

Numerical detection and reduction of non-uniqueness in nonlinear inverse problems

Emanuel Winterfors^{1,2} and Andrew Curtis^{2,3}

¹ Université Pierre et Marie Curie, Laboratoire Jacques-Louis Lions, 75252 Paris Cedex 05, France

² Edinburgh University, School of GeoScience, Grant Institute, West Mains Road, Edinburgh EH9 3JW, UK

³ ECOSSE (Edinburgh Collaborative of Subsurface Science and Engineering), Edinburgh, UK

E-mail: winterfors@ann.jussieu.fr

Received 5 October 2007, in final form 21 January 2008

Published 21 February 2008

Online at stacks.iop.org/IP/24/025016

Abstract

We present a novel approach to analyze uniqueness in nonlinear inverse problems, using a novel bifocal Newtonian algorithm for identifying pairs of non-unique solutions for any potential data set, prior to any data collection. For the case when the shape of the forward function depends on control parameters that can be tuned to reduce non-uniqueness, we present a second algorithm which minimizes the sum of squared distances between each pair of non-unique solutions. Both algorithms are also relevant in the presence of uncertainty, which we demonstrate by applying them to a simple nonlinear location problem.

(Some figures in this article are in colour only in the electronic version)

1. Introduction

Efficient algorithms and inexpensive computing power have allowed even more complex inverse problems to be solved, often with forward functions that can only be approximated numerically. One problem that remains generally unsolved is, however, that there is no guarantee that the solution found is unique. Analytical results guaranteeing uniqueness exist for some particular classes of problems, for which (Ramm 2005) provides numerous examples, but these are rarely applicable to general nonlinear problems, particularly those in which the forward function has to be approximated numerically. This has led us to look at the problem from a different viewpoint: instead of demonstrating uniqueness we provide a numerical method to demonstrate non-uniqueness of potential solutions to inverse problems, only requiring the forward function to be continuous and differentiable.

The most important applications of inverse problems are in parameter estimation for investigation techniques—e.g. in measurement techniques, imaging, remote location, spectroscopy and many other techniques where parameters of a physical system are determined by observing quantities having a known relation to the parameters of interest (Tarantola and

Valette 1982a). Such applications are, therefore, the main focus of our work and have motivated most of the nomenclature used, even though the problem of non-uniqueness and the method we present are highly relevant to other applications of inverse problems as well.

In any investigation technique, there is always some uncertainty present in the directly observed variables (even though it is not always accounted for). This will imply some degree of non-uniqueness of all solutions of such an inverse problem. Our method applies also to such situations, by providing a way to identify possible non-unique solutions between which the difference is greater than some predefined ‘acceptable’ uncertainty.

Bayesian Monte Carlo methods to solve the inverse problem (Mosegaard and Tarantola 1995) can address some of the problems of having distant non-unique solutions, but—as discussed in section 5—the standard use of Markov chains to create the necessary random samples involved can give erroneous results in some cases that our methods aim to detect. In addition, in many applications, even if all potential solutions can be found, it is often difficult to use them in subsequent analysis if they are not all in some sense clustered together.

Often, the shape of the forward function is dependent on the values of parameters that can be chosen and controlled. Examples of such control parameters might be type and location of sensors used in acquiring data or characteristics of radiation sources (e.g. acoustic or electromagnetic) used to probe the object of study. Modifying control parameters will generally change the distribution and prevalence of non-uniqueness in solutions to the inverse problem. We, therefore, present a systematic approach in the form of an optimization algorithm to minimize a measure of non-uniqueness with respect to the control parameters.

This latter problem has been considered in many forms in the field of statistical experimental design. It is an established branch of statistics (Atkinson and Donev 1992, Pázman 1986), but has not yet been widely applied to any large range of scientific investigation techniques. There are a few exceptions, mainly in pharmacology (Mentré *et al* 1997) and geophysics (Rabinowitz and Steinberg 1990, Steinberg *et al* 1995, Curtis 1999b, Curtis 1999a, Curtis *et al* 2004). One reason for this is that existing theory almost exclusively considers linearized forward functions, thereby failing to account for non-uniqueness due to nonlinear effects. Our approach is different in that it fully accounts for nonlinearity in the forward function and aims to minimize the difference between multiple possible solutions.

The first part of this paper treats the problem of detecting non-uniqueness and can be applied independently of the second part. We begin by presenting the central concept of intersection point pairs (section 2.1), followed by computational strategies for detecting such intersections (section 2.2). We finally extend the concept to account for uncertainty (section 2.4).

In the second part, we present a strategy for reducing inverse problem non-uniqueness. In section 3.1 we use the previous methods to create a quantitative measure of the prevalence of intersection point pairs. Section 3.2 presents a numerical algorithm for the reduction of that quantity.

Section 4 illustrates the new method, applying it to a schematic seismic location problem, whereafter we discuss the results and compare our new method to other, previously described approaches (section 5).

2. Detecting non-uniqueness

2.1. Definitions and formalism

We consider a continuous and differentiable forward function Φ_{ξ} mapping some open space $\Theta \subseteq \mathbb{R}^m$ of parameters to be estimated onto a space $\Omega \subseteq \mathbb{R}^n$ of variables that can be directly observed or whose values can be known in some other way:

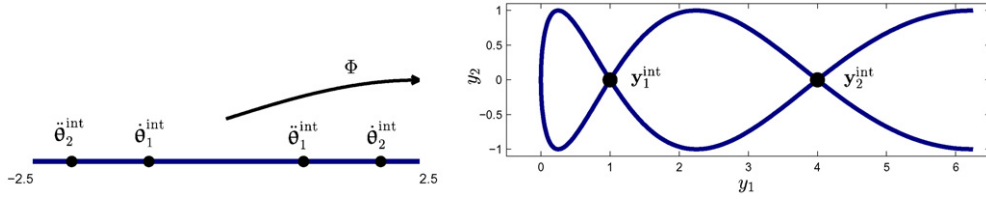


Figure 1. An example of a forward function defined by $y_1 = \theta^2$ and $y_2 = \sin(\theta)$ in the real interval $\Theta = [-2.5, 2.5]$ self-intersecting at $\mathbf{y}_1^{\text{int}} = \Phi(\dot{\theta}_1^{\text{int}}) = \Phi(\ddot{\theta}_1^{\text{int}})$ and $\mathbf{y}_2^{\text{int}} = \Phi(\dot{\theta}_2^{\text{int}}) = \Phi(\ddot{\theta}_2^{\text{int}})$.

