



Contents lists available at ScienceDirect

Computers & Geosciences

journal homepage: www.elsevier.com/locate/cageo

A free lunch in linearized experimental design?

Darrell Coles^{a,b,*}, Andrew Curtis^a^a School of Geoscience, University of Edinburgh, Grant Institute, Edinburgh EH9 3JW, UK^b Geoscience Research Centre, Total E&P UK Limited, Crawpeel Rd., Altens Industrial Estate, Aberdeen AB12 3FG, UK

ARTICLE INFO

Article history:

Received 3 February 2010

Received in revised form

14 September 2010

Accepted 15 September 2010

Available online 20 November 2010

Keywords:

Design criteria

Algorithms

Ellipse

Anisotropy

ABSTRACT

The No Free Lunch (NFL) theorems state that no single optimization algorithm is ideally suited for all objective functions and, conversely, that no single objective function is ideally suited for all optimization algorithms. This paper examines the influence of the NFL theorems on linearized statistical experimental design (SED). We consider four design algorithms with three different design objective functions to examine their interdependency. As a foundation for the study, we consider experimental designs for fitting ellipses to data, a problem pertinent to the study of transverse isotropy in many disciplines. Surprisingly, we find that the quality of optimized experiments, and the computational efficiency of their optimization, is generally independent of the criterion–algorithm pairing. We discuss differences in the performance of each design algorithm, providing a guideline for selecting design algorithms for other problems. As a by-product we demonstrate and discuss the principle of diminishing returns in SED, namely, that the value of experimental design decreases with experiment size. Another outcome of this study is a simple rule-of-thumb for prescribing optimal experiments for ellipse fitting, which bypasses the computational expense of SED. This is used to define a template for optimizing survey designs, under simple assumptions, for Amplitude Variations with Azimuth and Offset (AVAZ) seismics in the specialized problem of fracture characterization, such as is of interest in the petroleum industry. Finally, we discuss the scope of our conclusions for the NFL theorems as they apply to nonlinear and Bayesian SED.

© 2010 Elsevier Ltd. All rights reserved.

1. Introduction

Statistical experimental design (SED) is the theory and practice of optimizing experiments, normally with the goal of maximizing the expected information in collected data sets. Operationally, SED uses a *design criterion* – a quantitative metric of experiment quality – which is extremized with respect to the experiment by means of an optimization algorithm (*design algorithm*).

Several design criteria and algorithms have been introduced to the geosciences in the past decade (e.g., Curtis et al., 2004; Routh et al., 2005; Stummer et al., 2004). The few comparisons that have been conducted among these have focused either on comparing design criteria or design algorithms. No study has yet compared multiple design criteria and design algorithms together to examine the interplay between them.

The No Free Lunch (NFL) theorems (Wolpert and Macready, 1997) assert that no single optimization algorithm can perform optimally for all objective functions and that no single objective function is ideally suited for all optimization algorithms. If an optimization algorithm performs well for one class of objective functions then it must perform

* Corresponding author at: Geoscience Research Centre, Total E&P UK Limited, Crawpeel Rd., Altens Industrial Estate, Aberdeen AB12 3FG, UK.
Tel.: +44 122 423 3121; fax: +44 122 423 3026.

E-mail addresses: Darrell.coles@gmail.com (D. Coles), Andrew.curtis@ed.ac.uk (A. Curtis).

worse on average for the remaining classes of objective functions. These theorems are clearly relevant to SED. We would expect dependencies between design algorithms and design criteria for two reasons. First, design criteria are potentially strongly nonlinear with respect to the experimental designs they qualify. This gives rise to high-dimensional optimization problems whose complexity potentially interacts strongly with the design algorithm. Second, many design algorithms are heuristic and either have no convergence proofs or are known to globally converge only for special classes of optimization problems. This underscores the possibility that design criteria and algorithms are interdependent. When choosing design criterion–algorithm pairings, it is important to be aware of interdependencies that could benefit or hinder the design problem. Our aim is to determine whether such pairings for linearized problems mediate optimum experiments of significantly disparate quality and computational expense.

As a foundation for the study we design experiments for an ellipse fitting problem. This involves estimating the ellipticity and orientation of an ellipse from a limited number of samples of its error-contaminated surface. This problem arises in a variety of theoretical and practical situations in engineering and technology, including in the earth sciences (Grechka and Tsvankin, 1998; Freeze and Cherry, 1979), solid-state physics (Charles, 1996), medical imaging (Szabo, 2004), and material science (Newnham, 2005). The design problem is to optimize the azimuths where samples should be taken along the ellipse surface.

2. Background

We adopt the SED approach pioneered by Box and Lucas (1959) and Draper and Hunter (1967), wherein linear SED methods are applied to a linearization of a nonlinear model function about a prior reference parameterization. This is not as sophisticated as a nonlinear or generalized Bayesian design methods but our goal is to investigate criterion–algorithm interdependence, and the scope of this investigation requires efficient SED methods, for which linearized methods are preferable.

2.1. Posterior covariance

Consider

$$\mathbf{d} = f(\mathbf{m}, \xi), \quad (1)$$

where f is a theoretical function relating data vector \mathbf{d} to parameter vector \mathbf{m} with respect to experiment ξ . Adopting a Gauss–Newton approach to the inverse problem, the least squares objective function considered is

$$\Theta = \|\Sigma_d^{-1/2}(\Delta\mathbf{d} - \mathbf{G}\Delta\mathbf{m}_0)\|^2 + \|\Sigma_m^{-1/2}[\mathbf{m}_{ref} - (\mathbf{m}_0 + \Delta\mathbf{m}_0)]\|^2, \quad (2)$$

where $\mathbf{G} = \partial f / \partial \mathbf{m}$ at \mathbf{m}_0 and ξ , $\Delta\mathbf{d}$ is the data residual, $\Delta\mathbf{m}_0$ is the change in parameter vector \mathbf{m}_0 , Σ_d is the expected data noise covariance, and Σ_m is the expected prior model covariance about reference parameter vector \mathbf{m}_{ref} . The second term conditions the inverse problem to favor solutions near \mathbf{m}_{ref} (Chen et al., 2002). Θ is minimized when

$$\Delta\mathbf{m}_0 = [\mathbf{G}^T \Sigma_d^{-1} \mathbf{G} + \Sigma_m^{-1}]^{-1} [\mathbf{G}^T \Sigma_d^{-1} \Delta\mathbf{d}_0 + \Sigma_m^{-1} (\mathbf{m}_{ref} - \mathbf{m}_0)], \quad (3)$$

and the posterior parameter covariance matrix is

$$\Sigma_p = [\mathbf{G}^T \Sigma_d^{-1} \mathbf{G} + \Sigma_m^{-1}]^{-1}, \quad (4)$$

Σ_p is a linearized measure of uncertainty in post-inversion parameter estimates, and many experimental design criteria are based on it in some manner.

It is noteworthy that experimental designs optimized with respect to Σ_p are conditional on the assumed reference model \mathbf{m}_{ref} and its expected covariance Σ_m . Optimum designs will vary as this prior information is varied.

2.2. Design criteria

For practicality, a limited selection of design criteria is examined. Their mathematical details are in the appendices.

2.2.1. D-criterion and A-criterion

It is easiest to discuss the D-criterion and A-criterion together because they are part of the same family of design criteria (Kiefer, 1974, 1975). In linearized inverse theory, Σ_p forms an ellipsoidal confidence region about the parameter estimates, assuming the estimates are approximately multivariate Gaussian, and the lengths of its semiaxes correspond to the eigenvalues of Σ_p .

