Optimal Bayesian experimental design: methodologies and materials applications

Youssef Marzouk
joint work with Chi Feng, Raghav Aggarwal, Michael Demkowicz

Department of Aeronautics and Astronautics
Center for Computational Engineering
Statistics and Data Science Center

Massachusetts Institute of Technology
http://uqgroup.mit.edu

Support from AFOSR and DOE

17 January 2019

Optimal experimental design

- Given limited experimental resources, how to find the most **informative** data?
  - **What** to measure? **When** or **where** to do it? Which experimental **conditions** yield the most information?

Optimal experimental design

- Given limited experimental resources, how to find the most **informative** data?
  - **What** to measure? **When** or **where** to do it? Which experimental **conditions** yield the most information?
  - *Practical impact:* Experiments are expensive and time-consuming; better predictive models and understanding from **limited data**
  - Broad applications: chemical kinetics, environmental sensing, health monitoring, materials modeling, ...
- Approach: develop and optimize design criteria that employ a *model* of the experiment, that account for *uncertainties* in the model, and that adapt to *experimental goals*

More specific questions:

- What exactly is the experimental goal? How do we define "most informative?"
- Models contain many uncertain parameters; can we **focus** the design on a chosen subset?
- Can we handle uncertainty in model *structure*?

Computational challenge: efficient evaluation & optimization of design criteria in a fully nonlinear and non-Gaussian setting...

Bayes' rule (via probability density functions):

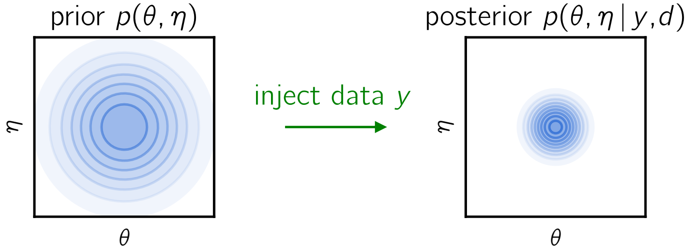
$$\overbrace{p(\boldsymbol{\theta}, \boldsymbol{\eta} | \boldsymbol{y}, \boldsymbol{d})}^{\text{posterior}} = \frac{\overbrace{p(\boldsymbol{y} | \boldsymbol{\theta}, \boldsymbol{\eta}, \boldsymbol{d})}^{\text{likelihood}} \overbrace{p(\boldsymbol{\theta}, \boldsymbol{\eta})}^{\text{prior}}}{\underbrace{p(\boldsymbol{y} | \boldsymbol{d})}_{\text{evidence}}}$$

$\boldsymbol{\theta}$ — parameters of interest

($\boldsymbol{\eta}$ — nuisance parameters)

$\boldsymbol{y}$ — data

$\boldsymbol{d}$ — design parameters



Decision theoretic approach

Choose experimental conditions by maximizing an **expected utility:**

- Expected utility of design d with a generic utility function u:

$$U(d) = \int_{\mathcal{Y}} \int_{\Theta} u(d, y, \theta)\, p(\theta, y | d)\, d\theta\, dy$$

Decision theoretic approach

Choose experimental conditions by maximizing an **expected utility:**

- Expected utility of design d with a generic utility function u:

$$U(d) = \int_{\mathcal{Y}} \int_{\Theta} u(d, y, \theta)\, p(\theta, y|d)\, d\theta\, dy$$

- Good choice for **parameter inference**: Kullback–Leibler (KL) divergence from prior to posterior

$$\begin{aligned} u(d, y, \theta) = u(d, y) &= D_{\mathrm{KL}}(p(\theta|y, d) \,\|\, p(\theta)) \\ &= \int_{\Theta} p(\theta|y, d) \log \frac{p(\theta|y, d)}{p(\theta)}\, d\theta \end{aligned}$$

- Resulting $U(d)$ is *expected information gain*; equivalent to *mutual information* between Y and θ

Optimal experimental design for prediction

Alternatively, choose experiments to improve **predictions** of some quantity Q:

- Prior predictive distribution

$$p(Q) = \int p(Q|\theta)p(\theta)d\theta$$

- Posterior predictive distribution

$$p(Q|y, d) = \int p(Q|\theta)p(\theta|y, d)d\theta$$

Optimal experimental design for prediction

Alternatively, choose experiments to improve **predictions** of some quantity Q:

- Prior predictive distribution

$$p(Q) = \int p(Q|\theta)p(\theta)d\theta$$

- Posterior predictive distribution

$$p(Q|y,d) = \int p(Q|\theta)p(\theta|y,d)d\theta$$

- Now choose utility function as:

$$u(d,y) = D_{\mathrm{KL}}(p(Q|y,d)\,\|\,p(Q))$$

- Automatically incorporates an *information theoretic sensitivity analysis:* only learn aspects of θ that are relevant to Q

Optimal experimental design for models

What about uncertainty in **models**? Consider optimal experimental design for **model discrimination**.