Table 1. Notation.

θ	parameters to be determined
Θ	parameter space, an open subset of \mathbb{R}^m containing all possible values of θ
$(\dot{\theta}, \ddot{\theta})$	pair of points in model space (dots used for indexing only)
\mathbf{y}	directly observed variables, or data
Ω	observation space, $\Omega \subseteq \mathbb{R}^n$, contains all possible values of \mathbf{y}
Φ_ξ	forward function, predicting an observation $\mathbf{y} = \Phi_\xi(\theta)$
ξ	control parameters that modify the shape of Φ_ξ and which can be controlled; particular values of ξ are also referred to as a designs
d	distance measure, e.g. $d(\dot{\theta}, \ddot{\theta}) = \text{distance between } \dot{\theta} \text{ and } \ddot{\theta}$
ε_θ	error bound on estimates of θ
ε_Ω	error bound on observations \mathbf{y}
$A_{\mathbf{y}}$	set of possible values of parameters θ given an observation \mathbf{y}
B_θ	set of possible observations \mathbf{y} given θ

$$\mathbf{y} = \Phi_\xi(\theta), \tag{1}$$

where $\theta \in \Theta$ and $\mathbf{y} \in \Omega$. Φ_ξ might also depend on some control parameters ξ whose values can be controlled and set arbitrarily. Typically, Φ_ξ is a mathematical model of the investigation technique predicting what data \mathbf{y} would be observed for any given θ . These and other central definitions made further on are summarized in table 1.

Particularly when Φ_ξ is nonlinear, there might be several points θ in Θ that will map onto the same observation \mathbf{y} . If we denote the range of Φ_ξ on Θ as $\Phi_\xi[\Theta] = \{\mathbf{y} \in \Omega : \mathbf{y} = \Phi_\xi(\theta), \forall \theta \in \Theta\}$, these points will correspond to self-intersections of $\Phi_\xi[\Theta]$ in Ω . We will, therefore, refer to pairs of points $\dot{\theta} \neq \ddot{\theta}$ (dots used for indexing only) for which $\Phi_\xi(\dot{\theta}) = \Phi_\xi(\ddot{\theta})$ as *intersection point pairs* ($\dot{\theta}^{\text{int}}, \ddot{\theta}^{\text{int}}$). An example with a model space $\Theta \subset \mathbb{R}$ and an observation space $\Omega \subset \mathbb{R}^2$ can be found in figure 1.

2.2. Finding intersection point pairs

Detecting and locating intersections is trivial in low-dimensional cases where Φ_ξ can be plotted, but is much more difficult in higher dimension. Assuming that Φ_ξ is continuous and (at least numerically) differentiable, it is however possible to search for intersection point pairs ($\dot{\theta}^{\text{int}}, \ddot{\theta}^{\text{int}}$) using a modified version of Newton’s method for finding roots of equations. Let $\dot{\theta}$ and $\ddot{\theta}$ be an initial guess of an intersection point pair. In order to find a self-intersection of the forward function Φ_ξ in the vicinity of this estimate, two linear approximations of Φ_ξ can be made, one around each of the points $\dot{\theta}$ and $\ddot{\theta}$:

$$\Phi_\xi(\dot{\theta} + \Delta\dot{\theta}) \approx \Phi_\xi(\dot{\theta}) + \mathbf{J}\Delta\dot{\theta}, \quad \mathbf{J} = \frac{\partial \Phi_\xi}{\partial \dot{\theta}}(\dot{\theta}) \tag{2}$$

$$\Phi_\xi(\ddot{\theta} + \Delta\ddot{\theta}) \approx \Phi_\xi(\ddot{\theta}) + \mathbf{J}\Delta\ddot{\theta}, \quad \mathbf{J} = \frac{\partial \Phi_\xi}{\partial \ddot{\theta}}(\ddot{\theta}). \tag{3}$$

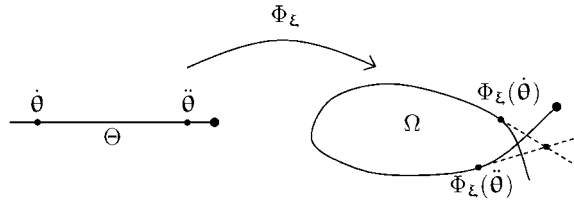


Figure 2. Example of linear approximations (dashed lines) of Φ_ξ around the points $\dot{\theta}$ and $\ddot{\theta}$, the intersection of which provides an estimate for the real intersection.

The following system can then be solved for a pair of increments $\Delta\dot{\theta}$ and $\Delta\ddot{\theta}$ to find the intersection between the two local linear approximations of Φ_ξ :

$$\Phi_\xi(\dot{\theta}) + \mathbf{J}\Delta\dot{\theta} = \Phi_\xi(\ddot{\theta}) + \mathbf{J}\Delta\ddot{\theta}. \quad (4)$$

This can be visualized as calculating the intersection point of the dashed lines in figure 2, an approximation to the true intersection point. If $\Phi_\xi(\theta)$ and θ are represented by column vectors, equation (4) may be rewritten,

$$\begin{bmatrix} \mathbf{J} & | & -\mathbf{J} \end{bmatrix} \begin{bmatrix} \Delta\dot{\theta} \\ \Delta\ddot{\theta} \end{bmatrix} = \Phi_\xi(\ddot{\theta}) - \Phi_\xi(\dot{\theta}). \quad (5)$$

The matrix $\begin{bmatrix} \mathbf{J} & | & -\mathbf{J} \end{bmatrix}$ may be non-square and/or singular, so in the general case equation (5) has to be solved using a pseudo-inverse which may be defined by performing a singular value decomposition,

$$\begin{bmatrix} \mathbf{J} & | & -\mathbf{J} \end{bmatrix} = \sum_{i=1}^k \mathbf{u}_i \sigma_i \mathbf{v}_i^T \quad (6)$$

where $\{\mathbf{u}_i\}$ and $\{\mathbf{v}_i\}$ are k orthonormal column vectors, in $\mathbb{R}^n \supseteq \Omega$ and $\mathbb{R}^{2m} \supseteq \Theta \times \Theta$ respectively, associated with k nonzero singular values $\{\sigma_i\}$. Denoting by $\{\mathbf{v}_i^{\text{null}}\}$ a set of orthonormal vectors spanning the null space of $\begin{bmatrix} \mathbf{J} & | & -\mathbf{J} \end{bmatrix}$, a complete orthonormal basis spanning $\Theta \times \Theta$ is formed by $\{\mathbf{v}_i\} \cup \{\mathbf{v}_i^{\text{null}}\}$.

