{A}'\mathbf{A}'^T) = \det(\mathbf{A}\mathbf{A}^T)\det\{\mathbf{B}[\mathbf{I} - \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{A}]\mathbf{B}^T\}. \tag{7}$$

Dividing by $\det(\mathbf{A}\mathbf{A}^T)$ and substituting $\mathbf{a}^T$ for $\mathbf{B}$ (i.e., $p = 1$) yields

$$\Theta_1 \equiv \frac{\det\left(\mathbf{A}'\mathbf{A}'^T\right)}{\det\left(\mathbf{A}\mathbf{A}^T\right)} = \mathbf{a}^T\left[\mathbf{I} - \mathbf{A}^T\left(\mathbf{A}\mathbf{A}^T\right)^{-1}\mathbf{A}\right]\mathbf{a}. \tag{8}$$

If $m + p > n$, the update for $\det \mathbf{A}'^T\mathbf{A}'$ is

$$\det(\mathbf{A}'^T\mathbf{A}') = \det(\mathbf{A}^T\mathbf{A})\det[\mathbf{I} + (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{B}^T\mathbf{B}]. \tag{9}$$

Dividing by $\det(\mathbf{A}^T\mathbf{A})$, substituting $\mathbf{a}^T$ for $\mathbf{B}$ and manipulating terms yields

$$\Theta_2 \equiv \frac{\det(\mathbf{A}'^T\mathbf{A}')}{\det(\mathbf{A}^T\mathbf{A})} = 1 + \mathbf{a}^T(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{a}. \tag{10}$$

Because $\det \mathbf{A}\mathbf{A}^T$ and $\det \mathbf{A}^T\mathbf{A}$ are assumed known and fixed, maximizing Θ_1 or $\Theta_2$with respect to $\mathbf{a}$ satisfies the design criterion in eq. (6). The utility of this fact is made clear in Section 2.3, when the design algorithm is developed. Importantly, eqs (8) and (10) involve only matrix–vector products, which are computationally more efficient than explicitly evaluating determinants and inverses.

2.1.2 The Sherman–Morrison formula

The *Sherman–Morrison formula* (SMF) is an identity relating the inverse of a matrix to the inverse of its rank-one update. Formally, the SMF states

$$(\mathbf{A} + \mathbf{u}\mathbf{v}^T)^{-1} = \mathbf{A}^{-1} - \frac{\mathbf{A}^{-1}\mathbf{u}\mathbf{v}^T\mathbf{A}^{-1}}{1 + \mathbf{v}^T\mathbf{A}^{-1}\mathbf{u}}. \tag{11}$$

The SMF provides a computationally inexpensive way to update a matrix inverse without resorting to a full inverse calculation. Let $\mathbf{A}' = [\,\mathbf{A}^T\ \mathbf{a}\,]^T \in \mathbb{R}^{m+1\times n}$ where $\mathbf{A} \in \mathbb{R}^{m\times n}$, $\mathbf{a} \in \mathbb{R}^n$ and $m + 1 \geq n$. If $\mathbf{A}^T\ \mathbf{A}$ is invertible, then

$$\begin{aligned}\left(\mathbf{A}'^T\mathbf{A}'\right)^{-1} &= (\mathbf{A}^T\mathbf{A} + \mathbf{a}\mathbf{a}^T)^{-1} = (\mathbf{A}^T\mathbf{A})^{-1} \\ &\quad - \frac{(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{a}\mathbf{a}^T(\mathbf{A}^T\mathbf{A})^{-1}}{1 + \mathbf{a}^T(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{a}}.\end{aligned} \tag{12}$$

Since $(\mathbf{A}^T\mathbf{A})^{-1}$ is assumed known, evaluating $(\mathbf{A}'^T\mathbf{A}')^{-1}$ requires only a few matrix–vector products, which is more efficient than performing the full numerical inverse.

The situation is more complicated for $\mathbf{A}'\mathbf{A}'^T$. If $\mathbf{A}\mathbf{A}^T$ is $m \times m$, then $\mathbf{A}'\mathbf{A}'^T$ is $(m + 1) \times (m + 1)$. The object is to relate the inverses of two matrices of different dimensions. It is straightforward to show that

$$\left(\mathbf{A}'\mathbf{A}'^T\right)^{-1} = \begin{pmatrix}\mathbf{A}\mathbf{A}^T & \mathbf{A}\mathbf{a} \\ \mathbf{a}^T\mathbf{A}^T & \mathbf{a}^T\mathbf{a}\end{pmatrix}^{-1} = \frac{1}{\nu}\begin{bmatrix}\nu(\mathbf{A}\mathbf{A}^T)^{-1} + \boldsymbol{\alpha}\boldsymbol{\alpha}^T & -\boldsymbol{\alpha} \\ -\boldsymbol{\alpha}^T & 1\end{bmatrix}, \tag{13}$$

where

$$\begin{aligned}\nu &= \mathbf{a}^T\left[\mathbf{I} - \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{A}\right]\mathbf{a} \\ \alpha &= (\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{A}\mathbf{a}.\end{aligned} \tag{14}$$

Eq. (13) provides a way to evaluate$(\mathbf{A}'\mathbf{A}'^T)^{-1}$ cheaply, by avoiding a full inverse computation.

Collectively, the Sherman–Morrison and determinant update formulae provide an efficient way to quickly find observations satisfying eq. (6) without explicitly evaluating inverses or determinants. These formulae will be critical for the speed of our design algorithms introduced below.

2.2 OED considerations for DC resistivity

In practice, four electrodes are needed to make any single DC resistivity measurement, two as the source dipole and two as the receiver dipole. Noel & Xu (1991) and others (Xu & Noel 1993; Daily *et al.* 2004; Stummer *et al.* 2004) have proven that, for any resistivity survey of N geoelectrodes, there are a total of

$$N\,(N-1)\,(N-2)\,(N-3)\,/8 \tag{15}$$

four-electrode observations available for experimentation, accounting for reciprocity and polarity switching. Noel & Xu (1991) have further shown that the maximum number of *independent* measurements is

$$S_N = \frac{N\,(N-3)}{2}. \tag{16}$$

That is, if one has identified a set of S_N independent resistivity observations, then any other observation can be expressed as a linear combination of these. If $N = 20$ geoelectrodes, for example, there are 14535 possible four-electrode measurements but only 170 of these are independent. This is one of the reasons why resistivity inversion is nearly always non-unique: the amount of potential information a survey can offer is far smaller than is suggested by the number of available observations.

Considering a generic linearization of the DC resistivity forward problem,

$$\mathbf{G}\Delta\mathbf{m} = \Delta\mathbf{d}, \tag{17}$$

the fact that there are at most S_N independent observations means that the *maximum attainable rank* (MAR) of $\mathbf{G}$ cannot exceed S_N:

$$\text{rank}\,(\mathbf{G}) \leq S_N = \frac{N\,(N-3)}{2}. \tag{18}$$

In other words, $\mathbf{G}$ can never have more than S_N non-zero singular values. This concept is examined in Section 2.3.1.

2.3 Basic OED algorithm

Before describing the new design algorithm, an important detail should be pointed out. The eigenvalues of $\mathbf{G}^T\mathbf{G}$ and $\mathbf{G}\mathbf{G}^T$ are identical up to zeros. Because the determinant operates on these eigenvalues (it equals their product), $\mathbf{G}^T\mathbf{G}$ and $\mathbf{G}\mathbf{G}^T$ effectively convey the same information about the quality of a survey, but in two different ways. The reason update formulae are given for inner and outer products is because in the early stages of the algorithm, there will be fewer observations than model parameters, so $\mathbf{G}\mathbf{G}^T$ will be full rank while $\mathbf{G}^T\mathbf{G}$ will not; in later stages, $\mathbf{G}^T\mathbf{G}$ will be full rank while $\mathbf{G}\mathbf{G}^T$ will not. Because their eigenvalues are the same, the trick will be to use $\mathbf{G}\mathbf{G}^T$ when the design problem is 'underdetermined' and $\mathbf{G}^T\mathbf{G}$ when it is 'overdetermined'. This avoids zero eigenvalues which could render a determinant-based objective function useless.

Geoelectrical inverse problems are almost never full rank, however, and special consideration is made to accommodate this in Section 2.3.1. Nevertheless, the purpose of this paper is to introduce a design technique useful for a wide class of problems, so the main algorithm is introduced first and specific modifications for geoelectrical problems are discussed afterward.

The following nomenclature is used in the basic OED algorithm:

$\Omega_{\rm M}$, $\mathbf{G}_{\rm M}$	The *master* set of all permitted observations and its corresponding Jacobian matrix.
$\Omega_{\rm b}$, $\mathbf{G}_{\rm b}$	The *base* set of observations in the designed survey and its corresponding Jacobian matrix.
$\mathbf{g}$	Sensitivity kernel of a single observation [see eq. (4)].
k	Iteration counter and the number of observations in $\Omega_{\rm b}$.
s	Number of observations in $\Omega_{\rm M}$.
$\mathcal{M}$	Number of model parameters.

