SAND2016-5224C

A Consistent Bayesian Approach for Stochastic Inverse Problems

Tim Wildey, John Jakeman
Optimization and UQ Department
Center for Computing Research
Sandia National Laboratories, Albuquerque, NM, USA

Troy Butler
Department of Mathematical and Statistical Sciences
University of Colorado Denver

European Congress on Computational Methods in Applied Sciences and Engineering (ECCOMAS)
June 5-10, 2016
Crete Island, Greece

Sandia National Laboratories is a multi program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

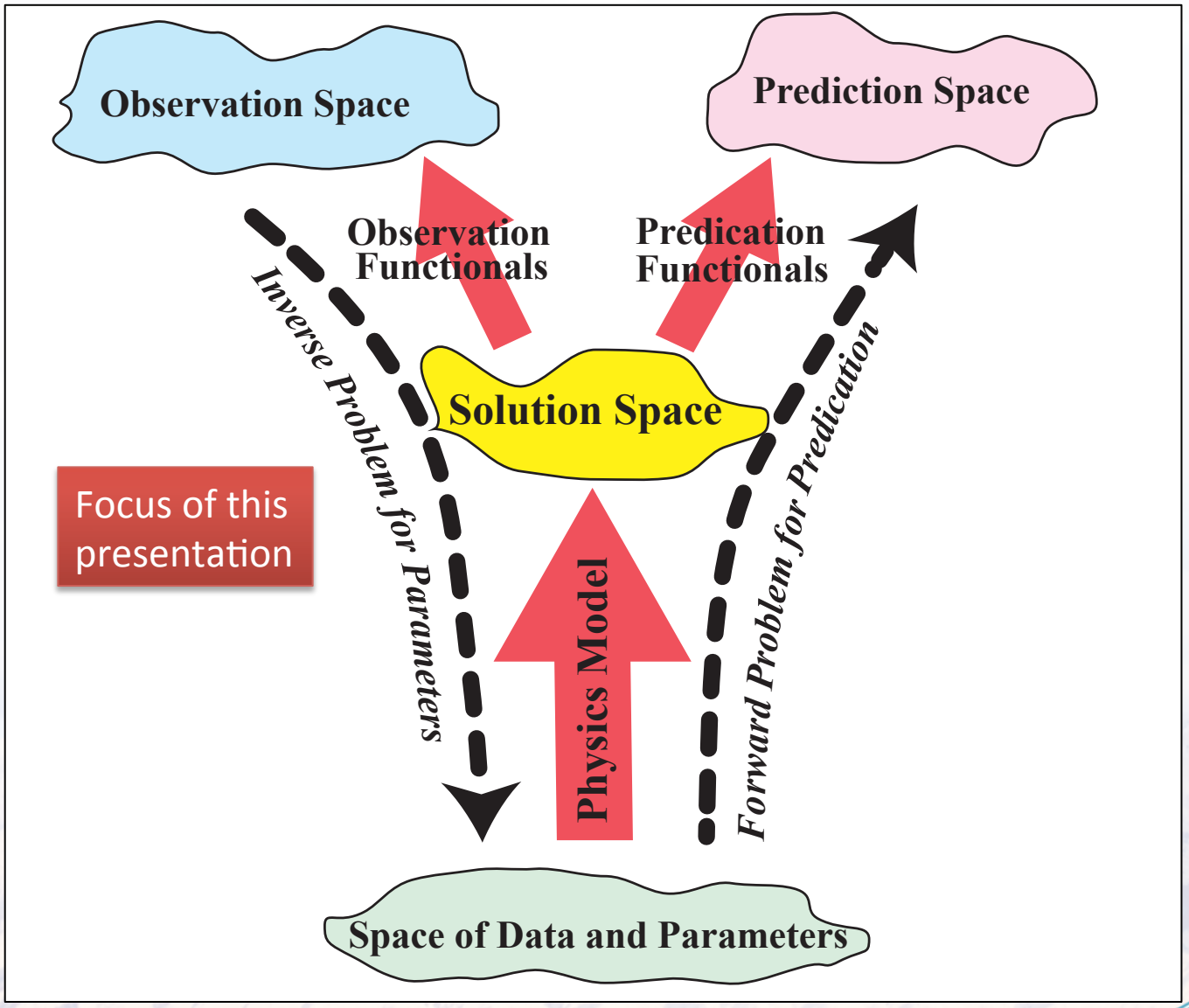
Outline

- **Deterministic and stochastic inverse problems**
- **A consistent Bayesian formulations**
  - Derivation
  - Mathematical properties
  - Numerical examples
- **Comparison with existing approaches**
  - Measure-theoretic inversion
  - Classical/statistical Bayesian inference
  - An analytical example
- **Computational Considerations**
  - Density estimation
  - Surrogate models and adaptivity
  - Numerical examples
- **Conclusions and future work**

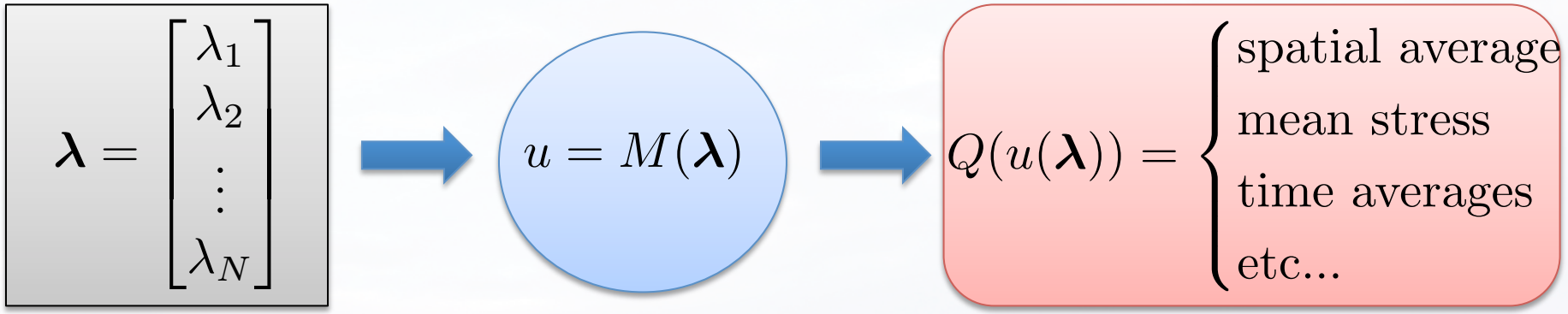
Deterministic and Stochastic Inverse Problems

Sandia National Laboratories

Inversion and Prediction



Defining the Problem



Compact parameter domain $\Lambda \subset \mathbb{R}^n$

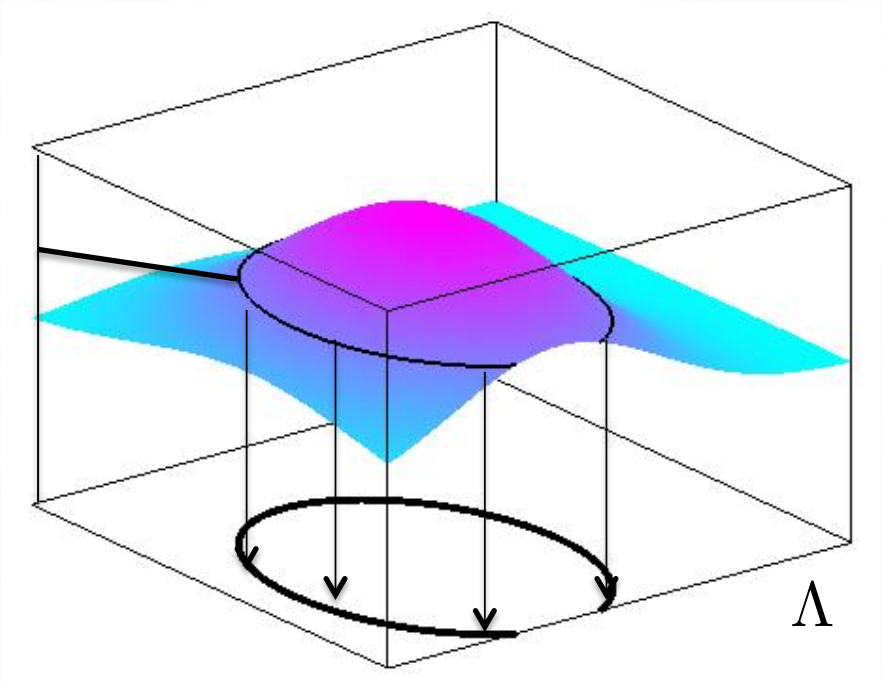
Model giving $u(\lambda)$

Quantities of interest (QoI)

- Assume $Q(\lambda)$ is a measurable mapping
- Define the space of QoI: $\mathcal{D} = Q(\lambda) \subset \mathbb{R}^d$

We can often observe $Q(\lambda)$ and need to use these observations to infer information about λ