A generalized solution to equation (5) can then be defined,

$$\begin{bmatrix} \Delta\dot{\theta} \\ \Delta\ddot{\theta} \end{bmatrix} = \sum_{i=1}^k \mathbf{v}_i \sigma_i^{-1} \mathbf{u}_i^T (\Phi_\xi(\ddot{\theta}) - \Phi_\xi(\dot{\theta})). \quad (7)$$

If equation (5) has no solution, equation (7) will give an approximate least-squares solution that minimizes $|\Phi_\xi(\dot{\theta}) + \mathbf{J}\Delta\dot{\theta} - (\Phi_\xi(\ddot{\theta}) + \mathbf{J}\Delta\ddot{\theta})|^2$. In the case of (5) having an infinite number of solutions, equation (7) will give the solution closest to the initial guess $(\dot{\theta}, \ddot{\theta})$.

According to equation (4), the approximate intersection point pair will be $(\dot{\theta}^{\text{int}}, \ddot{\theta}^{\text{int}}) \approx (\dot{\theta} + \Delta\dot{\theta}, \ddot{\theta} + \Delta\ddot{\theta})$. An iterative algorithm is easily created by taking this new point pair as the new initial guess and repeating the procedure from equation (5), repeating until an intersection is found ($d(\dot{\mathbf{y}}, \ddot{\mathbf{y}}) < \varepsilon_\Omega$), or some failure criterion is met (if $d(\dot{\theta}, \ddot{\theta}) < \varepsilon_\Theta$, if not $\dot{\theta}, \ddot{\theta} \in \Theta$ or if some threshold number of iterations has been exceeded). See algorithm 1 for details.

Since the algorithm considers two points in model space simultaneously, we refer to it as a *Bifocal Newtonian algorithm*. It belongs to the class of Newtonian optimization algorithms and has the same well-studied convergence properties (Dennis and Schnabel 1983) as the rest of the class. Newtonian algorithms are widely used in inverse problems to find a particular

solution θ given an observation \mathbf{y} , since they are efficient and simple to implement (Tarantola and Valette 1982b). Our algorithm does something completely different: it updates two points $\dot{\theta}$ and $\ddot{\theta}$ simultaneously in order to find non-unique solutions $\Phi_{\xi}(\dot{\theta}) = \Phi_{\xi}(\ddot{\theta})$, not in relation to one particular observation \mathbf{y} , but also to any possible observation. It is a tool for investigating the properties of the whole forward function Φ_{ξ} with respect to non-uniqueness, not just in relation to one particular observation, but even before any observation has been made.

The partial derivatives of $\partial\Phi_{\xi}/\partial\theta$ will typically be evaluated using some numerical differentiation scheme, the choice of which can influence the behavior of the algorithm. The choice of differentiation scheme can also influence whether or not non-smooth functions will be differentiable at all points.

Algorithm 1 : Bifocal Newtonian algorithm for finding intersections point pairs

Input :

$\dot{\theta}, \ddot{\theta}$ starting guess for model parameters
 ξ control parameters
 $\varepsilon_{\Theta}, \varepsilon_{\Omega}$ precision limits in Θ and Ω , respectively.

Output :

$\dot{\theta}, \ddot{\theta}$ point in $\Theta \times \Theta$ so that
 (a) $d(\dot{\theta}, \ddot{\theta}) < \varepsilon_{\Theta}$ (failure) or
 (b) $d(\Phi_{\xi}(\dot{\theta}), \Phi_{\xi}(\ddot{\theta})) < \varepsilon_{\Omega}$ (success)
 $\{\mathbf{v}_i^{\text{null}}\}$ orthonormal vectors in $\Theta \times \Theta$ spanning intersection space
 $\mathbf{J}, \ddot{\mathbf{J}}$ Jacobians at points $\dot{\theta}$ and $\ddot{\theta}$

while $d(\Phi_{\xi}(\dot{\theta}), \Phi_{\xi}(\ddot{\theta})) > \varepsilon_{\Omega}$ **do**

```

   $\dot{\mathbf{y}} \leftarrow \Phi_{\xi}(\dot{\theta})$ 
   $\ddot{\mathbf{y}} \leftarrow \Phi_{\xi}(\ddot{\theta})$ 
   $\mathbf{J} \leftarrow \frac{\partial\Phi_{\xi}}{\partial\dot{\theta}}(\dot{\theta})$ 
   $\ddot{\mathbf{J}} \leftarrow \frac{\partial\Phi_{\xi}}{\partial\ddot{\theta}}(\ddot{\theta})$ 
  // Perform Singular Value Decomposition of  $[\mathbf{J} \quad | \quad -\ddot{\mathbf{J}}]$ 
   $(\{\mathbf{u}_i\}, \{\sigma_i\}, \{\mathbf{v}_i\}, \{\mathbf{v}_i^{\text{null}}\}) \leftarrow \text{svd}([\mathbf{J} \quad | \quad -\ddot{\mathbf{J}}])$ 
   $\begin{bmatrix} \dot{\theta} \\ \ddot{\theta} \end{bmatrix} \leftarrow \begin{bmatrix} \dot{\theta} \\ \ddot{\theta} \end{bmatrix} + \sum_i \mathbf{v}_i \sigma_i^{-1} \mathbf{u}_i^T (\ddot{\mathbf{y}} - \dot{\mathbf{y}})$ 
  // End loop if  $d(\dot{\theta}, \ddot{\theta}) \rightarrow 0$  or if  $d(\dot{\mathbf{y}}, \ddot{\mathbf{y}})$  diverges
if  $(\dot{\theta}, \ddot{\theta}) \notin \Theta \times \Theta$  or  $d(\dot{\theta}, \ddot{\theta}) < \varepsilon_{\Theta}$ 
|  $\dot{\theta} \leftarrow \ddot{\theta}$ 