Basic sequential OED algorithm:

1. Set $k = 1$ and find the observation in $\Omega_{\rm M}$ whose sensitivity kernel, $\mathbf{g}$, is of maximum length (note: $\det \mathbf{g}^T\mathbf{g} = \mathbf{g}^T\mathbf{g}$, so eq. 6 is satisfied). Set $\mathbf{G}_{\rm b} = \mathbf{g}^T$.
2. If $k < \mathcal{M}$, update $(\mathbf{G}_{\rm b}\mathbf{G}_{\rm b}^T)^{-1}$ using SMF; if $k = \mathcal{M}$, evaluate $(\mathbf{G}_{\rm b}^T\mathbf{G}_{\rm b})^{-1}$ explicitly; else if $k > \mathcal{M}$, update $(\mathbf{G}_{\rm b}^T\mathbf{G}_{\rm b})^{-1}$ using SMF.
3. If $k < \mathcal{M}$, use eq. (8) to evaluate the *candidacy* of observations in $\Omega_{\rm M}$ for transferal to $\Omega_{\rm b}$. Eq. (8) can be efficiently executed by

$$\Theta_1(j) = \sum_{i=1}^{\mathcal{M}} \left[\left(\mathbf{I} - \mathbf{G}_{\rm b}^T\left(\mathbf{G}_{\rm b}\mathbf{G}_{\rm b}^T\right)^{-1}\mathbf{G}_{\rm b}\right)\mathbf{G}_{\rm M}^T\right]_{ij}^2.$$

   Else if $k \geq \mathcal{M}$, use eq. (11), which can be efficiently executed by

$$\Theta_2(j) = \sum_{i=1}^{k} \left[\mathbf{G}_{\rm b}\left(\mathbf{G}_{\rm b}^T\mathbf{G}_{\rm b}\right)^{-1}\mathbf{G}_{\rm M}^T\right]_{ij}^2.$$

   Note that $j = 1, \ldots, s$.
4. Identify the winning candidate in $\Omega_{\rm M}$ that maximizes either Θ_1 or Θ_2 and transfer it to $\Omega_{\rm b}$.
5. Set $\mathbf{G}_{\rm b} \leftarrow [\mathbf{G}^T\ \mathbf{g}]^T$ where $\mathbf{g}$ is the sensitivity kernel in $\mathbf{G}_{\rm M}$ corresponding to the winning candidate observation identified in step 4.
6. Increment $k \leftarrow k + 1$.
7. Stopping criterion, e.g. if k is less than some desired number of observations, go to 2; else exit.

The algorithm described above is plainly sequential; it proceeds by identifying the best observation to add to an experiment (conditional on the observations already in the experiment), adding it, and repeating until a stopping criterion is met. This formulation allows for repeat observations in the survey design, but it can easily be modified if this is undesirable. This is an example of a greedy algorithm (e.g. Papadimitriou & Steiglitz 1998), which is a method of optimization that proceeds by making locally optimal updates to a solution given the solution currently at hand. Only under strict circumstances can greedy optimization algorithms give rise to globally optimum solutions, which is largely due to the fact that they do not exhaustively search the entire space of possible solutions. The utility of such algorithms lies in the fact that they execute far more quickly than global optimization algorithms. The question is: does the value gained in computing efficiency offset the value lost in deviation from a globally optimum experiment? We look at this question later in our synthetic examples.

2.3.1 Maximum attainable rank

Section 2.2 raised the possibility that the inverse problem is inherently ill posed, meaning the model parameters cannot be fully resolved, no matter what is the number of data observations. The MAR, denoted ζ, defines the highest rank any row-deleted submatrix of $\mathbf{G}_{\rm M}$ can attain, which is plainly

$$\zeta \equiv \text{rank}\,(\mathbf{G}_{\rm M})\,. \tag{19}$$

Special care is needed in survey design when the inverse problem is inherently ill posed, which happens if $\zeta < \mathcal{M}$. In such cases, any row-deleted submatrix, $\mathbf{G}_{\rm b}$ of $\mathbf{G}_{\rm M}$, that has more rows than the MAR must also have non-trivial left and right null spaces and hence $\det \mathbf{G}_{\rm b}^T\mathbf{G}_{\rm b} = \det \mathbf{G}_{\rm b}\mathbf{G}_{\rm b}^T = 0$. This makes Θ_1 and Θ_2 both degenerate (they vanish or become indeterminate), rendering them useless in our algorithm.

There are a number of ways of dealing with this problem; here, we focus on a form of information compression via singular value decomposition (SVD). If it is unknown *a priori* that the inverse problem is inherently ill posed, it shall become apparent when Θ_1 goes to zero for all candidate observations before the number of observations in the base experiment reaches $\mathcal{M}$. This is because Θ_1 is maximized at each design step, so $\det \mathbf{G}_{\rm b}\mathbf{G}_{\rm b}^T$ must be non-negative and vanishes only if $\zeta < \mathcal{M}$. In fact, it should be clear Θ_1 automatically vanishes for all candidate observations at iteration $\zeta + 1$, when the number of observations exceeds the MAR by one.

To circumvent this problem, we use the SVD of $\mathbf{G}_{\rm M}$,

$$\mathbf{G}_{\rm M} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T, \tag{20}$$

where $\mathbf{G}_{\rm M}$ has ζ non-zero singular values. $\mathbf{G}_{\rm M}$ can be exactly expressed by its truncated singular value decomposition,

$$\mathbf{G}_{\rm M} = \tilde{\boldsymbol{U}}\tilde{\boldsymbol{\Sigma}}\tilde{\boldsymbol{V}}^T, \tag{21}$$

such that $\tilde{\boldsymbol{\Sigma}} \in \mathbb{R}^{\zeta\times\zeta}$, $\tilde{\mathbf{U}} \in \mathbb{R}^{\zeta\times\zeta}$ and $\tilde{\mathbf{V}} \in \mathbb{R}^{\mathcal{M}\times\zeta}$. We transform $\mathbf{G}_{\rm b}$ and $\mathbf{G}_{\rm M}$ by right multiplication with $\tilde{\mathbf{V}}$ to produce

$$\begin{aligned} \mathbf{G'}_{\rm b} &= \mathbf{G}_{\rm b}\tilde{\mathbf{V}} \\ \mathbf{G'}_{\rm M} &= \mathbf{G}_{\rm M}\tilde{\mathbf{V}}, \end{aligned} \tag{22}$$

so that the number of columns in $\mathbf{G'}_{\rm b}$ and $\mathbf{G'}_{\rm M}$ is now ζ ($\zeta < \mathcal{M}$). Critically, $\tilde{\mathbf{V}}$ is an isometric projection operator – it maps points in model space to a transformed domain while preserving relative distances and angles. Due to this property, the rows of $\mathbf{G'}_{\rm b}$ and $\mathbf{G}_{\rm b}$ have the same L_2 norms, and the angles between their row vectors are also preserved. This means

$$\det \mathbf{G'}_{\rm b}\mathbf{G'}_{\rm b}^T = \prod_{i=1}^{\zeta} \sigma_i^2, \tag{23}$$

where the σ_i's are the *non-zero* singular values of $\mathbf{G}_{\rm b}$ and *all* the singular values of $\mathbf{G}_{\rm b}'$. The same applies for $\mathbf{G'}_{\rm M}$ and $\mathbf{G}_{\rm M}$. Effectively, right multiplication by $\tilde{\mathbf{V}}$ eradicates the zero singular values in $\mathbf{G}_{\rm b}$ and $\mathbf{G}_{\rm M}$**,** compressing the available information into a lower dimensional space where Θ_1 and Θ_2 are non-degenerate.

The following modification to step 3 can be made:

Sequential OED Algorithm: Modification 1 (Diagnosis)

3′. If $k < \mathcal{M}$, evaluate candidacy via

$$\Theta_1\,(j) = \sum_{i=1}^{\mathcal{M}} \left[\left(\mathbf{I} - \mathbf{G}_{\rm b}^T\left(\mathbf{G}_{\rm b}\mathbf{G}_{\rm b}^T\right)^{-1}\mathbf{G}_{\rm b}\right)\mathbf{G}_{\rm M}^T\right]_{ij}^2.$$

.

If $\Theta_1(j) = 0$ for all j, the inverse problem is ill posed. Calculate the truncated SVD of $\mathbf{G}_{\rm M}$ ($\mathbf{G}_{\rm M} = \tilde{\mathbf{U}}\tilde{\boldsymbol{\Sigma}}\tilde{\mathbf{V}}^T$); evaluate $\mathbf{G'}_{\rm b}$ and $\mathbf{G'}_{\rm M}$, vis-à-vis (22);

Update $\mathbf{G}_\mathrm{b} \leftarrow \mathbf{G'}_\mathrm{b}$ $\mathbf{G}_\mathrm{M} \leftarrow \mathbf{G'}_\mathrm{M}$ and $\mathcal{M} \leftarrow \zeta$.

Else, if $k \geq \mathcal{M}$, evaluate candidacy via

$$\Theta_2(j) = \sum_{i=1}^{k} \left[\mathbf{G}_\mathrm{b}\left(\mathbf{G}_\mathrm{b}^T\mathbf{G}_\mathrm{b}\right)^{-1}\mathbf{G}_\mathrm{M}^T\right]_{ij}^2.$$

Note that $j = 1, \ldots, s$.

If, on the other hand, it is known *a priori* that the inverse problem is inherently ill posed, it is possible to perform the compression at the beginning of the algorithm, precluding the diagnostic modification above. The following step can simply be included at the beginning of the basic algorithm:

Sequential OED algorithm: Modification 2 (Foreknowledge)

0. If the problem is inherently ill posed, determine $\mathbf{G'}_\mathrm{M}$ by eq. (22) and update $\mathbf{G}_\mathrm{M} \leftarrow \mathbf{G'}_\mathrm{M}$.

The main detriment to modifications 1 and 2 is the singular value decomposition. It is worth noting though that the SVD and associated 'matrix compressions' need to be evaluated only once. Once the compression is done, neither $\Theta_1(j)$ nor $\Theta_2(j)$ will evaluate to zero for all j because the zero singular values of $\mathbf{G}_\mathrm{M}$ have been removed. This approach 'frontloads' the additional computational expense of the SVD and matrix compression. Nonetheless, if $\mathbf{G}_\mathrm{M}$ is large, it may be infeasible to calculate its SVD. In that case, it may be possible to perform the compression through modified Gram–Schmidt orthogonalization. In fact, any truncated orthogonalization of $\mathbf{G}_\mathrm{M}$ should suffice for the compression step.