Deterministic Inverse Problems

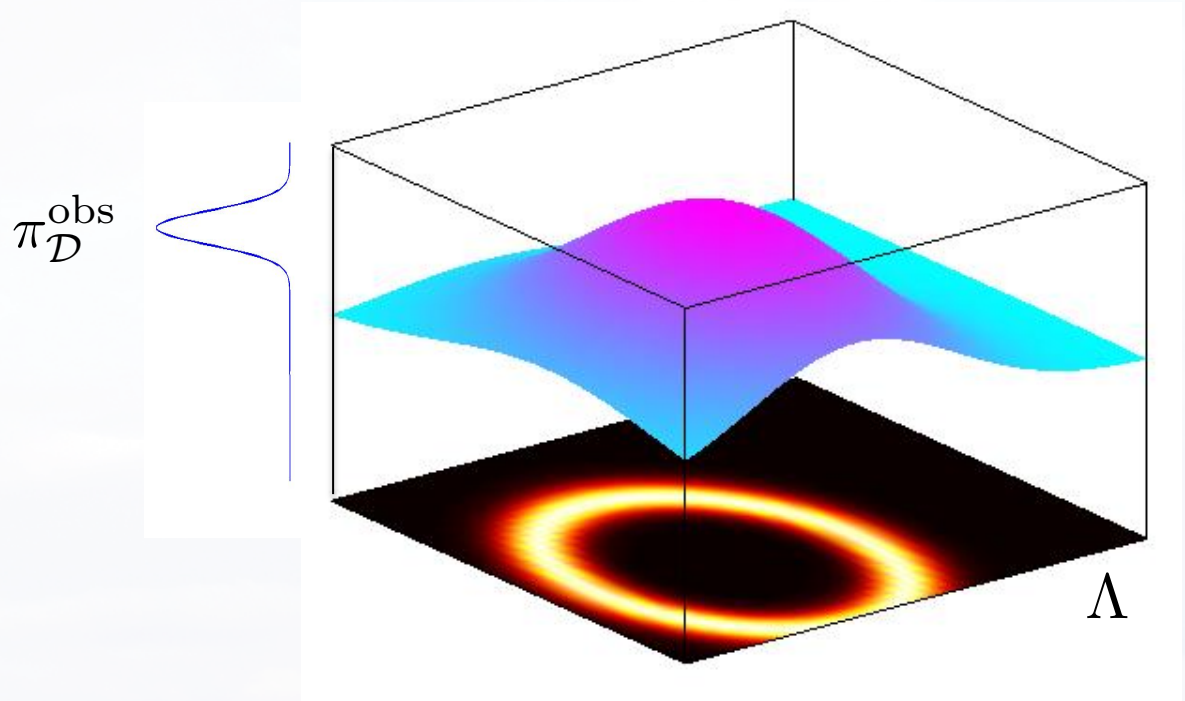


Given a vector of deterministic observations, $\hat{Q}$, find $\lambda \in \Lambda$ such that $Q(\lambda) = \hat{Q}$

Solutions are often non-unique

A unique solution can be found by imposing more structure, e.g. regularization

Stochastic Inverse Problems



Given a probability density on observations, $\pi_{\mathcal{D}}^{\text{obs}}$, find a probability density on Λ such that the propagation of the parameter density through the model matches the density on the observations.

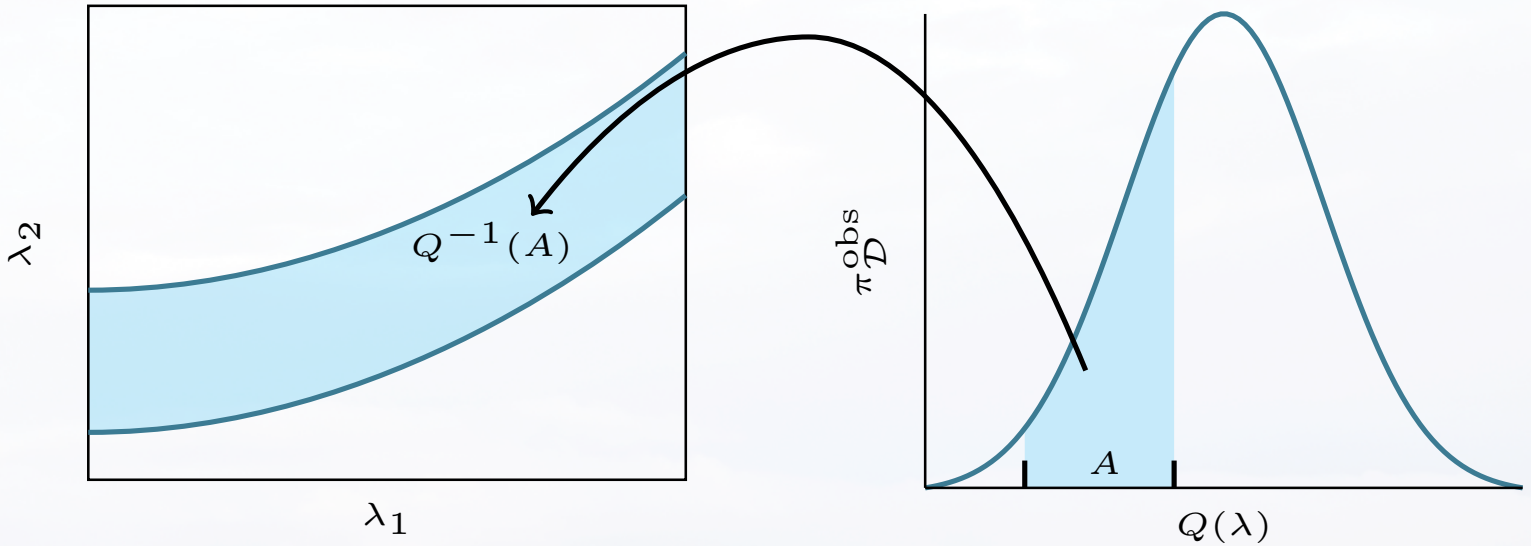
Solutions are often non-unique

A unique solution can be found by imposing more structure, e.g. prior information

A Consistent Bayesian Formulation

Stochastic Inverse Problems

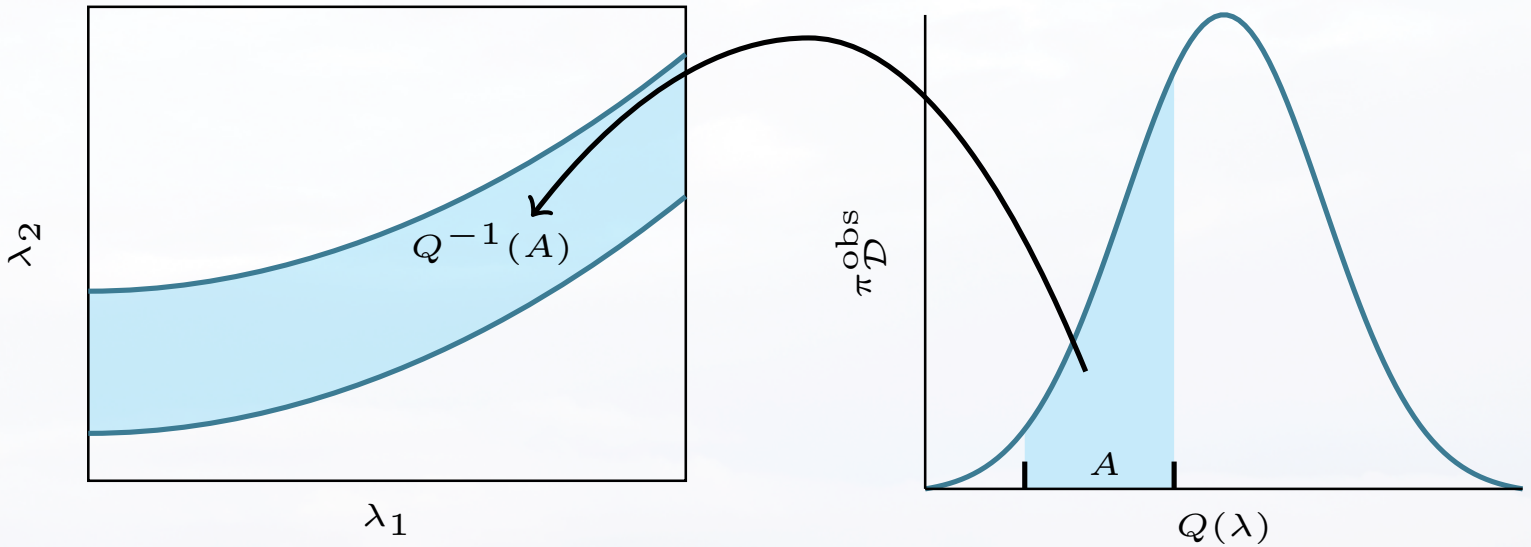
Natural to start with measurable sets



For measurable $A \in \mathcal{B}_{\mathcal{D}}$ (Borel σ-algebra on $\mathcal{D}$), define

