Data Assimilation and Model Inversion

for Decision-Making in a Data-Rich World

Mark Asch

Total (0.8)
Université de Picardie Jules Verne (0.2)

IPAM, UCLA - Sep'24-28th, 2018.

Motivation

[A. Tarantola]

- Using observations to infer the values of some parameters corresponds to solving an 'inverse problem'. Practitioners usually seek the 'best solution' implied by the data, but observations should only be used to falsify possible solutions, not to deduce any particular solution.
- FINDING A NEEDLE IN A HAYSTACK IS HARD IF THE HAYSTACK HAS HUNDREDS OF DIMENSIONS...
- Bayes is the (best) way to go!

Motivation

[A. Tarantola]

- Using observations to infer the values of some parameters corresponds to solving an 'inverse problem'. Practitioners usually seek the 'best solution' implied by the data, but observations should only be used to falsify possible solutions, not to deduce any particular solution.
- FINDING A NEEDLE IN A HAYSTACK IS HARD IF THE HAYSTACK HAS HUNDREDS OF DIMENSIONS...
- Bayes is the (best) way to go!

Motivation

[A. Tarantola]

- Using observations to infer the values of some parameters corresponds to solving an 'inverse problem'. Practitioners usually seek the 'best solution' implied by the data, but observations should only be used to falsify possible solutions, not to deduce any particular solution.
- FINDING A NEEDLE IN A HAYSTACK IS HARD IF THE HAYSTACK HAS HUNDREDS OF DIMENSIONS...
- Bayes is the (best) way to go!

Motivation

[A. Tarantola]

- Using observations to infer the values of some parameters corresponds to solving an 'inverse problem'. Practitioners usually seek the 'best solution' implied by the data, but observations should only be used to falsify possible solutions, not to deduce any particular solution.
- FINDING A NEEDLE IN A HAYSTACK IS HARD IF THE HAYSTACK HAS HUNDREDS OF DIMENSIONS...
- Bayes is the (best) way to go!

Abstract

We have a reasonable corpus of theory and experience in Model Inversion and Data Assimilation, but the abundance of Data that is available today is changing the paradigms and inciting us to study new approaches to these problems. In this talk we will begin by reviewing “classical” inverse problem and data assimilation theories and insist on their broad applicability. Then we will broach the potential of data-driven assimilation and inversion, and detail some of the most promising methods for this - specifically for time-series data. This will lead us to the challenging question: how can one couple model- and data-driven approaches to solve complex, real-world control problems in the domain of energy?

Abstract

We have a reasonable corpus of theory and experience in Model Inversion and Data Assimilation, but the abundance of Data that is available today is changing the paradigms and inciting us to study new approaches to these problems. In this talk we will begin by reviewing "classical" inverse problem and data assimilation theories and insist on their broad applicability. Then we will broach the potential of data-driven assimilation and inversion, and detail some of the most promising methods for this - specifically for time-series data. This will lead us to the challenging question: how can one couple model- and data-driven approaches to solve complex, real-world control problems in the domain of energy?

Plan of the Talk

1. Context and Challenges.
2. Worked examples: for a (realistic) toy problem, perform
    1. deterministic inversion,
    2. statistical inversion.
3. Conclusions and Open Problems.

Plan of the Talk

1. Context and Challenges.
2. Worked examples: for a (realistic) toy problem, perform
    1. deterministic inversion,
    2. statistical inversion.
3. Conclusions and Open Problems.

Plan of the Talk

1. Context and Challenges.
2. Worked examples: for a (realistic) toy problem, perform
   1. deterministic inversion,
   2. statistical inversion.
3. Conclusions and Open Problems.

Plan of the Talk

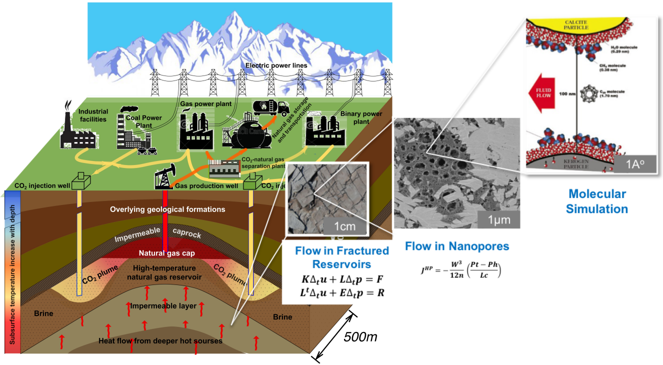
1. Context and Challenges.
2. Worked examples: for a (realistic) toy problem, perform
    1. deterministic inversion,
    2. statistical inversion.
3. Conclusions and Open Problems.

Plan of the Talk

1. Context and Challenges.
2. Worked examples: for a (realistic) toy problem, perform
    1. deterministic inversion,
    2. statistical inversion.
3. Conclusions and Open Problems.

Plan

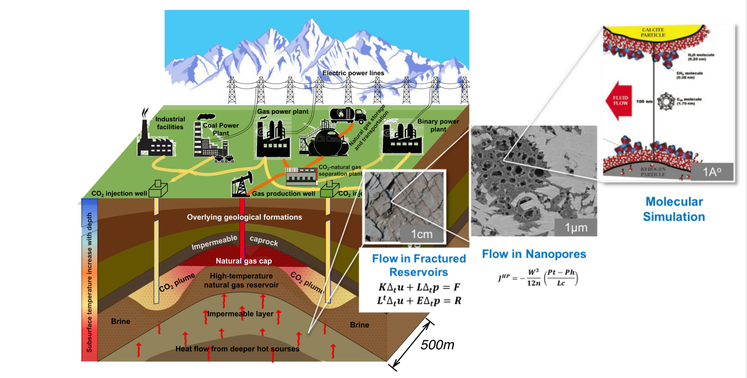
Forecasting Challenge



- multi-scale
- multi-physics
- nonlinear
- stochastic...
- data- and compute-intensive

Coupling of Carbon Capture and Storage (CCS) and geothermal energy extraction to enable permanent storage of CO2 in (sedimentary basin) geothermal reservoirs, while simultaneously extracting heat energy that can be used to generate clean, i.e., CO2-emission-free, baseload or dispatchable power. [Credit: M. Saar, ETH Zurich]

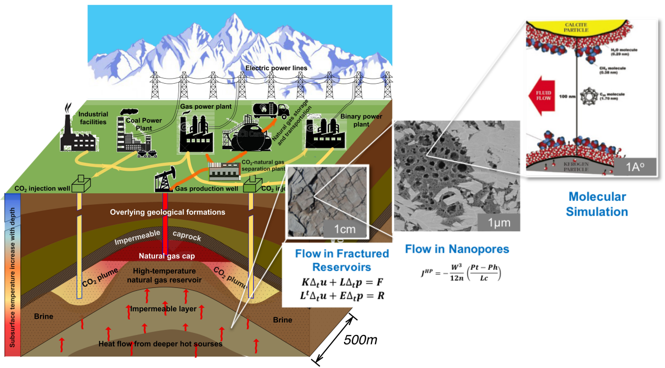
Forecasting Challenge



- multi-scale
- multi-physics
- nonlinear
- stochastic...
- data- and compute-intensive

Coupling of Carbon Capture and Storage (CCS) and geothermal energy extraction to enable permanent storage of CO2 in (sedimentary basin) geothermal reservoirs, while simultaneously extracting heat energy that can be used to generate clean, i.e., CO2-emission-free, baseload or dispatchable power. [Credit: M. Saar, ETH Zurich]

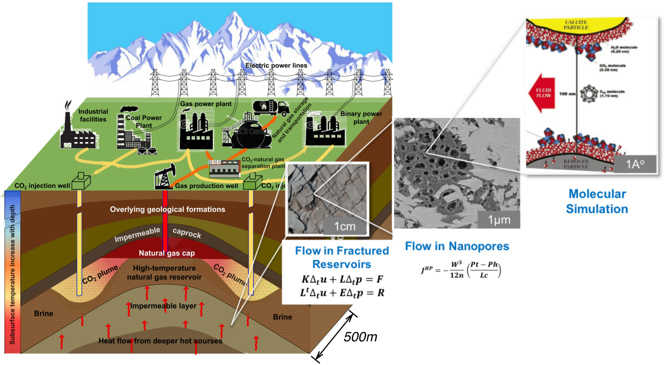
Forecasting Challenge



- multi-scale
- multi-physics
- nonlinear
- stochastic...
- data- and compute-intensive

Coupling of Carbon Capture and Storage (CCS) and geothermal energy extraction to enable permanent storage of CO2 in (sedimentary basin) geothermal reservoirs, while simultaneously extracting heat energy that can be used to generate clean, i.e., CO2-emission-free, baseload or dispatchable power. [Credit: M. Saar, ETH Zurich]

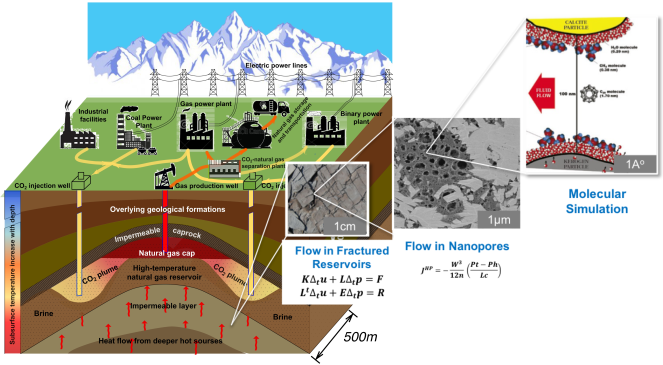
Forecasting Challenge



- multi-scale
- multi-physics
- nonlinear
- stochastic...
- data- and compute-intensive

Coupling of Carbon Capture and Storage (CCS) and geothermal energy extraction to enable permanent storage of CO2 in (sedimentary basin) geothermal reservoirs, while simultaneously extracting heat energy that can be used to generate clean, i.e., CO2-emission-free, baseload or dispatchable power. [Credit: M. Saar, ETH Zurich]