Another way to get around the problem of determinant-based survey designs for ill-posed inverse problems is to design with respect to the regularized inverse problem. Constraints usually force the inverse problem to be well posed, which would free the designer from the SVD approach proposed here. Note that if data noise is independent and identically distributed (iid), the posterior model covariance matrix is $\mathbf{C}_\mathrm{m} \equiv \sigma_\mathrm{d}^2 (\mathbf{G}^T\mathbf{G})^{-1}$. Clearly, then Θ_D (eq. (5)) is directly proportional to the determinant of $\mathbf{C}_\mathrm{m}$. For regularized inverse problems, the model covariance matrix is

$$\mathbf{C}_\mathrm{m} = \sigma_\mathrm{d}^2 \left[\mathbf{G}^T\mathbf{G} + \frac{\lambda}{1-\lambda}\mathbf{L}^T\mathbf{L}\right]^{-1} \mathbf{G}^T\mathbf{G} \left[\mathbf{G}^T\mathbf{G} + \frac{\lambda}{1-\lambda}\mathbf{L}^T\mathbf{L}\right]^{-1} \tag{24}$$

for iid data noise, where $\mathbf{L}$ is a generic regularization operator, and λ is parameter from 0 to 1 that governs the tradeoff between data rms error and the constraint imposed by $\mathbf{L}$. It is thus natural to replace Θ_D with

$$\Theta'_\mathrm{D} = \det\left[\mathbf{G}^T\mathbf{G} + \frac{\lambda}{1-\lambda}\mathbf{L}^T\mathbf{L}\right]^{-1} \mathbf{G}^T\mathbf{G} \left[\mathbf{G}^T\mathbf{G} + \frac{\lambda}{1-\lambda}\mathbf{L}^T\mathbf{L}\right]^{-1}. \tag{25}$$

The problem with this objective function – which is why it has not been adopted here – is that λ is a dynamic variable in the inversion, changing from iteration to iteration by the Levenberg–Marquardt algorithm. If λ were fixed, one could easily adapt the basic algorithm to accommodate Θ'_D, but different values of λ will mediate different survey designs. On the other hand, since this approach employs linearization, it effectively considers a single iteration of a non-linear inversion. It is therefore possible (and sometimes appropriate) to incorporate fixed λ in such cases, though we have not done so here. Haber *et al.* (2008) consider design methods for fixed λ.

2.3.2 MAR and experiment size

If the MAR is ζ, should the survey be deliberately designed to have exactly ζ design points, or is there added value in designing a larger survey?

This is answered in part by considering the well-posed case, where $\zeta = \mathcal{M}$. It is well known that the posterior model uncertainties decrease as data set size increases. The simplest form of posterior model covariance, mentioned in the last section, is $\mathbf{C}_\mathrm{m} \equiv \sigma_\mathrm{d}^2(\mathbf{G}^T\mathbf{G})^{-1}$, where σ_d^2 is the variance of iid data noise. If each observation were repeated once, creating an augmented Jacobian $\mathbf{G}' = [\mathbf{G}^T \vdots \mathbf{G}^T]^T$, the posterior model covariance would become $\mathbf{C}'_\mathrm{m} = \mathbf{C}_\mathrm{m}/2$. The posterior model variances are halved by doubling the size of the experiment. There is clearly added value in this case to design surveys with more than $\zeta = \mathcal{M}$ observations. As discussed previously, an explicit relationship exists between the posterior model covariance matrix and the design objective function in eq. (5). Minimizing eq. (5) minimizes a global measure on the covariance matrix, so there is a direct connection between the design criterion, posterior model uncertainties and survey size.

The situation is actually no different when the MAR is less than $\mathcal{M}$. If the inverse problem is inherently ill posed, model regularization is used to stabilize the inversion and to produce a unique solution. Taking as the posterior model covariance matrix the expression in eq. (25), if the experiment is again doubled, the posterior model covariance matrix is again approximately $\mathbf{C}'_\mathrm{m} \approx \mathbf{C}_\mathrm{m}/2$. The exact factor by which $\mathbf{C}_\mathrm{m}$ is divided is more complicated (and depends on λ), but the gist of the argument is unchanged. Increasing the survey size beyond the MAR reduces posterior model uncertainty, and this is true irrespective of whether the MAR is less than or equal to the number of model parameters.

2.4 Borehole resistivity modelling and inversion

The synthetic example we have chosen to test our OED method on is a simple, single-borehole DC resistivity problem. Researchers have examined various aspects of axially symmetric borehole DC resistivity modelling and inversion for more than two decades (e.g. Yang & Ward 1984; Zemanian & Anderson 1987; Liu & Shen 1991; Zhang & Xiao 2001; Spitzer & Chouteau 2003; Wang 2003). These methods model the earth with a set of rectangular prisms that are treated as being axially symmetric about the borehole. Axial symmetry imposes a strong assumption on the lithological structure of the earth, for there are only a limited number of scenarios where this symmetry attains, particularly horizontal layering. Despite the strong restriction, axially symmetric geoelectrical imaging is useful in settings where data can only be collected from a single borehole, despite the limitation that the azimuthal position of a discrete three-dimensional anomaly cannot be pinpointed because of the non-directionality of source and receiver electrodes.

The borehole problem is cast in cylindrical coordinates (Fig. 1), and the forward and inverse problems are discretized using a transmission network analogy (Swift 1971; Madden 1972; Zhang *et al.* 1995; Shi 1998) adapted to cylindrical coordinates. The inversion is executed by non-linear least squares, using an L_1-norm regularization constraint on the gradient of the model parameters. Formally, the inversion objective function is

$$\arg\min_{\mathbf{m}} \; (1-\lambda)\,\|\mathbf{d} - \mathbf{G}\mathbf{m}\|_2^2 + \lambda\,\|\nabla\mathbf{m}\|_1\,, \tag{26}$$

where $\mathbf{d}$ is the data vector, $\mathbf{G}$ is the Jacobian matrix and $\mathbf{m}$ is the model vector. The Levenberg–Marquardt algorithm (Levenberg

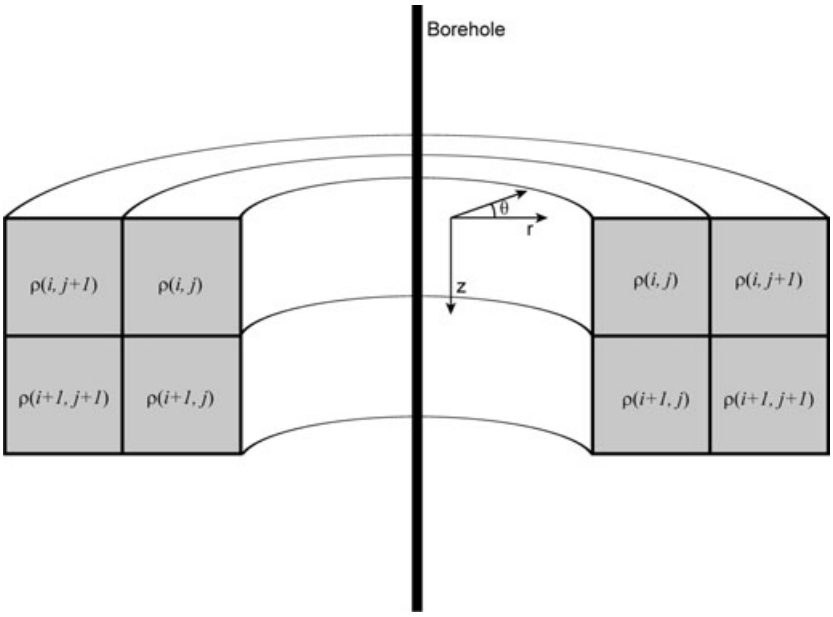

Figure 1. Cartoon of the borehole model. With a single borehole, resistivity is treated as being azimuthally invariant, reducing the problem from 3-D to 2.5-D.

1944; Marquardt 1963) is used to regulate λ. Expanded details of the inversion can be found in the Appendix.

2.5 Synthetic trials

The discretized model in Fig. 2 was used for testing the design method. The model depicts a 100-Ωm background with 20- and 500-Ωm embedded anomalies. The resistive anomaly has been deliberately placed partially in the 'outer space' region (Maurer & Friedel 2006) of the survey electrodes. It is widely recognized that the sensitivity of resistivity arrays drops off rapidly outside their 'inner space', the rectangular region adjacent to the array and bounded by the outermost electrodes. This is why, for example, dipole–dipole surveys produce geometric distortions of the inversion image at tomogram edges (Shi 1998). The model therefore simulates any situation where a survey has not, or cannot, incorporate anomalies in its inner space, or simply where no prior information was available to guide electrode placement. We have deliberately placed an anomaly straddling this 'inner/outer space' to establish whether the design method can provide any improvement there. To be clear, the purpose of the design criterion is to improve overall imaging accuracy through minimization of expected model uncertainties; it is not simply to improve image resolution in the 'outer space'. The design technique might indeed improve resolution in this region, but this should be understood as a tangential benefit mediated by the true objective of the design criterion as just stated.

Ten borehole electrodes were simulated from the surface to a depth of 9 m at equispaced 1-m intervals. With 10 electrodes, eq. (15) indicates that there are 630 possible unique four-electrode observations. A 26 × 16 irregular mesh was used (including boundary condition cells), with cell sizes increasing proportionally as their distance from the borehole array (Fig. 3), so there are 416 model parameters in the inverse problem.

Eq. (16) tells us that the MAR for 10 electrodes used in four-electrode setups is 35. Since there are 416 model parameters, the un-regularized inverse problem is ill posed. The design algorithm thus employs modification 2 in Section 2.3.1, compressing the dimensions of $\mathbf{G}_{\mathrm{M}}$ from 630 × 416 to 630 × 35. This eliminates zero singular values and ensures that the design algorithm will operate properly for experiments of more than ζ observations. This dimension reduction has the added advantage of making the design algorithm faster.