```

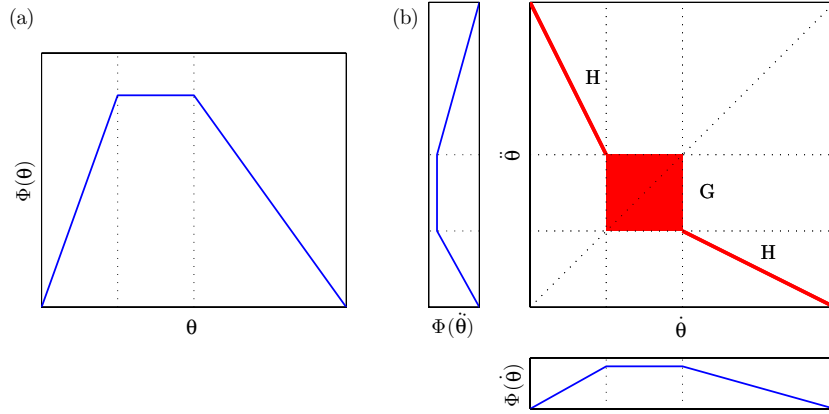


Figure 3. Example of different types of non-uniqueness. (a) A scalar forward function Φ with a plateau region where the Jacobian is singular. (b) The space $\Theta \times \Theta$ of all point pairs $(\hat{\theta}, \tilde{\theta})$ along with compressed versions of (a) aligned with the two axes. The tinted area consists of intersection point pairs $(\hat{\theta}^{\text{int}}, \tilde{\theta}^{\text{int}})$. The square G corresponds to the ‘plateau’ and the line H intersection point pairs with one point on each of the two ‘slopes’ of Φ .

2.3. Relations to locally singular Jacobians

If applied to a linear forward function Φ , algorithm 1 in a single iteration will give an intersection point pair separated by a vector in the null space of Φ , or fail finding an intersection point pair if no null space is present.

Local singularities of the Jacobian \mathbf{J} is a one possible cause for the presence of intersection point pairs—but not the only one, as we have seen above. Figure 3 shows how different causes may contribute to non-uniqueness in a simple one-dimensional example.

Taking the analysis of figure 3 a bit further, one can note that the intersection points pairs outside the square here form a straight unidimensional manifold in the two-dimensional space $\Theta \times \Theta$ (and not simply discrete points as in figure 1). Such manifolds will generally not be straight, but their local tangent plane in $\Theta \times \Theta$ will be spanned by the null space $\{\mathbf{v}_i^{\text{null}}\}$ of the matrix $[\mathbf{J} \quad | \quad -\mathbf{J}]$ at any point $(\hat{\theta}^{\text{int}}, \tilde{\theta}^{\text{int}})$ on the manifold, a fact that can be used to improve the efficiency of sampling such manifolds in higher dimension. A different way of visualizing an intersection manifold in a 2-dimensional model space Θ can be found in figure 6.

2.4. Uncertainty in observations and parameters

If there is uncertainty present in the observation \mathbf{s} , it can be taken into account by adding an additional term ϵ to the right-hand side of equation (1):

$$\mathbf{y} = \Phi_{\xi}(\theta) + \epsilon, \tag{8}$$

where ϵ is a random error that lies within some nonnegative error bound $\epsilon_{\Omega}/2$ so that $\|\epsilon\| < \epsilon_{\Omega}/2$. This implies that the actual observation \mathbf{y} will lie within some neighborhood B_{θ} of $\Phi_{\xi}(\theta)$ within the error bound $\epsilon_{\Omega}/2$:

$$B_{\theta} = \left\{ \mathbf{y} \in \Omega : d(\mathbf{y}, \Phi_{\xi}(\theta)) < \frac{\epsilon_{\Omega}}{2} \right\}, \tag{9}$$

where $d(\mathbf{y}, \Phi_{\xi}(\theta))$ is the distance between measured and predicted observations in Ω .

Having made an observation \mathbf{y} , there will be a set $A_{\mathbf{y}} \subset \Theta$ of possible values of the parameters θ that could have given rise to the observation:

$$A_{\mathbf{y}} = \{\theta \in \Theta : \mathbf{y} \in B_{\theta}\} \tag{10}$$

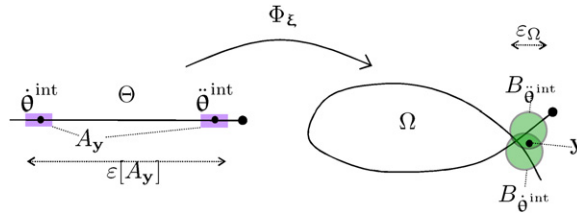


Figure 4. An example of an intersection point pair $(\hat{\theta}^{\text{int}}, \check{\theta}^{\text{int}})$ with sets of possible observations $B_{\hat{\theta}^{\text{int}}}$ and $B_{\check{\theta}^{\text{int}}}$, both with an error bound ε_{Ω} . For one ambiguous observation $\mathbf{y} \in B_{\hat{\theta}^{\text{int}}} \cap B_{\check{\theta}^{\text{int}}}$, the set $A_{\mathbf{y}}$ of possible values of the parameters is depicted along with its error bound $\varepsilon[A_{\mathbf{y}}]$.

The error bound $\varepsilon[A_{\mathbf{y}}]$ on estimates of parameter θ given such an observation will then be the maximum distance between any two points in $A_{\mathbf{y}}$:

$$\varepsilon[A_{\mathbf{y}}] = \text{diam}[A_{\mathbf{y}}] = \sup_{\hat{\theta}, \check{\theta} \in A_{\mathbf{y}}} d(\hat{\theta}, \check{\theta}). \tag{11}$$

It is convenient to set an *acceptable error bound* $\varepsilon_{\Theta} > 0$ for parameter estimates, so that only observations \mathbf{y} for which $\varepsilon[A_{\mathbf{y}}] > \varepsilon_{\Theta}$ have to be considered.

For every pair of points $\hat{\theta}, \check{\theta} \in A_{\mathbf{y}}$ their respective sets of possible observations $B_{\hat{\theta}}$ and $B_{\check{\theta}}$ will intersect so that

$$B_{\hat{\theta}} \cap B_{\check{\theta}} \neq \emptyset. \tag{12}$$

We will therefore refer to pairs of points $(\hat{\theta}, \check{\theta})$ for which $d(\hat{\theta}, \check{\theta}) > \varepsilon_{\Theta}$ and $B_{\hat{\theta}} \cap B_{\check{\theta}} \neq \emptyset$ as *{intersection point pairs}* $(\hat{\theta}^{\text{int}}, \check{\theta}^{\text{int}})$, by analogy with the deterministic case described in section 2.1. As a consequence of equation (9), $d(\Phi_{\xi}(\hat{\theta}^{\text{int}}), \Phi_{\xi}(\check{\theta}^{\text{int}})) < \varepsilon_{\Omega}$. An illustration can be found in figure 4.

