Experiment Design

Daniel Alexander

Overview

- What is experiment design?
- Fisher information and optimality criteria
- Linear and non-linear models
- Examples
- Variations of the problem
- The modelling pipeline

Problem statement

- Minimize the bias and uncertainty in the parameter estimates $\tilde{x}_1,\ldots,\tilde{x}_N$ with respect to the choice of sample points, $\xi = \{\mathbf{y}_1, \ldots, \mathbf{y}_K\}$.
- Eg, for a single unknown parameter, x, minimize the standard deviation $\sigma(\tilde{x})$ with respect to ξ, subject to $E(\tilde{x} - x_T) = 0$ where x_T is the true parameter value, and $|\xi| = K$.

Broader context

- Classical experiment design
  - Randomization
  - Grouping
  - Making the right decision
- Related problems
  - Active learning
  - Optimal decision theory

General strategy

- Assume an unbiased estimator.
- Quantify or estimate $\sigma(\tilde{x})$ as a function of $\mathbf{y}_1, \ldots, \mathbf{y}_K$.
- Use optimization to find the $\mathbf{y}_1, \ldots, \mathbf{y}_K$ that minimize the estimate of $\sigma(\tilde{x})$

Fisher information

- The general Fisher matrix definition is
$$F = -E\left(\frac{d^2}{d\mathbf{x}^2}\log p(\mathbf{A} \mid \mathbf{x})\right)$$
- For Gaussian noise, this reduces to
$$F_{ij} = \sum_{k=1}^{K} \frac{1}{\sigma_k^2} \frac{\partial S(\mathbf{x},\mathbf{y}_k)}{\partial x_i} \frac{\partial S(\mathbf{x},\mathbf{y}_k)}{\partial x_j}$$
- F^{-1} is the expected covariance on $\mathbf{x}$.

Linear models

- For the simple linear least squares problem $\mathbf{A}=G\mathbf{x}$.
- The Fisher matrix $F = G^T G$.

Alphabetical optimality criteria

- **A-optimality** minimizes the trace of F^{-1}.
- **D-optimality** minimizes the determinant of F^{-1}.
- **E-optimality** maximizes the minimum eigenvalue of F.
- **T-optimality** maximizes the trace of F.

Weighted linear least squares

- Information matrix becomes $F=G^T WG$
- Where $W_{ii} = \sigma_i^{-2}$

Non-linear estimation

- Unlike linear models, F now depends on $\mathbf{x}$.
- Average over a representative set of $\mathbf{x}$.

Practical considerations

- Scale and units of parameters
  - Choose units to equalize numerical scale
  - Normalize F by $\mathbf{x}\mathbf{x}^T$.
- Prior choice of parameter settings
- Optimization algorithm

Example 1

- Experiment design for dynamic contrast enhanced computed tomography.
  - Prevost et al ISBI 2010.
- Inject tracer
- Image repeatedly to monitor uptake of tracer in tumour; depends on vessel permeability.
- Permeability informs cancer grade.

The signal model

$$v_1 \frac{dC_1}{dt} = \phi(C_a(t) - C_1(t)) - PS(C_1(t) - C_2(t))$$

$$v_2 \frac{dC_2}{dt} = PS(C_1(t) - C_2(t))$$

$$R(t) = A\exp(\alpha t) + (1-A)\exp(\beta t)$$

$$I(t) = \phi \int_0^t R(\tau) C_a(t-\tau)\, d\tau$$

A, α and β depend on the parameters of interest: PS, v_1, v_2 and ϕ.

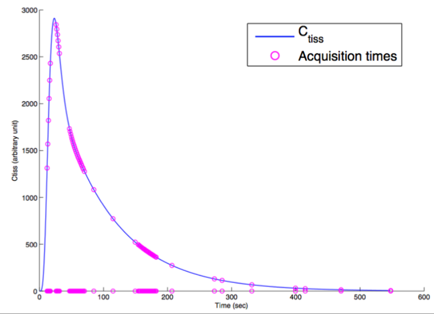
General optimized experiment design

- 70 time points
- Min 10s between consecutive acquisitions
- Find 70 values of t that minimize

$$V = \sum_{i=1}^{4} p_i^{-2} [F^{-1}]_{ii}, \text{ where } p_1 = v_1, p_2 = v_2, p_3 = \phi, p_4 = PS.$$

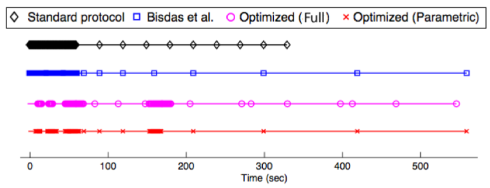
- A-priori settings:
  - PS=17 ml/min/100g, v_1=5 ml/100g, v_2=20 ml/100g and ϕ=70 ml/min/100g.
  - Typical for malignant tumour.
- Optimize using SOMA

General optimized experiment design



Parametrized optimization

- Four groups of consecutive measurements
- One group of increasingly spaces measurements.