- Bayes' rule for models:

$$P(M_i|y,d) = \frac{p(y|M_i,d)P(M_i)}{p(y|d)},$$

with "marginal likelihood"

$$p(y|M_i,d) = \int_{\Theta_i} p(y|\theta_i,d,M_i)p(\theta_i|M_i)d\theta_i.$$

  - For each model M_i, average the likelihood over the prior distribution of the model's parameters $\theta_i \in \Theta_i$
  - Yields an automatic Occam's razor...

Optimal experimental design for model discrimination

Continue same principle: maximize expected KL divergence from *{prior over models}* to *{posterior over models}*

- Utility function:

$$u(d, y) = \sum_i P(M_i|y, d) \log \frac{P(M_i|y, d)}{P(M_i)}.$$

- Expected utility of design d:

$$U(d) = \int_{\mathcal{Y}} u(d, y) p(y|d) dy$$

where

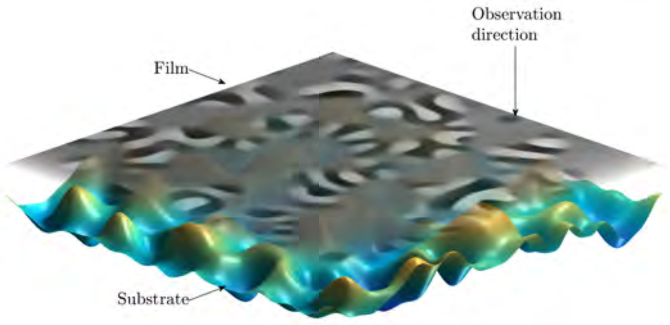
$$p(y|d) = \sum_i p(y|M_i, d) P(M_i)$$

  - Data y may result from any of the competing models under consideration!
- Resulting designs seek to concentrate the posterior onto *fewer* models

Example: OED for parameter inference

Thin film evolving on a heterogeneous substrate:

- Heterogeneity in temperature, local chemistry, topography, etc.



- **Goal:** learn about substrate properties from behavior of the film

Example: OED for parameter inference

- **Problem setup:**
  - Film behavior governed by Cahn-Hilliard equation

$$\frac{\partial c}{\partial t} = \Delta\left(\frac{\partial g}{\partial c} - \epsilon^2 \Delta c\right)$$

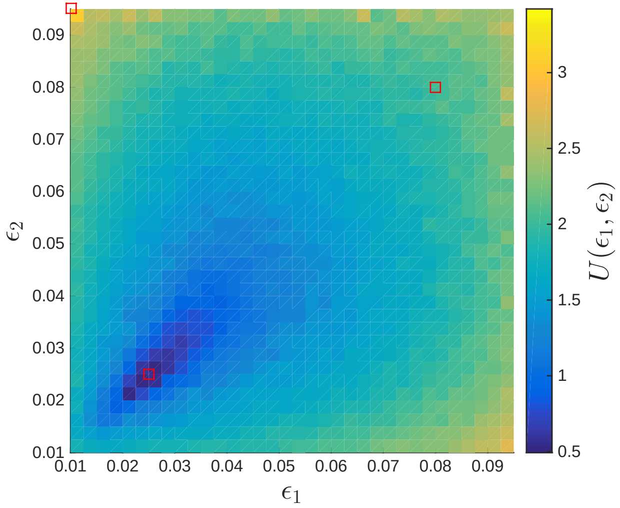
  - Substrate-dependent energy potential g

$$g(c, \mathcal{T}(x,y)) = \frac{c^4}{4} + \mathcal{T}(x,y)\frac{c^2}{2}$$

  - Substrate property $\mathcal{T}(x,y)$ is a Gaussian random field with correlation length ℓ_s
  - Forward model is thus *stochastic*
- **Goal:** infer ℓ_s from observations of film length scale as $t \to \infty$, Λ_∞
- **Design parameters:** film properties ϵ