The ‘precision limits’ in algorithm 1 are equivalent to error bounds when accounting for uncertainty. The algorithm will return any examples it can find of intersection point pairs with $d(\hat{\theta}^{\text{int}}, \check{\theta}^{\text{int}})$ that exceeds the acceptable error bound ε_{Θ} . If, for a particular design ξ , no two points $\hat{\theta}$ and $\check{\theta}$ further apart than ε_{Θ} give observations $\hat{\mathbf{y}}$ and $\check{\mathbf{y}}$ closer than ε_{Ω} , the algorithm will not find any intersection points.

3. Quantifying and reducing non-uniqueness

3.1. Quantifying non-unique solutions

Every intersection point pair will be separated by a distance $d(\hat{\theta}^{\text{int}}, \check{\theta}^{\text{int}})$. The collection of all intersection point pairs $(\hat{\theta}^{\text{int}}, \check{\theta}^{\text{int}})$ will form a subset Ψ of $\Theta \times \Theta$. One way of measuring the total amount of ambiguity and how different, on average, the ambiguous solutions are is to take the integral of some positive, strictly increasing function f of d over the whole subset Ψ of intersection point pairs:

$$V[\Psi] = \int_{\hat{\theta} \in \Theta} \int_{\check{\theta} \in \Theta} f(d(\hat{\theta}, \check{\theta})) i(\hat{\theta}, \check{\theta}) dp_{\Theta} dp_{\Theta} \tag{13}$$

where $i(\hat{\theta}, \check{\theta}) = 1$ when $(\hat{\theta}, \check{\theta})$ is an intersection point pair, 0 otherwise, and dp_{Θ} is some probability measure representing prior knowledge on Θ so that if $P(X)$ denotes the probability of event X , then $P(\theta \in \Theta) = 1$ and $P(\theta \in A) = \int_{\theta \in A} dp_{\Theta}$ for any measurable subset $A \subset \Theta$.

The integral in equation (13) can be estimated using Monte Carlo methods by,

$$V[\Psi] \approx \frac{1}{N} \sum_{l=1}^N f(d(\hat{\theta}_l, \check{\theta}_l)) i(\hat{\theta}_l, \check{\theta}_l) \tag{14}$$

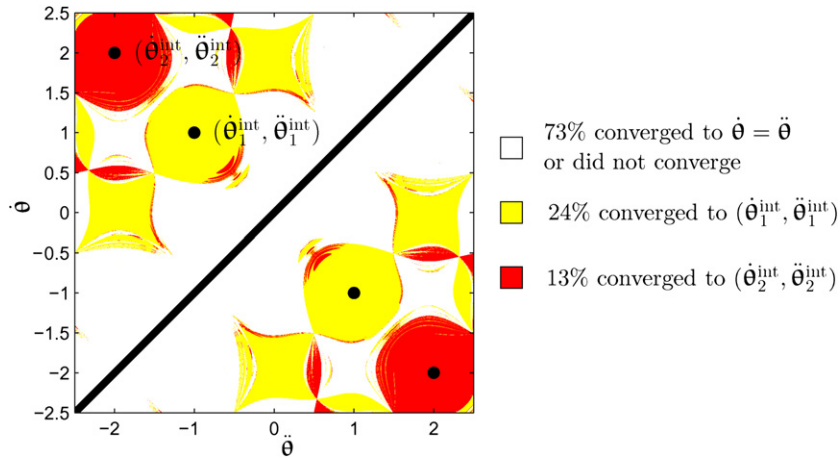


Figure 5. Analysis of convergence of the Bifocal Newtonian algorithm as a function of starting point pair, applied to the forward function in figure 1. The graph shows the set $\Theta \times \Theta$ of all possible point pairs. The sizes of the respective convergence regions give a 99.9% chance of finding both intersections in figure 1 by repeating the algorithm 50 times, each time using a different uniformly random starting point pair.

where $\{\hat{\theta}_l\}$ and $\{\tilde{\theta}_l\}$ are sampled according to the prior probability measure dp_Θ . This may, however, be inefficient since Ψ usually covers a tiny fraction of $\Theta \times \Theta$ so that $i(\hat{\theta}_l, \tilde{\theta}_l) = 0$ for almost all pairs $(\hat{\theta}_l, \tilde{\theta}_l)$ selected according to the prior probability distribution p_Θ .

On the other hand, a similar quality measure that can be evaluated with great efficiency is

$$T[\Psi] = \int_{\hat{\theta} \in \Theta} \int_{\tilde{\theta} \in \Theta} f(d(\hat{\theta}^{\text{int}}, \tilde{\theta}^{\text{int}})) j(\hat{\theta}, \tilde{\theta}) dp_\Theta dp_\Theta, \tag{15}$$

where $j(\hat{\theta}, \tilde{\theta}) = 1$ when the Bifocal Newtonian algorithm, starting at $(\hat{\theta}, \tilde{\theta})$, converges to an intersection point pair $(\hat{\theta}^{\text{int}}, \tilde{\theta}^{\text{int}})$ and $j(\hat{\theta}, \tilde{\theta}) = 0$ otherwise.

Compared to equation (13), each intersection point pair $(\hat{\theta}^{\text{int}}, \tilde{\theta}^{\text{int}})$ in equation (15) will be ‘over sampled’ by a factor proportional to the size of the subset of $\Theta \times \Theta$ where the Bifocal Newtonian algorithm converges to $(\hat{\theta}^{\text{int}}, \tilde{\theta}^{\text{int}})$. This is illustrated in figure 5.

The integral in equation (15) can also be estimated by Monte Carlo methods

$$T[\Psi] \approx \frac{1}{N} \sum_{l=1}^M f(d(\hat{\theta}_l^{\text{int}}, \tilde{\theta}_l^{\text{int}})), \tag{16}$$

where $\{(\hat{\theta}_l^{\text{int}}, \tilde{\theta}_l^{\text{int}})\}$ is the set of M intersection point pairs that were found from $N \geq M$ initial starting point pairs $\{(\hat{\theta}_l, \tilde{\theta}_l)\}$, the latter again sampled according to the prior probability measure dp_Θ .

