Computers & Geosciences 37 (2011) 1026–1034

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.

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 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 interdepend. When choosing design criterion–algorithm pairings, it is important to be aware of inter-dependencies 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.

* 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).

0098-3004/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.cageo.2010.09.012

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

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

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

$$\Theta = \|\boldsymbol{\Sigma}_d^{-1/2}(\Delta\mathbf{d} - \mathbf{G}\Delta\mathbf{m}_0)\|^2 + \|\boldsymbol{\Sigma}_m^{-1/2}[\mathbf{m}_{\text{ref}} - (\mathbf{m}_0 + \Delta\mathbf{m}_0)]\|^2, \tag{2}$$

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 $\mathbf{m}_0$, $\boldsymbol{\Sigma}_d$ is the expected data noise covariance, and $\boldsymbol{\Sigma}_m$ is the expected prior model covariance about reference parameter vector $\mathbf{m}_{ref}$. The second term conditions the inverse problem to favor solutions near $\mathbf{m}_{\text{ref}}$ (Chen et al., 2002). Θ is minimized when

$$\Delta\mathbf{m}_0 = [\mathbf{G}^T\boldsymbol{\Sigma}_d^{-1}\mathbf{G} + \boldsymbol{\Sigma}_m^{-1}]^{-1}[\mathbf{G}^T\boldsymbol{\Sigma}_d^{-1}\Delta\mathbf{d}_0 + \boldsymbol{\Sigma}_m^{-1}(\mathbf{m}_{\text{ref}} - \mathbf{m}_0)], \tag{3}$$

and the posterior parameter covariance matrix is

$$\boldsymbol{\Sigma}_p = \left[\mathbf{G}^T\boldsymbol{\Sigma}_d^{-1}\mathbf{G} + \boldsymbol{\Sigma}_m^{-1}\right]^{-1}, \tag{4}$$

$\boldsymbol{\Sigma}_p$ 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 $\boldsymbol{\Sigma}_p$ are conditional on the assumed reference model $\mathbf{m}_{ref}$ and its expected covariance $\boldsymbol{\Sigma}_m$. 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 $\boldsymbol{\Sigma}_p$. The D-criterion is proportional to the geometric mean of the eigenvalues and is evaluated by $\Phi_D = \det \boldsymbol{\Sigma}_p$. 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 $\Phi_D = \text{tr}\ \boldsymbol{\Sigma}_p$, 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