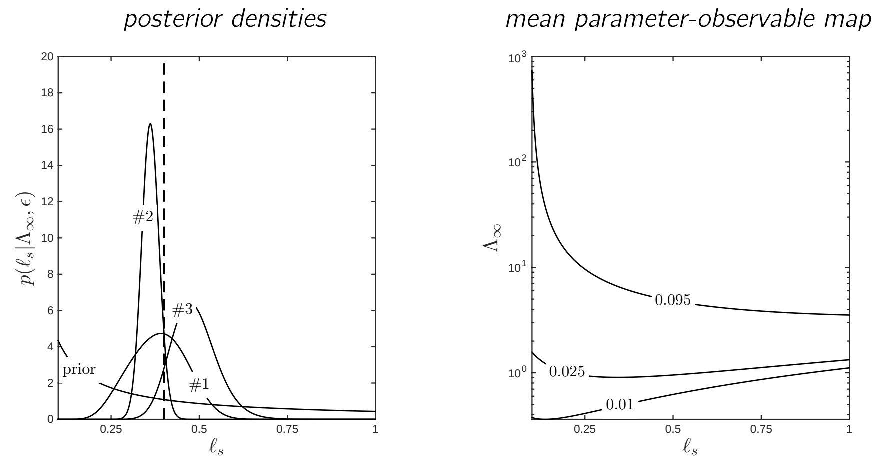
[Details in Aggarwal, Demkowicz, & M 2016]

Map of expected information gain $U(\epsilon_1, \epsilon_2)$ in the substrate length scale ℓ_s



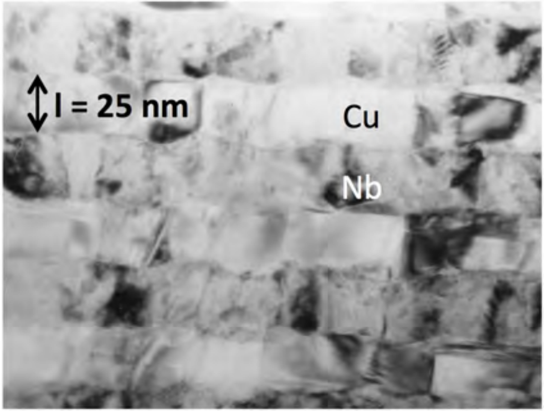
Example: OED for parameter inference

Results of inference:

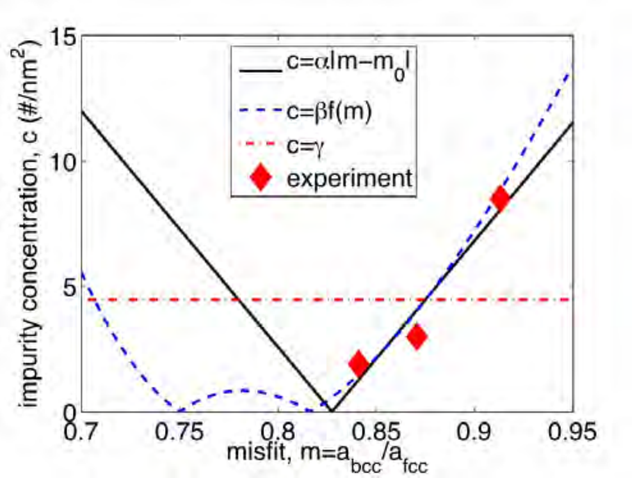


Example: OED for model discrimination

Layered composite of Cu and Nb, created by physical vapor deposition; use to study heterophase interfaces



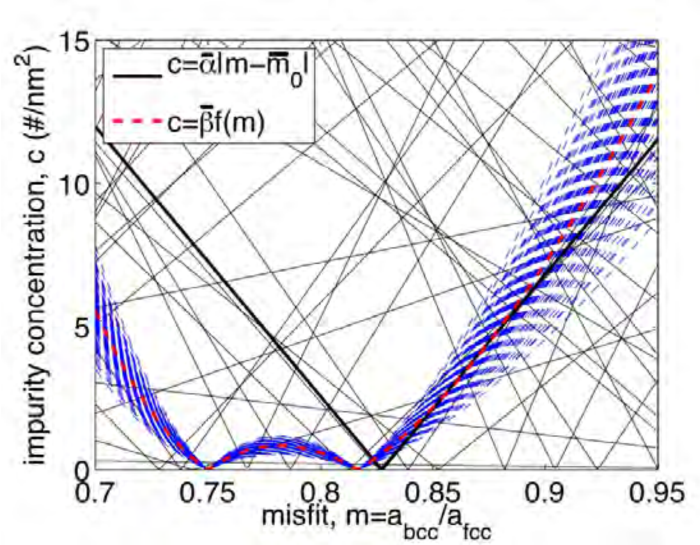
Competing models for He trapping at Cu-Nb interfaces



[Details in Aggarwal, Demkowicz, & M 2016]