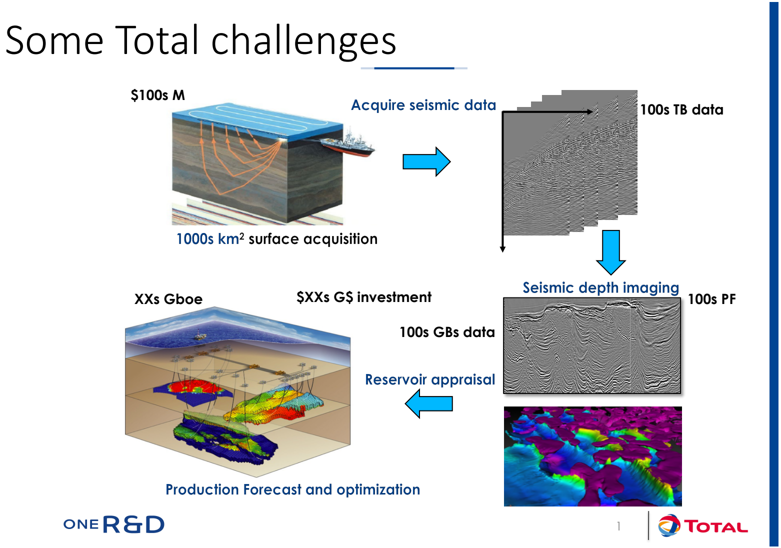
Forecasting Challenge



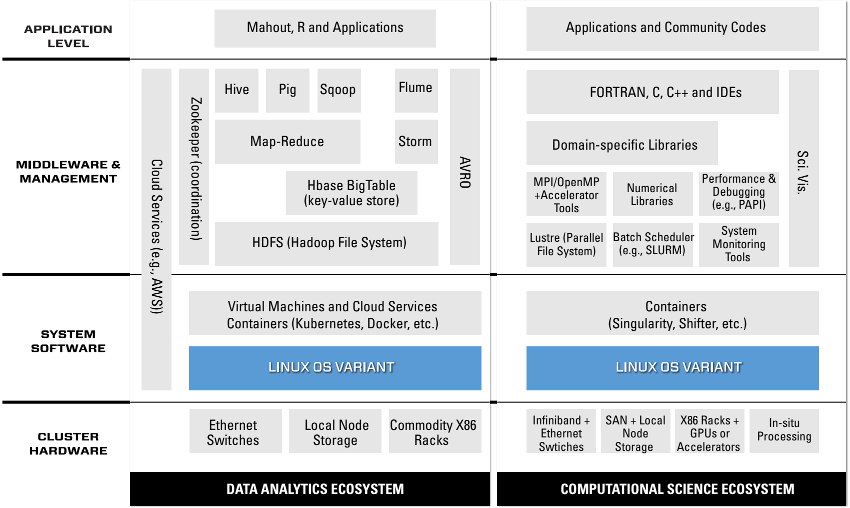
- multi-scale
- multi-physics
- nonlinear
- stochastic...
- data- and compute-intensive

Coupling of Carbon Capture and Storage (CCS) and geothermal energy extraction to enable permanent storage of CO2 in (sedimentary basin) geothermal reservoirs, while simultaneously extracting heat energy that can be used to generate clean, i.e., CO2-emission-free, baseload or dispatchable power. [Credit: M. Saar, ETH Zurich]

CSE Challenges



The split

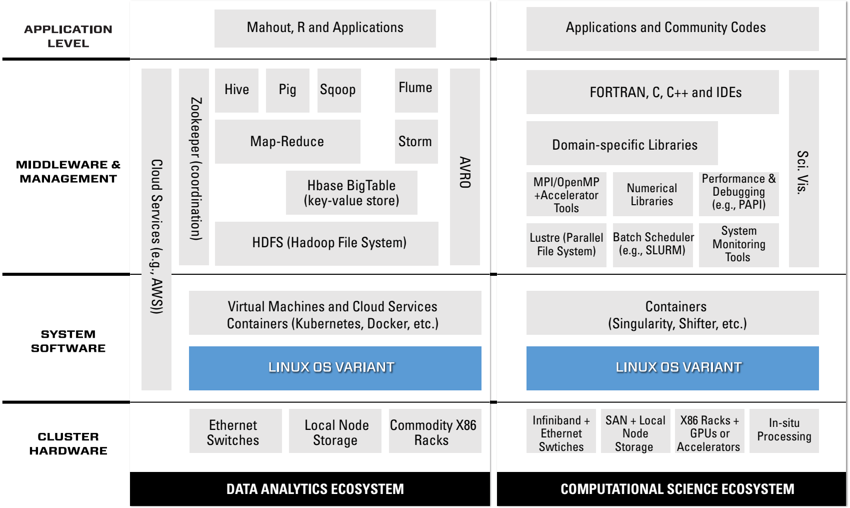


We'll go on living

separate lives...

[P. Collins. 1985]

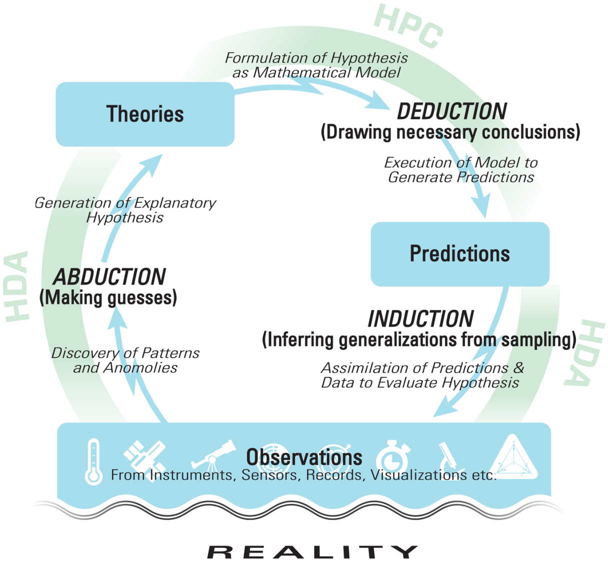
The split



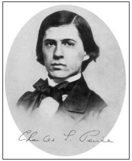
We'll go on living

separate lives...

[P. Collins. 1985]



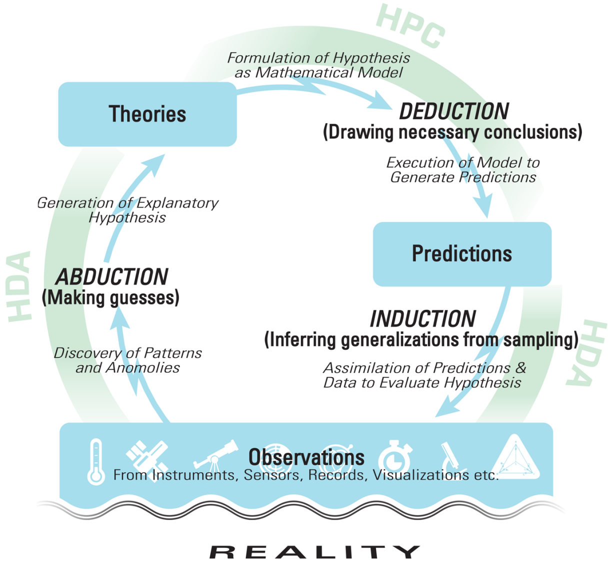
The inference cycle. [Asch, Moore, et al. IJHPCA. 2018]



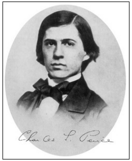
Come together, right now...
[T. Beatles. 1969]

To move from reality to decision-making, we must go back to Peirce's Logic (1903)

- Abductive (explore data, suggest hypothesis - effect to cause)
- Deductive (refine hypothesis - cause to effect)
- Inductive (empirical substantiation - specific to general)



The inference cycle. [Asch, Moore, et al. IJHPCA. 2018]



Come together, right now...
[T. Beatles. 1969]

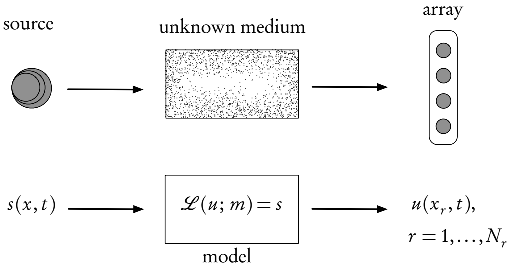
To move from reality to decision-making, we must go back to Peirce's Logic (1903)

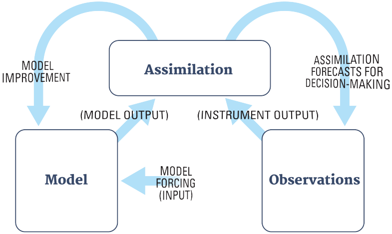
- Abductive (explore data, suggest hypothesis - effect to cause)
- Deductive (refine hypothesis - cause to effect)
- Inductive (empirical substantiation - specific to general)

Inverse Problems - Introduction

- definitions + cartoons
- different strokes...
  - optimal control (initial, boundary)
  - data assimilation
  - parameter identification
  - history matching
- 3 spaces:
  - space of data $\mathcal{D}$
  - space of models $\mathcal{M}$
  - space of observations $\mathcal{O}$
- Q1: how to invert in $\mathcal{M} \cap \mathcal{D}$? (“classical” model-space inversion)
- Q2: how to invert in $\mathcal{D} \cap \mathcal{O}$? (“new” data-space inversion)

Inverse Problems - Definitions





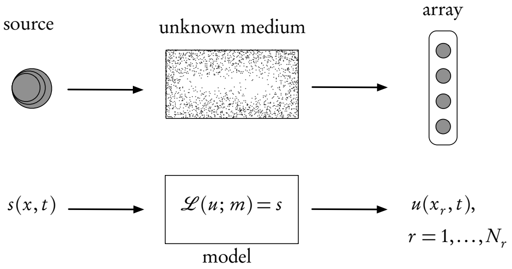
Take a parameter-dependent dynamical system, $\mathrm{d}u/\mathrm{d}t = s(t, u; m)$, $\quad u(t_0) = u_0$, with s a known source, $m \in \mathcal{M}$, $u(t) \in \mathbb{R}^k$.