The D- and A-criteria measure the size of this ellipsoid through generalized means of the semiaxes lengths or the eigenvalues of Σ_p . The D-criterion is proportional to the geometric mean of the eigenvalues and is evaluated by $\Phi_D = \det \Sigma_p$. The volume of an ellipsoid is proportional to the product of its semiaxes (Abramowitz and Stegun, 1974), so the D-criterion is proportional to the volume of the confidence region. The A-criterion is proportional to the arithmetic mean of the eigenvalues and is evaluated by $\Phi_A = \text{tr} \Sigma_p$, so it is proportional to the mean length of the semiaxes of the confidence ellipsoid. Thus, minimizing Φ_D or Φ_A overall permissible experiments is equivalent to minimizing the size of the post-experimental uncertainty region in two different senses.

Experimental designs minimizing these two measures are respectively called *D-optimal* and *A-optimal*.

2.2.2. Linear dependence reduction.

Linear dependence reduction (LDR) is a method suggested for data decimation by Sabatier (1977) and introduced to SED by Curtis et al. (2004). Rather than measuring the size of uncertainty regions, LDR focuses on the linearized relationship between data and model parameters. The gist of the method is that a strongly linearly dependent row in \mathbf{G} signifies a datum that is related to the parameters in a manner similar to other data observations in the experiment. Strongly linearly dependent rows and their corresponding data should thus be removed.

Experimental designs minimizing this linear dependence measure are called *LD-optimal*.

2.3. Design algorithms

Global optimization strategies such as the genetic and simulated annealing algorithms are dependable for solving combinatorial optimization problems like those sometimes found in experimental design, but they are prohibitively expensive for large problems. An alternative is to use greedy algorithms such as the sequential design algorithms described below. An algorithm is greedy if it makes locally – rather than globally – optimal updates to the solution. Such methods typically converge more quickly than global methods though often to suboptimal solutions (Cormen, 2009).

Three sequential design algorithms and one global one have been considered, with the latter being treated as a control. The sequential algorithms are called the *construction*, *decimation*, and *exchange algorithms*, and the global one is the genetic algorithm (GA). Each is detailed in the appendices. The appendices also give update formulae that expedite sequential design algorithms. These help avoid the explicit calculation of matrix inverses, determinants, etc. implicit in the design criteria listed in Section 2.2.

2.3.1. Construction

Sequential design by construction, also sometimes called *iterative construction*, constructs an experiment by adding observation points (one at a time or in groups) until some desired stopping criterion is met. Each added observation point is conditional on the set of observations already assigned to the experiment. Stummer et al. (2004), Wilkinson et al. (2006), Coles and Morgan (2009), Guest and Curtis (2009), and Khodja et al. (2010) have all demonstrated constructive SED methods for geoscientific experiment optimization.

2.3.2. Decimation

Sequential design by decimation, sometimes also called *iterative destruction*, is the converse of construction. This approach deletes observation points from an experiment. Each deleted point is conditional on the current experiment (similar to construction), not on the set of deleted points. Curtis et al. (2004) have demonstrated this technique.

2.3.3. Exchange

Sequential design by exchange is a hybrid, having both constructive and destructive components. It is motivated by the fact that greedy algorithms do not generally guarantee global solutions. Exchange addresses this by cycling through the observation points in the current experiment and performing a test replacement with all permissible observation points. Each point in the experiment is so treated, and the test replacement that minimizes the objective value of the experiment is exchanged for the current observation

point. Consequently, observation points can be exchanged with themselves. Exchange still does not guarantee global optimality, but it often comes close, as will be seen. Mitchell (1974) introduced this algorithm to SED but it has not recently been demonstrated in a geoscientific context.

2.3.4. Global search

The GA is the global optimization method we use as the control in this study. It is the only algorithm in this study with convergence proofs guaranteeing global solutions (Rudolph, 1994). Geoscientific SED researchers have used the genetic algorithm and simulated annealing extensively in the past (Barth and Wunsch, 1990; Curtis, 1999a,b; Curtis and Snieder, 1997; Furman et al., 2004; Routh et al., 2005).

3. Methods

To evaluate the interdependency between design algorithms and criteria, we have selected ellipse fitting as our demonstration problem. Sets of azimuthal observation positions (the experiment) are optimized using various pairings of design criteria and algorithms and the results are compared. We chose this problem because it offers a simple model of transverse isotropy (also *azimuthal anisotropy*), the phenomenon in which observation of a physical property varies with the direction of observation. Crampin (1986) showed that when this form of anisotropy is not too strong, azimuthal observations in the anisotropy plane closely describe an elliptical pattern. Because of the varied physical domains in which anisotropy arises, the ellipse fitting problem has been kept general to make this work accessible to a range of disciplines.

In polar coordinates, the radial distance from the center to any point on the surface of an ellipse is

$$f(\theta) = \sqrt{\frac{\sec^2(\theta - \omega)}{b^2 + a^2 \tan^2(\theta - \omega)}} \quad (5)$$

where a and b , respectively, denote the major and minor semi-axes, ω is the strike of a in the chosen reference frame, and θ is an observation azimuth (polar angle) in that frame (Fig. 1). Let $\mathbf{m} \equiv [a, b, \omega]^T$ be the vector of model parameters; let Ξ be the set of all permissible observation azimuths; and let ξ denote an experimental design comprising a subset of observation points from Ξ .

Experimental designs were created with respect to parameters following the multivariate normal distribution $N(\mathbf{m}_{ref}, \Sigma_m)$, where $\mathbf{m} \equiv [5, 4, \pi/2]^T$ is the prior reference parameter vector and $\Sigma_m \equiv \text{diag}([2.14, 2.14, 1.48])$ is the prior parameter covariance matrix about \mathbf{m}_{ref} , which in this case conveys the a priori belief that the true model parameters are $a = 5 \pm 4$, $b = 4 \pm 4$ and $\omega = \pi/2 \pm \pi$ with 99% confidence. The prior data covariance was set to $\Sigma_d = 0.45\mathbf{I}$, corresponding to data noise of 15% of the root-mean-square of the expected data. Because of the point symmetry of the ellipse, we limited Ξ to the interval $[-\pi/2, \pi/2]$, which we discretized into 1800 azimuths, permitting experimental designs with azimuthal resolution to 0.1° . For each criterion–algorithm pairing we designed experiments of $n = 3, 6, 9, 12$ and 15 azimuths, for a total of 60 separate optimized experiments.

We used the expected weighted relative error in parameter estimates (WRME) as an independent measure of experimental quality. This was calculated in two stages: first, for each experiment, by performing 10^4 Gauss–Newton inversions with contaminated data, from which we estimated posterior parameter uncertainties. Second, we took the weighted average of the relative model errors by normalizing for the size of the experiment. This

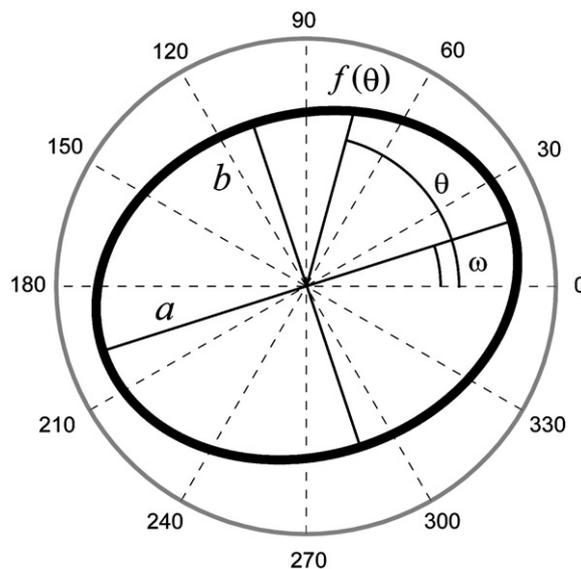


Fig. 1. Ellipse in polar coordinates. Radial length, $f(\theta)$; major axis, a ; minor axis, b ; strike, ω ; observation azimuth, θ .

eliminates the expected $1/\sqrt{n}$ behavior in relative model errors associated with an experiment's size, which would otherwise obfuscate our analysis of the interplay between design criteria and algorithms. Each inversion used a random initial guess for \mathbf{m}_0 from the joint Gaussian $N(\mathbf{m}_{ref}, \Sigma_m)$ and a unique realization of data noise from $N(\mathbf{0}, \Sigma_d)$, with $\Sigma_d = 0.45\mathbf{I}$.