$$Q^{-1}(A) = \{\lambda \in \Lambda \mid Q(\lambda) \in A\}$$

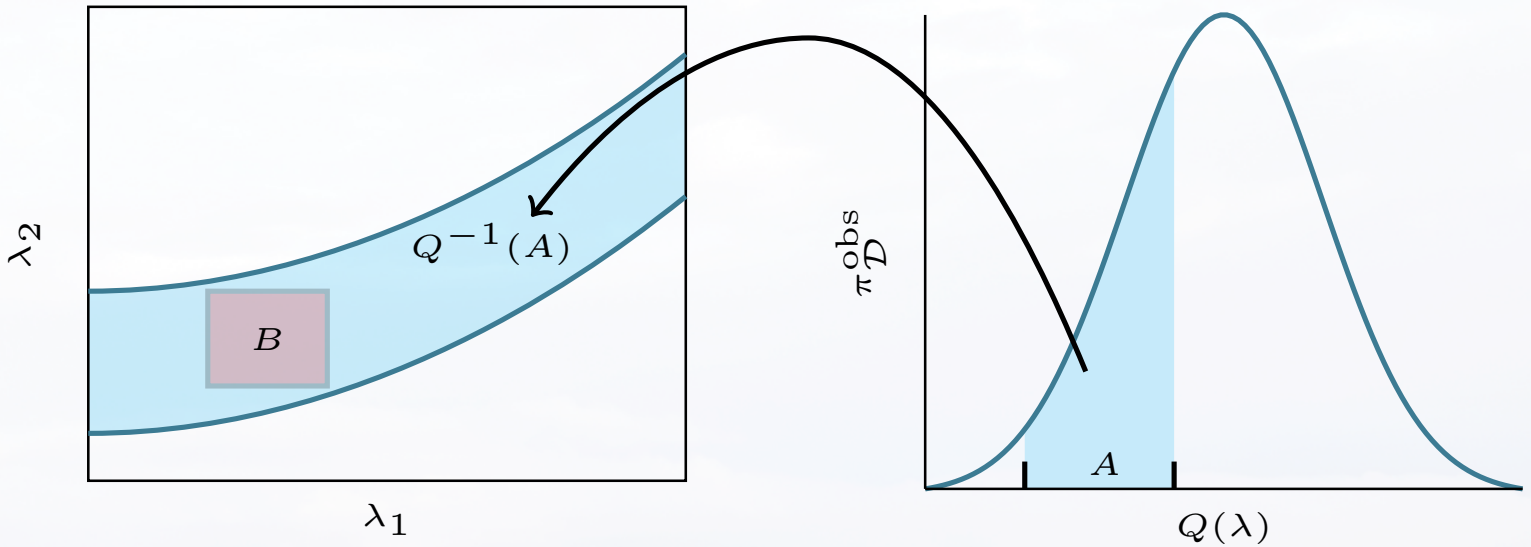
Stochastic Inverse Problems



We need to define a probability measure, P_Λ, such that

$$P_\Lambda(Q^{-1}(A)) = P_{\mathcal{D}}^{\text{obs}}(A)$$

Stochastic Inverse Problems

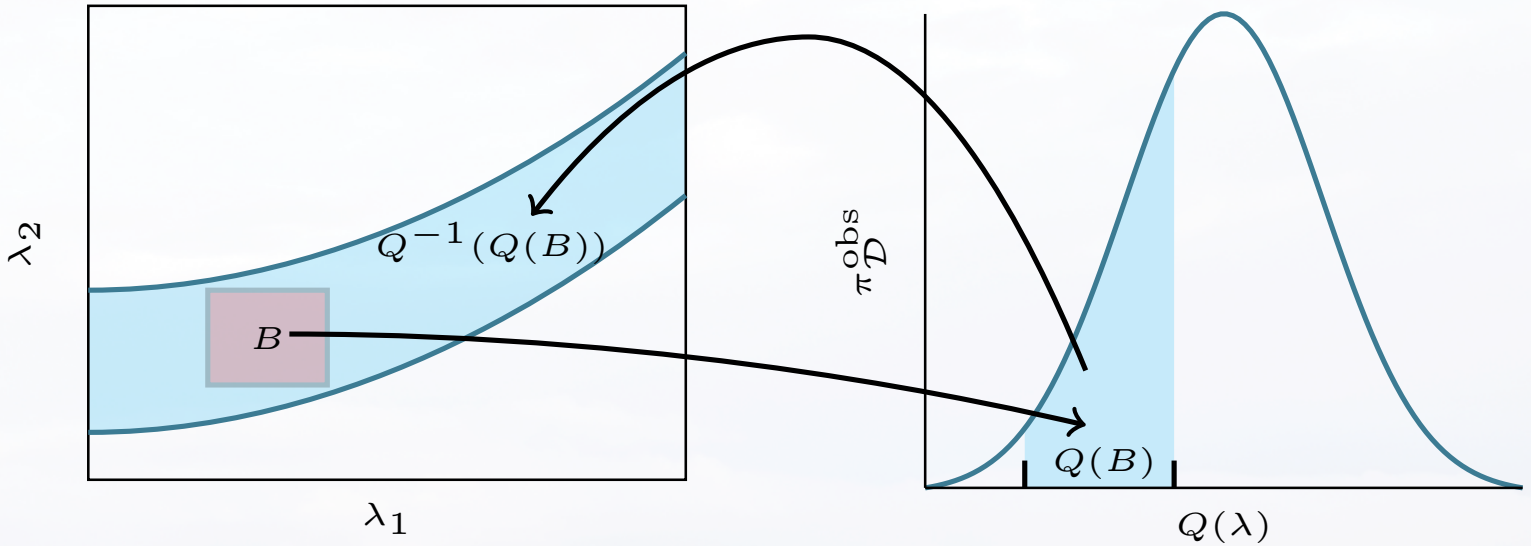


We need to define a probability measure, P_Λ, such that

$$P_\Lambda(Q^{-1}(A)) = P_{\mathcal{D}}^{\text{obs}}(A)$$

But how should we define $P_\Lambda(B)$ for $B \in \mathcal{B}_\Lambda$?

Stochastic Inverse Problems

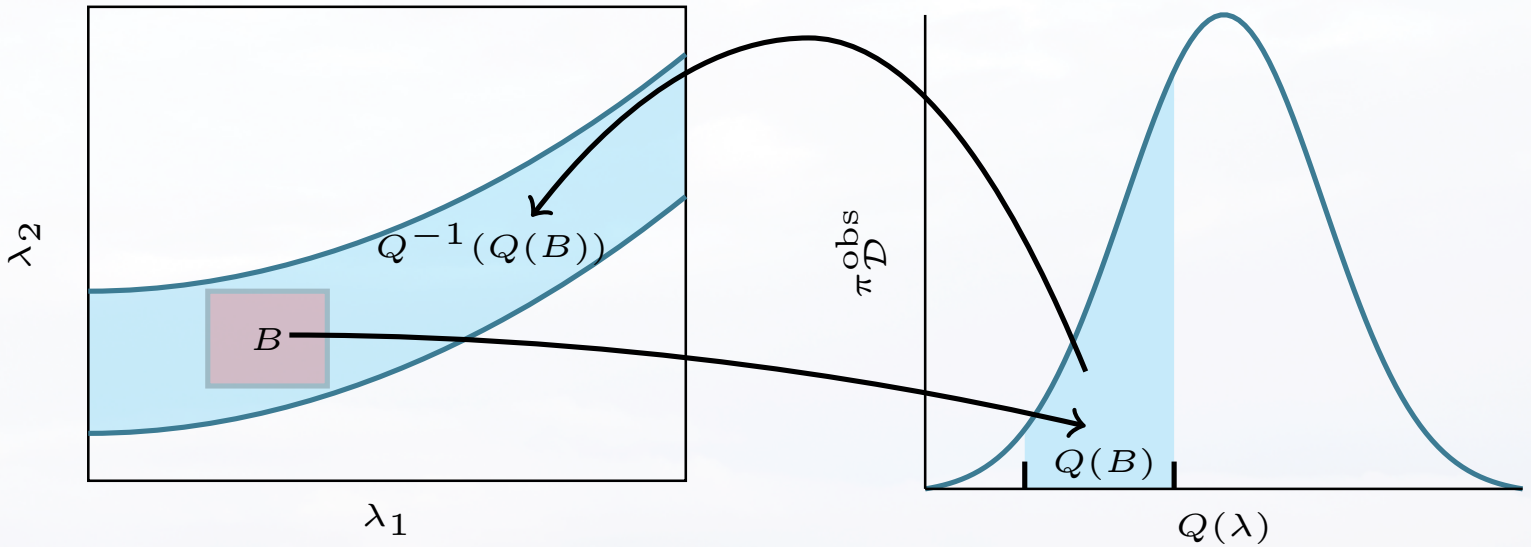


We define a *prior* probability measure: $P_{\Lambda}^{\text{prior}}$

We apportion probability to B based on $P_{\Lambda}^{\text{prior}}$

$$P_{\Lambda}^{\text{post}}(B) = P_{\mathcal{D}}^{\text{obs}}(Q(B)) \frac{P_{\Lambda}^{\text{prior}}(B)}{P_{\Lambda}^{\text{prior}}(Q^{-1}(Q(B))}$$

Stochastic Inverse Problems



More concisely,

$$P_{\Lambda}^{\text{post}}(B) = P_{\Lambda}^{\text{prior}}(B)\frac{P_{\mathcal{D}}^{\text{obs}}(Q(B))}{P_{\mathcal{D}}^{Q(\text{prior})}(Q(B))}$$

where $P_{\mathcal{D}}^{Q(\text{prior})}$ is the push-forward measure induced by the prior and the model

An Assumption on the Prior

We need to make an assumption on the prior to prove that the posterior solves the stochastic inverse problem.

We assume that the push-forward measure induced by the model and the prior *dominates* the observed measure:

More precisely, there exists $C > 0$ such that

$$0 \leq P_{\mathcal{D}}^{\text{obs}}(A) \leq C P_{\mathcal{D}}^{Q(\text{prior})}(A), \quad \forall A \in \mathcal{B}_{\mathcal{D}}$$

In other words, we assume that if we observe an event with positive probability, then we can also predict that event with positive probability.

Since $\mathcal{D} = Q(\mathbf{\Lambda})$, this is an assumption only on the prior.

Analogous to choosing the prior to be as general as possible

Mathematical Properties

Theorem

Assume $Q : \mathbf{\Lambda} \to \mathcal{D}$ is measurable. Given probability measures $P_{\mathbf{\Lambda}}^{\text{prior}}$ and $P_{\mathcal{D}}^{\text{obs}}$ that are absolutely continuous with respect to the volume measures $\mu_{\mathbf{\Lambda}}$ and $\mu_{\mathcal{D}}$ respectively, the *posterior* measure, $P_{\mathbf{\Lambda}}^{\text{post}}$, is a probability measure.

Most of the requirements are trivial to show.

Just need to prove $P_{\mathbf{\Lambda}}^{\text{post}}(\mathbf{\Lambda}) = 1$.

This follows easily from the fact that we define $\mathcal{D} = Q(\mathbf{\Lambda})$.

Mathematical Properties

Theorem

Assume $Q : \mathbf{\Lambda} \to \mathcal{D}$ is measurable. Given probability measures $P_{\mathbf{\Lambda}}^{\text{prior}}$ and $P_{\mathcal{D}}^{\text{obs}}$ that are absolutely continuous with respect to $\mu_{\mathbf{\Lambda}}$ and $\mu_{\mathcal{D}}$ respectively, the *posterior* probability density satisfies the consistency requirement.

Recall that this means for any $A \in \mathcal{B}_{\mathcal{D}}$ we have

$$P_{\mathbf{\Lambda}}^{\text{post}}(Q^{-1}(A)) = P_{\mathcal{D}}^{\text{obs}}(A)$$

i.e., propagating the posterior though the model gives the observed data.

We also show that the posterior is absolutely continuous with respect to $\mu_{\mathbf{\Lambda}}$ which implies there exists a density, $\pi_{\mathbf{\Lambda}}^{\text{post}}$.

Mathematical Properties

Theorem

Assume $Q : \mathbf{\Lambda} \to \mathcal{D}$ is measurable and piecewise continuous. Given probability measures $P_{\mathbf{\Lambda}}^{\text{prior}}$ and $P_{\mathcal{D}}^{\text{obs}}$ that are absolutely continuous with respect to $\mu_{\mathbf{\Lambda}}$ and $\mu_{\mathcal{D}}$ respectively (and these volume measures are absolutely continuous with respect to Lebesgue measures). Then wherever the *posterior* probability density is non-zero, it is given by

$$\pi_{\mathbf{\Lambda}}^{\text{post}}(\lambda) = \pi_{\mathbf{\Lambda}}^{\text{prior}}(\lambda) \frac{\pi_{\mathcal{D}}^{\text{obs}}(Q(\lambda))}{\pi_{\mathcal{D}}^{Q(\text{prior})}(Q(\lambda))}, \quad \lambda \in \Lambda$$

Both $\pi_{\mathbf{\Lambda}}^{\text{prior}}$ and $\pi_{\mathcal{D}}^{\text{obs}}$ are assumed to be given.