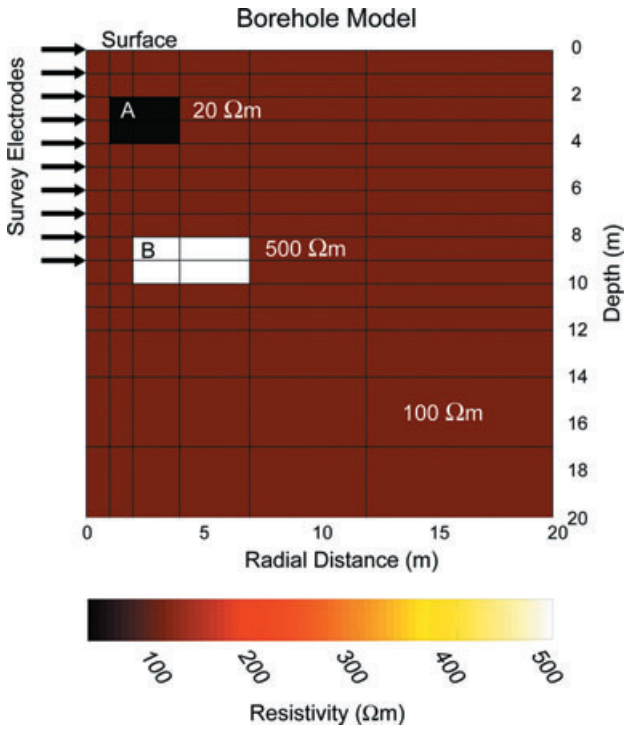

Figure 2. Resistivity model. Ten electrodes (arrows) are placed at equispaced intervals of 1 m along the borehole, from the surface to a depth of 9 m. Background resistivity is 100 Ωm and anomalies A and B are 20 and 500 Ωm, respectively. The discretization is an irregular mesh, as suggested by the grid overlay, with cell size increasing as the distance from the survey. The mesh extends beyond what is shown here to accommodate boundary blocks needed for modelling accuracy (see Fig. 3). A total of 416 cells (model parameters) are used.

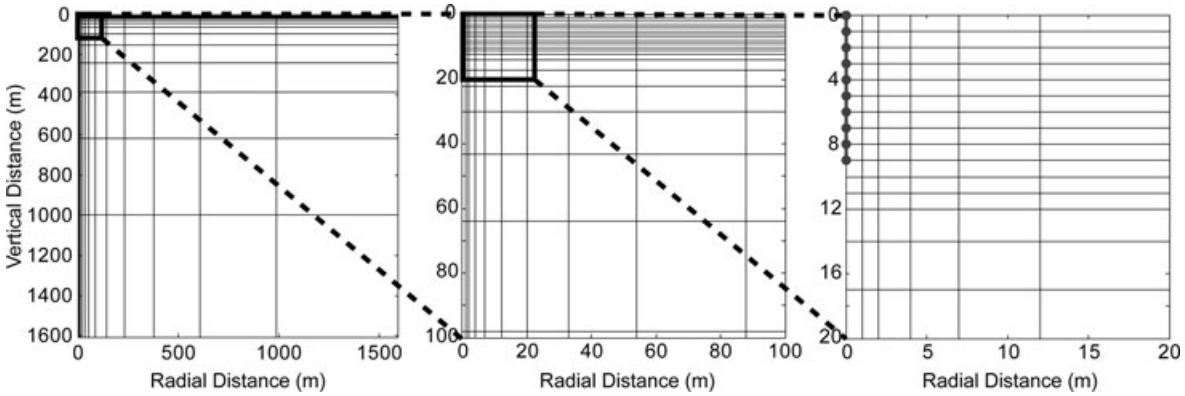

Figure 3. The 26 × 16 discretized mesh, at three different magnifications, that is used for forward and inverse modelling. The 10 borehole electrodes are shown in the right-hand panel.

2.5.1 Random, standard and designed experiments with no prior information

Designed experiments are compared with randomly generated ones in a Monte Carlo simulation. Experiments were designed for a homogeneous earth model but deployed on the heterogeneous model in Fig. 2, which is one way of designing surveys when no prior geological information is available. Survey performances were evaluated for increasing numbers of observations, from 28 to 140. For each experiment size, 25 random experiments were generated and their data were synthesized and inverted. A dipole–dipole survey of 28 observations and the ERL survey (Earth Resources Lab, M.I.T.) with 140 observations were also compared as standard surveys (see Appendix for details).

2.5.2 Two-stage design with prior information

An approach to non-linear design problems is to create adaptive survey strategies that take advantage of linearized design. We propose an adaptive design strategy that executes in two stages. In stage one, a standard (or optimized) data set is collected and inverted to produce a reference model. The first stage can take advantage of prior information if any is available, but such information is neither needed nor assumed. In stage two, a survey is tailored to the reference model and a second data set is collected and inverted. The reference model serves as prior information in stage two.

The technique is demonstrated by taking as a reference the inversion result from a dipole–dipole data set of 28 observations (stage one). In stage two, experiments of 28 and 140 observations were designed with respect to the reference model and were compared with dipole–dipole and ERL surveys, both of which were also used as second-stage surveys. Additionally, a Monte Carlo study was conducted using 100 random 28-observation surveys, which were compared as second-stage surveys with a designed experiment of the same size.

2.5.3 Greedy design

The primary contribution this paper makes to geoscientific survey design is not a new objective function – as noted earlier, determinant-based design objective functions have been around a long while – but rather a greedy optimization approach for finding stationary points of this objective function. To demonstrate the computational savings of our design methodology and also to demonstrate the disadvantages of greedy optimization, we compare it with a genetic algorithm, which produces superior experiments (according to a determinant design criterion) but which takes longer to execute.

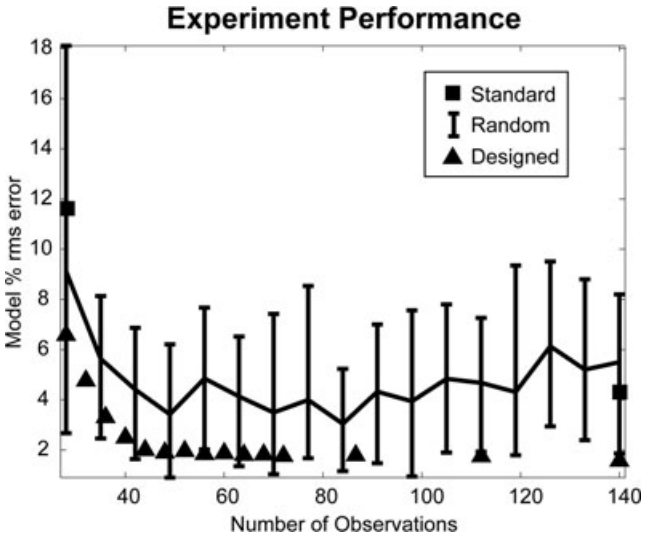

Figure 4. Experiment performances, measured by post-inversion model per cent rms error for the model in Fig. 2. Random, designed and standardized surveys are represented. The 'random' curve shows the approximate p_{25}, p_{50} and p_{75} for inversions done with 25 randomly generated surveys (25 for each experiment size). Survey designs were generated with respect to a homogeneous reference model but deployed on the model in Fig. 1. Data were noiseless.

3 RESULTS AND DISCUSSION

3.1 Random, standard and designed experiments

Fig. 4 shows Monte Carlo experiment performance curves comparing random, standard and designed experiments of different sizes. Three important features are apparent. First, designed experiments produce smaller model per cent rms errors than standard surveys. Second, designed experiments consistently produce smaller relative modelling errors than the p_{50} for random experiments and are generally as small, or smaller, than the p_{25} (meaning designed surveys are ∼75% likely to be superior to random surveys).

Third, modelling errors asymptote for designed surveys after about 45 observations. This is probably associated with the MAR, which is 35 in this case. Modelling errors naturally decrease as

experimental size increases, and the rate of decrease will be most rapid when experiments have fewer observations than the MAR, because each additional observation potentially reduces the solution null space by one dimension. However, the design performance curve does not asymptote exactly at 35 observations, which would be expected for noiseless data. This is probably because the conditioning of the Jacobian is initially poor at 35 observations, even though it has reached the MAR. Additional observations beyond this will further increase the determinant of $\mathbf{G}^T\mathbf{G}$, thereby improving the conditioning and quickly causing errors to asymptote.

While designed surveys produce smaller model per cent rms errors than most random surveys, they are evidently not superior to all such surveys. It is important to remember that the sequential OED algorithm is crafted to be expeditious, and it consequently produces experiments that are not necessarily globally optimal. This means that there is some probability that randomly generated experiments can outperform ones designed by this method. Still, the odds are in favour of the design methodology, as is clear in Fig. 4, and it is worth stressing that the designed surveys outperform both standard experiments in this demonstration. At the end of the day, globality can only be guaranteed by global search algorithms, but their computational complexity is high. We revisit this tradeoff in Section 3.3.

Another important point to keep in mind is that the surveys designed in this exercise were optimized with respect to a homogeneous earth, not the heterogeneous model in Fig. 2 on which they were deployed. It is a revelation that geoelectrical experiments designed for a homogeneous half space can produce such superior results for a heterogeneous earth. This is an indication of how much room for improvement there is in geoelectrical data acquisition.

3.2 Two-stage design with prior information

The preceding surveys were designed without prior information, but the design technique can easily be set up for adaptive experimental design, where surveys are adapted according to (prior) information that becomes available over time.

Fig. 5 shows results for two-stage design technique. With prior information this adaptive strategy performs well. The 28-observation design (panel a) reduced model per cent rms errors by over an order of magnitude compared to the initial survey. Importantly, the designed survey captured the correct shape of the 500-Ωm resistive anomaly, which was misrepresented in the original image. Compared with the 28-observation examples in Fig. 4, this small, adapted experiment performs exceptionally. Panel (b) shows the result for a 140-observation designed experiment. The model per cent rms error for this designed experiment is better than the result in panel (a) but only marginally, suggesting that the smaller experiment would practically suffice in view of financial considerations. Both designed experiments outperform the ERL survey shown at right in panel (b).

These results are encouraging: small, adapted experiments, using our sequential design methodology, can significantly improve data quality without resorting to computationally expensive non-linear design techniques. In panel (a), only 56 observations were made (28 for the dipole–dipole and 28 for the designed survey). By comparison, the ERL survey used 140 observations without producing a better image. The conclusion is plain: rather than collect a large data set, which might present a significant financial obstacle, this simple two-stage technique can be employed to produce superior imaging capabilities at a fraction of the cost.

To determine whether the 28-observation design produced a statistically significant improvement in the acquisition strategy, it was compared with 100 random experiments of 28 observations employed as second stage surveys. The histogram in Fig. 6 shows the result; the experiment designed with prior information from the reference model outperformed more than 95% of random surveys, showing that the technique produces experiments whose information content is statistically significantly superior.