Minimizing $T[\Psi]$ in equation (15) is equivalent to minimizing the sum of f applied to each of the distances between the intersection point pairs $\{(\hat{\theta}_l^{\text{int}}, \tilde{\theta}_l^{\text{int}})\}$ that have been found. This goal is similar to the strategy of (Curtis and Spencer 1999), but our new algorithm for finding the self-intersection points make this method more computationally tractable.

3.2. Optimizing control parameters to reduce non-uniqueness

We now show how to minimize the uncertainty measure defined in equation (16) with respect to the design ξ for the case of $f(d(\hat{\theta}, \tilde{\theta})) = |\hat{\theta} - \tilde{\theta}|^2$, that is $f(x) = x^2$. Say, $\{(\hat{\theta}_l^{\text{int}}, \tilde{\theta}_l^{\text{int}})\}$

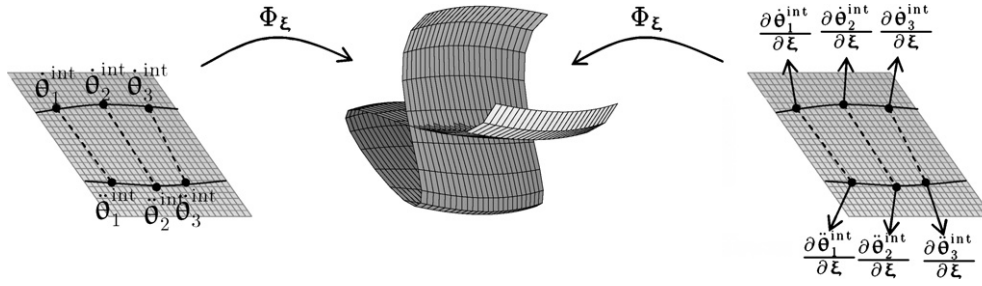


Figure 6. Left: points on intersection manifold; dotted lines connect points that map on to the same point in observation space. Right: evolution direction of intersection points with a small perturbation of ξ .

is a set of M detected intersection point pairs on N attempts, we want to find a design ξ that minimizes equation (16), in this case,

$$T[\Psi] \approx \frac{1}{N} \sum_{i=1}^M |\hat{\theta}_i^{int} - \check{\theta}_i^{int}|^2. \quad (17)$$

Making a linear extrapolation according to $\partial \theta^{int} / \partial \xi$ (how the intersection point pairs move under perturbations of ξ), equation (17) may be rewritten

$$T[\Psi] \approx \frac{1}{N} \sum_{i=1}^M |\mathbf{H}_i \Delta \xi + \Delta \theta_i^{int}|^2, \quad (18)$$

where $\Delta \xi$ represents a change in control parameters ξ , $\Delta \theta^{int} = \hat{\theta}^{int} - \check{\theta}^{int}$, and $\mathbf{H}_i = \frac{\partial}{\partial \xi} \Delta \theta_i^{int}$.

Equating the derivative of equation (18) with respect to $\Delta \xi$ to zero (corresponding to finding an extremum of $T[\Psi]$ with respect to design ξ), one obtains,

$$\sum_{i=1}^M \mathbf{H}_i^T \mathbf{H}_i \Delta \xi = - \sum_{i=1}^M \mathbf{H}_i^T \Delta \theta_i^{int} \quad (19)$$

from which $\Delta \xi$ that minimizes (18) can be calculated:

$$\Delta \xi = - \left[\sum_{i=1}^M \mathbf{H}_i^T \mathbf{H}_i \right]^{-1} \sum_{i=1}^M \mathbf{H}_i^T \Delta \theta_i^{int}. \quad (20)$$

The matrix inverse $^{-1}$ can be a pseudo-inverse in the case that the matrix is singular; it is, however, unlikely to be so as long as the number of control parameters does not approach the number of sampled intersection point pairs.

After a new design has been calculated by $\xi_{\text{new}} = \xi_{\text{old}} + \Delta \xi$, the set of intersection points $\{(\hat{\theta}_i^{int}, \check{\theta}_i^{int})\}$ have to be updated using the Bifocal Newtonian algorithm. If they have not all been eliminated, a new iteration of equation (20) can be performed. Algorithm 2 is an unconstrained iterative optimization algorithm to find a locally optimal design based on this approach.

4. Application: seismic location

In a seismic location experiment, the waves from a seismic event propagate through some inhomogeneous medium for which the structure is assumed to be known. The waves are recorded by detectors at the surface or in boreholes. The seismic event will generate both

pressure (P) and shear (S) waves that travel at different speeds through the medium. It is, therefore, possible to deduce the distance from the detector to the source by recording the difference in arrival times for the two wave types (Tarantola and Valette 1982b). In isolation, the data from each detector will thus yield a manifold of possible source locations, with size dependent on the difference between arrival times for pressure and shear waves (the travel time difference contours in figure 7). Combining the data from several detectors rules out all locations but those on the intersection between the manifolds.

Algorithm 2 : Find locally optimal design ξ

Input :

ξ initial guess for design ξ
 j_{init} number of intersection searches to perform initially
 $j_{\text{incr.}}$ number of new intersection searches after each change in ξ
 $d\xi$ vector of smallest finite differentiation steps in Ξ
 $d\xi^{\text{max}}$ vector of largest finite differentiation steps in Ξ
 $\varepsilon_{\Theta}, \varepsilon_{\Omega}$ precision limits in Θ and Ω respectively

Output : ξ locally optimal design

$\Upsilon \leftarrow \{j_{\text{init}} \text{ random point pairs } (\dot{\theta}, \ddot{\theta}) \sim p_{\Theta}\}$

repeat

```

M ← 0 · ξξT
v ← 0 · ξ
foreach (θ̇, θ̈) ∈ Υ do
  Υ ← Υ − {(θ̇, θ̈)}
  (θ̇, θ̈) ← Bifocal Newtonian algorithm(θ̇, θ̈, ξ, εΘ, εΩ) // (algo. 1)
  if θ̇ ≠ θ̈ then Υ ← Υ + {(θ̇, θ̈)}
  for i ← 1 to k do // Differentiate with respect to ξ
    dξ' ← 0 · ξ
    dξ'_i ← dξ_i/2
    repeat // Assure dξ' is big enough to change (θ̇, θ̈)
      dξ' ← 2dξ' (algo. 1):
      (θ̇', θ̈') ← Bifocal Newtonian algorithm(θ̇, θ̈, ξ + dξ', εΘ, εΩ)
    until d(θ̇, θ̇')2 + d(θ̈, θ̈')2 > 2εΘ2 or dξ'_i > dξ_imax
    g_i ←  $\frac{1}{d\xi'_i}(\thetȧ' - \thetä' - \thetȧ + \thetä)$ 
  for i ← 1 to k do
    for j ← 1 to k do Mij ← Mij + g_iT g_j
    v_i ← v_i + g_iT(θ̇ - θ̈)
  ξ ← ξ - M-1v
  Υ ← Υ ∪ {jincr. random point pairs (θ̇, θ̈) ∼ pΘ}
until convergence or divergence of ξ is assured