We only need to calculate $\pi_{\mathcal{D}}^{Q(\text{prior})}$.

This is the **forward propagation** of the prior through the model.

Algorithms

Algorithm 1 Computing the Push-Forward of the Prior

Given a set of samples from the prior density: $\lambda_i, \quad i = 1, \ldots, M$
Evaluate the model and compute the QoIs: $q_i = Q(\lambda_i)$
Use the set of QoIs and use a standard technique, such as a kernel density method, to estimate $\pi_{\mathcal{D}}^{Q(\text{prior})}(q)$

Algorithm 2 Contructing the Posterior Density

Given the $\pi_{\mathcal{D}}^{Q(\text{prior})}(q)$ from Algorithm 1
Select a set of samples in $\mathbf{\Lambda}$: $\lambda_p, \quad p = 1, \ldots, P$
Evaluate the model and compute the QoIs: $q_p = Q(\lambda_p)$
Compute the posterior density at each λ_p:

$$\pi_{\mathbf{\Lambda}}^{\text{post}}(\lambda_p) = \pi_{\mathbf{\Lambda}}^{\text{prior}}(\lambda_p) \frac{\pi_{\mathcal{D}}^{\text{obs}}(q_p)}{\pi_{\mathcal{D}}^{Q(\text{prior})}(q_p)}.$$

Use $\pi_{\mathbf{\Lambda}}^{\text{post}}(\lambda_p)$ to construct an approximation of the posterior density

Algorithms

Algorithm 3: Generating samples from the posterior

Given the $\pi_{\mathcal{D}}^{Q(\text{prior})}(q)$ from Algorithm 1;
Select a set of samples in $\mathbf{\Lambda}$: $\lambda_p, \quad p = 1, \ldots, P$;
Evaluate the model and compute the QoIs: $q_p = Q(\lambda_p)$;
Compute the posterior density at each λ_p:

$$\pi_{\mathbf{\Lambda}}^{\text{post}}(\lambda_p) = \pi_{\mathbf{\Lambda}}^{\text{prior}}(\lambda_p) \frac{\pi_{\mathcal{D}}^{\text{obs}}(q_p)}{\pi_{\mathcal{D}}^{Q(\text{prior})}(q_p)}.$$

Estimate $M = \max_{\mathbf{\Lambda}} \left(\pi_{\mathbf{\Lambda}}^{\text{prior}}(\lambda) / \pi_{\mathbf{\Lambda}}^{\text{post}}(\lambda) \right)$;
for $p = 1, \ldots, P$ **do**
  Generate a random number, ξ_p, from a uniform distribution on $[0, 1]$;
  Compute the ratio: $\eta_p = M \frac{\pi_{\mathbf{\Lambda}}^{\text{post}}(\lambda_p)}{\pi_{\mathbf{\Lambda}}^{\text{prior}}(\lambda_p)}$;
  if $\eta_p > \xi_p$ **then**
    Accept λ_p;
  else
    Reject λ_p;
  end
end

A Simple Example

Consider a 2-component nonlinear system of equations with two parameters

$$\begin{aligned} \lambda_1 x_1^2 + x_2^2 &= 1 \\ x_1^2 - \lambda_2 x_2^2 &= 1 \end{aligned}$$

The quantity of interest is the second component, i.e., $q(\lambda) = x_2$.

The parameter ranges are given by

$$\lambda_1 \in [0.79, 0.99], \quad \lambda_2 \in [1 - 4.5\sqrt{0.1}, 1 + 4.5\sqrt{0.1}]$$

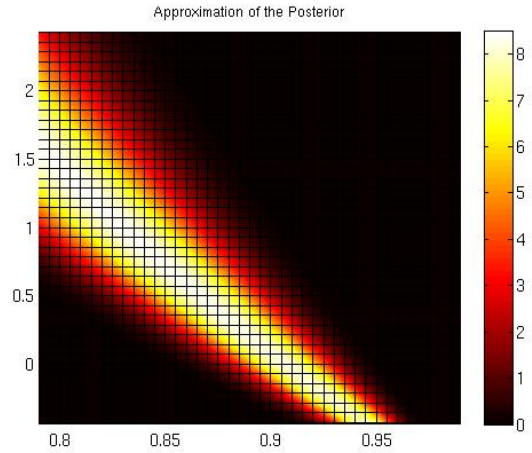
Assume that we observe $q \sim N(0.3, 0.025^2)$

We consider two different priors on $\mathbf{\Lambda}$, uniform and Beta$(2, 5)$.

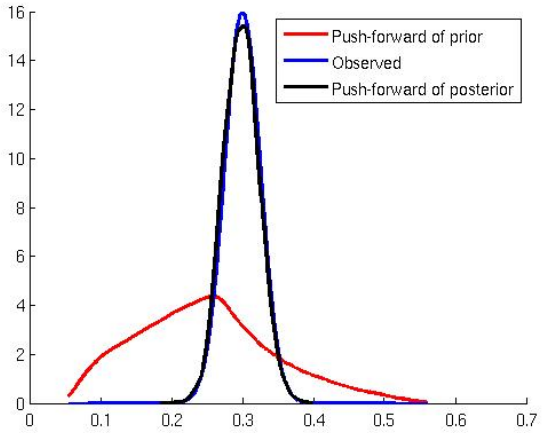
We show that both the push-forward densities induced by the posteriors match the observed density.

We use 100,000 samples from each prior and a standard kernel density approach to compute the push-forward densities

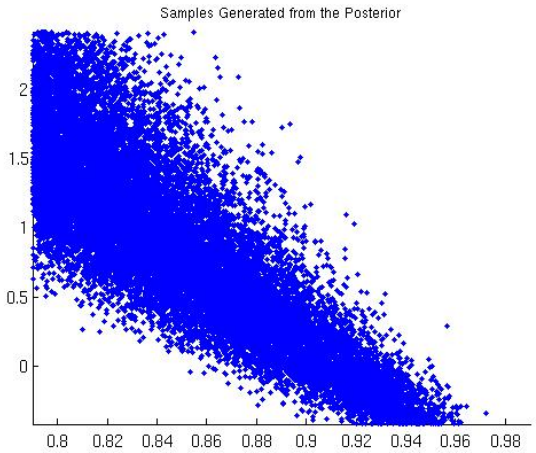
A Simple Example: Uniform Prior



The posterior $(\pi_{\mathbf{\Lambda}}^{\text{post}})$

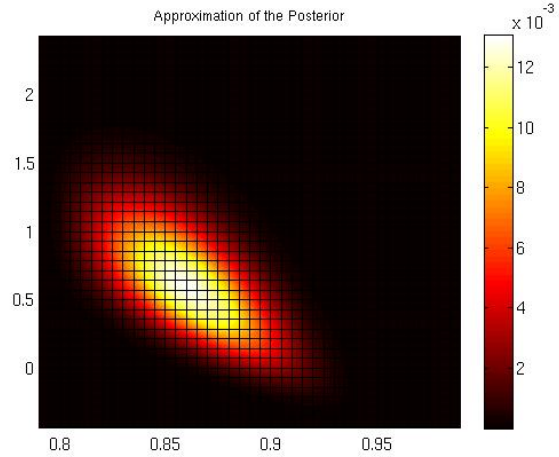


The three distributions on $\mathcal{D}$ $(\pi_{\mathcal{D}}^{\text{obs}}, \pi_{\mathcal{D}}^{Q(\text{prior})}$ and $\pi_{\mathcal{D}}^{Q(\text{post})})$

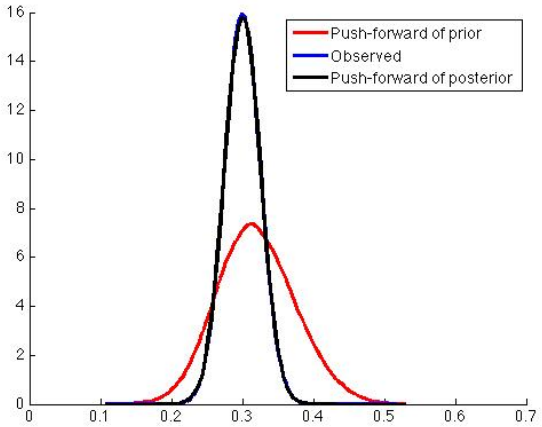


Samples from the posterior $(\pi_{\mathbf{\Lambda}}^{\text{post}})$

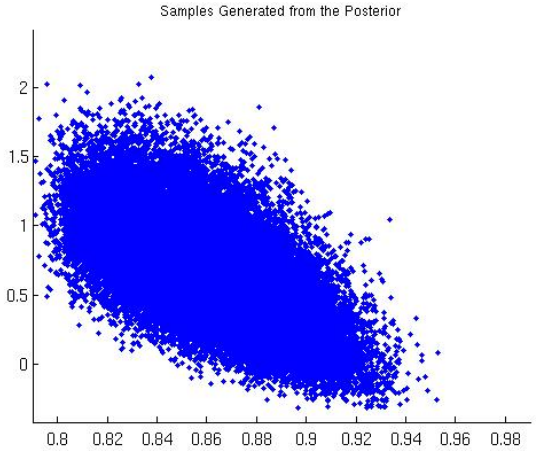
A Simple Example: Beta Prior



The posterior ($\pi_{\mathbf{\Lambda}}^{\text{post}}$)



The three distributions on $\mathcal{D}$ ($\pi_{\mathcal{D}}^{\text{obs}}$, $\pi_{\mathcal{D}}^{Q(\text{prior})}$ and $\pi_{\mathcal{D}}^{Q(\text{post})}$)