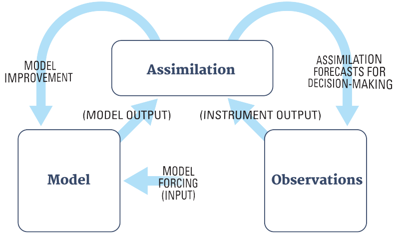
Forward: Given m, u_0, find $u(t)$ for $t \geq t_0$.

Inverse: Given $u(t)$ for $t \geq t_0$, find $m \in \mathcal{M}$.

- Observation equation: $f(t, m) = \mathcal{H}z(t, m)$, where $\mathcal{H}$ is the observation operator.
- Usually we have a finite number of discrete (space-time) observations $\{\tilde{y}_j\}_{j=1}^{n}$, where $\tilde{y}_j = f(t_j, m) + \epsilon_j$ and ϵ_j represents the measurement error - an additive (Gaussian) noise model.

Inverse Problems - Definitions





Take a parameter-dependent dynamical system, $\mathrm{d}u/\mathrm{d}t = s(t, u; m), \qquad u(t_0) = u_0,$ with s a known source, $m \in \mathcal{M}$, $u(t) \in \mathbb{R}^k$.

Forward: Given m, u_0, find $u(t)$ for $t \geq t_0$.

Inverse: Given $u(t)$ for $t \geq t_0$, find $m \in \mathcal{M}$.

- Observation equation: $f(t, m) = \mathcal{H}z(t, m)$, where $\mathcal{H}$ is the observation operator.
- Usually we have a finite number of discrete (space-time) observations $\{\tilde{y}_j\}_{j=1}^n$, where $\tilde{y}_j = f(t_j, m) + \epsilon_j$ and ϵ_j represents the measurement error - an additive (Gaussian) noise model.

Classical Inverse Problem

Definition

In a typical inverse problem, we seek to determine plausible values of model parameters given inexact (uncertain) data and an assumed theoretical model relating the observed data to the model.

Examples

These are difficult problems that can be found in diverse areas of science, engineering, and medicine. Examples from geophysics include (1) forecasting production performance of geothermal reservoirs, (2) extracting oil and gas, (3) estimating pathways of subsurface contaminant transport, (4) hydrology and groundwater management, and many others.

Classical Inverse Problem

Definition

In a typical inverse problem, we seek to determine plausible values of model parameters given inexact (uncertain) data and an assumed theoretical model relating the observed data to the model.

Examples

These are difficult problems that can be found in diverse areas of science, engineering, and medicine. Examples from geophysics include (1) forecasting production performance of geothermal reservoirs, (2) extracting oil and gas, (3) estimating pathways of subsurface contaminant transport, (4) hydrology and groundwater management, and many others.

Well-posedness of IPs

Let X and Y be two normed spaces and let $K : X \longrightarrow Y$ be a map between the two. The problem of finding x given y, such that $K(x) = y$, is *well-posed* if the following 3 properties hold:

1. Existence: for every $y \in Y$ there is (at least) one solution $x \in X$ such that $Kx = y$.
2. Uniqueness: for every $y \in Y$ there is at most one $x \in X$ such that $Kx = y$.
3. Continuous dependence of solutions on observations (stability) : the solution x depends continuously on the data y in that for every sequence $\{x_n\} \subset X$ with $Kx_n \longrightarrow Kx$ as $n \longrightarrow \infty$, we have that $x_n \longrightarrow x$ as $n \longrightarrow \infty$.

- The existence and uniqueness together are also known as "identifiability".
- The continuous dependence is related to the "stability" of the inverse problem.

Warning:

All (interesting) inverse problems are ill-posed.

Well-posedness of IPs

Let X and Y be two normed spaces and let $K : X \to Y$ be a map between the two. The problem of finding x given y, such that $K(x) = y$, is *well-posed* if the following 3 properties hold:

1. Existence: for every $y \in Y$ there is (at least) one solution $x \in X$ such that $Kx = y$.
2. Uniqueness: for every $y \in Y$ there is at most one $x \in X$ such that $Kx = y$.
3. Continuous dependence of solutions on observations (stability) : the solution x depends continuously on the data y in that for every sequence $\{x_n\} \subset X$ with $Kx_n \to Kx$ as $n \to \infty$, we have that $x_n \to x$ as $n \to \infty$.

- The existence and uniqueness together are also known as "identifiability".
- The continuous dependence is related to the "stability" of the inverse problem.

Warning:

All (interesting) inverse problems are ill-posed.

Well-posedness of IPs

Let X and Y be two normed spaces and let $K : X \longrightarrow Y$ be a map between the two. The problem of finding x given y, such that $K(x) = y$, is *well-posed* if the following 3 properties hold:

1. Existence: for every $y \in Y$ there is (at least) one solution $x \in X$ such that $Kx = y$.
2. Uniqueness: for every $y \in Y$ there is at most one $x \in X$ such that $Kx = y$.
3. Continuous dependence of solutions on observations (stability) : the solution x depends continuously on the data y in that for every sequence $\{x_n\} \subset X$ with $Kx_n \longrightarrow Kx$ as $n \longrightarrow \infty$, we have that $x_n \longrightarrow x$ as $n \longrightarrow \infty$.

- The existence and uniqueness together are also known as "identifiability".
- The continuous dependence is related to the "stability" of the inverse problem.

Warning:

All (interesting) inverse problems are ill-posed.

Inverse Problem: a simple example

To recap (in finite-dimension):

- $\mathbf{d}$ is the data (vector, $\dim(\mathbf{d}) = n$),
- $\mathbf{m}$ is the model parameter (vector, $\dim(\mathbf{m}) = p$),
- we have an equation, $G(\mathbf{m}) = \mathbf{d}$, relating the two.
  - Forward problem: find $\mathbf{d}$ given $\mathbf{m}$, by solving the ode/pde/integral equation
  - Inverse problem: find $\mathbf{m}$ given $\mathbf{d}$, by "inverting" the ode/pde/integral equation
- But due to ill-posedness, $(n \ll p)$, we have to minimize a (regularized) cost function to achieve this,

$$\min_{\mathbf{m}} \|G(\mathbf{m}) - \mathbf{d}\|_X^2 + \alpha^2 \|L\mathbf{m}\|_X^2 ,$$

where α is a regularization coefficient and L is a (discretized) gradient operator.

Inverse Problem: a simple example II

Consider the boundary value problem, for f a given function and b, c unknown parameters:

$$\begin{cases} -b(x)u''(x) + c(x)u'(x) = f(x), & 0 < x < 1, \\ u(0) = 0, \;\; u(1) = 0. \end{cases}$$

- Discretize the bvp by finite differences, with $h = x_{i+1} - x_i$, $0 = x_0, x_1, \ldots, x_N = 1$

$$\begin{cases} -b_i \frac{u_{i+1} - 2u_i + u_{i+1}}{h^2} + c_i \frac{u_{i+1} - u_i}{h} = f_i, & i = 1, \ldots, N-1, \\ u_0 = 0, \;\; u_N = 0, \end{cases}$$

- Rewrite as a linear system, $a_i^{(1)} = -b_i/h^2 + c_i/h$, $a_i^{(2)} = 2b/h^2 - c_i/h$, $a_i^{(3)} = b_i/h^2$,

$$\begin{bmatrix} 1 & 0 & \cdots & \cdots & \cdots & 0 \\ a_1^{(1)} & a_1^{(2)} & a_1^{(3)} & & & \\ & \ddots & \ddots & \ddots & & \\ & & a_i^{(1)} & a_i^{(2)} & a_i^{(3)} & \\ & & & \ddots & \ddots & \ddots \\ 0 & \cdots & \cdots & \cdots & 0 & 1 \end{bmatrix} \begin{bmatrix} u_0 \\ \vdots \\ u_i \\ \vdots \\ u_N \end{bmatrix} = \begin{bmatrix} f_0 \\ \vdots \\ f_i \\ \vdots \\ f_N \end{bmatrix} \Leftrightarrow A\mathbf{u} = \mathbf{f}$$

Inverse Problem: a simple example II

Consider the boundary value problem, for f a given function and b, c unknown parameters:

$$\begin{cases} -b(x)u''(x) + c(x)u'(x) = f(x), & 0 < x < 1, \\ u(0) = 0, \; u(1) = 0. \end{cases}$$

- Discretize the bvp by finite differences, with $h = x_{i+1} - x_i$, $0 = x_0, x_1, \ldots, x_N = 1$

$$\begin{cases} -b_i \frac{u_{i+1} - 2u_i + u_{i+1}}{h^2} + c_i \frac{u_{i+1} - u_i}{h} = f_i, & i = 1, \ldots, N-1, \\ u_0 = 0, \; u_N = 0, \end{cases}$$

- Rewrite as a linear system, $a_i^{(1)} = -b_i/h^2 + c_i/h$, $a_i^{(2)} = 2b/h^2 - c_i/h$, $a_i^{(3)} = b_i/h^2$,

$$\begin{bmatrix} 1 & 0 & \cdots & \cdots & \cdots & 0 \\ a_1^{(1)} & a_1^{(2)} & a_1^{(3)} & & & \\ & \ddots & \ddots & \ddots & & \\ & & a_i^{(1)} & a_i^{(2)} & a_i^{(3)} & \\ & & & \ddots & \ddots & \ddots \\ 0 & \cdots & \cdots & \cdots & 0 & 1 \end{bmatrix} \begin{bmatrix} u_0 \\ \vdots \\ u_i \\ \vdots \\ u_N \end{bmatrix} = \begin{bmatrix} f_0 \\ \vdots \\ f_i \\ \vdots \\ f_N \end{bmatrix} \Leftrightarrow A\mathbf{u} = \mathbf{f}$$

Inverse Problem: a simple example III

- For given parameters, we can now solve the linear system (at will, but note that is nonlinear in the parameters b and c)

$$\begin{aligned}\mathbf{u} &= A^{-1}\mathbf{f}\\ &\doteq G(\mathbf{m}).\end{aligned}$$

- For given observations, d, choose α and solve the optimization problem

$$m_* = \arg\min_{\mathbf{m}} \left\|G(\mathbf{m}) - \mathbf{d}\right\|_2^2 + \alpha^2 \left\|L\mathbf{m}\right\|_2^2$$

directly by a quasi-Newton method of choice, or by an adjoint method when possible - see [1].