```

With an inhomogeneous velocity distribution in the medium, the manifolds are irregular and might intersect at several different locations. In that case there will be several different candidate locations for the true source. It is of great interest to be able to analyze the geometry of the problem to detect such combinations of arrival times that give rise to ambiguous locations, as well as to quantify how distant the multiple possible locations might be from each other. The detector locations might then be chosen to minimize the expectation of this distance, post-experiment. Similar location problems are found in medical imaging and source tracking, general imaging and source location problems in exploration seismology (e.g. for scatterer location or to locate subsurface well trajectories), GPS, emergency beacons, etc.

To illustrate our method clearly, we applied the methods above to design a schematic 2D location problem. The Bifocal Newtonian algorithm was applied repeatedly with different starting points to find pairs of locations $(\hat{\theta}^{\text{int}}, \tilde{\theta}^{\text{int}})$ that would yield the same detected time differences \mathbf{y} between pressure and shear wave arrival time at the detectors. The design algorithm was then applied to find the optimal location of the detectors on the ground surface.

The velocity model of the medium used was a background linear increase in velocity difference $\Delta v = v_p - v_s$ between P and S wave velocity with respect to depth z , plus a significant perturbing random variation ϵ of up to $\pm 1500 \text{ ms}^{-1}$ (figure 7(d)),

$$\Delta v = 2500 [\text{ms}^{-1}] + 1.1z + \epsilon. \quad (21)$$

The S–P travel time difference was calculated using a 2D numerical solution of the Eikonal equation in an elastic medium.

Figure 7(a) shows an example of a typical result of algorithm 1 (repeated many times with different random starting point pairs) for a setup using two receivers separated by 2048 m. Contours of travel time difference for each receiver are shown by curved lines. Straight lines are drawn between each pair of locations that were found to give rise to the same P–S arrival time differences (i.e., if an event occurred at one end of any line, the travel time difference would be indistinguishable from the case where an event occurred at the other end). The precision limits were set to $\epsilon_{\Theta} = 80 \text{ m}$ so that point pairs closer than 80 m are considered to represent ‘acceptable’ ambiguity and $\epsilon_{\Omega} = 5 \text{ ms}$ corresponding to about one per cent uncertainty in arrival time detection. The result is a large number of point pairs in a fan shape between the two detectors, forming an ordered structure of the direction of the maximum uncertainty created by this experimental design.

If a third detector is added in the middle of the existing ones, the ambiguous points are almost completely eliminated, with only a few relatively close point pairs remaining (figure 7(b)).

In order to find the optimal design of an experiment using only two detectors, the design algorithm described above was applied to search for the value of the distance ξ between two detectors that minimizes the sum of the squared distances between all intersection point pairs $(\hat{\theta}^{\text{int}}, \tilde{\theta}^{\text{int}})$ found.

The algorithm converged to a value of $\xi = 1280 \text{ m}$, which gave the distribution of ambiguous location pairs shown in figure 7(c). The straight lines are shorter than those in figure 7(a), showing that if any one location is found post-experiment, although ambiguity still exists the alternative possible location will always be relatively proximal no matter where the true event occurs in the section of the earth considered.

It is important to note that the optimal distance ξ is dependent on the velocity profile of the medium. The optimal design must, therefore, be computed for each velocity profile individually.

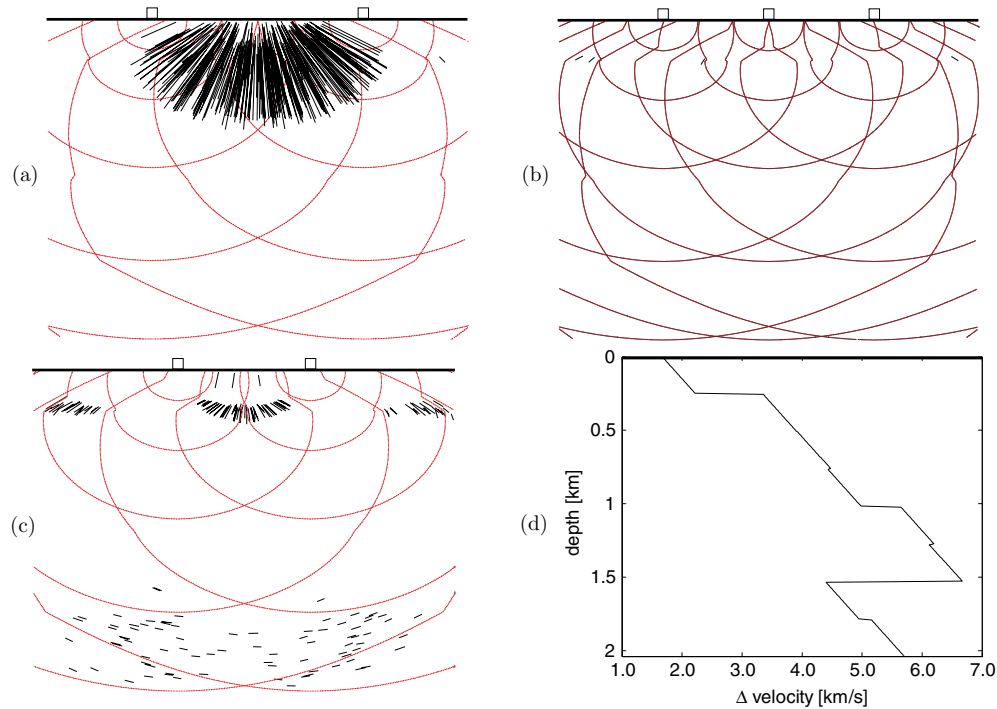


Figure 7. Surface detectors (squares) for a source location problem, each with P–S travel time difference contours (light lines). Bold straight lines connect pairs of ambiguously resolved seismic source locations, sampled randomly using the Bifocal Newtonian algorithm. (a) Initial two-detector configuration. (b) Three-detector configuration, eliminating almost all location ambiguity. (c) Optimized two-detector configuration, eliminating most of the location ambiguity. (d) Velocity profile of the medium.