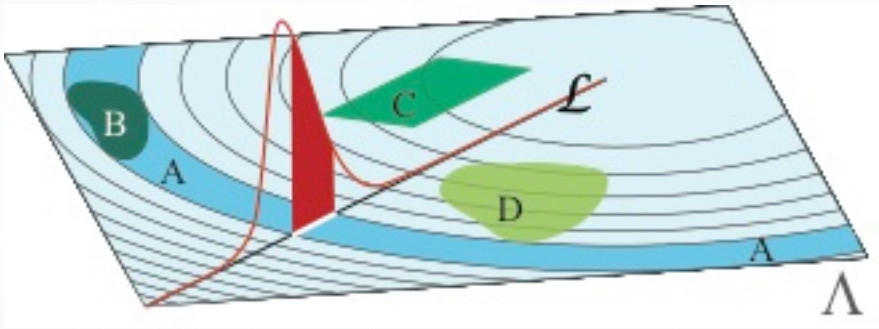


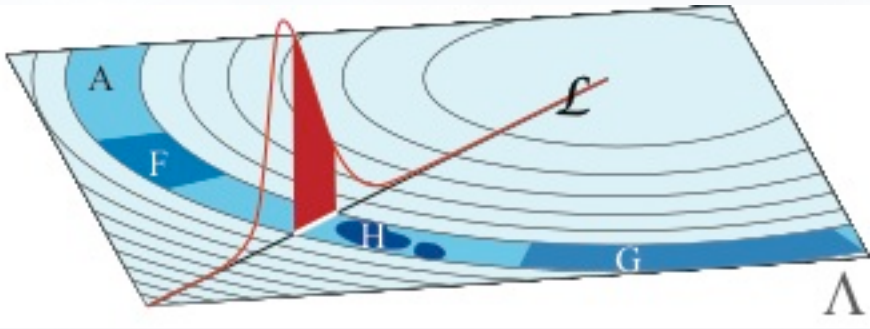
Samples from the posterior ($\pi_{\mathbf{\Lambda}}^{\text{post}}$)

Comparison with Existing Approaches

Comparison With Existing Approaches

Measure-Theoretic Approach [Butler et al SINUM 2012]





- Both approaches are consistent with the model and the data.
- Make different assumptions and produce different densities
  - unless the solution is unique
- Different computational considerations
- With our consistent Bayesian approach, it is easier to
  - Leverage advanced approaches for forward propagation of uncertainty
  - Generate samples from the posterior

Comparison With Existing Approaches

Classical/Statistical Bayesian Approach [Kennedy and O'Hagan, JRSS 2001]

Assume that observational data, $q \in \mathcal{D}$ are noisy

$$q = Q(\lambda) + \eta,$$

where η is a random variable with mean zero and density ρ.

Gives the likelihood function

$$\pi(q|\lambda) = \rho(q - Q(\lambda)).$$

Using Bayes theorem we can express the posterior density as

$$\tilde{\pi}^{\text{post}}_{\mathbf{\Lambda}}(\lambda|q) = \pi^{\text{prior}}_{\mathbf{\Lambda}}(\lambda)\frac{\pi(q|\lambda)}{\int_{\Lambda}\pi(q|\lambda)\pi^{\text{prior}}_{\mathbf{\Lambda}}d\mu_{\mathbf{\Lambda}}},$$

The posterior density is **not guaranteed to be consistent**

An Analytical Example

Let $\mathbf{\Lambda} = [0, 1]$ and consider the simple nonlinear map

$$q(\lambda) = \lambda^p, \quad p > 0.$$

Here, p, is not uncertain and is used to vary the nonlinearity of the map.

We set $\mathcal{D} = Q(\mathbf{\Lambda}) = [0, 1]$ and assume the observed density is given by

$$\pi_{\mathcal{D}}^{\text{obs}}(q) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(q-\hat{q})^2}{2\sigma^2}},$$

where $\hat{q}$ and σ are also assumed to be given.

Regardless of the choice of p and the prior density, we can easily calculate

$$\pi_{\mathcal{D}}^{Q(\text{prior})}(q) = \frac{1}{p} q^{1/p-1} \pi_{\mathbf{\Lambda}}^{\text{prior}}(q^{1/p}).$$

An Analytical Example

We can also calculate the posterior density

$$\pi_{\mathbf{\Lambda}}^{\text{post}}(\lambda) = p\lambda^{p-1}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(\lambda^p-\hat{q})^2}{2\sigma^2}}.$$

and the push-forward of the posterior density

$$\pi_{\mathcal{D}}^{Q(\text{post})}(q) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(q-\hat{q})^2}{2\sigma^2}},$$

which is identical to the observed density on the data.

An Analytical Example

For the classical/statistical Bayesian approach, we use an observed value of $\hat{q}$ and assume $\eta \sim N(0, \sigma^2)$.

We can calculate the posterior density from the classical Bayesian approach

$$\tilde{\pi}^{\text{post}}_{\mathbf{\Lambda}}(\lambda) = \pi^{\text{prior}}_{\mathbf{\Lambda}}(\lambda)\frac{1}{C}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(\lambda^p-\hat{q})^2}{2\sigma^2}},$$

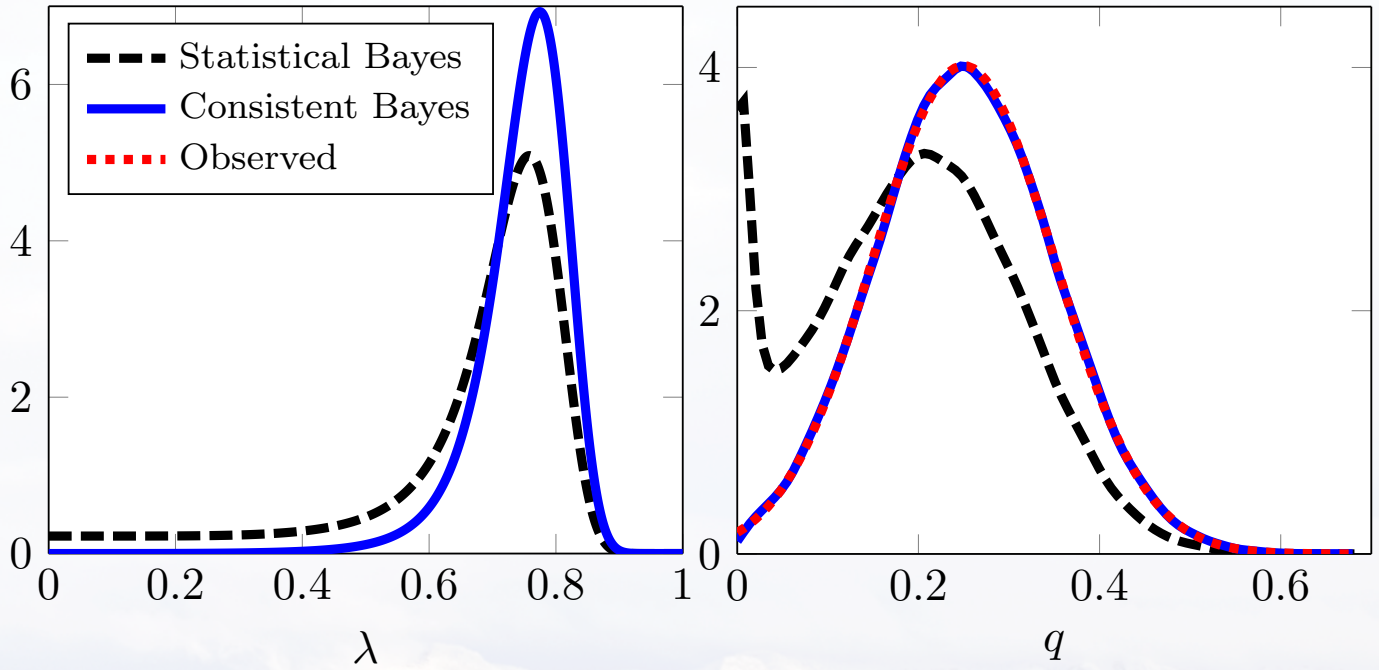
$$C = \int_\Lambda \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(\lambda^p-\hat{q})^2}{2\sigma^2}}\pi^{\text{prior}}_{\mathbf{\Lambda}}(\lambda)d\mu_{\mathbf{\Lambda}}.$$

Compare this with the consistent posterior

$$\pi^{\text{post}}_{\mathbf{\Lambda}}(\lambda) = p\lambda^{p-1}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(\lambda^p-\hat{q})^2}{2\sigma^2}}.$$

The two posteriors are identical if $\pi^{\text{prior}}_{\mathbf{\Lambda}} = 1$ and $p = 1$ (linear map)

An Analytical Example



For $p = 5$ $\bar{q} = 0.25$ and $\sigma = 0.1$, we plot the two posterior densities (left) and compare the push-forward densities from the posteriors with the observed density (right).

Computational Considerations

Density Estimation

Our approach relies estimating $\pi_{\mathcal{D}}^{Q(\text{prior})}$.

Various approaches for multi-variate density estimation exist, but all suffer from the *curse of dimensionality* in some way.

For example, the error in kernel density estimators decays at

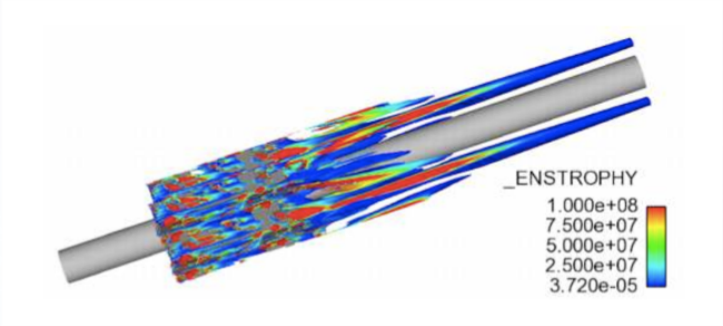
$$\mathcal{O}(N^{-1/(2\nu+d)})$$

where d is the dimension of $\mathcal{D}$ and ν is the order of the kernel (typically 2)

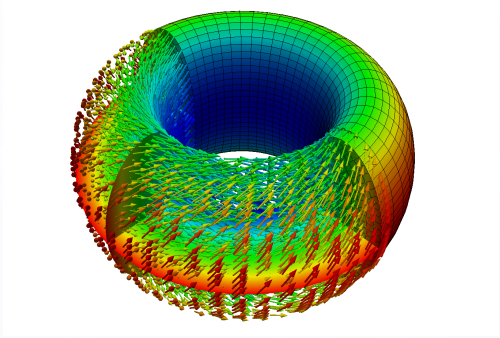
Leaves us with three options:

- Don't worry about accuracy (not very appealing)
- Come up with a better approach for density estimation (open research question)
- Take a large number of samples