Inverse Problem: a simple example III

- For given parameters, we can now solve the linear system (at will, but note that is nonlinear in the parameters b and c)

$$\begin{aligned}\mathbf{u} &= A^{-1}\mathbf{f} \\ &\doteq G(\mathbf{m}).\end{aligned}$$

- For given observations, $\mathbf{d}$, choose α and solve the optimization problem

$$m_* = \arg\min_{\mathbf{m}} \|G(\mathbf{m}) - \mathbf{d}\|_2^2 + \alpha^2 \|L\mathbf{m}\|_2^2$$

directly by a quasi-Newton method of choice, or by an adjoint method when possible - see [1].

Model-Space Inversion

Inversion

A major part of scientific discovery and research deals with questions of the nature: what can be said about the value of an unknown, or badly known variable m, that represents the parameters of the system, if we have some measured data $\mathcal{D}$ and a model $\mathcal{M}$ of the underlying mechanism that generated the data?

Model-Space Inversion

Inversion

A major part of scientific discovery and research deals with questions of the nature: what can be said about the value of an unknown, or badly known variable m, that represents the parameters of the system, if we have some measured data $\mathcal{D}$ and a model $\mathcal{M}$ of the underlying mechanism that generated the data?

Challenge

The current grand challenge in computational science and engineering is the solution of large-scale statistical inverse problems governed by PDEs that involve large amounts of (observational) data - see the Inference Cycle.

Bayesian Inversion

But solving the IP is precisely the Bayesian context, where we seek a quantification of the uncertainty in our knowledge of the parameters that according to Bayes' Theorem takes the form,

$$p(m \mid \mathcal{D}) = \frac{p(\mathcal{D} \mid m)p(m)}{p(\mathcal{D})} = \frac{p(\mathcal{D} \mid m)p(m)}{\int_m p(\mathcal{D} \mid m)p(m)}.$$

- Physical model is represented by the conditional probability (also known as the *likelihood*) $p(\mathcal{D} \mid m)$,
- Prior knowledge of the system by the term $p(m)$.
- The denominator is considered as a normalizing factor and represents the total probability of $\mathcal{D}$.
- From these we can then calculate the resulting posterior probability, $p(m \mid \mathcal{D})$.

Bayesian Inversion

But solving the IP is precisely the Bayesian context, where we seek a quantification of the uncertainty in our knowledge of the parameters that according to Bayes' Theorem takes the form,

$$p(m \mid \mathcal{D}) = \frac{p(\mathcal{D} \mid m)p(m)}{p(\mathcal{D})} = \frac{p(\mathcal{D} \mid m)p(m)}{\int_m p(\mathcal{D} \mid m)p(m)}.$$

- Physical model is represented by the conditional probability (also known as the *likelihood*) $p(\mathcal{D} \mid m)$,
- Prior knowledge of the system by the term $p(m)$.
- The denominator is considered as a normalizing factor and represents the total probability of $\mathcal{D}$.
- From these we can then calculate the resulting posterior probability, $p(m \mid \mathcal{D})$.

Bayesian Inversion

But solving the IP is precisely the Bayesian context, where we seek a quantification of the uncertainty in our knowledge of the parameters that according to Bayes' Theorem takes the form,

$$p(m \mid \mathcal{D}) = \frac{p(\mathcal{D} \mid m)p(m)}{p(\mathcal{D})} = \frac{p(\mathcal{D} \mid m)p(m)}{\int_m p(\mathcal{D} \mid m)p(m)}.$$

- Physical model is represented by the conditional probability (also known as the *likelihood*) $p(\mathcal{D} \mid m)$,
- Prior knowledge of the system by the term $p(m)$.
- The denominator is considered as a normalizing factor and represents the total probability of $\mathcal{D}$.
- From these we can then calculate the resulting posterior probability, $p(m \mid \mathcal{D})$.

Bayesian Inversion

But solving the IP is precisely the Bayesian context, where we seek a quantification of the uncertainty in our knowledge of the parameters that according to Bayes' Theorem takes the form,

$$p(m \mid \mathcal{D}) = \frac{p(\mathcal{D} \mid m)p(m)}{p(\mathcal{D})} = \frac{p(\mathcal{D} \mid m)p(m)}{\int_m p(\mathcal{D} \mid m)p(m)}.$$

- Physical model is represented by the conditional probability (also known as the *likelihood*) $p(\mathcal{D} \mid m)$,
- Prior knowledge of the system by the term $p(m)$.
- The denominator is considered as a normalizing factor and represents the total probability of $\mathcal{D}$.
- From these we can then calculate the resulting posterior probability, $p(m \mid \mathcal{D})$.

Bayesian Inversion

But solving the IP is precisely the Bayesian context, where we seek a quantification of the uncertainty in our knowledge of the parameters that according to Bayes' Theorem takes the form,

$$p(m \mid \mathcal{D}) = \frac{p(\mathcal{D} \mid m)p(m)}{p(\mathcal{D})} = \frac{p(\mathcal{D} \mid m)p(m)}{\int_m p(\mathcal{D} \mid m)p(m)}.$$

- Physical model is represented by the conditional probability (also known as the *likelihood*) $p(\mathcal{D} \mid m)$,
- Prior knowledge of the system by the term $p(m)$.
- The denominator is considered as a normalizing factor and represents the total probability of $\mathcal{D}$.
- From these we can then calculate the resulting posterior probability, $p(m \mid \mathcal{D})$.

MAP, MLE and co.

Definition

The most probable estimator, called the maximum a posterior (MAP) estimator, is the value of the parameter/model that maximizes the posterior probability, $m_* = \arg\max_m p(m \mid \mathcal{D})$.

MLE, LS

Note that for a flat, or uninformative prior, $p(m)$, the MAP is just the maximum likelihood estimator (MLE), the value of m that maximizes the likelihood $p(\mathcal{D} \mid m)$ of the model that generated the data-in this case neither $p(m)$ nor the denominator play a role in the optimization. When the model is linear and noise is Gaussian, we obtain the least squares estimate.

MAP, MLE and co.

Definition

The most probable estimator, called the maximum a posterior (MAP) estimator, is the value of the parameter/model that maximizes the posterior probability, $m_* = \arg\max_m p(m \mid \mathcal{D})$.

MLE, LS

Note that for a flat, or uninformative prior, $p(m)$, the MAP is just the maximum likelihood estimator (MLE), the value of m that maximizes the likelihood $p(\mathcal{D} \mid m)$ of the model that generated the data-in this case neither $p(m)$ nor the denominator play a role in the optimization. When the model is linear and noise is Gaussian, we obtain the least squares estimate.

Challenge

A point estimator is often insufficient. So the current grand challenge has now become: "how to explore the (high-dimensional, non-Gaussian) posterior probability?", a task that is notoriously intractable for real problems of interest.

UQ for Model Inversion - exploring the posterior

- Monte-Carlo: generate a large number of realizations, then compute statistics of the resulting distribution.
- Monte-Carlo Markov Chain: use a more sophisticated sampling scheme to generate the posterior.
- Ensemble Kalman Filters and Smoothers: widely used in weather forecasting...

Cost

These approaches are far too expensive to compute if we want high resolution models and reliable UQ.

Solutions

A huge number of "remedies" exist [Ghanem, et al. Springer Handbook on UQ, 2053 pp. (2017)]: replace forward problem with a reduced order model in both parameter and state space; approximate parameter-to-observable map, or the posterior with a Gaussian process response surface; employ a polynomial chaos approximation of the forward problem; use two-stage "delayed acceptance" MCMC method (multifidelity approach) eg. AMFMC; employ gradient information (of the negative log posterior) to accelerate sampling; exploit Riemannian geometry of parameter space to accelerate sampling; use sparse grid methods; create stochastic Newton MCMC method that uses local gradient and low-rank Hessian information of the negative log posterior to construct a local Gaussian approximation; employ a data-space inversion strategy (see below), etc.

UQ for Model Inversion - exploring the posterior

- Monte-Carlo: generate a large number of realizations, then compute statistics of the resulting distribution.
- Monte-Carlo Markov Chain: use a more sophisticated sampling scheme to generate the posterior.
- Ensemble Kalman Filters and Smoothers: widely used in weather forecasting...

Cost

These approaches are far too expensive to compute if we want high resolution models and reliable UQ.

Solutions

A huge number of "remedies" exist [Ghanem, et al. Springer Handbook on UQ, 2053 pp. (2017)]: replace forward problem with a reduced order model in both parameter and state space; approximate parameter-to-observable map, or the posterior with a Gaussian process response surface; employ a polynomial chaos approximation of the forward problem; use two-stage "delayed acceptance" MCMC method (multifidelity approach) eg. AMFMC; employ gradient information (of the negative log posterior) to accelerate sampling; exploit Riemannian geometry of parameter space to accelerate sampling; use sparse grid methods; create stochastic Newton MCMC method that uses local gradient and low-rank Hessian information of the negative log posterior to construct a local Gaussian approximation; employ a data-space inversion strategy (see below), etc.

UQ for Model Inversion - exploring the posterior

- Monte-Carlo: generate a large number of realizations, then compute statistics of the resulting distribution.
- Monte-Carlo Markov Chain: use a more sophisticated sampling scheme to generate the posterior.
- Ensemble Kalman Filters and Smoothers: widely used in weather forecasting…

Cost

These approaches are far too expensive to compute if we want high resolution models and reliable UQ.

Solutions

A huge number of “remedies” exist [Ghanem, et al. Springer Handbook on UQ, 2053 pp. (2017)]: replace forward problem with a reduced order model in both parameter and state space; approximate parameter-to-observable map, or the posterior with a Gaussian process response surface; employ a polynomial chaos approximation of the forward problem; use two-stage “delayed acceptance" MCMC method (multifidelity approach) eg. AMFMC; employ gradient information (of the negative log posterior) to accelerate sampling; exploit Riemannian geometry of parameter space to accelerate sampling; use sparse grid methods; create stochastic Newton MCMC method that uses local gradient and low-rank Hessian information of the negative log posterior to construct a local Gaussian approximation; employ a data-space inversion strategy (see below), etc.

