Optimal Experimental Design for Nonlinear Models with Non-Gaussian Uncertainties

Emanuel Winterfors[1,2] and Andrew Curtis[2]

(1) Université Pierre et Marie Curie
Laboratoire Jacques-Louis Lions
75252 Paris Cedex 05, FRANCE
winterfors@ann.jussieu.fr

(2) Edinburgh University
Department of Geology and Geophysics
Grant Institute, West Mains Road
Edinburgh EH9 3JW, UK75252

Table 1. Notation

θ	parameters
y	observation (data)
ξ	design
$p(\theta)$	prior probability density
$p(y\,\vert\,\theta,\xi)$	conditional observation density given θ and ξ
$p(\theta\,\vert\,y,\xi)$	posterior probability density, given observation y and design ξ
$d^2(\dot\theta,\ddot\theta)$	squared distance between $\dot\theta$ and $\ddot\theta$, dots used for indexing only

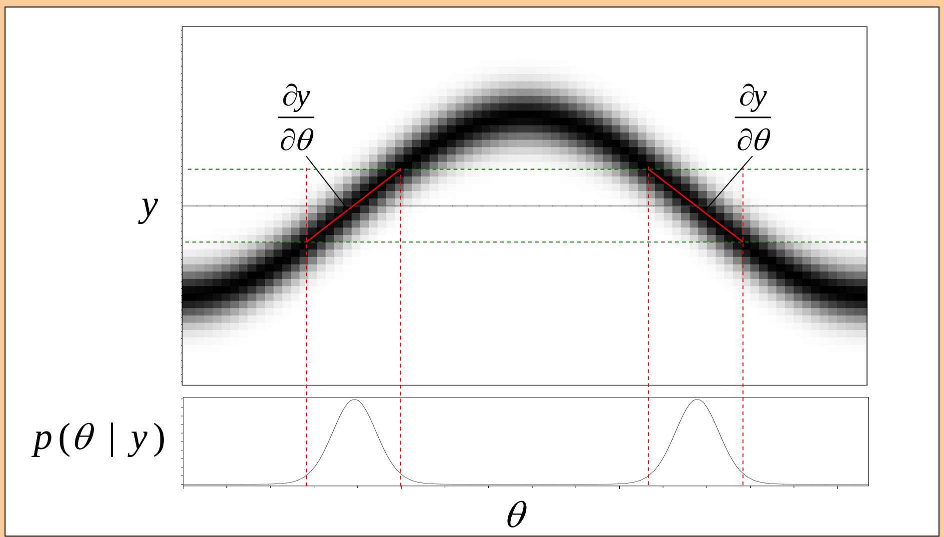
The Problem with Information Matrices in Nonlinear Problems

The most common approach to nonlinear experimental design is to use an optimality criterion (D, A or E etc.) on the derivative-based information matrix, which is averaged over a prior probability distribution $p(\theta)$ on the parameter space.

Information matrices may however fail to describe the uncertainty in parameter estimates. In nonlinear problems. Consider a simple nonlinear observation-parameter

$$y = \eta(\theta,\xi) = \sin(\theta) + \varepsilon$$

where ε is a Gaussian observational uncertainty.



In the figure above, shades of gray represent probability density over the joint parameter-observation space. The lower curve represents a posterior probability density $p(\theta\,|\,y)$ over the range of parameter θ, given an observation y.

Information matrices will give a good approximation of the width of each of the two peaks in $p(\theta\,|\,y)$ *separately*. However, the *total* width of $p(\theta\,|\,y)$ is orders of magnitude greater than that of each of the peaks (Curtis, 2004). Thus, **information-matrix based methods may fail in predicting the true uncertainty of parameter estimates in nonlinear problems**.