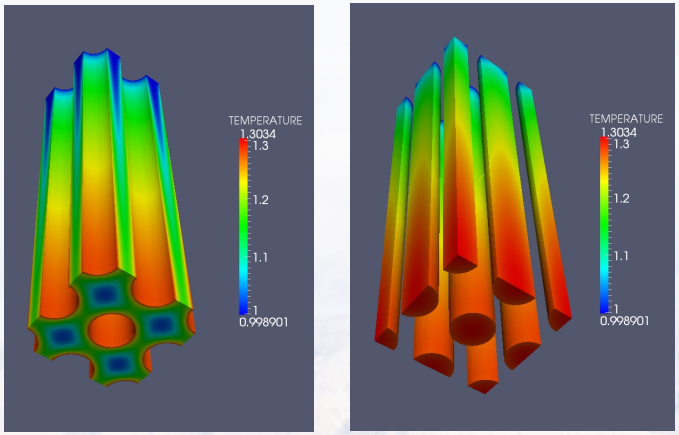
Large Scale Nonlinear Multi-Physics Simulations



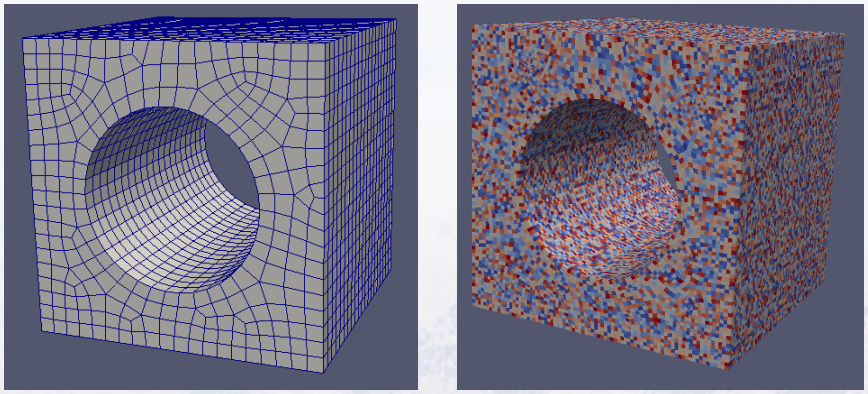
Flow in Nuclear Reactor (Turbulent CFD)



Tokamak Equilibrium (MHD)



Fully-coupled conjugate heat transfer



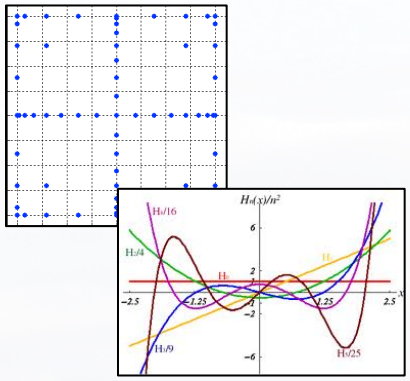
Multi-scale materials modeling

The number of high-fidelity model evaluations is limited.

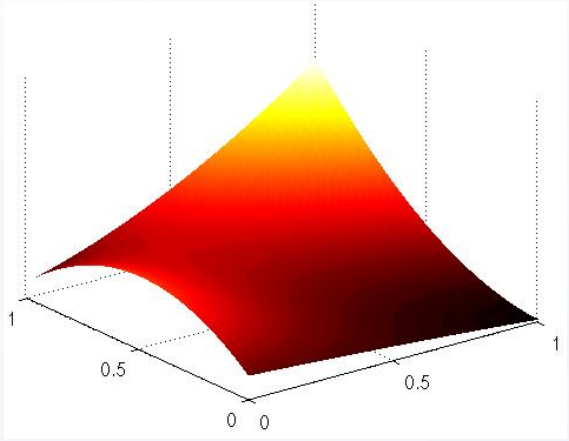
Sandia National Laboratories

Utilizing Surrogate Models

We use a small number of model evaluations to approximate the response surface.



$$u = M(\boldsymbol{\lambda})$$



Examples include regression, polynomial chaos, stochastic collocation, pseudospectral projection, Gaussian processes, etc.

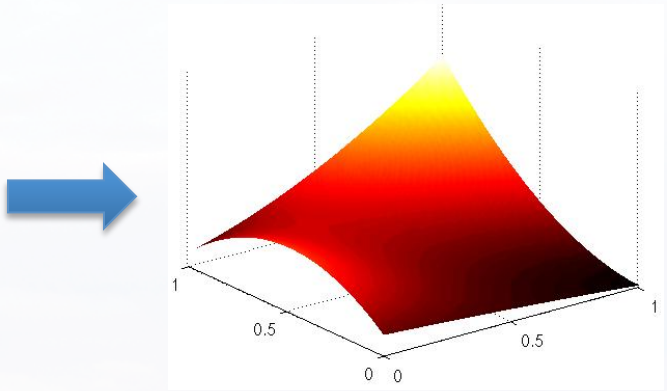
Utilizing Surrogate Models

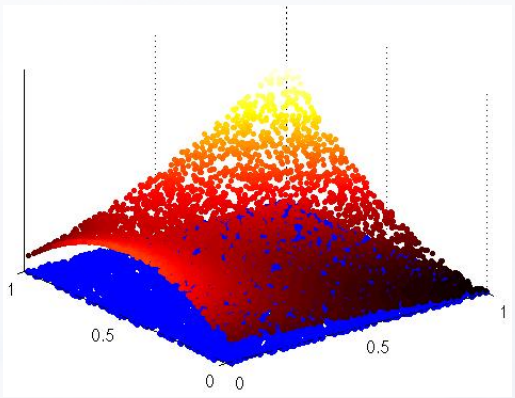
We can estimate the push-forward density using samples of the surrogate

$(\Omega, \mathcal{F}, P)$

$\boldsymbol{\lambda} : \Omega \rightarrow \Pi_{i=1}^{N} I_i$

$F_{\boldsymbol{\Lambda}}(\boldsymbol{\lambda})$





Assuming the number of samples of the surrogate is not limited, the accuracy of the posterior is dominated by the accuracy of the surrogate model.

Let $Q_s(\lambda)$ denote the surrogate approximation of the QoI

An Assumption on the Prior and the Surrogate

We need to make an assumption on **the prior and the surrogate** to prove that the posterior solves the stochastic inverse problem.

We assume that the push-forward measure induced by the surrogate model and the prior *dominates* the observed measure:

More precisely, there exists $C > 0$ such that

$$0 \leq P_{\mathcal{D}}^{\text{obs}}(A) \leq C P_{\mathcal{D}}^{Q_s(\text{prior})}(A), \quad \forall A \in \mathcal{B}_{\mathcal{D}}$$

In other words, we assume that if we observe an event with positive probability, then we can also predict that event **(using the surrogate model)** with positive probability.

Since $\mathcal{D} = Q(\mathbf{\Lambda}) \neq Q_s(\mathbf{\Lambda})$, this is an assumption on the **prior and the surrogate.**

Utilizing Surrogate Models

Theorem

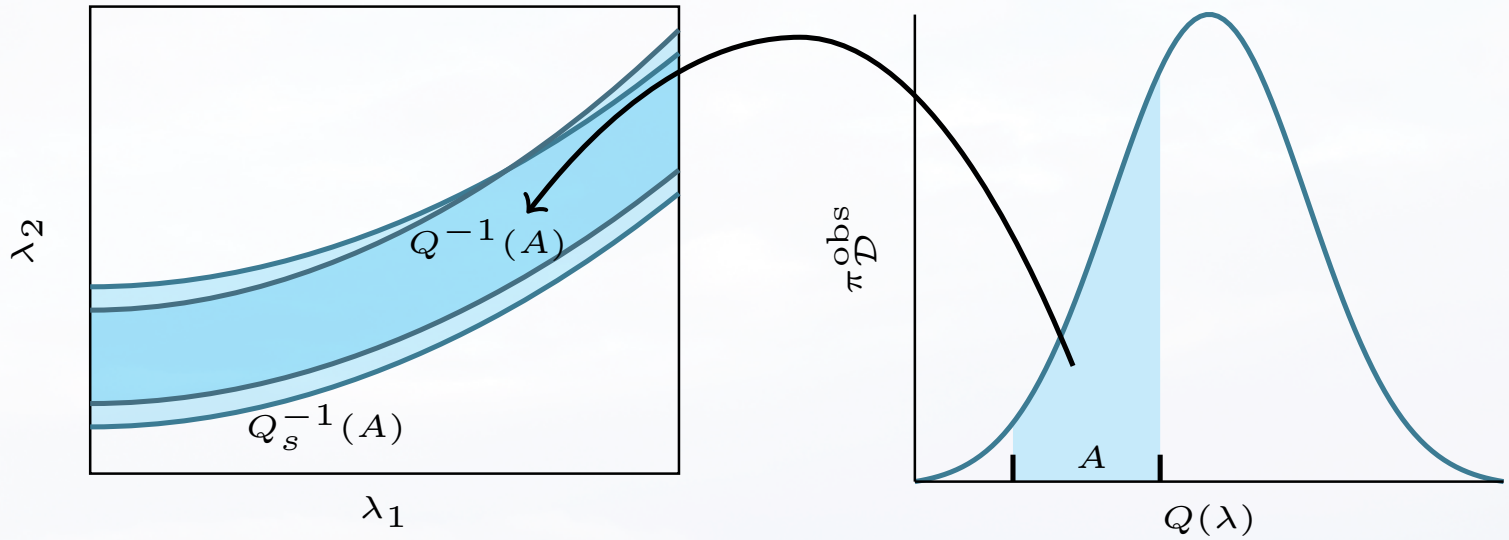
Assume $Q_s : \mathbf{\Lambda} \to \mathcal{D}$ is measurable. Given probability measures $P_{\mathbf{\Lambda}}^{\text{prior}}$ and $P_{\mathcal{D}}^{\text{obs}}$ that are absolutely continuous with respect to $\mu_{\mathbf{\Lambda}}$ and $\mu_{\mathcal{D}}$ respectively, the *posterior* probability density calculated from the surrogate model satisfies the consistency requirement.

Recall that this means for any $A \in \mathcal{B}_{\mathcal{D}}$ we have

$$P_{\mathbf{\Lambda}}^{\text{post}}(Q_s^{-1}(A)) = P_{\mathcal{D}}^{\text{obs}}(A)$$

i.e., propagating the posterior though the surrogate model gives the observed data density.