We also looked for interdependencies with respect to computational effort, a practical consideration for industrial SED applications. We computed the computational efficiency for each experimental design relative to the genetic algorithm, which was expected to be least efficient in this study.

4. Results

4.1. Main results

Fig. 2 offers a few examples of our optimized experimental designs. The main takeaway is that the distribution of optimal observation azimuths varies with the design criterion–algorithm pairing as well as the number of observation points. Notably, too, the genetic and exchange algorithms produce identical experiments for the D-criterion.

We expected the GA to have the longest convergence times, which is generally confirmed in Fig. 3. The sequential algorithms generally converge much more rapidly, approaching time-savings of three orders of magnitude. Decimation and construction actually became more efficient with increased experimental size, regardless of the design criterion. The relative efficiencies for the exchange algorithm are usually greater than 100%, and while they increase with experimental size for the LD-criterion, the same is not true for the A- and D-criteria. The salient points are these: (i) the exchange algorithm is the only algorithm that exhibits any apparent criterion-dependence with regard to computational efficiency; (ii) sequential design algorithms are computationally less expensive than global algorithms; and (iii) the relative computational efficiency of the construction and decimation algorithms actually increases with experimental size.

The WRMEs in Fig. 4 evidently cluster for each criterion–algorithm pairing; hence, though the geometry of optimum experiments varies with design algorithm (see Fig. 2), optimal experiments are still of

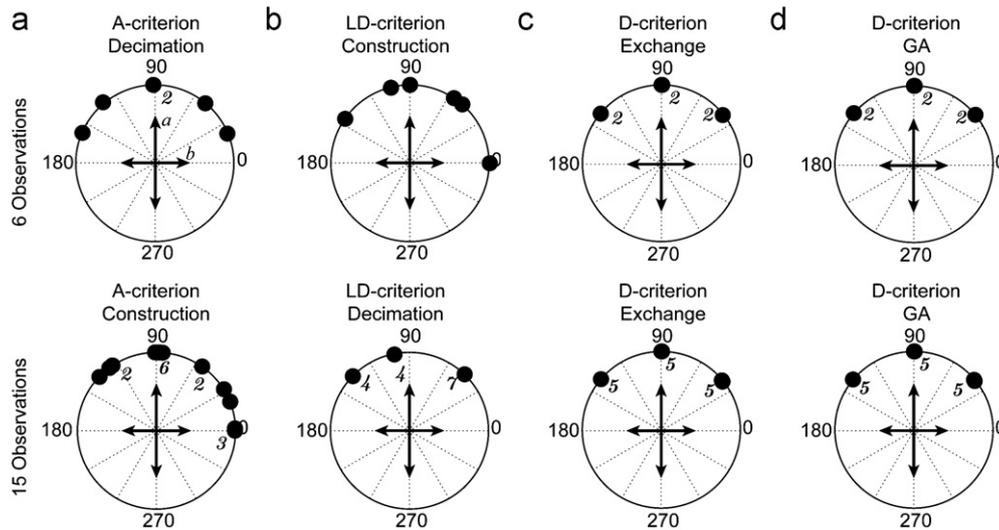


Fig. 2. Optimized azimuthal experiments. Dots: observations points; script numerals: repeated observations at an azimuth; cross arrows: *a* and *b* axes directions.

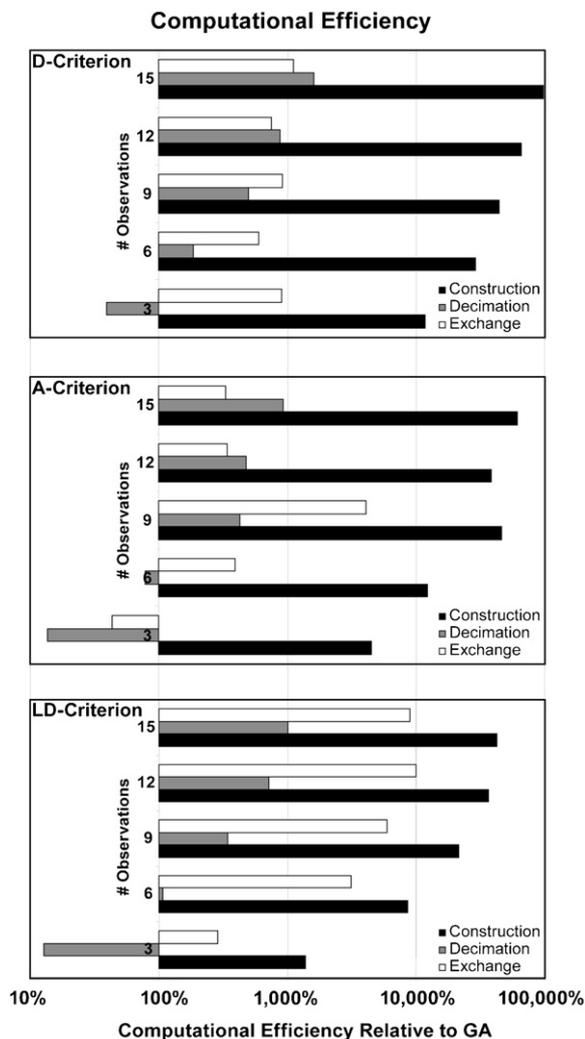


Fig. 3. Relative computational efficiencies. Bars report the ratio of the CPU time of each algorithm to that of the GA. Less than 100% means the GA was faster, and vice versa. The percentage axis is log scaled.

similar quality for a fixed design criterion. The construction algorithm is the only exception. This algorithm does not exhibit criterion dependence so much as it behaves like an outlier with respect to

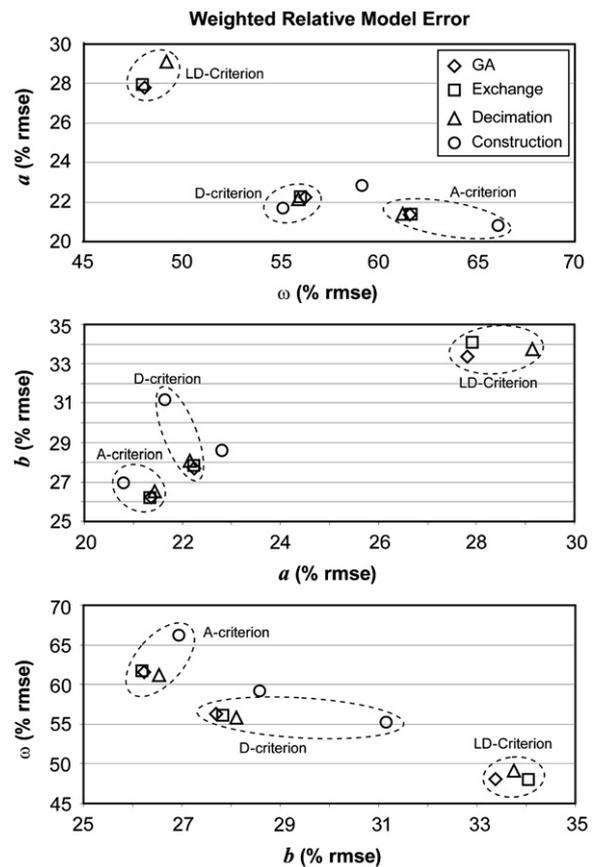


Fig. 4. WRME cross-plots for *a*, *b* and ω . Dashed regions group WRMEs of the same design criterion.

all criteria. The main points are these: (i) the decimation, exchange, and genetic algorithms do not show significant criterion-dependency when measured in terms WRME while (ii) the construction algorithm exhibits ‘outlier’ behavior.