3.3 Greedy design

Fig. 7 shows the CPU time for experimental designs of increasing numbers of observations. All computations were carried out on an HP laptop with dual 2 GHz processors and 2 GB RAM. The key point is that the sequential OED methodology executes in a matter of seconds. This is in contrast to traditional global search methods that must evaluate the objective function in eq. (5) thousands or millions of times. Importantly, global search methods must explicitly evaluate eq. (5) for each trial experiment, without the benefit of numeric tricks like the update formulae in Section 2.1. Consequently, global search methods require orders of magnitude more CPU time than are needed for our fast, sequential design algorithm.

The demonstration problem in this paper is small – there are only 630 possible observations and 416 model parameters, making $\mathbf{G}_\mathrm{M}$ manageably small. Larger 2.5 D geoelectrical problems can easily have tens of thousands of permitted observations and thousands of model parameters. However, this should pose little problem for this design procedure because it relies on update formulae, an approach also adopted in part by Wilkinson *et al.* (2006b). Explicit evaluation of a determinant or matrix inverse is an $O(n^3)$ operation, but these are replaced by matrix–vector multiplications, which are $O(n^2)$ operations. The computational complexity is therefore probably comparable to the sequential design procedures outlined by others (Curtis *et al.* 2004; Stummer *et al.* 2004), though it is claimed that Stummer *et al.*'s procedure is actually $O(n)$ (Wilkinson *et al.* 2006b). Moreover, that resistivity data sets are generally information-limited, as discussed in Section 2.2, means that the master Jacobian, $\mathbf{G}_\mathrm{M}$, can be usefully compressed to have only $N(N-3)/2$ columns, as per eq. (16). As noted elsewhere, this operation needs to be performed only once, the consequence of which is that the dimensions of $\mathbf{G}_\mathrm{M}$ can be losslessly reduced, making the design algorithm faster to execute. In this paper, we have used the truncated SVD to effect this compression but recognize that the SVD itself may be intractable for large problems. We have suggested that the compression operation can equally be effected through use of the modified Gram–Schmidt method or any other technique that rapidly finds an orthonormal basis spanning the range of $\mathbf{G}_\mathrm{M}$.

Fig. 8 compares the CPU time of our sequential design method with that for a genetic algorithm (a global search algorithm) maximizing the same design objective function. The same earth model in Fig. 2 was used in both design cases. The objective value ($\det \mathbf{G}^T\mathbf{G}$) of each survey design is plotted against the number of observations.

The conclusions to be drawn from Fig. 8 are clear. First, the slope of the sequential design curve is steeper than the slope for the global design curve. This means that the quality of sequentially designed surveys increases more rapidly per unit of computation time than that for globally designed surveys. Second, the disparity between the quality of sequentially and globally designed surveys decreases as the number of observations increase. In effect, the larger the experiment, the closer to global optimality comes the sequential design algorithm. Hence, the sequential design produces

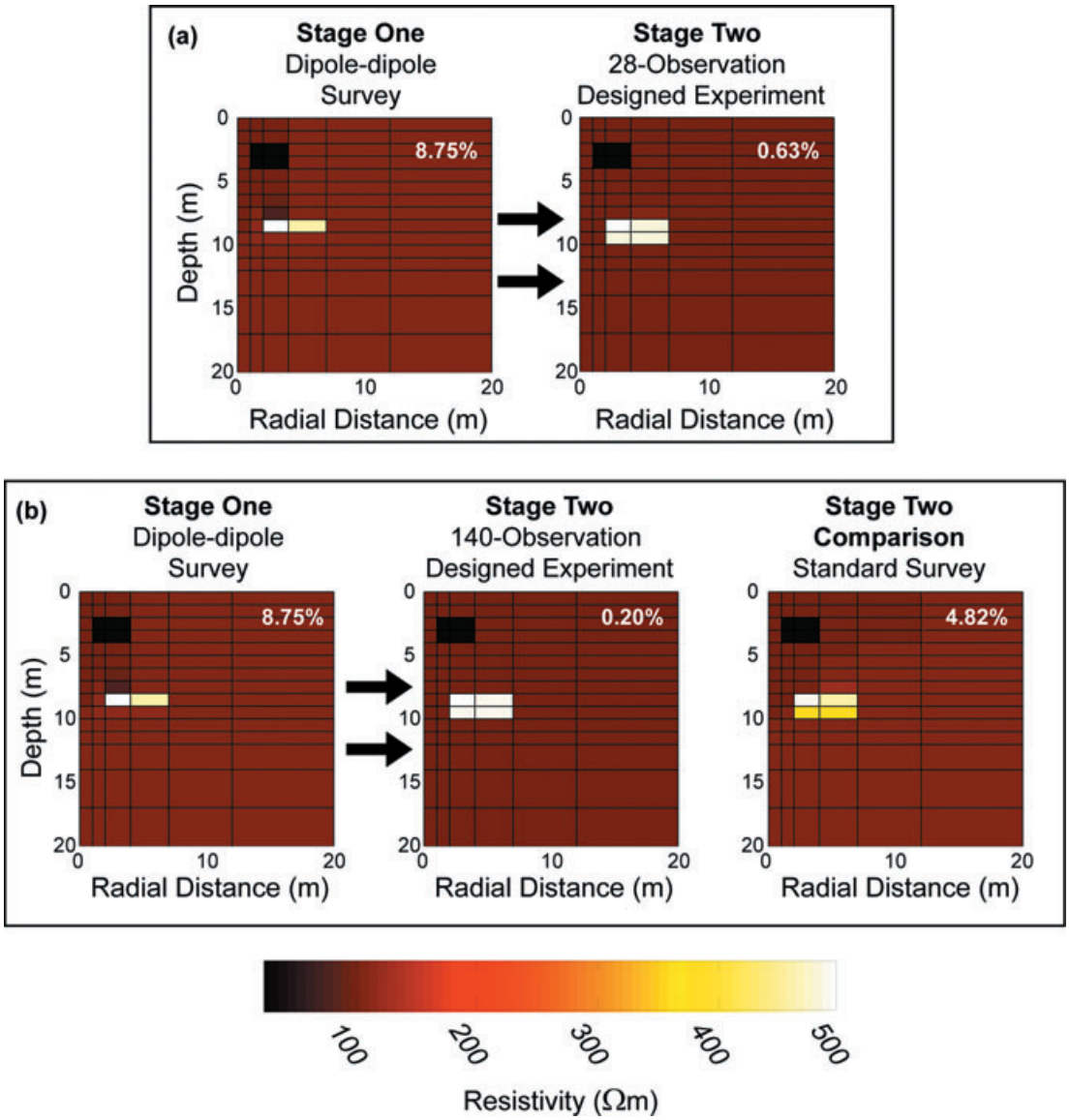

Figure 5. Results of two two-stage, adaptive designs. Post-inversion model per cent rms errors are shown in the top right corner of all panels. Data were noiseless. The stage-one survey in both examples is the 28-observation dipole–dipole survey (left-hand panels). (a) Results for a 28/28 design (28 observations for the dipole–dipole survey, 28 for the designed experiment). (b) Results for a 28/140 design (middle panel) and those for a 28/140 'standard' survey (right-hand panel; stage one: 28-observation dipole–dipole survey; stage two: 140-observation ERL survey).

suboptimal surveys that approach optimality as the size of the experiment increases and at rate significantly faster than global optimization strategies.

4 CONCLUSION

A novel algorithm for linear(ized) sequential experimental design was developed and demonstrated. The design objective function was $\det \mathbf{G}^T\mathbf{G}$, which is common to design theory. This objective function is known to be a proxy for global model uncertainty, the maximization of which is a common experimental design objective. Determinant-based design objective functions are computationally expensive to evaluate, making them impractical for real-world application. Critically, we introduced update formulae that minimize this computational expense in a greedy, sequential algorithmic framework.

Several demonstrative examples were provided, showing the utility of the design technique for a simple 2.5-D borehole example. The first of these was a design problem with no prior information wherein experiments were designed with respect to a homogeneous earth model but deployed on a heterogeneous target. Designed experiments in this investigation were shown to be statistically superior to random or standard surveys deployed on the same target.

A novel, adaptive experimental design procedure was also introduced that avails itself of prior information and which proceeds by using computationally efficient linearized design theory. Results demonstrated that this adaptive sequential design method can be

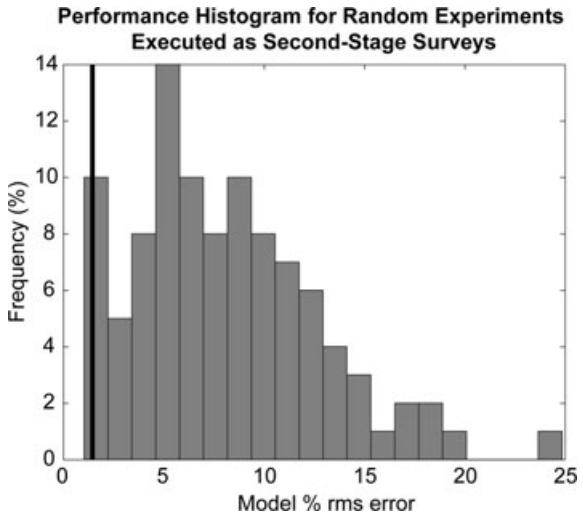

Figure 6. Histogram showing the frequency distribution of model per cent rms errors for random experiments of 28 observations executed as second-stage surveys, compared with an adapted experiment also executed as a second-stage survey (black line). The data were noiseless. The input model for all inversions was the stage-one dipole–dipole inversion in Fig. 5. The black line at left shows the model per cent rms error attained for the adaptively designed experiment of 28 observations. The model per cent rms error of the designed experiment is in the lowest 5th percentile, meaning it outperformed 95% of the random experiments.