Utilizing Surrogate Models



The only issue is that typically

$$Q_s^{-1}(A) \neq Q^{-1}(A)$$

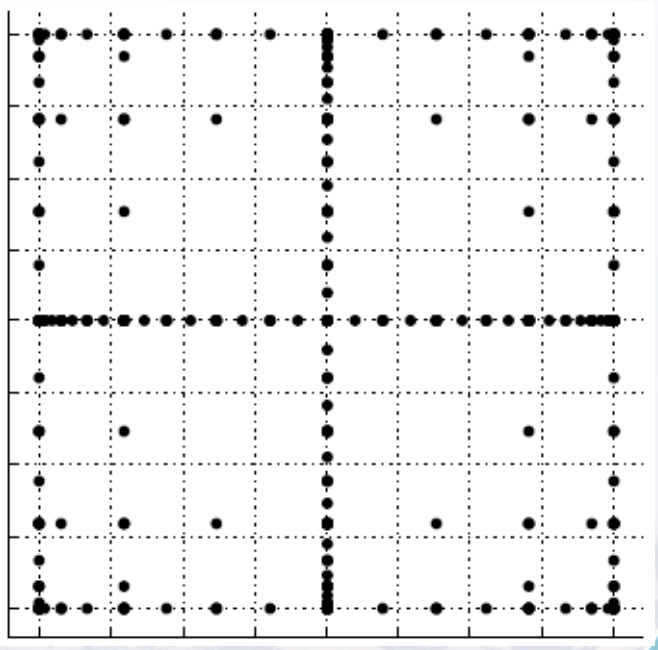
The surrogate model needs to be accurate while utilizing a minimal number of model evaluations

Generalized Sparse Grid Approximations

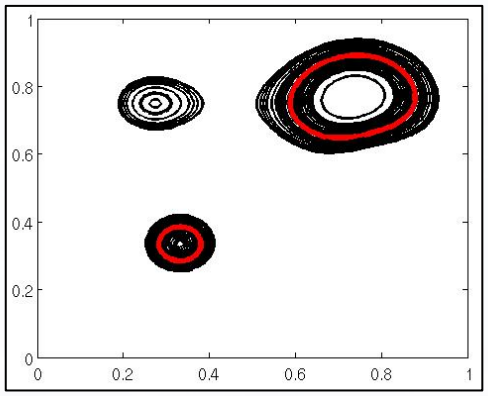
Polynomial chaos using sparse grid approximations has proven to be very effective for a range of problems.

Several appealing properties

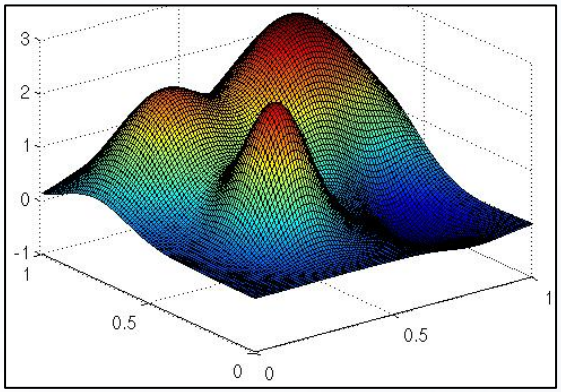
- Comparable accuracy with tensor product grids
- Number of points grows slower with dimension
- Enables either local or dimensional adaptivity
- Easy to incorporate goal-oriented error indicators [Jakeman, W. JCP 2015]



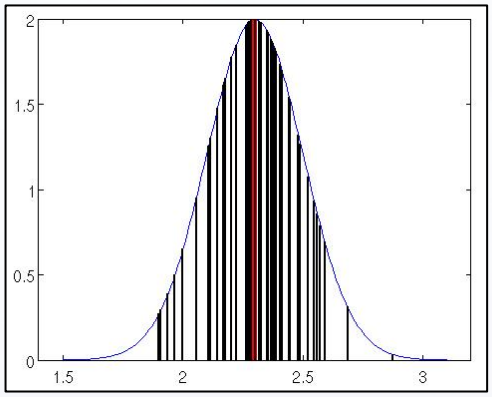
Example: Gaussian Peaks Problem



Contours in input space corr. to samples



Peaks function



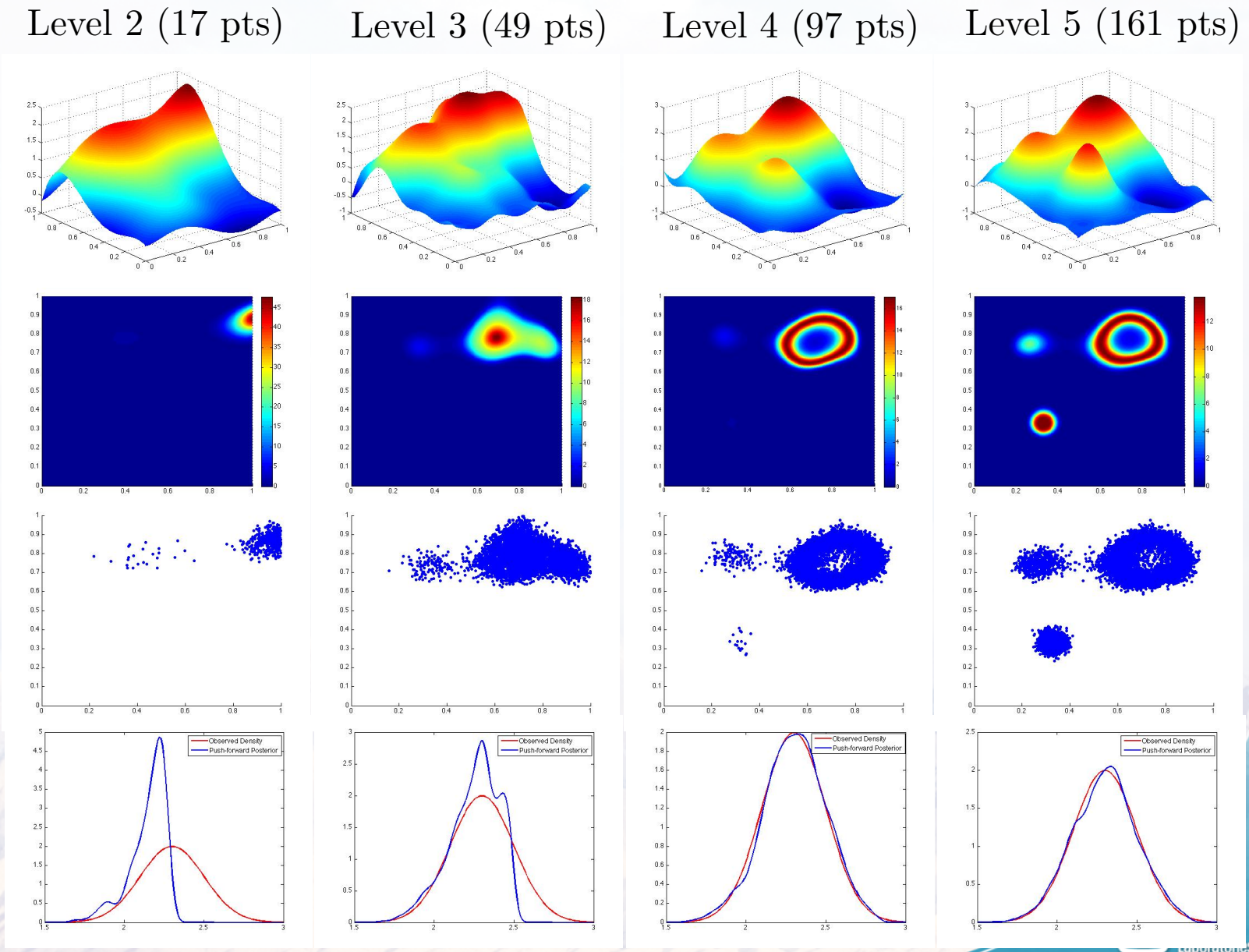
Distribution on data with samples

We let $\mathbf{\Lambda} = [0, 1]^2$ and consider a map composed of the sum of 4 Gaussian peaks

The prior is uniform over $\mathbf{\Lambda}$.

The goal is to investigate how the accuracy of the surrogate model affects the posterior

Example: Gaussian Peaks Problem



Adaptive Surrogates: Peaks Problem

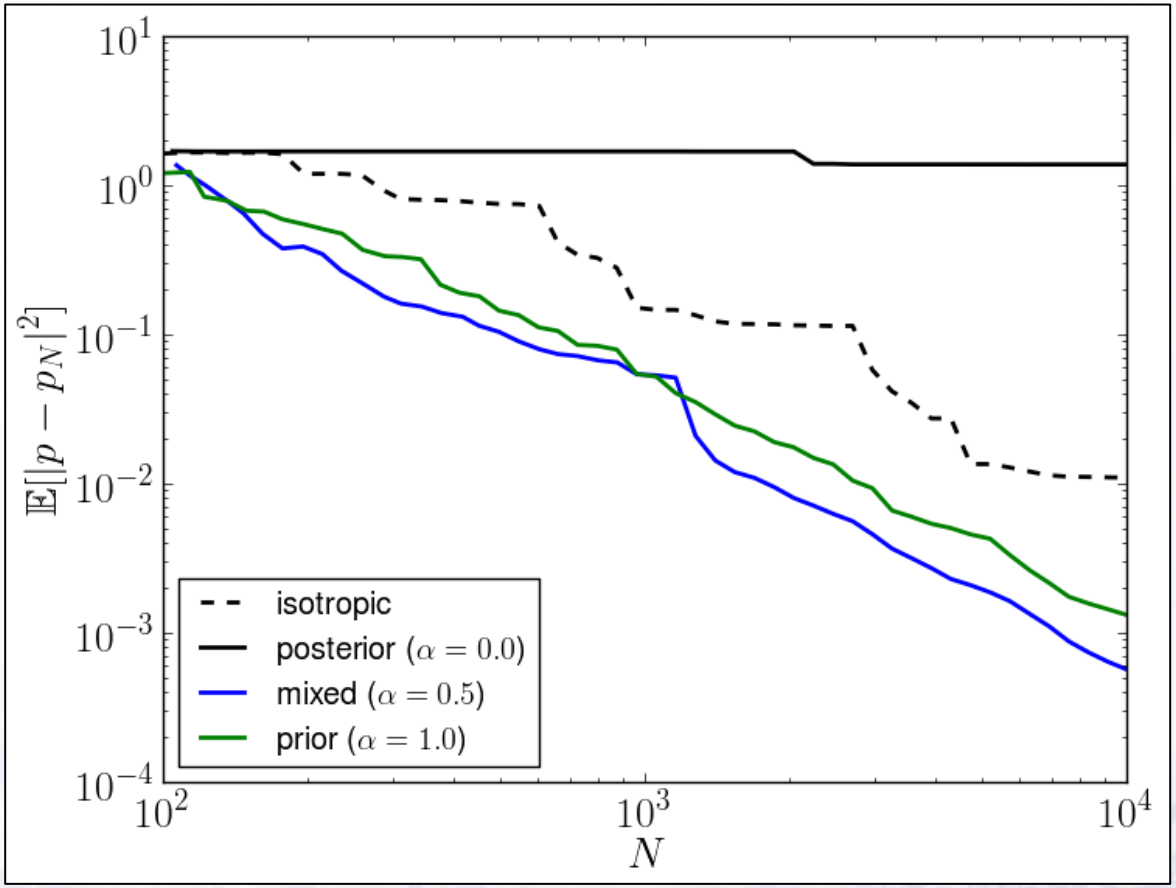


Figure: Convergence of the posterior to the "true" posterior computed using an overkill solution. We compare isotropic sparse grid refinement with local adaptivity driven by the hierarchical surplus for the forward map.

