Optimal Experimental Design (Survey)

Sudelfeld, Germany, March 2018

Andrew Curtis

University of Edinburgh
Guest Professor ETH Zurich
Survey and Experimental Design Methods

- Can not avoid: potentially one of the most common tasks
- Applied in many other fields of work and research
- Many geophysical surveys are designed in large part using tried and tested rules of thumb – heuristics
- Heuristics are generally robust, but not optimal: far more sophisticated theory exists and is used in other fields – *Statistical Experimental Design*
Experimental Design Methods

- Linear(ised) m-d relationships
  - Unfocussed design (cross-well tomography)
  - Focussed design (cross-well tomography)
  - Sequential design (microseismic location)

- Nonlinear m-d relationships
  - Bayesian design (microseismic location)
  - Maximum entropy design (AVO – for reflection data)

→ Curtis (2004 a,b) - *The Leading Edge - Tutorial*

→ Maurer, Curtis & Boerner (2010) – Review
Designing Experiments to Constrain Parameters

Experiments should be designed such that:

– They can be conducted in practice

– *Expected* post-experimental model parameter uncertainties are minimised

– Costs are minimised/constrained
How does experimental design work?

Designs the *relationship* between a set of parameters and some data

\[ F(m) = d \]
Linear Experimental Design

How fast do waves propagate through this homogeneous block of material?
Linear Experimental Design

How fast do waves propagate through this homogeneous block of material?
Linear Experimental Design

Is this region of the Earth heterogeneous?
Linear Experimental Design
Linear Experimental Design

\[ S_1 \quad S_3 \]
\[ S_2 \quad S_4 \]
Linear Experimental Design

\[
\begin{array}{c|c|}
S_1 & S_3 \\
\hline
S_2 & S_4 \\
\end{array}
\]
Linear Experimental Design

$S_1$, $S_2$, $S_3$, $S_4$

Normalised Eigenvalues

Eigenvalue Number
Linear Experimental Design

(a)

S

1

S

3

S

2

S

4

Normalised Eigenvalues

0 1 2 3 4

Eigenvalue Number
Linear Experimental Design

(a)  

(b)  

Normalised Eigenvalues

Eigenvalue Number
Linear Experimental Design
The Survey on the Right…

- Obtains twice as many independent pieces of information as the survey on the left, using the same number of data
- Is nominally carried out at half the cost (2 sources + 2 receivers)

General Points

- Eigenvalues specify precisely how many pieces of information can be constrained \textit{in principle} (no data uncertainties yet!)
- For each e-value, corresponding e-vector describes precisely the associated \textit{independent} piece of information
  
  ➔ E-system allows us to create measures of design quality…
Linear Experimental Design

(a)

(b)

\[
\Phi(S) = \sum_{i} \frac{\lambda_i}{\lambda_1} \quad \text{or} \quad \Phi(S) = \prod_{i} \lambda_i
\]
Linear Experimental Design

Eigenvalue = gradient squared
Unfocused Crosswell Example

Parameters = cell slownesses
Unfocussed Crosswell Example
Unfocussed Crosswell Example
Unfocussed Crosswell Example
Unfocussed Crosswell Example

Eigenvalue Spectra

(100 cells, 400 paths)

- Normalised Log Eigenvalue
- Eigenvalue Rank

- Regular Survey
- Optimum Survey
Unfocussed Crosswell Example

All long paths: all data average >9 s_i

Greatest path density in centre

Haven’t used cheap freedom at surface

Some short paths

Increasing density with depth
(since only longer paths in centre)

But how dense? Exactly where?
Linear Experimental Design

Eigenvalues: how much information

Eigenvectors: specifically what information

Two Further Possibilities:
Hence, can design model parameterisation using eigenvalues or, can focus information on a model subspace by only maximising eigenvalues of eigenvectors (information) spanning that subspace.
Linear Experimental Design

Eigenvalues: *how much* information

Eigenvectors: specifically *what* information

Two Further Possibilities:

… can design model parameterisation using eigenvalues

or, can *focus* information on a model subspace by only maximising eigenvalues of eigenvectors (information) spanning that subspace.
Linear Experimental Design

Is this region of the Earth heterogeneous?
Linear Experimental Design

Optimise Data Acquisition

Optimise Model Parameterisation
Focussed Crosswell Example

**Shading** shows diagonal elements of Resolution matrix (max. possible = 1)

**Red crosses** mark cells spanning model subspace of interest.

How dense? **Exactly** where?