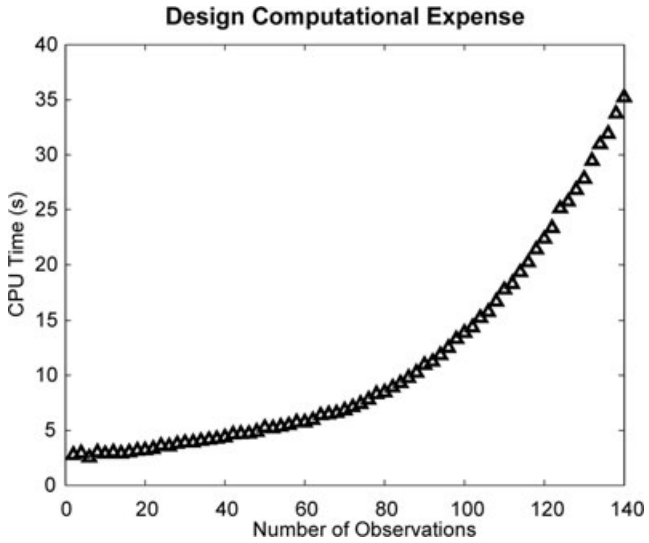

Figure 7. Computational expense for the sequential design algorithm.

usefully applied in situations where incomplete or inaccurate prior information is available. Inversions using adaptively designed surveys produced statistically significant improvements in model errors as compared with random and standard surveys, suggesting that the approach can serve as an effective alternative to computationally expensive non-linear design techniques.

Of primary significance is the development of a greedy algorithm methodology that designs surveys by sequential, local optimization of the determinant-based objective function. This methodology is further enhanced by update formulae that reduce computational complexity from $O(n^3)$ to $O(n^2)$. In contrast, it was shown that a comparable global search algorithm takes orders of magnitude more time to produce an OED. On this fact alone, it is easy to

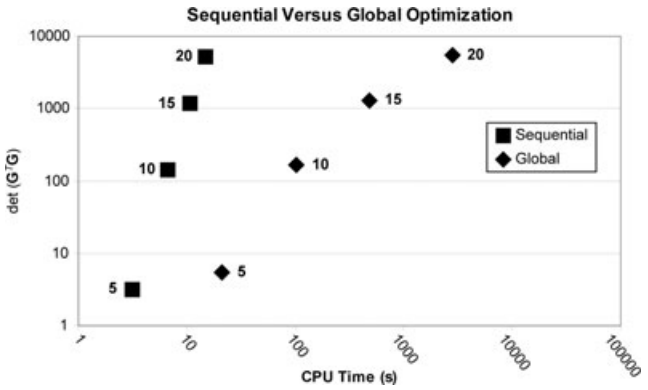

Figure 8. Sequential design versus global optimization. The design objective function, $\det \mathbf{G}^T\mathbf{G}$, is plotted against CPU time for designs produced by the sequential design algorithm (squares) and a genetic algorithm (diamonds). Numbers next to the 'Sequential' and 'Global' icons indicate the number of observations for which the experiments had been designed.

appreciate that this design technique can realize substantial time savings compared with global methods.

Lastly, it was seen that the quality of sequentially designed experiments approaches global optimality as the size of the experiment increases. Thus, the new algorithm not only requires a fraction of the computation time but it also approaches global optimality as the size of the experiment increases.

REFERENCES

Atkinson, A.C., Donev, A.N. & Tobias, R.D., 2007. *Optimum Experimental Designs, with SAS* Vol.34, Oxford University Press, Oxford.

Barth, N. & Wunsch, C., 1990. Oceanographic Experiment Design by Simulated Annealing, *J. Phys. Oceanogr.*, **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 Technical Program Expanded Abstracts,* **17,** 797–800.

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

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

Curtis, A. & Maurer, H., 2000. Optimizing the design of geophysical experiments: is it worthwhile? *Lead. Edge,* **19,** 1058–1062.

Curtis, A. & Spencer, C., 1999c. Survey design strategies for linearized nonlinear inversion, *SEG Technical Program Expanded Abstracts,* **18,** 1775–1778.

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

Daily, W., Ramirez, A., Binley, A. & LeBrecque, D., 2004. The meter reader—electrical resistance tomography, *Lead. Edge,* **23,** 438–442.

Fedorov, V.V., 1972. *Theory of Optimal Experiments,* Academic Press, New York.

Friedel, S., 2003. Resolution, stability and efficiency of resistivity tomography estimated from a generalized inverse approach, *Geophys. J. Int.,* **153,** 305–316.

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

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

Golub, G.H. & Van Loan, C.F., 1996. *Matrix Computations,* 3rd edn, Johns Hopkins University Press, Baltimore.

Haber, E., Horesh, L. & Tenorio, L., 2008. Numerical methods for experimental design of large-scale linear ill-posed inverse problems, *Inverse Problems,* **24**, 05512.

Kijko, A., 1977. Algorithm for optimum distribution of a regional seismic network .1., *Pure Appl. geophys.,* **115,** 999–1009.

Levenberg, K., 1944. A method for the solution of certain non-linear problems in least squares, *Q. Appl. Math.,* **2,** 164–168.

Liu, C. & Shen, L.C., 1991. Response of electromagnetic-pulse logging sonde in axially symmetrical formation, *IEEE T Geosci. Remote,* **29,** 214–221.

Madden, T.R., 1972. Transmission systems and network analogies to geophysical forward and inverse problems, *in ONR Technical Report* no. 72–73, Massachusetts Institute of Technology, Cambridge, MA.

Marquardt, D.W., 1963. An algorithm for least-squares estimation of nonlinear parameters, *J. Soc. Ind. Appl. Math.,* **11,** 431–441.

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

Maurer, H. & Boerner, D.E., 1998b. Optimized design of geophysical experiments, *Lead. Edge,* **17,** 1119.

Maurer, H. & Friedel, S., 2006. Outer-space sensitivities in geoelectrical tomography, *Geophysics,* **71,** G93–G96.

Maurer, H., Boerner, D.E. & Curtis, A., 2000. Design strategies for electromagnetic geophysical surveys, *Inverse Problems,* **16,** 1097–1117.

Narayanan, C., Rao, V.N.R. & Kaihatu, J.M., 2004. Model parameterization and experimental design issues in nearshore bathymetry inversion, *J. geophys. Res. Oceans,* **109**.

Noel, M. & Xu, B., 1991. Archaeological investigation by electrical resistivity tomography: a preliminary study, *Geophys. J. Int.,* **107,** 95–102.

Papadimitriou, C.H. & Steiglitz, K., 1998. *Combinatorial Optimization : Algorithms and Complexity,* Dover Publications, Mineola, NY.

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

Routh, P.S., Oldenborger, G.A. & Oldenburg, D.W., 2005. Optimal survey design using the point spread function measure of resolution, *SEG Technical Program Expanded Abstracts,* **24,** 1033–1036.

Rudin, L.I., Osher, S. & Fatemi, E., 1992. Nonlinear total variation based noise removal algorithms, *Physica D,* **60,** 259–268.

Scales, J.A. & Gersztenkorn, A., 1988. Robust methods in inverse theory, *Inverse Problems,* **4,** 1071–1091.

Sebastiani, P. & Wynn, H.P., 2000. Maximum entropy sampling and optimal Bayesian experimental design, *J. R. Stat. Soc.: Ser. B,* **62,** 145–157.

Shi, W., 1998. *Advanced Modeling and Inversion Techniques for Three-Dimensional Geoelectrical Surveys,* Massachusetts Institute of Technology, Cambridge, MA.

Spitzer, K. & Chouteau, M., 2003. A DC resistivity and IP borehole survey at the Casa Berardi gold mine in northwestern Quebec, *Geophysics,* **68,** 453–463.

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

Stummer, P., Maurer, H., Horstmeyer, H. & Green, A.G., 2002. Optimization of DC resistivity data acquisition: real-time experimental design and a new multielectrode system, *IEEE T Geosci. Remote,* **40,** 2727–2735.

Stummer, P., Maurer, H. & Green, A.G., 2004. Experimental design: electrical resistivity data sets that provide optimum subsurface information, *Geophysics,* **69,** 120–139.

Swift, C.M., 1971. Theoretical magnetotelluric and turam response from two-dimensional inhomogeneities, *Geophysics,* **36,** 38–52.

Van Den Berg, J., Curtis, A. & Trampert, J., 2003. Optimal nonlinear Bayesian experimental design: an application to amplitude versus offset experiments, *Geophys. J. Int.,* **155,** 411–421.

Wang, H.N., 2003. Simultaneous reconstruction of geometric parameter and resistivity around borehole in horizontally stratified formation from multiarray induction logging data, *IEEE T Geosci. Remote,* **41,** 81–89.

Wilkinson, P.B., Chambers, J.E., Meldrum, P.I., Ogilvy, R.D. & Caunt, S., 2006a. Optimization of array configurations and panel combinations for the detection and imaging of abandoned mineshafts using 3D cross-hole electrical resistivity tomography, *J Environ. Eng. geophys.,* **11,** 213–221.

Wilkinson, P.B., Meldrum, P.I., Chambers, J.E., Kuras, O. & Ogilvy, R.D., 2006b. Improved strategies for the automatic selection of optimized sets of electrical resistivity tomography measurement configurations, *Geophys. J. Int.,* **167,** 1119–1126.

Winterfors, E. & Curtis, A., 2008. Numerical detection and reduction of non-uniqueness in nonlinear inverse problems, *Inverse Problems*, **24**, 025016.

Wynn, H.P., 1970. The sequential generation of D-optimum experimental designs, *Ann. Math. Stat.,* **41,** 1655–1664.

Xu, B. & Noel, M., 1993. On the completeness of data sets with multi-electrode systems for electrical resistivity survey, *Geophys. Prospect.,* **41,** 791–801.

Yang, F.W. & Ward, S.H., 1984. Inversion of borehole normal resistivity logs, *Geophysics,* **49,** 1541–1548.

Zemanian, A.H. & Anderson, B., 1987. Modeling of borehole resistivity measurements using infinite electrical grids, *Geophysics,* **52,** 1525–1534.

Zhang, J., Mackie, R.L. & Madden, T.R., 1995. 3-D resistivity forward modeling and inversion using conjugate gradients, *Geophysics,* **60,** 1313–1325.

Zhang, Z. & Xiao, J., 2001. Inversions of surface and borehole data from large-loop transient electromagnetic system over a 1-D earth, *Geophysics,* **66,** 1090–1096.

APPENDIX A: DIPOLE–DIPOLE SURVEY

The dipole–dipole survey is tabulated below. A and B designate the positive and negative transmitter electrodes, respectively; M and N designate the positive and negative receiver electrodes, respectively.