Summary and Future Work

- We have developed a **Bayesian** approach to solve **stochastic inverse problems**
- Our approach is **consistent** with the model and the data
  - The push-forward of the posterior probability density through the model produces the observed density on the QoI
- Provides a unification of Bayesian methodologies, measure-theoretic approaches, and forward UQ.
- We can leverage advanced approaches for forward UQ such as adaptive response surface approximations
- Current/future work
  - Better adaptive strategies based on the information gained in the posterior
  - Optimal experimental design
  - Inference for prediction

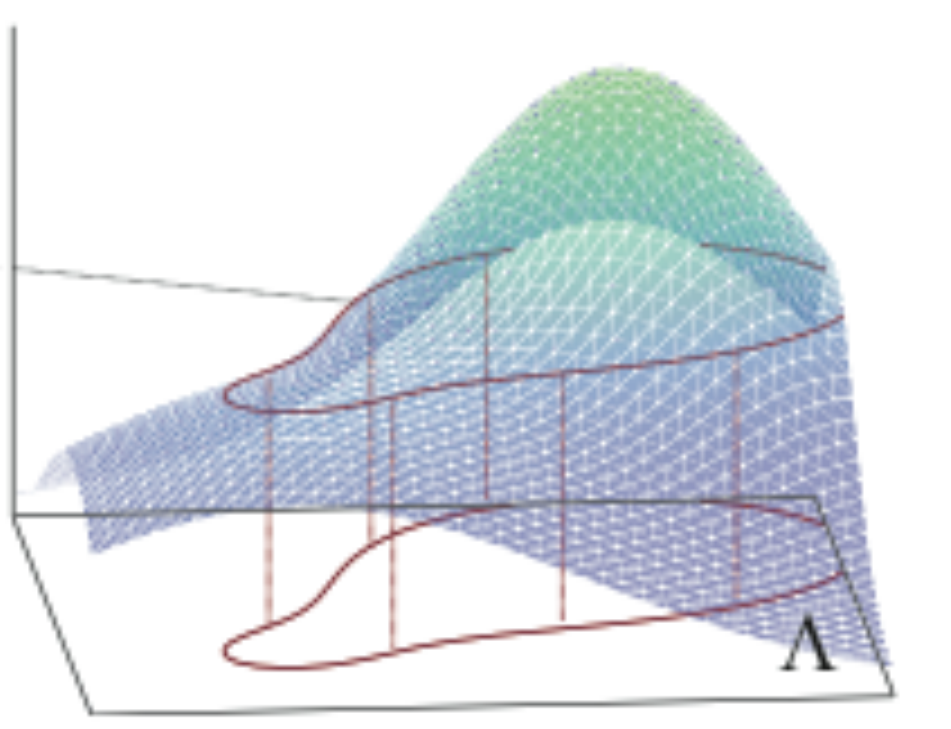
Thank you for your attention!
Questions?

Email: tmwilde@sandia.gov

Sandia National Laboratories

Extra Slides

Deterministic Inverse Problems

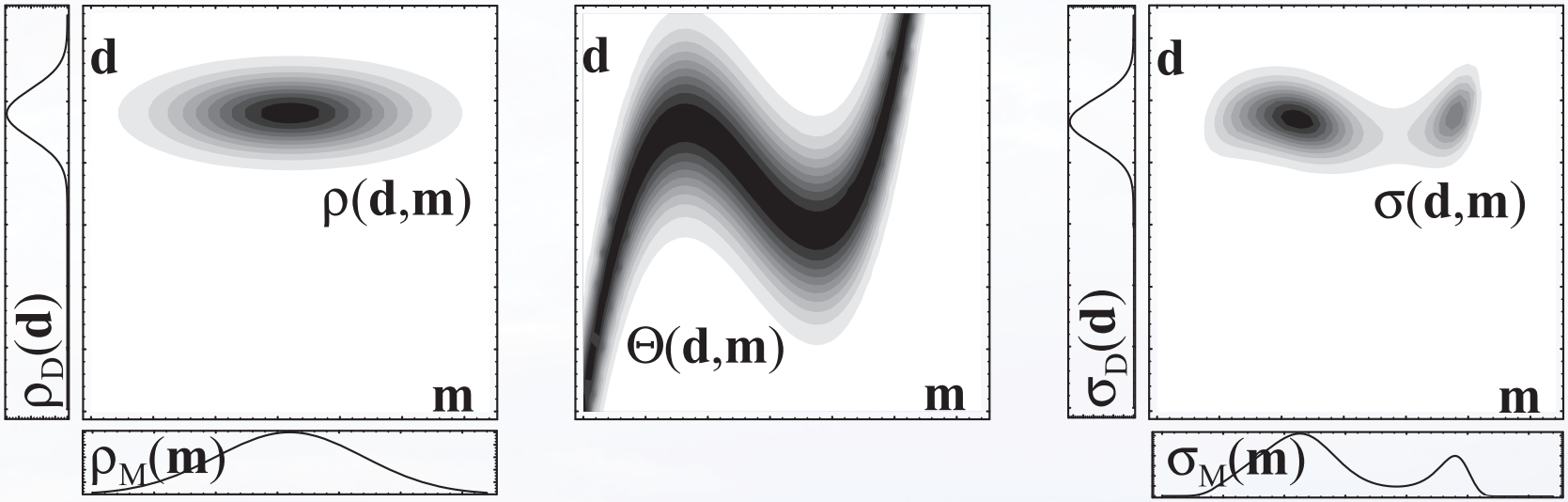


Given a vector of deterministic observations, $\hat{Q}$, find $\lambda \in \Lambda$ such that $Q(\lambda) = \hat{Q}$

Solutions are often non-unique

A unique solution can be found by imposing more structure, e.g. regularization

Stochastic Inverse Problems



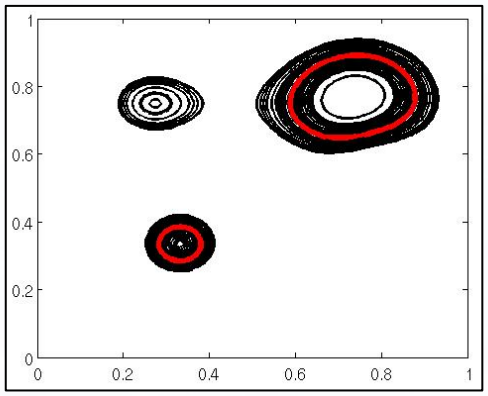
Given a vector of stochastic observations, $P_{\mathcal{D}}^{\text{obs}}$, find $P_{\mathbf{\Lambda}}^{\text{post}}$ such that $P_{\mathcal{D}}^{Q(\text{post})}(B) = P_{\mathcal{D}}^{\text{obs}}(B)$ for any $B \in \mathcal{B}_{\mathcal{D}}$

Solutions are often non-unique

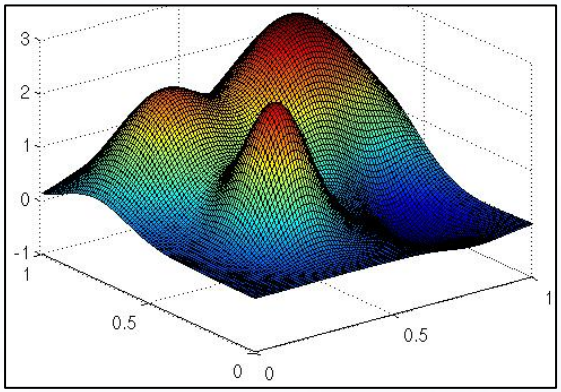
A unique solution can be found by imposing more structure, e.g. a prior

Sandia National Laboratories

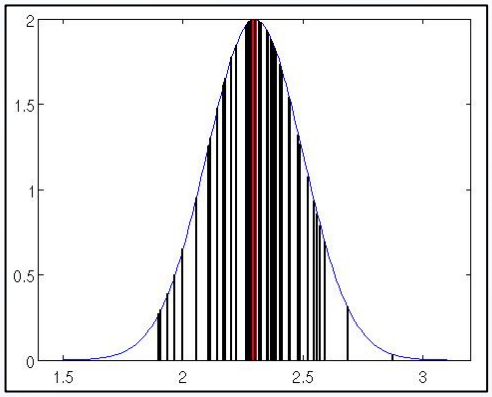
Stochastic Inverse Problems



Contours in input space corr. to samples



Peaks function



Distribution on data with samples

Given a vector of stochastic observations, $P_{\mathcal{D}}^{\text{obs}}$, find $P_{\mathbf{\Lambda}}^{\text{post}}$ such that $P_{\mathcal{D}}^{Q(\text{post})}(B) = P_{\mathcal{D}}^{\text{obs}}(B)$ for any $B \in \mathcal{B}_{\mathcal{D}}$

Solutions are often non-unique

A unique solution can be found by imposing more structure, e.g. a prior

Sandia National Laboratories