Solution: A Non-approximative Approach

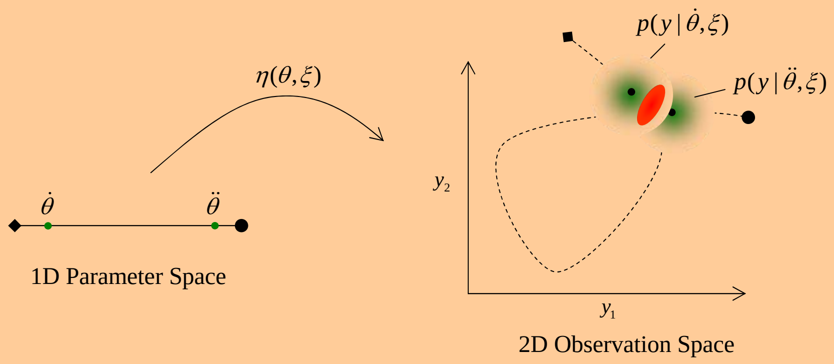
The situation to avoid when designing an experiment is to have very **different parameter values that can give rise to the same observations**. One measure of this capacity for the parameter values $\dot\theta$ and $\ddot\theta$ is

$$R(\dot\theta,\ddot\theta) \;=\; d^2(\dot\theta,\ddot\theta)\int_{y\in\Omega} p(y\,|\,\dot\theta,\xi)\,p(y\,|\,\ddot\theta,\xi)\,dy \quad ,$$

with $d^2(\dot\theta,\ddot\theta)$ and $p(y\,|\,\theta,\xi)$ as defined in table 1. The integral can be evaluated analytically, as long as the observational uncertainty is known in closed form. An example of a set of observations that can result from both $\dot\theta$ and $\ddot\theta$ is marked in red in the figure below.

In order to access the overall ambiguity $Q(\xi)$ in parameter estimates, it is necessary to compute the expectation of $R(\dot\theta,\ddot\theta)$ over all pairs of points $(\dot\theta,\ddot\theta)$ in parameter space

$$Q(\xi) \;=\; \int_{\dot\theta\in\Theta}\int_{\ddot\theta\in\Theta} p(\dot\theta)\,p(\ddot\theta)\,R(\dot\theta,\ddot\theta)\,d\dot\theta\,d\ddot\theta \quad .$$



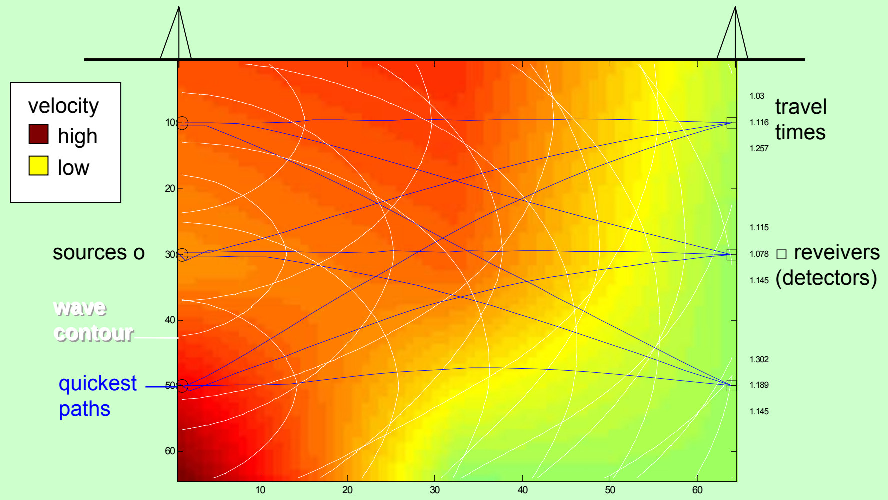
A simple interpretation of $Q(\xi)$ can be obtained by applying Bayes' rule and changing the order of integration:

$$Q(\xi) \;=\; 2\int_{y\in\Omega} p^2(y\,|\,\xi)\,\mathrm{Var}\big[\Theta\,|\,y,\xi\big]\,dy \quad ,$$

where $\mathrm{Var}\big[\Theta\,|\,y,\xi\big]$ is the variance of the posterior probability density $p(\theta\,|\,y,\xi)$. $Q(\xi)$ is thus the expected variance of the posterior, weighted by the marginal distribution $p(y\,|\,\xi)$ in observation space.