Not possible to design using intuition alone

→ Need to solve **Optimisation** Problems…

→ Above, a *genetic algorithm*, next *sequential design*…
Microseismic (Isotropic, velocity gradient)

- Method to locate micro-seismic events
- Acoustic or vector sensors in well or on ground surface detect wave energy
- Use arrival times or waveform data to locate/characterise events
- Useful to monitor micro-fracturing of rock due to fluid injection (e.g., geothermal, reservoir stimulation, CO$_2$ storage) or due to production
**Microseismic (Isotropic velocity gradient)**

- **Method:** Linear-dependence reduction

---

**Constant velocity gradient**
Microseismic (Isotropic velocity gradient)

- **Method**: Linear-dependence reduction

1. Begin with all possible sensor (o) and representative event (*) locations
Microseismic (Isotropic velocity gradient)

- **Method: Linear-dependence reduction**

1. Begin with all possible sensor (o) and representative event (*) locations

2. Sensor that adds least information (most linearly dependent on other rows) is removed, sequentially

3. Repeat 2. until required number of sensors remain
Sensor Rankings Focused on Event Offset, with Attenuation
Summary of Lecture (so far)

- Linearised design methods allow eigenvalues (gradients) to be maximised – **NEED NOT CALCULATE E-VALUES**
- Linearised algorithms are efficient and well understood
  - Dependence-reduction JAVA algorithm available
- **Linearised** design methods may be inadequate for significantly **nonlinear** problems (the *linearisation error*)
- Possible strategies:
  - Vary design to ‘linearise’ the problem!
  - Fully nonlinear quality measure (e.g., max. entropy)
Entropy-Based Design

\[
\text{Ent}(D, M) = \text{Ent}(D | \pi) + E_D[\text{Ent}(M | D, \pi)]
\]

Parameters M Data D Design = \pi

To maximise expected information, minimise w.r.t. \( \pi \):

\[
E_D[\text{Ent}(M | D, \pi)]
\]

If the left hand side is constant w.r.t. \( \pi \), then identity above shows that equivalently we can vary \( \pi \) to maximise:

\[
\text{Ent}(D | \pi)
\]

This is true for nonlinear regression with errors indep. of \( \pi \):

\[
D | m, \pi = f(m, \pi) + \varepsilon
\]
Generalising Linear Design

**Experiment $\pi_1$**

**Experiment $\pi_2$**
Generalising Linear Design

$$E_M[\text{Ent}(M \mid D, \pi)]$$

Experiment $\pi_1$

Experiment $\pi_2$
Generalising Linear Design

\[ \text{Ent}(D \mid \pi) \]

Experiment \( \pi_1 \)  
\[ P(d \mid \pi_1) \]
\[ P(m) \]

Experiment \( \pi_2 \)  
\[ P(d \mid \pi_2) \]
\[ P(m) \]

\[ \text{Ent}(D \mid \pi_1) < \text{Ent}(D \mid \pi_2) \]
Generalising Linear Design

\[ \text{Ent}(D \mid \pi) \]

Experiment \( \pi_1 \)

\[ P(d \mid \pi_1) \]

\[ P(m) \]

\[ \text{Ent}(D \mid \pi_1) < \text{Ent}(D \mid \pi_2) \]

Experiment \( \pi_2 \)

\[ P(d \mid \pi_2) \]

\[ P(m) \]
Microseismic receiver design

$m_k \sim p(m)$
Heuristic (contractor)

$\log_{10}$ expected post-inversion variance in hypocentral estimates
Current / Future Research
Inverse Theory

Model $m$

Parameters $\theta$

Data $y$

Sample Space $Y$
Inverse + Design Theory

Model $m$

Parameters $\theta$

Data $y$

Sample Space $Y$

Design of Data Acquisition
Inverse + Design Theory

Our Questions!

Model \( m \)
Parameters \( \theta \)

Data \( y \)
Sample Space \( Y \)

Design of Parametrization

Design of Data Acquisition

What is fixed??

Where are they?
Interrogation Theory
Interrogation Theory

- **Models** $m$
- **Parameters** $\theta$
- **Experiment** or **Elicitation**
- **Recorded Data** $y_d$
- **Forward Problem**
- **Inverse Problem**
- **Decision Problem**
- **Experimental Design** $d$
- **Sample Space** $Y$
- **QUESTIONS Q**
- **ANSWERS A**
Problems to Solve – Tonight!
Problems to Solve – Tonight!

- **Surface arrays**… what spacing for Gradiometry, Noise,…?
Problems to Solve – Tonight!

- **Surface arrays**… what spacing for Gradiometry, Noise,…?

- Nested arrays: what spacings? Which locations?
Problems to Solve – Tonight!

- **Collocated** seismometers + **rotations**: now what spacing?

- Nested arrays: what spacings? Which locations?
Problems to Solve – Tonight!

- **Collocated** seismometers + rotations: now what spacing?

- Nested arrays: what spacings? Which locations?

- What about 3D arrays (pyramids, boreholes, etc.)?
Problems to Solve – Tonight!

- Collocated seismometers + rotations: now what spacing?

- If we separate seismometers and rotations, irregularly
  – what relative spacings/locations maximise information?

⇒ What Questions do we want to ask, and answer?
Optimal Experimental Design (Survey)

Andrew.Curtis@ed.ac.uk

www.geos.ed.ac.uk/homes/acurtis

Curtis (2004 a,b) – Tutorial (Linear + Nonlinear)