The objective values of experiments produced by the GA were the lowest for all criteria and experimental sizes. The GA experiments therefore come closest to global optimality in all cases, as expected. It is noteworthy that the WRMEs mediated by the

exchange algorithm cluster very closely with those for the GA (Fig. 4), and, of equal importance, the exchange algorithm produces experiments nearly geometrically identical to those produced by the GA (Fig. 2c–d), regardless of design criterion (we have only shown D-optimal experiments Fig. 2). Of the sequential algorithms then, the exchange algorithm therefore mediates experimental designs closest to globally optimal.

4.2. Tangents

We assume relative parameter errors go as $1/\sqrt{n}$, where n is the number of observations. The three experiments in Fig. 5 validate this assumption, which thereby affirms our use of a WRME metric. More importantly, it illustrates that SED is subject to a law of diminishing returns. The number of design observations one must add to an experiment to realize a unit decrement in expected posterior parameter uncertainty increases geometrically with the number of observations already included.

We conducted this study with respect to a single reference model. However, we also investigated the geometry of optimized experiments with respect to different reference models. We found that the relative position of optimized azimuthal observations is invariant with the strike of the ellipse ω . This is intuitive since varying the strike is equivalent to a rotation of the coordinate system, which should leave the relative positions of optimal observation azimuths unchanged.

Interestingly, we found that the D-criterion yields predictable experiment geometries. First we noted that for fixed a , b , and ω , the D-criterion (optimized by the GA) always mediated experiments with repeated observations at the same three azimuthal points, no matter the experimental size. This was not observed with any other design criteria. Moreover, each optimal azimuth received equal weight in the experiment, i.e., if 15 observations were desired then each of the three azimuthal positions was repeated 5 times. Second, as the ratio a/b varies (this corresponds to the coefficient of anisotropy (e.g., Telford et al., 1990, Freeze and Cherry, 1979), D-optimal experiments always have one azimuthal observation along the strike angle ω . The other two are symmetrically displaced a predictable azimuth from ω . Fig. 6 graphs this relative azimuthal displacement for varying a/b ratios and shows a fitted curve given by $60b/a$.

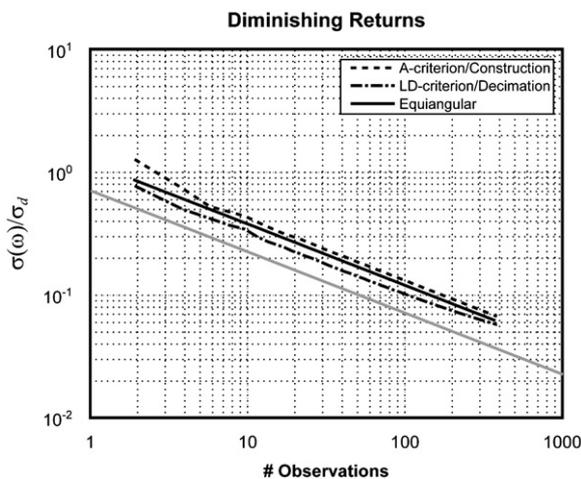


Fig. 5. Uncertainties in the estimates of ω as a function of experimental size, expressed as $\sigma(\omega)/\sigma_d$, where $\sigma(\omega)$ is the posterior standard deviation of ω estimated by simulating 10^4 inversions of noisy data with $\sigma_d(=0.2)$ the data standard deviation. The grey line is $1/\sqrt{n}$.

4.3. Application to amplitude variations with angle and Azimuth

This last result has immediate application for geoscientific anisotropy problems, such as seismic amplitude variations with angle and azimuth (AVAZ). Consider a two-layer earth model in which Layer 1 is homogenous, isotropic and Layer 2 is homogeneous, azimuthally anisotropic (Fig. 7).

This describes, for example, a situation in which the overburden is relatively uniform and overlies a system of sub-vertical, sub-parallel fractures, which are now known to be common (Crampin, 1985). There are other scenarios in which the Earth may be modeled as a homogeneous, isotropic overburden atop (effectively) anisotropic bedrock, such as due to varying lithology, bedding and folds in the bedrock. The purpose here is to draw attention to one useful application of the result from the preceding section. The petroleum industry, for example, is interested in fractured reservoir characterization through anisotropy observations made via AVO/A methods (e.g., Hall and Kendall, 2003). The amplitude of a P-wave propagating through Layer 1 and reflected off Layer 2 is approximated by

$$R_p(i, \phi) = A + (B^{iso} + B^{ani} \cos^2 \phi) \sin^2 i + (C^{iso} + C^{ani} \cos^4 \phi + C^{ani} \sin^2 \phi \cos^2 \phi) \sin^2 i \tan^2 i, \tag{6}$$

where i and ϕ are, respectively, the polar incidence angle and azimuth, and the coefficients are parameters of (an)isotropy, which depend on elastic properties, including the vertical P- and S-wave

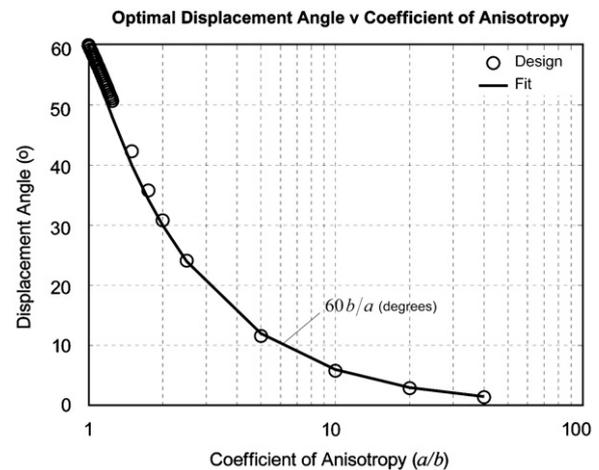


Fig. 6. Optimal relative displacement azimuth (\pm) from ω for D-optimal azimuthal experiments and fit.

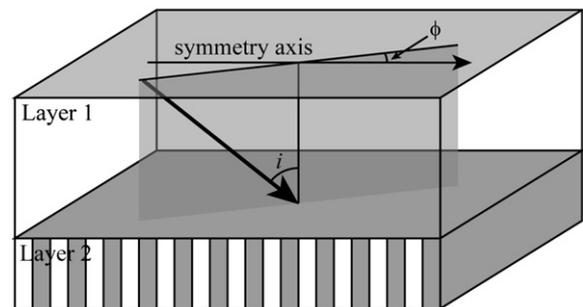


Fig. 7. Cartoon of an isotropic overburden atop an azimuthally anisotropic system of vertically aligned fractures. The reflection response depends on the polar incidence angle i and the azimuth angle ϕ defined with respect to the symmetry axis ϕ_0 (after Ruger and Tsvankin, 1997).

velocities, density, and Thomsen parameters. The anisotropy symmetry axis, ϕ_0 , can be introduced as a free variable by considering $R_p(i, \phi - \phi_0)$.