Data-Space Inversion

- Pure data-space inversion, based on Machine Learning, can potentially obviate the need for model-based approach, if:
  - the examples/data are representative enough of the underlying information-theoretic composition, and
  - the learning algorithm is agile enough to capture the subtleties of the task.

- However:
  - we do not (in general) have enough data (realizations);
  - ML does not (easily) provide UQ;
  - coupling ML and other methods might be an option?

Data-Space Inversion

- Pure data-space inversion, based on Machine Learning, can potentially obviate the need for model-based approach, if:
  - the examples/data are representative enough of the underlying information-theoretic composition, and
  - the learning algorithm is agile enough to capture the subtleties of the task.

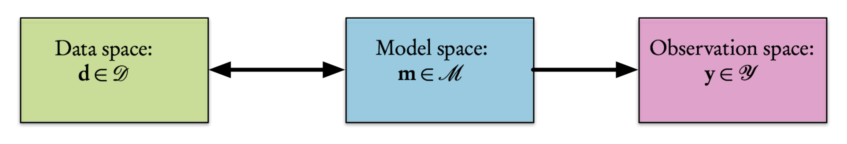
- However:
  - we do not (in general) have enough data (realizations);
  - ML does not (easily) provide UQ;
  - coupling ML and other methods might be an option?

Combined Data- Model-Space Inversions

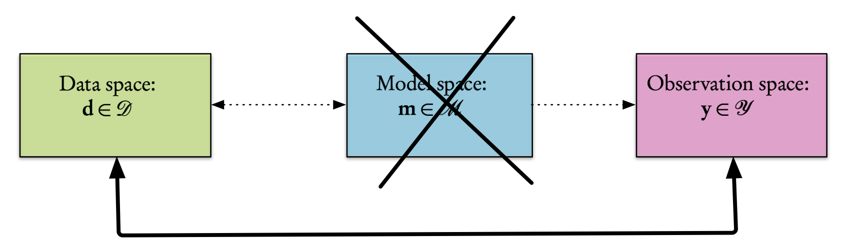
Idea

Reduce the number of model solutions by using mappings from data- to observation-space.

- We want to replace:



- by this:

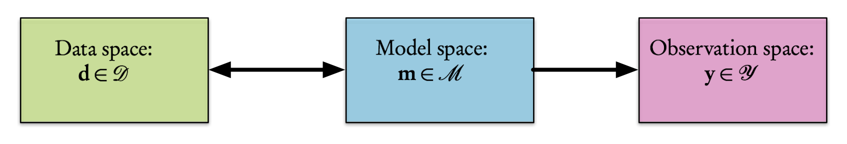


Combined Data- Model-Space Inversions

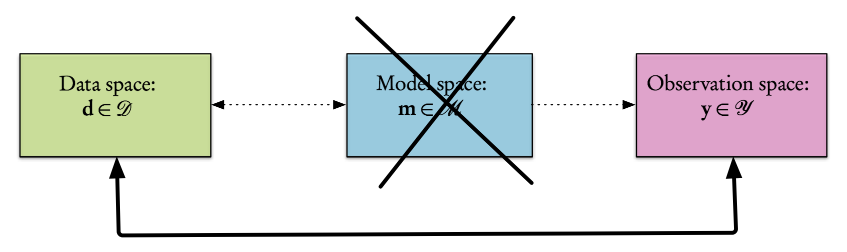
Idea

Reduce the number of model solutions by using mappings from data- to observation-space.

- We want to replace:



- by this:

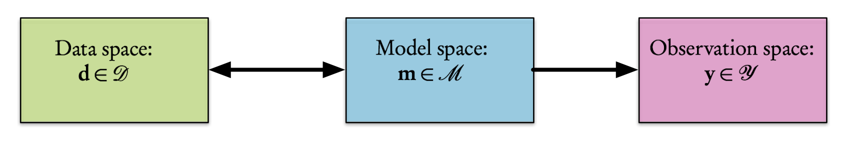


Combined Data- Model-Space Inversions

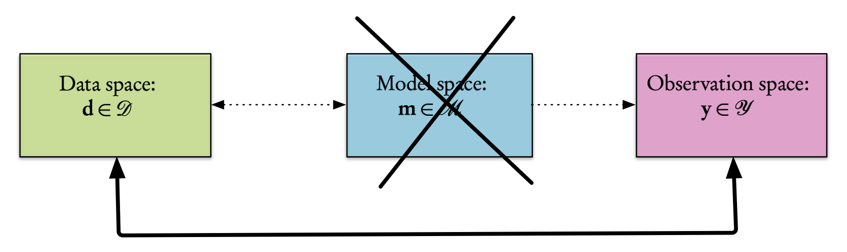
Idea

Reduce the number of model solutions by using mappings from data- to observation-space.

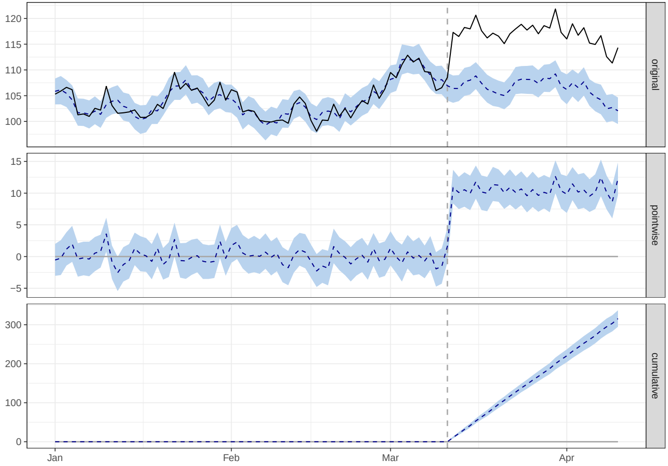
- We want to replace:



- by this:



Goal: Inference for Time Series + UQ



[Brodersen KH, Gallusser F, Koehler J, Remy N, Scott SL. Inferring causal impact using Bayesian structural time-series models. *Annals of Applied Statistics*, 2015, Vol. 9, No. 1, 247-274.]

Prediction Focused Analysis

In PFA we consider, in addition to data variables and model variables, the intended prediction, or outcome, variables - i.e. the intended purpose of the study, in the optic of a decision-making process.

- $\mathrm{d} \in \mathcal{D}$ is the data (a vector in the finite-dimensional case) that represents the (time-varying) data variables of the underlying (physical) problem;
- $\mathrm{m} \in \mathcal{M}$ is the (physical) model;
- $\mathrm{d} = g_d(\mathrm{m})$ is the (generally nonlinear) mapping from the model space $\mathcal{M}$ into the data space $\mathcal{D}$ (in our case, g is a PDE that is expensive to solve);
- finally, $\mathrm{y} = g_h(\mathrm{d})$ is the observed/measured (or predicted) data in the observation space $\mathcal{Y}$.

We are interested here in cases where the $\dim(y) \ll \dim(d)$,leading to so-called evidential learning (as opposed to causal learning), where we aim to *learn the relationship between the data variables (the evidence) and the decision variables (the decision hypothesis) via a statistical model.*

Prediction Focused Analysis

In PFA we consider, in addition to data variables and model variables, the intended prediction, or outcome, variables - i.e. the intended purpose of the study, in the optic of a decision-making process.

- $\mathbf{d} \in \mathcal{D}$ is the data (a vector in the finite-dimensional case) that represents the (time-varying) data variables of the underlying (physical) problem;
- $\mathbf{m} \in \mathcal{M}$ is the (physical) model;
- $\mathbf{d} = g_d(\mathbf{m})$ is the (generally nonlinear) mapping from the model space $\mathcal{M}$ into the data space $\mathcal{D}$ (in our case, g is a PDE that is expensive to solve);
- finally, $\mathbf{y} = g_h(\mathbf{d})$ is the observed/measured (or predicted) data in the observation space $\mathcal{Y}$.

We are interested here in cases where the $\dim(y) \ll \dim(d)$,leading to so-called evidential learning (as opposed to causal learning), where we aim to *learn the relationship between the data variables (the evidence) and the decision variables (the decision hypothesis) via a statistical model.*

Prediction Focused Analysis

In PFA we consider, in addition to data variables and model variables, the intended prediction, or outcome, variables - i.e. the intended purpose of the study, in the optic of a decision-making process.

- $\mathbf{d} \in \mathcal{D}$ is the data (a vector in the finite-dimensional case) that represents the (time-varying) data variables of the underlying (physical) problem;
- $\mathbf{m} \in \mathcal{M}$ is the (physical) model;
- $\mathbf{d} = g_d(\mathbf{m})$ is the (generally nonlinear) mapping from the model space $\mathcal{M}$ into the data space $\mathcal{D}$ (in our case, g is a PDE that is expensive to solve);
- finally, $\mathbf{y} = g_h(\mathbf{d})$ is the observed/measured (or predicted) data in the observation space $\mathcal{Y}$.

We are interested here in cases where the $\dim(y) \ll \dim(d)$,leading to so-called evidential learning (as opposed to causal learning), where we aim to *learn the relationship between the data variables (the evidence) and the decision variables (the decision hypothesis) via a statistical model.*

PFA: Rationale

- We would like to quantify posterior uncertainty on the forecast without history matching/inverting/assimilating individual models. Because:
  - direct modeling is complex;
  - models are extremely high dimensional and despite this, still remain a simplified representation of the actual subsurface geological and fluid complexity;
  - any sparse representation or dimension reduction method would further simplify an already simplified reality.
- On the other hand,
  - production data and forecast variables are simple time-series, on which
  - statistical dimension reduction techniques as well multivariate modeling can be readily applied,
  - without much loss of information.

PFA: Rationale

- We would like to quantify posterior uncertainty on the forecast without history matching/inverting/assimilating individual models. Because:
  - direct modeling is complex;
  - models are extremely high dimensional and despite this, still remain a simplified representation of the actual subsurface geological and fluid complexity;
  - any sparse representation or dimension reduction method would further simplify an already simplified reality.
- On the other hand,
  - production data and forecast variables are simple time-series, on which
  - statistical dimension reduction techniques as well multivariate modeling can be readily applied,
  - without much loss of information.