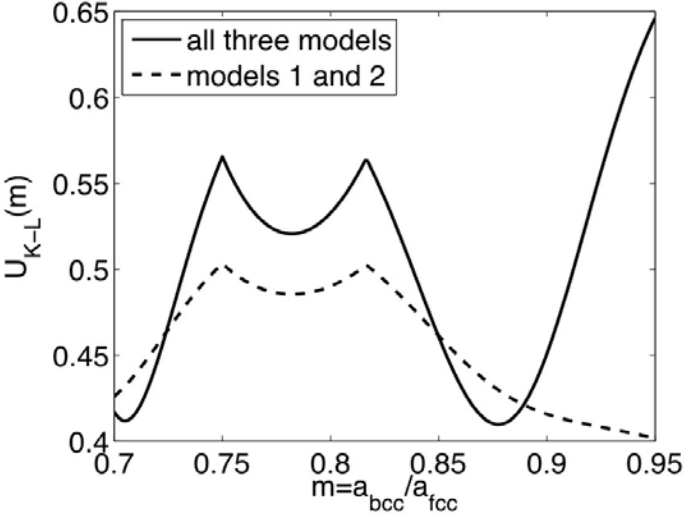
Example: OED for model discrimination

Realizations of models, given prior parameter uncertainties:



Example: OED for model discrimination

Result of OED: **expected information gain** in posterior distribution over models, as a function of misfit m



Under the hood: computational approaches

Experiment described by a (computational) model, e.g.,

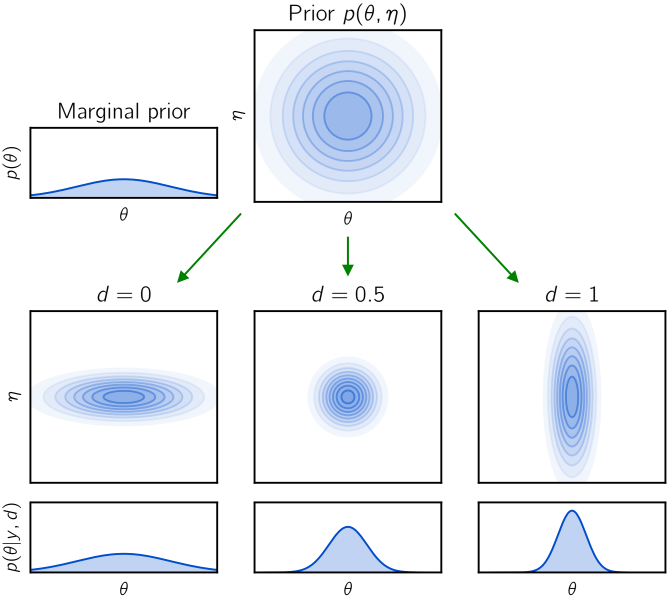
$$\boldsymbol{y} = \boldsymbol{G}(\boldsymbol{\theta}, \boldsymbol{\eta}, \boldsymbol{d}) + \boldsymbol{\epsilon}$$

- $\boldsymbol{y}$ – observations predicted by $\boldsymbol{G}(\,\cdot\,)$ with additive noise
- $\boldsymbol{\epsilon}$ – noise/error random variable

Focused inference on a *subset* of model parameters:

- $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots)$ – parameters of interest
- $\boldsymbol{\eta} = (\eta_1, \eta_2, \ldots)$ – nuisance parameters
- $\boldsymbol{d} = (d_1, d_2, \ldots)$ – design parameters

Bayesian experimental design for focused inference



Expected information gain in marginals

Consider the marginal posterior and marginal prior

$$p(\boldsymbol{\theta}|\boldsymbol{y},\boldsymbol{d}) = \int_H \underbrace{p(\boldsymbol{\theta},\boldsymbol{\eta}|\boldsymbol{y},\boldsymbol{d})}_{\text{joint posterior}}\,\mathrm{d}\boldsymbol{\eta}, \quad p(\boldsymbol{\theta}) = \int_H \underbrace{p(\boldsymbol{\theta},\boldsymbol{\eta})}_{\text{joint prior}}\,\mathrm{d}\boldsymbol{\eta}$$

Information gain in $\boldsymbol{\theta}$ from a single observation $\boldsymbol{y}$:

$$u_{\boldsymbol{\theta}}(\boldsymbol{y},\boldsymbol{d}) = \mathrm{D}_{\mathrm{KL}}\left[\overbrace{p(\boldsymbol{\theta}|\boldsymbol{y},\boldsymbol{d})}^{\substack{\text{marginal}\\\text{posterior}}} \,\middle\|\, \overbrace{p(\boldsymbol{\theta})}^{\substack{\text{marginal}\\\text{prior}}}\right]$$

Expected information gain – experimental outcome is unknown beforehand:

$$U_{\boldsymbol{\theta}}(\boldsymbol{d}) = \mathbb{E}_{\boldsymbol{y}}[u_{\boldsymbol{\theta}}(\boldsymbol{y},\boldsymbol{d})] = \int_Y u_{\boldsymbol{\theta}}(\boldsymbol{y},\boldsymbol{d})\,p(\boldsymbol{y}|\boldsymbol{d})\,\mathrm{d}\boldsymbol{y}$$

Expectation is taken over the prior predictive of the data