Eq. (6) is a second order approximation of the solution of Zoeppritz's equations for azimuthally anisotropic media (Ruger and Tsvankin, 1997; Ruger, 1997; Ruger, 1998). It describes seismic AVAZ, the variation of seismic reflection amplitude with polar incidence angle and azimuth. The approximation error increases with incidence angle but is reasonably small for incidence angles up to a few tens of degrees. Eq. (6) does not technically define an ellipse, but it is closely described by one in many realistic geoscience settings (Fig. 8).

We can apply the empirical formula for D-optimum experiments from Section 4.2 to problems in AVAZ by assuming that R_p is effectively elliptical and goes through its maximum when $\phi = \phi_0$ and minimum when $\phi = \phi_0 \pm 90^\circ$. Assuming ϕ_0 is known (albeit imperfectly), the displacement azimuth, $\Delta\phi$, at which additional D-optimum observation points should be taken about ϕ_0 is

$$\Delta\phi \equiv \frac{R_p(i, \phi = \phi_0 + 90^\circ)}{R_p(i, \phi = \phi_0)} \times 60^\circ, \quad (7)$$

which effectively prescribes D-optimal AVAZ experiments as a function of incidence angle.

To illustrate, we take Ruger and Tsvankin's (1997) 'wet crack' earth model, which features an isotropic layer over a system of fluid-filled, vertically aligned fractures. This model has the following properties. Layer 1: $\alpha_1 = 3.670$ km/s; $\beta_1 = 2.000$ km/s; $\rho_1 = 2.41$ g/cm³. Layer 2: $\alpha_2 = 4.498$ km/s; $\beta_2 = 2.530$ km/s; $\rho_2 = 2.80$ g/cm³; $\gamma_2 = 0.085$; $\delta_2 = -0.088$; $\epsilon_2 = -0.003$. The parameters γ_2 , δ_2 , and ϵ_2 are Thomsen parameters of anisotropy in Layer 2 and, respectively, describe the shear wave splitting parameter, the anisotropic response of a P-wave propagating parallel to the symmetry axis of anisotropy (ϕ_0), and the anisotropic response of a P-wave propagating perpendicular to ϕ_0 (Ruger and Tsvankin, 1997; Brittan et al., 1995). The coefficients in Eq. (6) implicitly depend on these elastic properties, and the anisotropic behavior we see in Fig. 8 actually derives from this wet crack model.

Fig. 9 illustrates the template of D-optimum azimuth/offset positions prescribed by Eq. (7) for the wet crack model, up to a maximum incidence angle of 54° (incidence angle has been converted to offset by the formula $x = 2h \tan i$, where h is the thickness of Layer 1 and x is the source-receiver distance). To use the template, pick an offset distance and select all azimuthal observations (represented by the bold lines) that intersect the circle defining that offset distance on the polar graph. Multiple offset distances can of course be selected, and the set of all azimuthal intersection points thus chosen prescribes a 'pseudo' D-optimal AVAZ experiment.

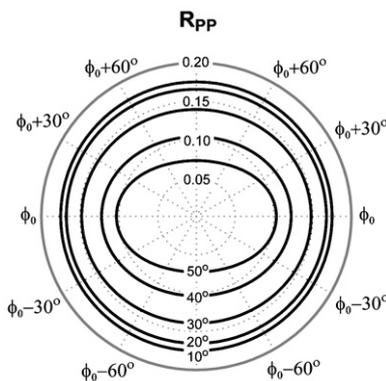


Fig. 8. P-wave reflection coefficient for varying incidence angles (10° – 50° lines) and azimuths (relative to the anisotropy symmetry axis ϕ_0). Example follows Ruger and Tsvankin's wet crack model (1997), which depicts an isotropic layer above an azimuthally anisotropic system of vertically aligned, fluid-filled fractures.

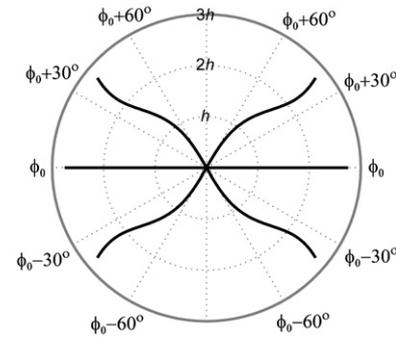


Fig. 9. Template for a 'pseudo' D-optimum AVAZ experiment to query the wet crack model. To use the template, pick an offset distance (expressed as a fraction of the Layer 1 thickness, h) and select all azimuthal observations (bold lines) that intersect the circle defining that offset distance on the polar graph. The maximum offset distance shown in the template corresponds to an angle of incidence of 54° , just less than the critical angle of 54.7° .

5. Discussion

5.1. NFL theorems

Except for the construction algorithm, whose optimized experiments were categorically inconsistent with the other algorithms, the sequential algorithms we have looked at produce experiments whose WRMEs cluster closely with (nearly) globally optimal GA experiments, regardless of design criterion. In addition, the computational behavior of all but one algorithm (exchange) was independent of design criterion. These bits of evidence are surprising given that we expected clear dependencies as predicted by the No Free Lunch theorems. While the No Free Lunch theorems assert that such dependency must exist, the design algorithms and criteria we have examined are generally robust to the criterion–algorithm pairing (again, excepting the construction algorithm and to a lesser extent the exchange algorithm). It may be that the design algorithms are similar enough to each other that their performances are roughly similar, or equally that the design criteria are similar enough to get these results. To be fair, the WRMEs mediated by the sequential algorithms do differ from those of the GA. We therefore conclude that there is a weak but general dependency between criteria and algorithms for linear SED methods. In any event, it is good news for linearized SED practitioners that the NFL Theorems only weakly influence their enterprise. One can freely select a criterion–algorithm pair without worrying overmuch about inefficiencies inherent to the pairing.

To be clear, however, this weak dependency apparently relates to design algorithms rather than to specific criterion–algorithm pairings. The computational efficiency of the exchange algorithm is unpredictable not because of the chosen design criterion but because the algorithm iterates as long as experimental quality increases on each pass through the experiment. There is no way to know a priori how many iterations this will take, wherein lies the unpredictability. Likewise, the inconsistency of experiments optimized by the construction algorithm is due to the algorithm. Discrete experimental optimization is combinatorial and the space of possible solutions can be vast. There are more than 10^{30} combinations of 12 observation points possible from a set of 1800, for example. Construction conducts by far the least number of trials compared to the other design algorithms (Table 1). Since many combinatorial optimization problems are NP-complete or -hard (Papadimitriou and Steiglitz, 1998), it is a matter of course that the quality of experiments optimized by construction would be inferior to those of the other design algorithms.

Though we have not shown it, the suboptimality of experiments designed by construction is prominent for small experiments but

Table 1

Algorithm ratings based on convergence times (speed) and expected WRME (quality). 1=best, 4=worst. Also shown are the number of trials each algorithm executes, where n is the number of desired observations points and n_{\pm} is the total number of permissible observations.