5. Discussion and conclusions

We have presented two novel algorithms: the first one can be used to sample the set of non-unique solutions to an inverse problem prior to any data being acquired, and the second to reduce the prevalence of these non-unique solutions. These algorithms are applicable not only to location problems, but to a wide range of investigation techniques in physics, chemistry and biology.

The algorithms are designed to optimize methods of parameter estimation, where the result will be represented as a best fit value of the parameters θ with error bound ε_{θ} . The rather crude error bound representation of uncertainty is a necessity for the intersection approach to work. If one wishes to account for full probability distributions when designing an investigation technique, more traditional Monte Carlo methods are necessary but their numerical efficiency is orders of magnitude lower than the methods we present.

In the field of statistical experimental design, approaches for minimizing the same squared distance between the most distant ambiguous solutions to an inverse problem have been developed for the case when the forward function Φ_{ξ} is linear. This is generally referred to as *E*-optimality criterion (Pázman 1986, p 90).

There is some literature on what is generally referred to as Bayesian experimental design considering nonlinear models (Chaloner and Verdinelli 1995), which one might be tempted to believe addresses the problem of parameter estimation for nonlinear models in general. This

is unfortunately not the case, since it is exclusively based on the assumption that there are no self-intersections of Φ_ξ , and thus no increase in uncertainty due to nonlinearity (Curtis and Spencer 1999, van den Berg *et al* 2003).

A limited amount of work has been done that is applicable to the full nonlinear case, optimizing designs with respect to the expected gain in Shannon information—an approach introduced by (Lindley 1956). Such a measure of non-uniqueness does not take into account the distance between multiple solutions, but rather relates only to how many there are (or how large total volume they occupy in model space). Algorithms for the numerical evaluation of expected gains in Shannon information in the fully nonlinear case have been studied by (Müller 1999, van den Berg *et al* 2003), as well as (Ryan 2003). If applied to the location problem of section 4, such a measure would lead to reduction of uncertainty in the location of each point of the point pairs in figure 7(a), rather than reducing the inter-pair distance as our method achieves in figure 7(c).

For probabilistic inverse problems, Markov Chain Monte Carlo (MCMC) methods for sampling the posterior probability distribution are nowadays widely used beside more traditional methods. If the samples created by these methods truly follows the posterior distribution, they will pick up any distant and separate probability peaks—corresponding to non-uniqueness in the situations discussed above. The way most MCMC sample generators work will, however, not generate a ‘true’ sample in these cases, since each sample generated in the sequence is usually in some sense close to the previous sample. Thus, if there is a large gap between different probability peaks, there is relatively little chance that the sampler will ‘jump’ from one peak to the other within a reasonable sampling time. So, even if MCMC methods are to be used to solve particular instances of an inverse problem, it might be useful to try to detect and reduce the occurrence of distant ambiguous parameter values in order to make the MCMC algorithm both efficient and more accurate within a finite computing effort.

Optimal designs using our new method naturally tradeoff estimated uncertainty due to self-intersections of Φ_ξ and uncertainty due to the local Jacobian of Φ_ξ becoming poorly conditioned. Figure 7(c) provides an example of this: while the intersection-related ambiguous pairs close to the surface get closer when moving the detectors closer together, new point pairs appear deeper down due to the travel time contours becoming more and more parallel.

In conclusion, the numerical efficiency of the algorithm allows for its application to far more complex problems than the 2D-location example presented here. We believe it has a great potential to help in designing experiments using other investigation techniques of higher dimension, especially since more self-intersections are likely to appear the higher the dimensionality of the problem, thus requiring a design method that takes their effects into account.

Acknowledgments

Parts of this work were carried out under the HPC-Europa project (RII3-CT-2003-506079), with the support of the European Community-Research Infrastructure Action under the FP6 ‘Structuring the European Research Area’ Programme. The Eikonal solver was provided by Enru Liu of the British Geological Survey.

References

- Atkinson A C and Donev A N 1992 *Optimum Experimental Designs* (Oxford: Oxford University Press)
- Chaloner K and Verdinelli I 1995 *Stat. Sci.* **10** 273–304

- Curtis A 1999a *Geophys. J. Int.* **139** 205–15
- Curtis A 1999b *Geophys. J. Int.* **136** 637–50
- Curtis A, Michelini A, Leslie D and Lomax A 2004 *Geophys. J. Int.* **157** 595–606
- Curtis A and Spencer C 1999 ‘69th Ann. Int. Meeting’ Expanded Abstracts Society of Exploration Geophysics pp 1775–8
- Dennis J and Schnabel R 1983 *Numerical Methods for Unconstrained Optimization and Nonlinear Equations* (Englewood Cliffs, NJ: Prentice-Hall)
- Lindley D V 1956 *Ann. Math. Stat.* **27** 986–1005
- Mentré F, Mallet A and Baccar D 1997 *Biometrika* **84** 429–42
- Mosegaard K and Tarantola A 1995 *J. Geophys. Research-Solid Earth* **100** 12431–47
- Müller P 1999 *Bayesian Statistics, 6 (Alcober, 1998)* (New York: Oxford University Press) pp 459–74
- Pázman A 1986 *Foundations of Optimum Experimental Design (Mathematics and its Applications (East European Series) vol 14)* (Dordrecht: D. Reidel Publishing Co.)
- Rabinowitz N and Steinberg D 1990 *Bull. Seismol. Soc. Am.* **80** 187–96
- Ramm A G 2005 *Inverse Problems: Mathematical and Analytical Techniques with Applications to Engineering* (Boston: Springer)
- Ryan K J 2003 *Journal Comput. and Graphical Statistics* **12** 585–603
- Steinberg D, Rabinowitz N, Shimshoni Y and Mizrahi D 1995 *Bull. Seis. Soc. Am.* **85** 1847–57
- Tarantola A and Valette B 1982a *Rev. Geophysics* **20** 219–32
- Tarantola A and Valette B 1982b *J. Geophys. Z. Geophys.* **50** 159–70
- van den Berg J A, Curtis A and Trampert J 2003 *Geophys. J. Int.* **55** 411–21