PFA: Rationale

- We would like to quantify posterior uncertainty on the forecast without history matching/inverting/assimilating individual models. Because:
  - direct modeling is complex;
  - models are extremely high dimensional and despite this, still remain a simplified representation of the actual subsurface geological and fluid complexity;
  - any sparse representation or dimension reduction method would further simplify an already simplified reality.
- On the other hand,
  - production data and forecast variables are simple time-series, on which
  - statistical dimension reduction techniques as well multivariate modeling can be readily applied,
  - without much loss of information.

PFA: Rationale

- We would like to quantify posterior uncertainty on the forecast without history matching/inverting/assimilating individual models. Because:
  - direct modeling is complex;
  - models are extremely high dimensional and despite this, still remain a simplified representation of the actual subsurface geological and fluid complexity;
  - any sparse representation or dimension reduction method would further simplify an already simplified reality.
- On the other hand,
  - production data and forecast variables are simple time-series, on which
  - statistical dimension reduction techniques as well multivariate modeling can be readily applied,
  - without much loss of information.

PFA: Rationale

- We would like to quantify posterior uncertainty on the forecast without history matching/inverting/assimilating individual models. Because:
  - direct modeling is complex;
  - models are extremely high dimensional and despite this, still remain a simplified representation of the actual subsurface geological and fluid complexity;
  - any sparse representation or dimension reduction method would further simplify an already simplified reality.
- On the other hand,
  - production data and forecast variables are simple time-series, on which
  - statistical dimension reduction techniques as well multivariate modeling can be readily applied,
  - without much loss of information.

PFA: Rationale

- We would like to quantify posterior uncertainty on the forecast without history matching/inverting/assimilating individual models. Because:
  - direct modeling is complex;
  - models are extremely high dimensional and despite this, still remain a simplified representation of the actual subsurface geological and fluid complexity;
  - any sparse representation or dimension reduction method would further simplify an already simplified reality.
- On the other hand,
  - production data and forecast variables are simple time-series, on which
  - statistical dimension reduction techniques as well multivariate modeling can be readily applied,
  - without much loss of information.

PFA: Rationale

- We would like to quantify posterior uncertainty on the forecast without history matching/inverting/assimilating individual models. Because:
  - direct modeling is complex;
  - models are extremely high dimensional and despite this, still remain a simplified representation of the actual subsurface geological and fluid complexity;
  - any sparse representation or dimension reduction method would further simplify an already simplified reality.
- On the other hand,
  - production data and forecast variables are simple time-series, on which
  - statistical dimension reduction techniques as well multivariate modeling can be readily applied,
  - without much loss of information.

PFA: Rationale

- We would like to quantify posterior uncertainty on the forecast without history matching/inverting/assimilating individual models. Because:
  - direct modeling is complex;
  - models are extremely high dimensional and despite this, still remain a simplified representation of the actual subsurface geological and fluid complexity;
  - any sparse representation or dimension reduction method would further simplify an already simplified reality.
- On the other hand,
  - production data and forecast variables are simple time-series, on which
  - statistical dimension reduction techniques as well multivariate modeling can be readily applied,
  - without much loss of information.

Algorithm

PFA proceeds as follows:

1. obtain/use a known prior $\pi(\mathbf{m})$ on $\mathbf{m}$;
2. generate a set of N prior models $\{\mathbf{m}_1, \ldots, \mathbf{m}_N\}$ by some MC/sampling algorithm;
3. evaluate the models by solving the forward problem $\mathbf{d} = g(\mathbf{m})$ and compute the forecast $\mathbf{y} = g_h(\mathbf{d})$ that can include measurement errors, but is basically a dimension reduction mechanism - these form the *training set* for statistical learning;
4. use a statistical dimension reduction method (bijective FPCA, FCCA) to generate reduced dimension vectors $\mathbf{d}^*$ and $\mathbf{y}^*$;
5. construct the multivariate distribution of $\mathbf{d}^*$ and $\mathbf{y}^*$ (by KDE, GPR, etc.);
6. back transform to obtain the posterior distribution $\pi(\mathbf{y}(\mathbf{m})|\mathbf{d}_{\text{obs}})$ - learn the statistical relationship.

Algorithm

PFA proceeds as follows:

1. obtain/use a known prior $\pi(\mathbf{m})$ on $\mathbf{m}$;
2. generate a set of N prior models $\{\mathbf{m}_1, \ldots, \mathbf{m}_N\}$ by some MC/sampling algorithm;
3. evaluate the models by solving the forward problem $\mathbf{d} = g(\mathbf{m})$ and compute the forecast $\mathbf{y} = g_h(\mathbf{d})$ that can include measurement errors, but is basically a dimension reduction mechanism - these form the *training set* for statistical learning;
4. use a statistical dimension reduction method (bijective FPCA, FCCA) to generate reduced dimension vectors $\mathbf{d}^*$ and $\mathbf{y}^*$;
5. construct the multivariate distribution of $\mathbf{d}^*$ and $\mathbf{y}^*$ (by KDE, GPR, etc.);
6. back transform to obtain the posterior distribution $\pi(\mathbf{y}(\mathbf{m})|\mathbf{d}_{\mathrm{obs}})$ - learn the statistical relationship.

Algorithm

PFA proceeds as follows:

1. obtain/use a known prior $\pi(\mathbf{m})$ on $\mathbf{m}$;
2. generate a set of N prior models $\{\mathbf{m}_1, \ldots, \mathbf{m}_N\}$ by some MC/sampling algorithm;
3. evaluate the models by solving the forward problem $\mathbf{d} = g(\mathbf{m})$ and compute the forecast $\mathbf{y} = g_h(\mathbf{d})$ that can include measurement errors, but is basically a dimension reduction mechanism - these form the *training set* for statistical learning;
4. use a statistical dimension reduction method (bijective FPCA, FCCA) to generate reduced dimension vectors d^* and y^*;
5. construct the multivariate distribution of d^* and y^* (by KDE, GPR, etc.);
6. back transform to obtain the posterior distribution $\pi(\mathrm{y}(\mathrm{m})|\mathrm{d}_{\mathrm{obs}})$ - learn the statistical relationship.

Algorithm

PFA proceeds as follows:

1. obtain/use a known prior $\pi(\mathbf{m})$ on $\mathbf{m}$;
2. generate a set of N prior models $\{\mathbf{m}_1, \ldots, \mathbf{m}_N\}$ by some MC/sampling algorithm;
3. evaluate the models by solving the forward problem $\mathbf{d} = g(\mathbf{m})$ and compute the forecast $\mathbf{y} = g_h(\mathbf{d})$ that can include measurement errors, but is basically a dimension reduction mechanism - these form the *training set* for statistical learning;
4. use a statistical dimension reduction method (bijective FPCA, FCCA) to generate reduced dimension vectors $\mathbf{d}^*$ and $\mathbf{y}^*$;
5. construct the multivariate distribution of $\mathbf{d}^*$ and $\mathbf{y}^*$ (by KDE, GPR, etc.);
6. back transform to obtain the posterior distribution $\pi(\mathbf{y}(\mathbf{m})|\mathbf{d}_{\mathrm{obs}})$ - learn the statistical relationship.

Algorithm

PFA proceeds as follows:

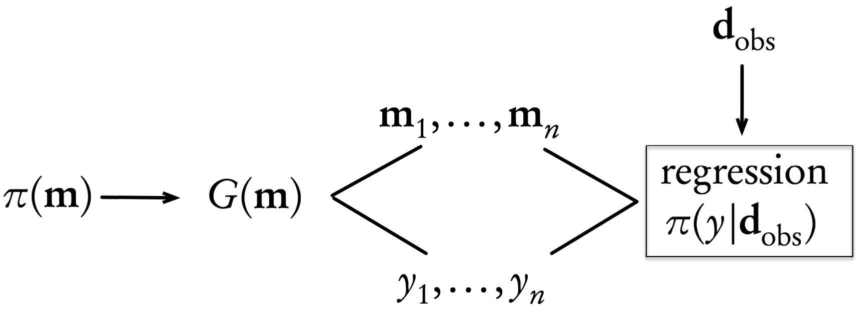
1. obtain/use a known prior $\pi(\mathbf{m})$ on $\mathbf{m}$;
2. generate a set of N prior models $\{\mathbf{m}_1, \ldots, \mathbf{m}_N\}$ by some MC/sampling algorithm;
3. evaluate the models by solving the forward problem $\mathbf{d} = g(\mathbf{m})$ and compute the forecast $\mathbf{y} = g_h(\mathbf{d})$ that can include measurement errors, but is basically a dimension reduction mechanism - these form the *training set* for statistical learning;
4. use a statistical dimension reduction method (bijective FPCA, FCCA) to generate reduced dimension vectors $\mathbf{d}^*$ and $\mathbf{y}^*$;
5. construct the multivariate distribution of $\mathbf{d}^*$ and $\mathbf{y}^*$ (by KDE, GPR, etc.);
6. back transform to obtain the posterior distribution $\pi(\mathbf{y}(\mathbf{m})|\mathbf{d}_{\mathrm{obs}})$ - learn the statistical relationship.