Algorithm	Quality	Speed	No. of trials
Construction	4	1	$n n_{\pm}$
Decimation	3	3	$n_{\pm}!/n!$
Exchange	2	2	$\geq (n+1)n_{\pm}$
GA	1	4	Unpredictable

diminishes as experimental size increases. Consequently, construction is a good candidate for design problems where a large number of observations are likely to be required, especially given its computational efficiency. The exchange algorithm always produced nearly globally optimal experiments; consequently, despite unpredictable convergence times, we suggest that exchange offers the best trade-off between optimality and computation time. The decimation algorithm falls somewhere between the construction and exchange algorithms; the quality of its experiments clusters more tightly with those of the GA than do those of construction but not as tightly as those of the exchange algorithm. Further, the convergence time of decimation is proportional to the number of permissible observation points (Table 1) and if this is large, decimation is impractical. In short, decimation is more reliable than construction and less reliable than exchange for experiment optimality, and it is more reliable than exchange and less reliable than construction for computational efficiency. Table 1 offers a relative rating of the design algorithms on these points.

This paper considers linear SED for linearized models but not for generalized Bayesian or nonlinear SED. However, both generalized Bayesian and nonlinear SED use similar design criteria and algorithms (e.g., Chaloner and Verdinelli, 1995; Guest and Curtis, 2009), and we anticipate that our linearized results are sufficiently robust to be relevant in those contexts.

5.2. Tangents

5.2.1. Diminishing returns

Diminishing returns is almost certainly a ubiquitous property of experimental design irrespective of the design algorithm, design criterion, or whether linear, nonlinear, or Bayesian SED methods are employed. It is a manifestation of the sampling property that the uncertainty of a first-order statistic formed from n samples goes as $1/\sqrt{n}$. Diminishing returns implies that the relative utility of optimizing experiments decreases geometrically as the desired size of an optimized experiment increases (and this does not even account for the concomitant increase in computational expense). For this reason, SED is most practical for the design of compact experiments, where data sets with a high information-to-cost ratio are desired.

5.2.2. Rule of thumb for ellipse fitting

The 60b/a D-optimality 'rule' was unexpected but potentially useful. As the ratio a/b increases, the optimal relative displacement azimuth of the two offset observations approaches ω ; as the ratio approaches unity, the optimum azimuths approach $\omega \pm 60^\circ$. This obviates the need for future design optimization for ellipse fitting. Moreover, it could be used to adapt experiments to the azimuthal anisotropy at a particular site in real time. Or, by averaging several D-optimum designs over an assumed prior parameter distribution, a relative distribution of observation azimuths could be generated, producing a pseudo-Bayesian D-optimal experiment without the need for significant computation. We emphasize: this 60b/a rule is general in that it pertains to any scientific problem involving ellipse fitting.

5.2.3. AVAZ

The AVAZ example we use to demonstrate the above D-optimality rule is also potentially useful. 'Pseudo' D-optimal AVAZ experiments can be generated without the need of any SED computation, assuming we have some basic prior information on the pertinent elastic parameters. Importantly, the optimized data from such seismic experiments can be inverted for the generic (an)isotropy coefficients in Eq. (6) and from these the symmetry axis and other elastic parameters of interest can be derived (cf. Ruger and Tsvankin, 1997).

The "tulip" shape of the D-optimum template in Fig. 9 is unusual. For small offsets, it prescribes nearly equiangular azimuths, which makes sense since the anisotropic effect is very small at small offsets, causing R_p to be nearly circular in azimuth. As offset increases, the ellipticity of R_p increases, so D-optimum azimuths move closer to the anisotropy symmetry axis, in accordance with Fig. 6. However, there is an inflection point for offsets around $2h$, where the recommended azimuths start to increase, suggesting that the eccentricity of the R_p ellipse is decreasing. There are two possible reasons for this. First, as stated previously, there is an error of approximation in Eq. (6) that increases with incidence angle, and it is possible that the "tulip" shape is due to this. Alternately, the azimuthal variation of R_p can sometimes be quite non-elliptical, especially at large incidence angles when R_p can become "peanut shaped" in a polar plot. This disagrees with our ellipticity assumption on R_p and therefore it would be inappropriate to use the empirical D-optimal ellipse fitting rule. To address the former, we would need a better approximation of the solution to Zoeppritz's equations for azimuthal anisotropy. To address the latter, we should restrict the prescription of optimal experiments to the range of incidence angles for which the ellipticity assumption holds.

One further point: though Eq. (7) offers a way to create optimal AVAZ experiments without further SED computation, the experiments are only 'pseudo' optimal. To properly optimize an AVAZ experiment, azimuth and offset should be jointly optimized as free variables in the SED exercise; whereas, in this article, we require specification of the offsets. Moreover, a proper AVAZ experimental design would also account for variable signal to noise as a function of offset.

Taking all these issues into account, our pseudo optimal design formula is nonetheless a handy tool for quickly designing AVAZ surveys and is still arguably preferable to the ad hoc design of experiments.

6. Summary

We have examined the interdependence between experimental design criteria and algorithms for linearized SED, implicit from the No Free Lunch theorems. The examination was carried out on an ellipse-fitting problem relevant to the study of azimuthal anisotropy. We found that there is only weak interdependence between design criteria and algorithms, which is a useful result for applied SED as it indicates that we can freely choose whichever criterion–algorithm suits our needs. We hypothesize that our results also apply to nonlinear SED and generalized Bayesian SED as well, as these techniques employ similar design criteria and algorithms.

We also noted that some algorithm behavior was not attributable to the NFL theorems. The construction algorithm is fast but yields significantly suboptimal experiments when small numbers of observations are used, and the exchange algorithm has unpredictable but relatively fast convergence times while producing (nearly) globally optimal experiments. Consequently, we recommend the exchange algorithm as the best overall choice for the trade-off between computational efficiency and experimental quality. We recommend the decimation algorithm as an intermediate choice, and construction for large designs where computational efficiency is vital. However, if experimental optimality is paramount and computation time is irrelevant, a global search algorithm is always the best choice.

We also discovered an empirical formula for prescribing D-optimal experimental designs that could be practical for designing pseudo-Bayesian experimental designs or for adapting real-time experiment involving transverse isotropy. We demonstrated one use of this formula for optimizing an AVAZ experiment with respect to a simple, azimuthally anisotropic earth model. As much as this simplification is valid, this development can be used to quickly design optimum surveys for the specialized problem of characterizing fractured petroleum reservoirs, for example. It is thus possible to rapidly optimize AVAZ experiments with a modicum of prior information, though some caution must be exercised in selecting suitably optimum azimuths. Such surveys can be used to infer the symmetry axis of anisotropy and a number of related elastic properties.

Finally, we observed that linearized SED is subject to a law of diminishing returns, from which we conclude that SED is most practical for the design of compact, information-dense experiments.

Appendices

Sequential design algorithms

The following notation, definitions, and conventions are used: Φ denotes the experimental design criterion or objective function; Ξ is the discrete ordered set of permissible observation points; there are n_{Ξ} points in Ξ ; n is the number of desired observations points; an experiment will be equivalently expressed in the vector form ξ ($\in \Xi^n$) or in the ordered set form ζ ($\subseteq \Xi$); ζ_b denotes the *base experiment*, the set of observation points currently assigned to an experiment; bracket notation is used to index elements in an ordered set, e.g., $\zeta[j]$ is the j th observation point in ζ ; the *set difference* is denoted $BA \equiv \{x \in B | x \notin A\}$ and is defined as the elements of B that are not in A ; algorithms are notated using the pseudocode conventions of Cormen (2009).