	A	B	M	N
1	1	2	3	4
2	1	2	4	5
3	1	2	5	6
4	1	2	6	7
5	1	2	7	8
6	1	2	8	9
7	1	2	9	10
8	2	3	4	5
9	2	3	5	6
10	2	3	6	7
11	2	3	7	8
12	2	3	8	9
13	2	3	9	10
14	3	4	5	6
15	3	4	6	7
16	3	4	7	8
17	3	4	8	9
18	3	4	9	10
19	4	5	6	7
20	4	5	7	8
21	4	5	8	9
22	4	5	9	10
23	5	6	7	8
24	5	6	8	9
25	5	6	9	10
26	6	7	8	9
27	6	7	9	10
28	7	8	9	10

APPENDIX B: ERL SURVEY

	A	B	M	N
1	1	2	3	4
2	1	2	4	5
3	1	2	5	6
4	1	2	6	7
5	1	2	7	8
6	1	2	8	9
7	1	2	9	10
8	2	3	4	5
9	2	3	5	6
10	2	3	6	7
11	2	3	7	8
12	2	3	8	9
13	2	3	9	10
14	3	4	5	6
15	3	4	6	7
16	3	4	7	8
17	3	4	8	9
18	3	4	9	10
19	4	5	6	7
20	4	5	7	8
21	4	5	8	9
22	4	5	9	10
23	5	6	7	8
24	5	6	8	9
25	5	6	9	10
26	6	7	8	9
27	6	7	9	10
28	7	8	9	10
29	1	2	3	4
30	1	2	4	5
31	1	2	5	6
32	1	2	6	7
33	1	2	7	8
34	1	2	8	9
35	1	2	9	10
36	1	3	4	5
37	1	3	5	6
38	1	3	6	7
39	1	3	7	8
40	1	3	8	9
41	1	3	9	10
42	1	4	2	3
43	1	4	5	6
44	1	4	6	7
45	1	4	7	8

	A	B	M	N
46	1	4	8	9
47	1	4	9	10
48	1	5	2	3
49	1	5	3	4
50	1	5	6	7
51	1	5	7	8
52	1	5	8	9
53	1	5	9	10
54	1	6	2	3
55	1	6	3	4
56	1	6	4	5
57	1	6	7	8
58	1	6	8	9
59	1	6	9	10
60	1	7	2	3
61	1	7	3	4
62	1	7	4	5
63	1	7	5	6
64	1	7	8	9
65	1	7	9	10
66	1	8	2	3
67	1	8	3	4
68	1	8	4	5
69	1	8	5	6
70	1	8	6	7
71	1	8	9	10
72	1	9	2	3
73	1	9	3	4
74	1	9	4	5
75	1	9	5	6
76	1	9	6	7
77	1	9	7	8
78	1	10	2	3
79	1	10	3	4
80	1	10	4	5
81	1	10	5	6
82	1	10	6	7
83	1	10	7	8
84	1	10	8	9
85	9	10	1	2
86	9	10	2	3
87	9	10	3	4
88	9	10	4	5
89	9	10	5	6
90	9	10	6	7

	A	B	M	N
91	9	10	7	8
92	8	10	1	2
93	8	10	2	3
94	8	10	3	4
95	8	10	4	5
96	8	10	5	6
97	8	10	6	7
98	7	10	1	2
99	7	10	2	3
100	7	10	3	4
101	7	10	4	5
102	7	10	5	6
103	7	10	8	9
104	6	10	1	2
105	6	10	2	3
106	6	10	3	4
107	6	10	4	5
108	6	10	7	8
109	6	10	8	9
110	5	10	1	2
111	5	10	2	3
112	5	10	3	4
113	5	10	6	7
114	5	10	7	8
115	5	10	8	9
116	4	10	1	2
117	4	10	2	3
118	4	10	5	6
119	4	10	6	7
120	4	10	7	8
121	4	10	8	9
122	3	10	1	2
123	3	10	4	5
124	3	10	5	6
125	3	10	6	7
126	3	10	7	8
127	3	10	8	9
128	2	10	3	4
129	2	10	4	5
130	2	10	5	6
131	2	10	6	7
132	2	10	7	8
133	2	10	8	9
134	1	10	2	3
135	1	10	3	4

	A	B	M	N
136	1	10	4	5
137	1	10	5	6
138	1	10	6	7
139	1	10	7	8
140	1	10	8	9

APPENDIX C: PROOFS FOR UPDATES TO $\det \mathbf{A}\mathbf{A}^T$

Given a block matrix $\mathbf{A}' = [\mathbf{A}^T\ \mathbf{B}^T]^T \in \mathbb{R}^{(m+p)\times n}$, where $\mathbf{A} \in \mathbb{R}^{m\times n}$, $\mathbf{B} \in {}^{p\times n}$ and $m + p \leq n$, the determinant of $\mathbf{A}'\mathbf{A}'^T$ is given by

$$\det(\mathbf{A}'\mathbf{A}'^T) = \det(\mathbf{A}\mathbf{A}^T)\det\{\mathbf{B}[\mathbf{I} - \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{A}]\mathbf{B}^T\}. \tag{C1}$$

Proof: The auto-outer product of $\mathbf{A}'$is

$$\mathbf{A}'\mathbf{A}'^T = \begin{bmatrix} \mathbf{A}\mathbf{A}^T & \mathbf{A}\mathbf{B}^T \\ \mathbf{B}\mathbf{A}^T & \mathbf{B}\mathbf{B}^T \end{bmatrix}. \tag{C2}$$

Transform (C2) by the following right multiplication:

$$\begin{bmatrix} \mathbf{A}\mathbf{A}^T & \mathbf{A}\mathbf{B}^T \\ \mathbf{B}\mathbf{A}^T & \mathbf{B}\mathbf{B}^T \end{bmatrix}\begin{bmatrix} \mathbf{I} & -(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{A}\mathbf{B}^T \\ \mathbf{0} & \mathbf{I} \end{bmatrix} = \begin{bmatrix} \mathbf{A}\mathbf{A}^T & \mathbf{0} \\ \mathbf{B}\mathbf{A}^T & \mathbf{B}(\mathbf{I} - \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{A})\mathbf{B}^T \end{bmatrix}. \tag{C3}$$

Taking determinant produces

$$\det\left(\mathbf{A}'\mathbf{A}'^T\right)\det\left(\begin{bmatrix} \mathbf{I} & -(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{A}\mathbf{B}^T \\ \mathbf{0} & \mathbf{I} \end{bmatrix}\right) = \det\left(\begin{bmatrix} \mathbf{A}\mathbf{A}^T & \mathbf{0} \\ \mathbf{B}\mathbf{A}^T & \mathbf{B}(\mathbf{I} - \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{A})\mathbf{B}^T \end{bmatrix}\right), \tag{C4}$$

and taking advantage of the fact that the determinant of a block upper (or lower) triangular matrix is the product of the determinants of its diagonal blocks (Golub & Van Loan 1996), the second and third determinants in eq. (C4) reduce to

$$\det\left(\begin{bmatrix} \mathbf{I} & -(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{A}\mathbf{B}^T \\ \mathbf{0} & \mathbf{I} \end{bmatrix}\right) = 1$$

$$\det\left(\begin{bmatrix} \mathbf{A}\mathbf{A}^T & \mathbf{0} \\ \mathbf{B}\mathbf{A}^T & \mathbf{B}(\mathbf{I} - \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{A})\mathbf{B}^T \end{bmatrix}\right) = \det(\mathbf{A}\mathbf{A}^T)\det(\mathbf{B}(\mathbf{I} - \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{A})\mathbf{B}^T). \tag{C5}$$

Substituting (C5) into (C4) completes the proof.

Dividing (C1) by $\det(\mathbf{A}\mathbf{A}^T)$ and substituting the vector $\mathbf{a}^T$ for the matrix $\mathbf{B}$, we have

$$\Theta_1 \equiv \frac{\det(\mathbf{A}'\mathbf{A}'^T)}{\det(\mathbf{A}\mathbf{A}^T)} = \mathbf{a}^T(\mathbf{I} - \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{A})\mathbf{a}. \tag{C6}$$

Because $\mathbf{I} - \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{A}$ is idempotent, Θ_1 can also be expressed as

$$\Theta_1 = \|(\mathbf{I} - \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{A})\mathbf{a}\|^2. \tag{C7}$$

Tangentially, eqs (C6) and (C7) express the squared norm of the projection of $\mathbf{a}$ onto the row null space of $\mathbf{A}$ (because $\mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{A}$ is the projection matrix of $\mathbf{A}$).