Algorithm

PFA proceeds as follows:

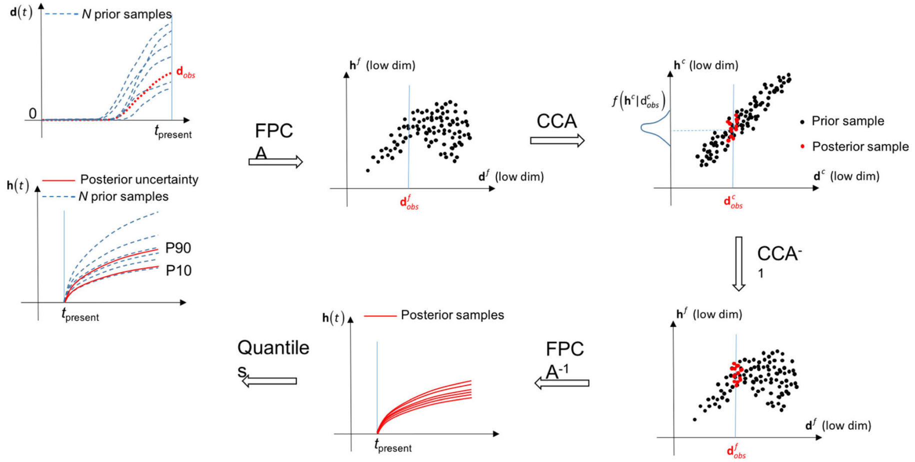
1. obtain/use a known prior $\pi(\mathbf{m})$ on $\mathbf{m}$;
2. generate a set of N prior models $\{\mathbf{m}_1, \ldots, \mathbf{m}_N\}$ by some MC/sampling algorithm;
3. evaluate the models by solving the forward problem $\mathbf{d} = g(\mathbf{m})$ and compute the forecast $\mathbf{y} = g_h(\mathbf{d})$ that can include measurement errors, but is basically a dimension reduction mechanism - these form the *training set* for statistical learning;
4. use a statistical dimension reduction method (bijective FPCA, FCCA) to generate reduced dimension vectors $\mathbf{d}^*$ and $\mathbf{y}^*$;
5. construct the multivariate distribution of $\mathbf{d}^*$ and $\mathbf{y}^*$ (by KDE, GPR, etc.);
6. back transform to obtain the posterior distribution $\pi(\mathbf{y}(\mathbf{m})|\mathbf{d}_{\text{obs}})$ - learn the statistical relationship.

Workflow 1



$$\boxed{\pi(y|\mathbf{d}_{\text{obs}}) = \frac{\pi(\mathbf{d}_{\text{obs}}|y)\pi(y)}{\pi(\mathbf{d}_{\text{obs}})}}$$

Workflow 2



[Credit: J. Caers, L. Li]

Priors

Statement of model parameterization and prior uncertainty

Parameter Name	Parameter code	Amount of parameters	Type of uncertainty	Established from	Fields of science
Local model lithology	ma	1	Scenario	Geophysics wells	geophysics, petrophysics, geostatistics
Regional and local permeability	Kh	22	Log-normal pdfs	Head data Well data	hydrogeology
pressure boundary conditions	ch	5	Uniform	Experience	hydrology engineering
River flows and conductance	riv	8	Conductance: log-normal DEM: uniform	Experience from previous studies	River science
Drain conductance	drn	8	Conductance: log-normal DEM: uniform	Experience from previous studies	hydrology
Aquifer Recharge	rch	1	Trapezoidal	Base-flow estimates	hydrology, meteorology, climate science

Prior distributions for the model data. [Credit: J. Caers]

Plan

Deterministic Inversion: Problem Formulation

Our toy problem is the estimation of an unknown (or badly known) coefficient, m, in an elliptic partial differential equation.

- Let $\Omega \subset \mathbb{R}^n$, $n \in \{1, 2, 3\}$ be an open, bounded domain, and consider the following inverse problem:

$$\min_m J(m) := \frac{1}{2}\int_\Omega (u - u_\mathrm{d})^2 \,\mathrm{d}x + \frac{\gamma}{2}\int_\Omega |\nabla m|^2 \,\mathrm{d}x, \quad (1)$$

where u is the solution of a Poisson, boundary-value problem,

$$\begin{aligned} -\nabla \cdot (\exp(m)\nabla u) &= f \text{ in } \Omega, \\ u &= 0 \text{ on } \partial\Omega. \end{aligned} \quad (2)$$

- $m \in \mathcal{M}_\mathrm{ad} := \{m \in H^1(\Omega) \bigcap L^\infty(\Omega)\}$ the unknown coefficient field,
- u_d denotes (possibly noisy) data,
- $f \in H^{-1}(\Omega)$ a given force, and
- $\gamma \geq 0$ the regularization parameter.

Note

This is a protoypical problem setting, and can be extended to a large number of other pde's, contexts and problems - see [1].

Deterministic Inversion: Problem Formulation

Our toy problem is the estimation of an unknown (or badly known) coefficient, m, in an elliptic partial differential equation.

- Let $\Omega \subset \mathbb{R}^n$, $n \in \{1, 2, 3\}$ be an open, bounded domain, and consider the following inverse problem:

$$\min_m J(m) := \frac{1}{2} \int_\Omega (u - u_\mathrm{d})^2 \, \mathrm{d}x + \frac{\gamma}{2} \int_\Omega |\nabla m|^2 \, \mathrm{d}x, \quad (1)$$

where u is the solution of a Poisson, boundary-value problem,

$$\begin{aligned} -\nabla \cdot (\exp(m) \nabla u) &= f \text{ in } \Omega, \\ u &= 0 \text{ on } \partial\Omega. \end{aligned} \quad (2)$$

- $m \in \mathcal{M}_\mathrm{ad} := \{m \in H^1(\Omega) \bigcap L^\infty(\Omega)\}$ the unknown coefficient field,
- u_d denotes (possibly noisy) data,
- $f \in H^{-1}(\Omega)$ a given force, and
- $\gamma \geq 0$ the regularization parameter.

Note

This is a protoypical problem setting, and can be extended to a large number of other pde's, contexts and problems - see [1].

Deterministic Inversion: Problem Formulation

Our toy problem is the estimation of an unknown (or badly known) coefficient, m, in an elliptic partial differential equation.

- Let $\Omega \subset \mathbb{R}^n$, $n \in \{1, 2, 3\}$ be an open, bounded domain, and consider the following inverse problem:

$$\min_m J(m) := \frac{1}{2} \int_\Omega (u - u_\mathrm{d})^2 \, \mathrm{d}x + \frac{\gamma}{2} \int_\Omega |\nabla m|^2 \, \mathrm{d}x, \qquad (1)$$

where u is the solution of a Poisson, boundary-value problem,

$$\begin{aligned} -\nabla \cdot (\exp(m) \nabla u) &= f \text{ in } \Omega, \\ u &= 0 \text{ on } \partial\Omega. \end{aligned} \qquad (2)$$

- $m \in \mathcal{M}_\mathrm{ad} := \{m \in H^1(\Omega) \bigcap L^\infty(\Omega)\}$ the unknown coefficient field,
- u_d denotes (possibly noisy) data,
- $f \in H^{-1}(\Omega)$ a given force, and
- $\gamma \geq 0$ the regularization parameter.

Note

This is a protoypical problem setting, and can be extended to a large number of other pde's, contexts and problems - see [1].

DIP: Numerical Solution

- We perform a series of increasingly large simulations to appreciate the computational cost of direct simulations, and of the deterministic inversion approach. This could, in principle, be used as "random differential equation" (a differential equation with random coefficients) approach, where the parameter is peturbed randomly, then for each perturbation we solve the inverse problem. Finally, we compute statistics of the ensemble. There are a number of caveats:
  - there is no propagation of uncertainty;
  - we have no guarantee that the ensemble is representative;
  - we require a very large number of simulations to obtain reliable (but, see above 2 points) statistics.
- Some advantages of the approach:
  - extremely simple and non-intrusive (requires no modification of the direct code);
  - can aid falsification (in a Bayesian approach) where we would like to ascertain whether realizations fall within a given distribution or not.

DIP: Computational Scaling

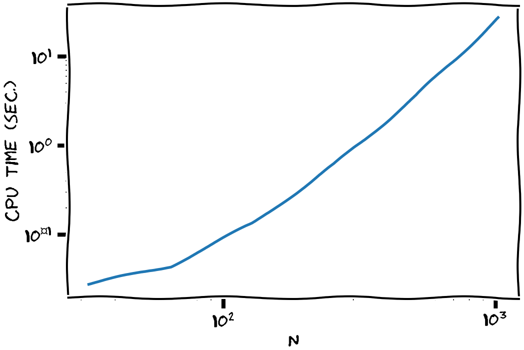
We perform a series of deterministic inversions to see how the computational cost evolves as a function of the problem size. The results (Table 1 and Figure 1) show a cubic (power of 3) exponential growth.

N	#dof	Time (sec.)
32	1024	0.0277
64	4096	0.0446
128	16384	0.1336
256	65536	0.6138
512	262144	3.6356
1024	1×10^6	27.112

N	#dof	Time (sec.)
32	1024	3.2
64	4096	10.9
128	16384	61.9
256	65536	368.1
512	262144	$\approx$ 3000

Table: CPU time for deterministic Poisson equation (left), deterministic inverse problem (right) - factor of 500X between direct and inverse. [Villa, U. and Petra, N. and Ghattas, O. *hIPPYlib: an Extensible Software Framework for Large-scale Deterministic and Bayesian Inversion.* 2016. http://hippylib.github.io]

DIP: Computational Scaling II



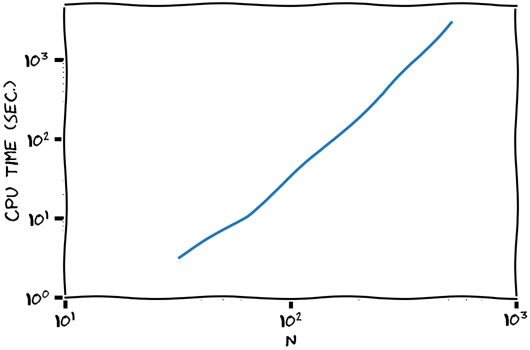


Figure: CPU time as function of problem size: direct (left), inverse problem (right).

DIP: Computational Scaling III

Observation

The computational cost for a direct solution scales as $\mathcal{O}\left(N^2\right)$ .The computational cost for an inverse problem solution is approximately $500\times$ that for the direct problem, i.e. $\mathcal{O}\left(N^3\right)$.

All computations performed in Ubuntu/Linux (Virtualbox) on MacBook Pro with 3.5 GHz Intel Core i7 and 16 GB of memory.

DIP: Computational Scaling III

Observation

The computational cost for a direct solution scales as $\mathcal{O}\left(N^2\right)$ .The computational cost for an inverse problem solution is approximately $500\times$ that for the direct problem, i.e. $\mathcal{O}\left(N^3\right)$.