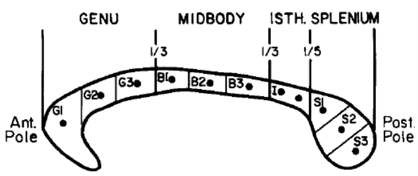
Standard protocol	Bisdas et al. protocol	Full optimized	Parametric optimized
9.44	10.9	9.03	6.81

Evaluation

Protocol & nb of scans (% of param. variation)	CV (Max/Mean)	Error (Max/Mean)
Standard 70 scans (0%)	17.8 %	2.3 %
Standard 70 scans (10%)	19.2% / 17.7%	2.8% / 2.3%
Standard 70 scans (20%)	21.3% / 17.9%	3.4% / 2.5%
Bisdas 70 scans (0%)	14.3%	1.8%
Bisdas 70 scans (10%)	14.7% / 13.8%	2.2% / 1.7%
Bisdas 70 scans (20%)	15.5% / 14%	2.0% / 1.7%
Optimized 70 scans (0%)	10.0%	0.8%
Optimized 70 scans (10%)	11.7% / 10.5%	1.1% / 0.9%
Optimized 70 scans (20%)	12.6% / 10.9%	1.6% / 1.0%
Optimized 55 scans (0%)	10.6%	0.8%
Optimized 55 scans (10%)	12.4% / 11.6%	1.0% / 1.0%
Optimized 55 scans (20%)	13.9% / 12.0%	1.6% / 1.1%
Optimized 35 scans (0%)	14.1%	1.3%
Optimized 35 scans (10%)	17.1% / 15.1%	1.4% / 1.1%
Optimized 35 scans (20%)	17.3% / 14.9%	2.0% / 1.4%

Example 2

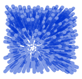
- Axon diameter estimation in diffusion MRI
  - Alexander MRM 2008
  - Drobnjak et al JMR 2010



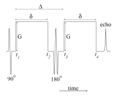
Aboitiz Brain Res. 92

Alexander MRM 2008

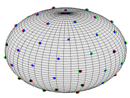
- Aim: to measure axon diameter with diffusion MRI.
- Model: zeppelin and cylinder



- Measurement: HARDI PGSE



- Constraints: 360 measurements (1 hour); unknown fibre orientation.

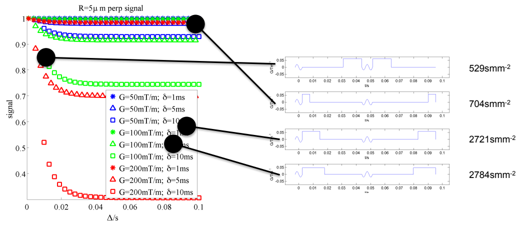


Optimal experiment design

- Take M shells of N directions, eg M=4, N=90

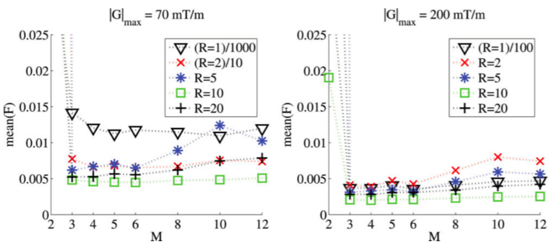
$$V = \sum_{i=1}^{4} p_i^{-2}[F^{-1}]_{ii}, \text{ where } p_1 = d_\parallel, p_2 = d_\perp, p_3 = f, p_4 = R.$$

- SOMA optimized (1 day for each of 5 runs)



Angular v radial resolution

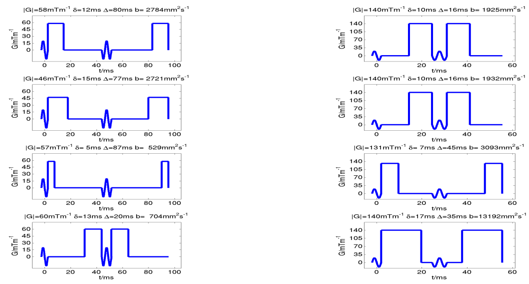
- V minimum for M=4, N=90.