APPENDIX D: PROOFS FOR UPDATES TO $\det \mathbf{A}^T\mathbf{A}$

Given a block matrix $\mathbf{A}' = [\mathbf{A}^T\ \mathbf{B}^T]^T \in \mathbb{R}^{(m+p)\times n}$, where $\mathbf{A} \in \mathbb{R}^{m\times n}$, $\mathbf{B} \in \mathbb{R}^{p\times n}$ and $m + p > n$, we can find the determinant of $\mathbf{A}'^T\mathbf{A}'$ by starting with the expression

$$\det\left(\begin{bmatrix} \mathbf{I} & \mathbf{B} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}\begin{bmatrix} \mathbf{I} & -\mathbf{B} \\ (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{B}^T & \mathbf{I} \end{bmatrix}\right) = \det\begin{bmatrix} \mathbf{I} + \mathbf{B}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{B}^T & \mathbf{0} \\ (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{B} & \mathbf{I} \end{bmatrix}, \tag{D1}$$

which expands to

$$\det\left(\begin{bmatrix} \mathbf{I} & \mathbf{B} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}\right)\det\left(\begin{bmatrix} \mathbf{I} & -\mathbf{B} \\ (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{B}^T & \mathbf{I} \end{bmatrix}\right) = \det\begin{bmatrix} \mathbf{I} + \mathbf{B}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{B}^T & \mathbf{0} \\ (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{B} & \mathbf{I} \end{bmatrix}. \tag{D2}$$

By virtue of the identity

$$\det\left(\begin{bmatrix} \mathbf{W} & \mathbf{X} \\ \mathbf{Y} & \mathbf{Z} \end{bmatrix}\right) = |\mathbf{W}|\ |\mathbf{Z} - \mathbf{Y}\mathbf{W}^{-1}\mathbf{X}|; \tag{D3}$$

it follows that

$$|\mathbf{I} + (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{B}^T\mathbf{B}| = |\mathbf{I} + \mathbf{B}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{B}^T|; \tag{D4}$$

hence,

$$|\mathbf{A}'^T\mathbf{A}'| = |\mathbf{A}^T\mathbf{A}|\ |\mathbf{I} + (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{B}^T\mathbf{B}| = |\mathbf{A}^T\mathbf{A}||\mathbf{I} + \mathbf{B}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{B}^T|. \tag{D5}$$

Defining $\mathbf{B} = \mathbf{a}^T$ yields

$$|\mathbf{A}'^T\mathbf{A}'| = |\mathbf{A}^T\mathbf{A}|\ |1 + \mathbf{a}^T(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{a}| = |\mathbf{A}^T\mathbf{A}|\ (1 + \mathbf{a}^T(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{a})\ . \tag{D6}$$

Note that

$$\mathbf{a}^T(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{a} = \mathbf{a}^T(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{a}, = \|\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{a}\|^2 \tag{D7}$$

so

$$|\mathbf{A}'^T\mathbf{A}'| = |\mathbf{A}^T\mathbf{A}|\ (1 + \|\mathbf{A}'(\mathbf{A}'^T\mathbf{A}')^{-1}\mathbf{b}\|^2). \tag{D8}$$

Dividing by $\det(\mathbf{A}^T\mathbf{A})$ yields

$$\Theta_2 \equiv \frac{\det(\mathbf{A}'^T\mathbf{A}')}{\det(\mathbf{A}^T\mathbf{A})} = 1 + \mathbf{a}^T(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{a} = 1 + \|\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{a}\|^2. \tag{D9}$$

This last expression is well-known in design theory and is used for designing so-called D-optimal experiments (e.g. Atkinson *et al.* 2007). It was independently published by Fedorov (1972) and Wynn (1970).

APPENDIX E: PROOFS FOR UPDATES TO $(\mathbf{A}\mathbf{A}^T)^{-1}$

Proofs of the SMF for updates to $(\mathbf{A}^T\mathbf{A})^{-1}$ can easily found in any good reference. However, updates to $(\mathbf{A}\mathbf{A}^T)^{-1}$ are not so straightforward, as the dimensions of $\mathbf{A}\mathbf{A}^T$ change if one or more rows are appended to $\mathbf{A}$.

Given a block matrix $\mathbf{A}' = [\mathbf{A}^T\ \mathbf{B}^T]^T \in \mathbb{R}^{(m+p)\times n}$, where $\mathbf{A} \in \mathbb{R}^{m\times n}$, $\mathbf{B} \in \mathbb{R}^{p\times n}$ and $m + p > n$, it is straightforward to algebraically derive the inverse of $\mathbf{A}'\mathbf{A}'^T$, which is

$$\left(\mathbf{A}'\mathbf{A}'^T\right)^{-1} = \begin{pmatrix} \mathbf{A}\mathbf{A}^T & \mathbf{A}\mathbf{B}^T \\ \mathbf{B}\mathbf{A}^T & \mathbf{B}\mathbf{B}^T \end{pmatrix}^{-1} = \begin{pmatrix} \left(\mathbf{A}\mathbf{A}^T\right)^{-1} + \mathbf{\Phi}\mathbf{\Gamma}^{-1}\mathbf{\Phi}^T & -\mathbf{\Phi}\mathbf{\Gamma}^{-1} \\ -\mathbf{\Gamma}^{-1}\mathbf{\Phi}^T & \mathbf{\Gamma}^{-1} \end{pmatrix}, \tag{E1}$$

where

$$\mathbf{\Phi} \equiv (\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{A}\mathbf{B}^T$$
$$\mathbf{\Gamma} \equiv \mathbf{B}[\mathbf{I} - \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{A}]\mathbf{B}^T. \tag{E2}$$

Defining $\mathbf{B} = \mathbf{a}^T$, eq. (E1) simplifies to

$$\left(\mathbf{A}'\mathbf{A}'^T\right)^{-1} = \begin{pmatrix} \mathbf{A}\mathbf{A}^T & \mathbf{A}\mathbf{a} \\ \mathbf{a}^T\mathbf{A}^T & \mathbf{a}^T\mathbf{a} \end{pmatrix}^{-1} = \frac{1}{\gamma}\begin{pmatrix} \gamma(\mathbf{A}\mathbf{A}^T)^{-1} + \boldsymbol{\varphi}\boldsymbol{\varphi}^T & -\boldsymbol{\varphi} \\ -\boldsymbol{\varphi}^T & 1 \end{pmatrix}, \tag{E3}$$

where

$$\begin{aligned} \boldsymbol{\varphi} &\equiv \left(\mathbf{A}\mathbf{A}^T\right)^{-1}\mathbf{A}\mathbf{a} \\ \boldsymbol{\gamma} &\equiv \mathbf{a}^T[\mathbf{I} - \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{A}]\mathbf{a}. \end{aligned} \tag{E4}$$

APPENDIX F: FORWARD AND INVERSE MODELLING

The model was discretized into an irregular mesh with cell size increasing as a function of distance from the borehole electrodes (Fig. 3). Resistivities were treated as azimuthally invariant, per Fig. 1. The large, distant blocks are needed to help accurately predict electrical potentials and electrode sensitivities near the borehole. Modelling was done through an adaptation of the transmission network analogy for cylindrical coordinates.

The inversion was done by non-linear least squares and is described below. The objective function combines data rmse and L_1 smoothness regularization:

$$\Phi(\mathbf{m}) = (1-\lambda)\,\|\mathbf{d} - \mathbf{G}\mathbf{m}\|_2^2 + \lambda\,\|\nabla\mathbf{m}\|_1\,, \tag{F1}$$

where $\mathbf{d}$ is the data vector, $\mathbf{G}$ is the Jacobian, $\mathbf{m}$ is the model parameter vector, $\nabla\mathbf{m}$ is the spatial gradient of the model, λ is the tradeoff parameter between data rmse and the model smoothness constraint. The tradeoff parameter varies dynamically from iteration to iteration and is controlled by the Levenberg–Marquardt algorithm.

Consider the 1-norm of $\nabla\mathbf{m}$ in eq. (F1); this is called *total variation regularization* in inverse theory and is inspired by methods of image reconstruction and denoising from digital imaging theory (Rudin *et al.* 1992). It has the nice property that it permits sharp boundaries in the solution of inverse problems (Rudin *et al.* 1992), which is why it has been adopted in this work.

In addition, the inversion has been further stabilized by bounding resistivity values to between 0 and 10^4 Ωm. This was effected through the simple parameter substitution,

$$m_i = \frac{10^4}{1 + \exp(\gamma_i)}, \tag{F2}$$

where m_i is the resistivity of the ith cell and $\gamma_i \in \mathbb{R}$ is the substitute variable for which the inversion formally solves. The chain rule is used to express the partials of the forward operator, call it g, with respect to γ. If $\mathbf{G} = \partial g/\partial\mathbf{m}$ and $\mathbf{S} = \partial\mathbf{m}/\partial\boldsymbol{\gamma}$ then

$$\mathbf{G}\mathbf{S} = \frac{\partial g}{\partial \mathbf{m}}\frac{\partial \mathbf{m}}{\partial \gamma} = \frac{\partial g}{\partial \gamma}. \tag{F3}$$

Non-linear updates in the inversion code therefore take the form

$$\begin{aligned} \Delta\gamma = &[(1-\lambda)\mathbf{S}^T\mathbf{G}^T\mathbf{G}\mathbf{S} + \lambda\tilde{\mathbf{L}}^T\tilde{\mathbf{L}}]^{-1} \\ &[(1-\lambda)\mathbf{S}^T\mathbf{G}^T\Delta\mathbf{d} - \lambda\tilde{\mathbf{L}}^T\tilde{\mathbf{L}}\gamma], \end{aligned} \tag{F4}$$

where $\Delta\mathbf{d}$ is the residual between observed and predicted data.

The matrix $\tilde{\mathbf{L}}^T\tilde{\mathbf{L}}$ relates to the L_1 norm of $\nabla\mathbf{m}$ and depends on a modification to the ordinary first difference matrix, $\mathbf{L}$. The 1-norm of the gradient of $\mathbf{m}$ is approximated by

$$\|\nabla\mathbf{m}\|_1 \cong \|\mathbf{L}\mathbf{m}\|_1 = |L_{ij}m_j|, \tag{F5}$$

using Einstein notation. Following the ideas of Scales & Gersztenkorn (1988), the partials of (F5) with respect to $\mathbf{m}$ are given by

$$|L_{ij}m_j|\frac{\partial}{\partial\mathbf{m}} = \mathbf{L}^T\mathbf{R}\mathbf{L}\mathbf{m}, \tag{F6}$$

where $\mathbf{R} \equiv \mathrm{diag}(|L_{ij}m_j| + \varepsilon)^{-1}$ and ε is a small pre-whitening term to prevent blow-up if any element of $L_{ij}m_j$ is zero. Letting $\mathbf{L}_r$ and $\mathbf{L}_z$ denote the ordinary 1st-difference operators in the radial and depth directions, respectively, we therefore define the partials of $\|\nabla\mathbf{m}\|_1$ as

$$\tilde{\mathbf{L}}^T\tilde{\mathbf{L}}\mathbf{m} \equiv \left(\mathbf{L}_r^T\mathbf{R}_r\mathbf{L}_r + \mathbf{L}_z^T\mathbf{R}_z\mathbf{L}_z\right)\mathbf{m} \tag{F7}$$

and hence

$$\tilde{\mathbf{L}}^T\tilde{\mathbf{L}} \equiv \mathbf{L}_r^T\mathbf{R}_r\mathbf{L}_r + \mathbf{L}_z^T\mathbf{R}_z\mathbf{L}_z. \tag{F8}$$