Construction

```

0    $\zeta_b \leftarrow \emptyset$ 
1   for  $i \leftarrow 1$  to  $n$ 
2     for  $j \leftarrow 1$  to  $n_{\Xi}$ 
3       do  $\Phi[j] = \Phi(\zeta_b \cup \Xi[j])$ 
4      $j^* \leftarrow \operatorname{argmin}_{j \in [1, n_{\Xi}]} \Phi[j]$ 
5      $\zeta_b \leftarrow \zeta_b \cup \Phi[j^*]$ 
6   return  $\zeta_b$ 
    
```

Decimation

```

0    $\zeta_b \leftarrow \Xi$ 
1   for  $i \leftarrow n_{\Xi}$  to  $n$ 
2     for  $j \leftarrow 1$  to  $i$ 
3       do  $\Phi[j] = \Phi(\zeta_b \setminus \zeta_b[j])$ 
4      $j^* \leftarrow \operatorname{argmin}_{j \in [1, i]} \Phi[j]$ 
5      $\zeta_b \leftarrow \zeta_b \setminus \zeta_b[j^*]$ 
6   return  $\zeta_b$ 
    
```

Exchange

```

0   find  $\zeta_b$  by construction; set  $\text{success} \leftarrow 1$ 
1   while  $\text{success} = 1$ 
2      $\text{success} \leftarrow 0$ 
3     for  $i \leftarrow 1$  to  $n$ 
4        $\zeta_b \leftarrow \zeta_b \setminus \zeta_b[i]$ 
5       for  $j \leftarrow 1$  to  $n_{\Xi}$ 
6         do  $\Phi[j] = \Phi(\zeta_b \cup \Xi[j])$ 
7        $j^* \leftarrow \operatorname{argmin}_{j \in [1, n_{\Xi}]} \Phi[j]$ 
8       if  $\Xi[j^*] \neq \zeta_b[i]$  then  $\text{success} \leftarrow 1$ 
9        $\zeta_b[i] \leftarrow \Xi[j^*]$ 
10  return  $\zeta_b$ 
    
```

Update formulae for matrix inverses, traces, and determinants

Let $\mathbf{A} \in \mathbf{R}^{p \times r}$ and $\mathbf{B} \in \mathbf{R}^{q \times r}$, and define

$$\mathbf{A}' = \begin{bmatrix} \mathbf{A} \\ \mathbf{B} \end{bmatrix} \quad (8)$$

The following are formulae for updating the matrix inverse, the determinant (D-criterion), and the trace of an inverse matrix (A-criterion) when observation points are added (equivalent to adding rows to \mathbf{A}) or deleted (equivalent to deleting rows from \mathbf{A}'). Many of these results can be found in the linear algebra and SED literature (e.g., Fedorov, 1972; Wynn, 1970; Dykstra, 1971; Golub and Van Loan, 1996; Coles and Morgan, 2009), but to our knowledge they have never been compiled in one place.

For $p \geq r$

$$(\mathbf{A}'^T \mathbf{A}')^{-1} = (\mathbf{A}^T \mathbf{A})^{-1} - \Phi(\mathbf{I} + \mathbf{B}\Phi)^{-1} \Phi^T, \quad (9)$$

$$(\mathbf{A}^T \mathbf{A})^{-1} = (\mathbf{A}'^T \mathbf{A}')^{-1} + \Phi(\mathbf{I} - \mathbf{B}\Phi)^{-1} \Phi^T, \quad (10)$$

$$\operatorname{tr}(\mathbf{A}'^T \mathbf{A}')^{-1} = \operatorname{tr}(\mathbf{A}^T \mathbf{A})^{-1} - \operatorname{tr} \Phi(\mathbf{I} + \mathbf{B}\Phi)^{-1} \Phi^T, \quad (11)$$

$$\operatorname{tr}(\mathbf{A}^T \mathbf{A})^{-1} = \operatorname{tr}(\mathbf{A}'^T \mathbf{A}')^{-1} + \operatorname{tr} \Phi(\mathbf{I} - \mathbf{B}\Phi)^{-1} \Phi^T, \quad (12)$$

$$\det(\mathbf{A}'^T \mathbf{A}') = \det(\mathbf{A}^T \mathbf{A}) \det(\mathbf{I} + \mathbf{B}(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{B}^T), \quad (13)$$

$$\det(\mathbf{A}^T \mathbf{A}) = \det(\mathbf{A}'^T \mathbf{A}') \det(\mathbf{I} - \mathbf{B}(\mathbf{A}'^T \mathbf{A}')^{-1} \mathbf{B}^T). \quad (14)$$

where $\Phi \equiv (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{B}^T$ and $\Phi \equiv (\mathbf{A}'^T \mathbf{A}')^{-1} \mathbf{B}^T$.

For $p < r$

$$(\mathbf{A}' \mathbf{A}'^T)^{-1} = \begin{pmatrix} (\mathbf{A} \mathbf{A}^T)^{-1} + \Phi \Gamma^{-1} \Phi^T & -\Phi \Gamma^{-1} \\ -\Gamma^{-1} \Phi^T & \Gamma^{-1} \end{pmatrix}, \quad (15)$$

$$(\mathbf{A} \mathbf{A}^T)^{-1} = \mathbf{Z}' - \mathbf{Y} \mathbf{X}'^{-1} \mathbf{Y}'^T, \quad (16)$$

$$\operatorname{tr}(\mathbf{A}' \mathbf{A}'^T)^{-1} = \operatorname{tr}(\mathbf{A} \mathbf{A}^T)^{-1} + \operatorname{tr} \Phi \Gamma^{-1} \Phi^T + \operatorname{tr} \Gamma^{-1}, \quad (17)$$

$$\operatorname{tr}(\mathbf{A} \mathbf{A}^T)^{-1} = \operatorname{tr} \mathbf{Z}' - \operatorname{tr} \mathbf{Y} \mathbf{X}'^{-1} \mathbf{Y}'^T, \quad (18)$$

$$\det(\mathbf{A}' \mathbf{A}'^T) = \det(\mathbf{A} \mathbf{A}^T) \det[\mathbf{B}(\mathbf{I} - \mathbf{A}^T (\mathbf{A} \mathbf{A}^T)^{-1} \mathbf{A}) \mathbf{B}^T], \quad (19)$$

$$\det(\mathbf{A} \mathbf{A}^T) = \det(\mathbf{A}' \mathbf{A}'^T) \det(\mathbf{X}'), \quad (20)$$

where $\Phi \equiv (\mathbf{A} \mathbf{A}^T)^{-1} \mathbf{A} \mathbf{B}^T$ and $\Gamma \equiv \mathbf{B}[\mathbf{I} - \mathbf{A}^T (\mathbf{A} \mathbf{A}^T)^{-1} \mathbf{A}] \mathbf{B}^T$, $\mathbf{Z}' \equiv \mathbf{X}(1:p, 1:p)$, $\mathbf{X}' \equiv \mathbf{X}(p+1:p+q, p+1:p+q)$, $\mathbf{Y}' \equiv \mathbf{X}(1:p, p+1:p+q)$ and $\mathbf{X} \equiv (\mathbf{A}' \mathbf{A}'^T)^{-1}$.

Update formula for the LD-criteria (Curtis et al., 2004)

Let \mathbf{g}_i^T be the i th row in \mathbf{G} corresponding to the i th datum in the design ξ . The degree of linear dependence between the i th and j th data is denoted ρ_{ij} and given by the normalized expression:

$$\rho_{ij} \equiv \frac{\mathbf{g}_i^T \Sigma_m \mathbf{g}_j}{(\mathbf{g}_i^T \Sigma_m \mathbf{g}_i)^{1/2} (\mathbf{g}_j^T \Sigma_m \mathbf{g}_j)^{1/2}}. \quad (21)$$

The experimental quality of ξ is

$$\Phi_{LD}(\xi) = - \sum_{i,j=1}^m (1 - \tau_{ij} |\rho_{ij}|), \quad (22)$$

where m is the number of observations currently in the experiment and τ_{ij} is a weight matrix defined by $\tau_{ij} \equiv \sigma_i \sigma_j / \sigma_{max}^2$, the product of the expected standard deviations of the i th and j th data normalized by the largest expected variance, σ_{max}^2 .