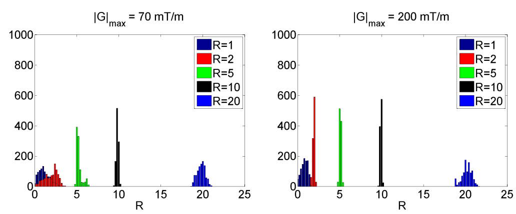
Optimised designs

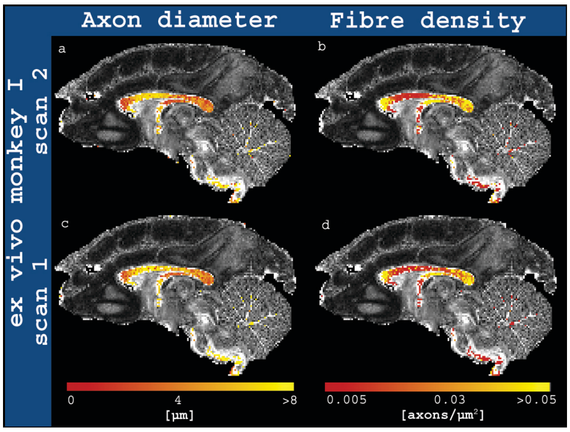
In-vivo: Gmax = 60mT/m

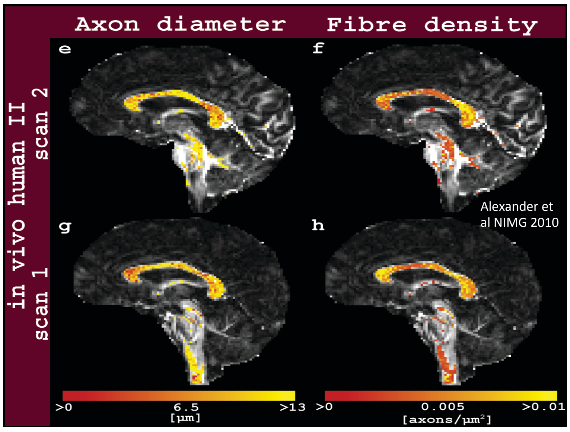
Fixed: Gmax = 140mT/m



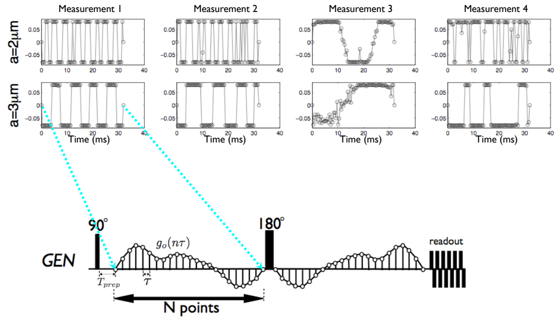
Posterior distributions on *R*





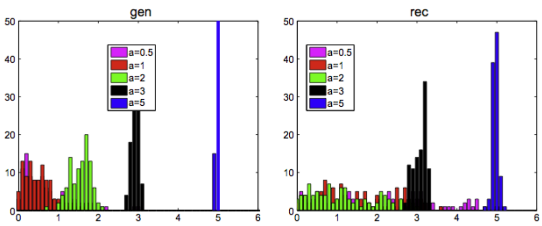


Drobnjak JMR 2010



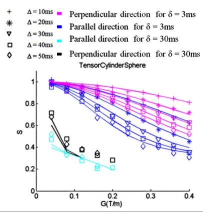
Drobnjak JMR 2010

MCMC posterior distributions on axon diameter. Gmax = 40mT/m; N=4.



Thoughts and observations

- Number of unique sample points
- Constraining the problem
- Model selection
  - Panagiotaki NIMG 2012 design:
- Rule of thumb: before the model is known sample diversely; once the model is known focus on informative points.



Variations of the problem

- Adaptive/sequential design
  - After acquiring measurements $A(\mathbf{y}_1), \ldots, A(\mathbf{y}_K)$, what choice of $\mathbf{y}_{K+1}$ makes $A(\mathbf{y}_{K+1})$ most informative?
- Design for prediction
  - What choice of $\mathbf{y}_1, \ldots, \mathbf{y}_K$ provides the model with the most accurate predictions of future $A(\mathbf{y} \mid \mathbf{x})$?
  - G, I and V-optimality.
- Design for model selection
  - What choice of $\mathbf{y}_1, \ldots, \mathbf{y}_K$ best informs a choice between two models M_1 and M_2?

Bayesian experiment design

- Defines a utility function $U(\mathbf{x},\mathbf{y},\xi)$ that encapsulates the aim of the experiment.
- Then
  $$U(\xi) = \iint U(\mathbf{x},\mathbf{y},\xi)\, p(\mathbf{x} \mid \mathbf{y},\xi)\, p(\mathbf{y} \mid \xi)\, d\mathbf{x}\, d\mathbf{y}$$
- We seek
  $$\hat{\xi} = \operatorname{argmax} U(\xi)$$
- For example, maximize Shannon entropy of posterior (similar to D-optimality):
  $$U(\xi) = \iint \log\left[p(\mathbf{x} \mid \mathbf{y},\xi)\right] p(\mathbf{y} \mid \xi)\, d\mathbf{x}\, d\mathbf{y}$$

Summary

- Optimal experiment design
- Fisher information matrix
- Optimality criteria

- Essential to evaluate capabilities of a system to measure a parameter.
- Can improve accuracy or reduce acquisition requirements
- Can give new insight into measurement strategy

Modelling pipeline

- Identify set of candidate models
- Acquire rich and diverse measurements
- Identify best model(s)
- Identify prior parameter distributions
- Identify acquisition constraints
- Optimize the experiment design
    - Unconstrained to identify types of measurement
    - Parametrized to find good solution
- Choose working protocol
- Validate parameter estimates