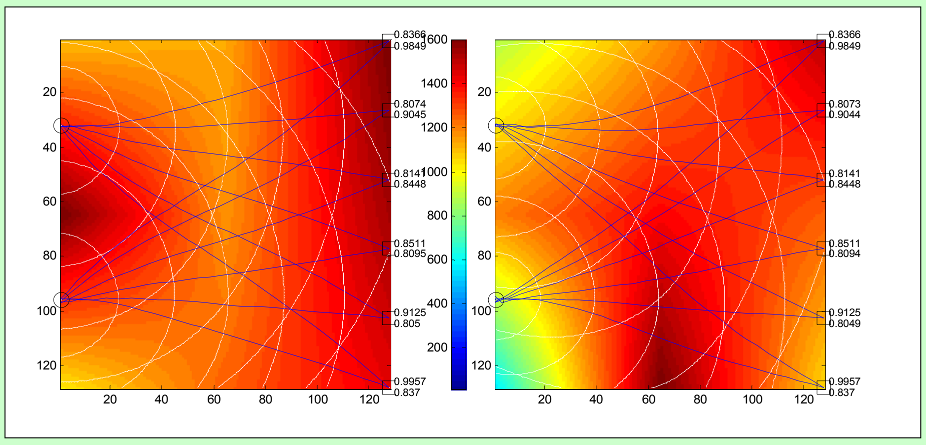
Application: Travel Time Tomography

Travel time tomography consists of calculating a velocity distribution in some heterogeneous medium by recording the travel times of pulses traversing the medium in different directions. The problem is high dimensional and also, due to refraction, strongly nonlinear. The figure below shows a typical application in geophysics, where one seeks to know the velocity profile of the subsurface section between two boreholes equipped with seismic wave sources and receivers.



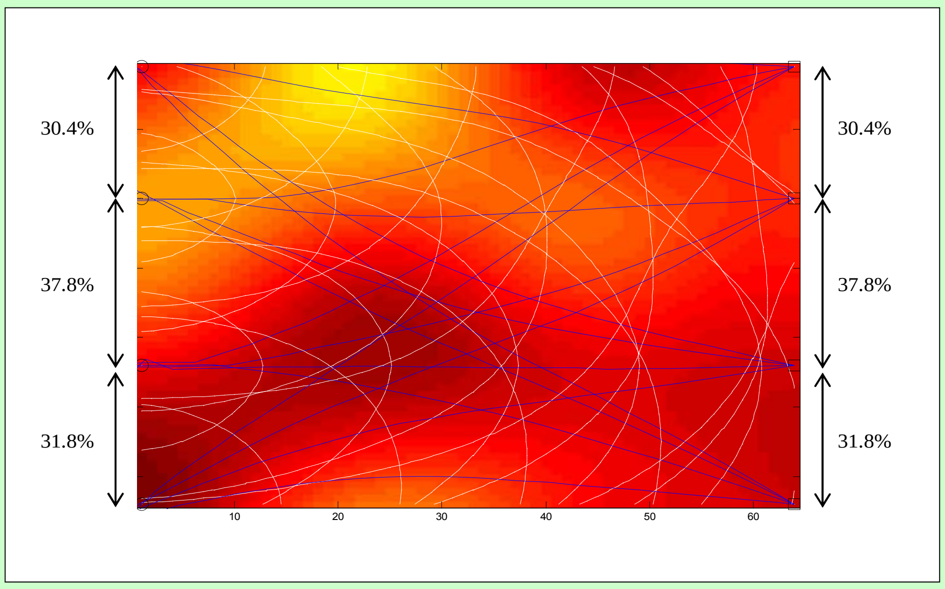
Nonlinearity gives multiple solutions to one set of travel times

Even though the information matrix is well-conditioned for all wave velocity profiles, there are multiple solutions to some sets of observed travel times (Winterfors and Curtis, 2008).



Optimal design for 4-source, 4-receiver setup

Estimating $Q(\xi)$ using Monte Carlo methods (similar to Müller, 1999), optimal source and receiver locations ξ were obtained using a simple steepest descent optimisation algorithm. The velocity profile θ was represented using 3×3 parameters.



References

Curtis A 2004. Theory of model-based geophysical survey and experimental design Part B - Nonlinear Problems *The Leading Edge* 23(10) pp 1112–1117

Müller P 1999. Simulation-Based Optimal Design *Bayesian Statistics 6 (eds Bernardo et al.)* pp 459–74

Winterfors E and Curtis A 2008. Numerical Detection and Reduction of Non-Uniqueness in Nonlinear Inverse Problems *Inverse Problems* 24(2) pp 025016