If the k th observation point $\xi[k]$ is to be deleted from the experiment, Φ_{LD} is updated by adding

$$\Delta\Phi_{LD} = \tau_{kk}|\rho_{kk}| - 1 + 2 \sum_{i=1}^m (1 - \tau_{ik}|\rho_{ik}|). \quad (23)$$

If $\xi[k]$ is added instead then subtract $\Delta\Phi_{LD}$. Care is needed in bookkeeping.

References

- Abramowitz, M., Stegun, I., 1974. Handbook of Mathematical Functions, with Formulas, Graphs, and Mathematical Tables. Dover Publications, Mineola, NY 1046 pp.
- Barth, N., Wunsch, C., 1990. Oceanographic experiment design by simulated annealing. *Journal of Physical Oceanography* 20, 1249–1263.
- Box, G.E.P., Lucas, H.L., 1959. Design of experiments in non-linear situations. *Biometrika* 46 (1/2), 77–90.
- Brittan, J., Warner, M., Pratt, G., 1995. Anisotropic parameters of layered media in terms of composite elastic properties. *Geophysics* 60 (4), 1243–1248.
- Chaloner, K., Verdinelli, I., 1995. Bayesian experimental design: a review. *Statistical Science* 10 (3), 273–304.
- Charles, K., 1996. Introduction to Solid State Physics. Wiley, New York, NY, 673pp.
- Chen, J.P., Haber, E., Oldenburg, D.W., 2002. Three-dimensional numerical modelling and inversion of magnetometric resistivity data. *Geophysical Journal International* 149 (3), 679–697.
- Coles, D., Morgan, F.D., 2009. A method of fast, sequential experimental design for linearized geophysical inverse problems. *Geophysical Journal International* 178 (1), 145–158.
- Cormen, T.H., 2009. Introduction to Algorithms. MIT Press, Cambridge, MA, 1216pp.
- Crampin, S., 1985. Evidence for aligned cracks in the Earth's crust. *First Break* 3 (3), 12–15.
- Crampin, S., 1986. Anisotropy and transverse isotropy. *Geophysical Prospecting* 34 (1), 94–99.
- Curtis, A., 1999a. Optimal design of focused experiments and surveys. *Geophysical Journal International* 139 (1), 205–215.
- Curtis, A., 1999b. Optimal experiment design: cross-borehole tomographic examples. *Geophysical Journal International* 136 (3), 637–650.
- Curtis, A., Michelini, A., Leslie, D., Lomax, A., 2004. A deterministic algorithm for experimental design applied to tomographic and microseismic monitoring surveys. *Geophysical Journal International* 157 (2), 595–606.
- Curtis, A., Snieder, R., 1997. Reconditioning inverse problems using the genetic algorithm and revised parameterization. *Geophysics* 62 (5), 1524–1532.
- Draper, N.R., Hunter, W.G., 1967. The use of prior distributions in the design of experiments for parameter estimation in non-linear situations. *Biometrika* 54 (1/2), 147–153.
- Dykstra, O., 1971. Augmentation of experimental data to maximize $|X'X|$. *Technometrics* 13 (3), 682–688.
- Fedorov, V.V., 1972. Theory of Optimal Experiments. Academic Press, New York, NY, 292 pp.
- Freeze, R.A., Cherry, J.A., 1979. Groundwater. Prentice-Hall, Englewood Cliffs, NJ, 604 pp.
- 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 Journal* 3 (4), 1230–1239.
- Golub, G.H., Van Loan, C.F., 1996. Matrix Computations. Johns Hopkins University Press, Baltimore, MD, 694 pp.
- Grechka, V., Tsvankin, I., 1998. 3-D description of normal moveout in anisotropic inhomogeneous media. *Geophysics* 63 (3), 1079–1092.
- Guest, T., Curtis, A., 2009. Iteratively constructive sequential design of experiments and surveys with nonlinear parameter–data relationships. *Journal of Geophysical Research-Solid Earth and Planets* 114, B04307.
- Hall, S.A., Kendall, J., 2003. Fracture characterization at Valhall: application of P-wave amplitude variation with offset and azimuth (AVOA) analysis to a 3D ocean-bottom data set. *Geophysics* 68 (4), 1150–1160.
- Khodja, M., Prange, M., Djikpesse, H., 2010. Guided Bayesian linearized optimal experimental design. *Inverse Problems* 26 (5), 055008.
- Kiefer, J., 1974. General equivalence theory for optimum designs (approximate theory). *Annals of Statistics* 2 (5), 849–879.
- Kiefer, J., 1975. Optimal design: variation in Structure and performance under change of criterion. *Biometrika* 62 (2), 277–288.
- Mitchell, T.J., 1974. Algorithm for construction of D-optimal experimental designs. *Technometrics* 16 (2), 203–210.
- Newnham, R.E., 2005. Properties of Materials : Anisotropy, Symmetry, Structure. Oxford University Press, Oxford, UK, 440 pp.
- Papadimitriou, C.H., Steiglitz, K., 1998. Combinatorial Optimization : Algorithms and Complexity. Dover Publications, Mineola, NY, 496 pp.
- 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 (1), 1033–1036.
- Rudolph, G., 1994. Convergence analysis of canonical genetic algorithms. *Neural Networks, IEEE Transactions on* 5 (1), 96–101.
- Ruger, A., 1997. P-wave reflection coefficients for transversely isotropic models with vertical and horizontal axis of symmetry. *Geophysics* 62 (3), 713–722.
- Ruger, A., 1998. Variation of P-wave reflectivity with offset and azimuth in anisotropic media. *Geophysics* 63 (3), 935–947.
- Ruger, A., Tsvankin, I., 1997. Using AVO for fracture detection: analytic basis and practical solutions. *The Leading Edge* 16 (10), 1429–1434.
- Sabatier, P.C., 1977. Geophysical inverse problems and constraints. *Journal of Geophysics* 43 (1–2), 115–137.
- Stummer, P., Maurer, H., Green, A.G., 2004. Experimental design: electrical resistivity data sets that provide optimum subsurface information. *Geophysics* 69 (1), 120–139.
- Szabo, T.L., 2004. Diagnostic Ultrasound Imaging : Inside Out.. Elsevier Academic Press, London, 549 pp.
- Telford, W.M., Geldart, L.P., Sheriff, R.E., 1990. Applied Geophysics. Cambridge University Press, Cambridge, UK, 770 pp.
- Wilkinson, P.B., Chambers, J.E., Meldrum, P.I., Ogilvy, R.D., Caunt, S., 2006. Optimization of array configurations and panel combinations for the detection and imaging of abandoned mineshafts using 3D cross-hole electrical resistivity tomography. *Journal of Environmental and Engineering Geophysics* 11 (3), 213–221.
- Wolpert, D.H., Macready, W.G., 1997. No free lunch theorems for optimization. *Evolutionary Computation, IEEE Transactions on* 1 (1), 67–82.
- Wynn, H.P., 1970. The sequential generation of D-optimum experimental designs. *The Annals of Mathematical Statistics* 41 (5), 1655–1664.