All computations performed in Ubuntu/Linux (Virtualbox) on MacBook Pro with 3.5 GHz Intel Core i7 and 16 GB of memory.

Bayesian Inversion: Problem Formulation

Definition

A Bayesian Inverse Problem (BIP) is defined as follows:

- Given:
  - observational data and their uncertainties,
  - a (possibly stochastic) forward model that maps model parameters to observations,
  - and a prior probability distribution on model parameters that encodes any prior knowledge or assumptions about the parameters.

Find:

- the posterior probability distribution of the parameters conditioned on the observational data.

Bayesian Inversion: Problem Formulation

Definition

A Bayesian Inverse Problem (BIP) is defined as follows:

- Given:
  - observational data and their uncertainties,
  - a (possibly stochastic) forward model that maps model parameters to observations,
  - and a prior probability distribution on model parameters that encodes any prior knowledge or assumptions about the parameters.

BIP

This probability density function (pdf) is defined as the Bayesian solution of the inverse problem. The posterior distribution assigns to any candidate set of parameter fields our belief (expressed as a probability) that a member of this candidate set is the “true” parameter field that gave rise to the observed data.

Inverse Permeability Problem

- Determining the permeability of an unknown medium (subsurface rock, battery cell, etc.) is enormously important in a range of different applications. Among these applications are:
  - the prediction of transport of radioactive waste from underground waste repositories,
  - the forecast of geothermal production,
  - the optimization of oil recovery from underground fields,
  - the electrochemical behaviour of battery cells.
- Darcy's law is an excellent model of the pressure field as a function of the permeability,

$$
\begin{aligned}
-\nabla \cdot (k \nabla p) &= 0, \quad x \in D, \\
p &= h, \quad x \in \partial D.
\end{aligned}
$$

- The inverse problem is: find the permeability k from observations of the pressure at points in the interior of D.

BIP: Theoretical Result

A. Stuart's formulation of the BIP [4]. A stochastic inverse problem formulation for Darcy flow (and other pde's...) is:

- Consider Darcy flow with (log) permeability $u \in X = L^\infty(D)$

$$
\begin{aligned}
-\nabla \cdot (\exp(u) \nabla p) &= 0, \quad x \in D \\
u &= g, \quad x \in \partial D
\end{aligned}
\tag{3}
$$

- Find $u \in X$, given noisy observations

$$
y_j = p(x_j) + \eta_j, \tag{4}
$$

  where $\eta \sim \mathcal{N}(0, \Gamma)$ and the prior, μ_0 is a Gaussian measure on u
- Abstractly the inverse problem is: for $\mathcal{G} : X \mapsto Y = R^J$, find u given noisy measurements,

$$
y = \mathcal{G}(u) + \eta, \text{ noise.} \tag{5}
$$

BIP: Existence Theorem (well-posedness)

Theorem (A. Stuart)

Consider the Bayesian Inverse Problem (BIP) for $u(x) = \ln k(x)$ subject to observations of the form (4) where p solves the flow equation (3). Suppose that we have a prior measure $\mu_0 = \mathcal{N}(0, \beta)$, then $\mu^y(\mathrm{d}u) = P(\mathrm{d}u \mid y)$ is absolutely continuous with respect to μ_0, with Radon-Nikodym derivative

$$\frac{\mathrm{d}\mu^y}{\mathrm{d}\mu_0}(u) \propto \exp\left(-\frac{1}{2}|y - \mathcal{G}(u)|_\Gamma^2\right)$$

with $\mathcal{G}$ given by (5). The expression $\mathrm{d}\mu^y/\mathrm{d}\mu_0$ is precisely the posterior probability density function expressed in terms of measures.

BIP: Explore Posterior by Rejection Sampling

- Objective: generate a set of samples of the model $\mathbf{m}$ that are distributed according to the posterior $\sigma(\mathbf{m}) \propto \pi(\mathbf{m}) L(\mathbf{m})$ - we do not attempt to compute σ explicitly!
- Algorithm [Ripley, 1987] in Bayesian context:

for $i = 1 : N$
  sample $\mathbf{m}$ from its prior $\pi(\mathbf{m})$
  sample ξ from $\mathcal{U}[0, 1]$
  compute likelihood $L(\mathbf{m})$
  if $\xi \leq L(\mathbf{m})/S_L$, where $S_L = \sup L$
    accept $\mathbf{m}$

BIP: Numerical Solution by Rejection Sampling

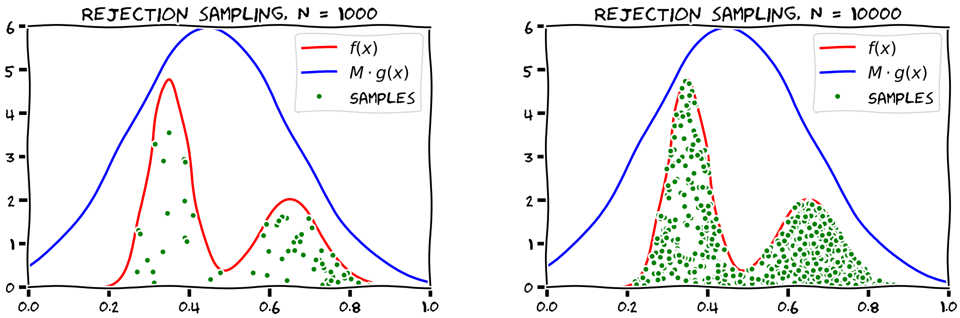


Figure: Rejection sampling example

But, the curse of dimensionality is fatal here... and leads to very low acceptance rates $\approx (\sigma_g/\sigma_f)^{-d}$. If, $\sigma_g/\sigma_f = 1.01$ in dimension $d = 1000$, then acceptance rate is $1/20000$.

BIP: Numerical Solution by Rejection Sampling

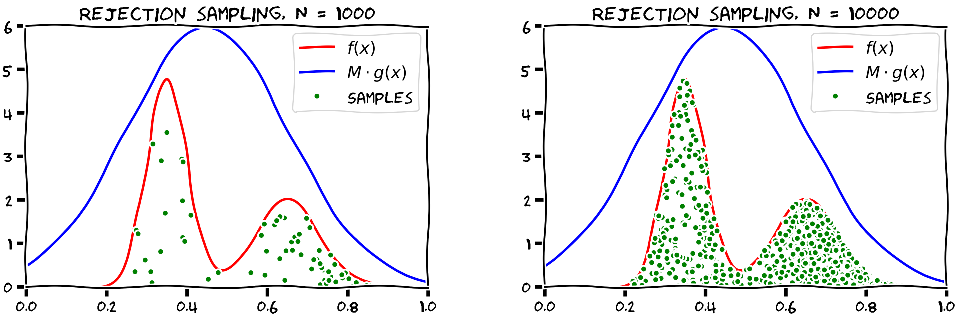


Figure: Rejection sampling example

But, the curse of dimensionality is fatal here... and leads to very low acceptance rates $\approx (\sigma_g/\sigma_f)^{-d}$. If, $\sigma_g/\sigma_f = 1.01$ in dimension $d = 1000$, then acceptance rate is $1/20000$.

BIP: Numerical Solution by MCMC and RMAP

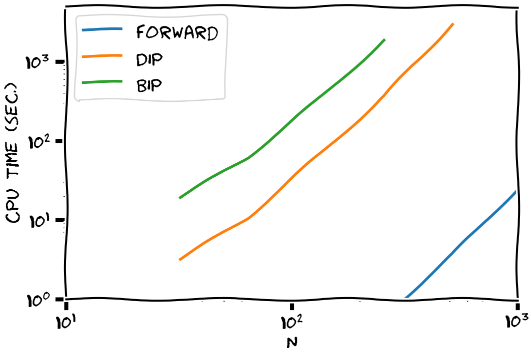
N	#dof	Time (sec.)
32	1024	19.3
64	4096	63.1
128	16384	317.0
256	65536	1865.8

Table: CPU times for MAP solution of Bayesian inverse problem

Method used here [Ghattas, et al] is state-of-the-art, scalable, adjoint-based algorithm, based on a stochastic Newton MCMC method with MAP-based Hessian.

BIP: Computational Scaling

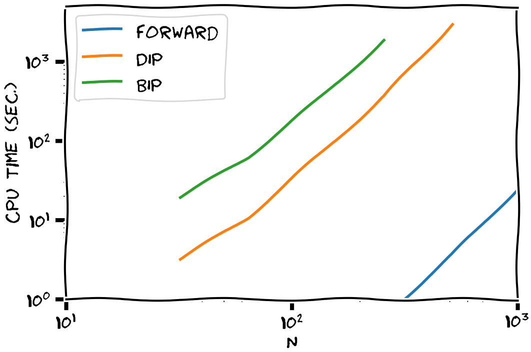
Comparison of CPU times for forward and inverse problems.



Remark: Cost of BIP is 5-50X DIP, which is 2 500-25 000X the direct problem (depending on the required posterior statistics).

BIP: Computational Scaling

Comparison of CPU times for forward and inverse problems.



Remark: Cost of BIP is 5-50X DIP, which is 2 500-25 000X the direct problem (depending on the required posterior statistics).

Plan

Some Open Research Questions

- How to take into account model error (insufficient numerical precision, incorrect assumptions of the physical model, simplifications in the simulator) ?
- How to deal with inconsistent priors: the major challenge is the specification of the prior and evaluating its consistency with the observed data?
- How to efficiently perform high-dimensional regression?

Some Open Research Questions

- How to take into account model error (insufficient numerical precision, incorrect assumptions of the physical model, simplifications in the simulator) ?
- How to deal with inconsistent priors: the major challenge is the specification of the prior and evaluating its consistency with the observed data?
- How to efficiently perform high-dimensional regression?

Some Open Research Questions

- How to take into account model error (insufficient numerical precision, incorrect assumptions of the physical model, simplifications in the simulator) ?
- How to deal with inconsistent priors: the major challenge is the specification of the prior and evaluating its consistency with the observed data?
- How to efficiently perform high-dimensional regression?

Thoughts and Conclusions

- Inverse problems are tough.
- UQ is even tougher.
- Tough problems require creative approaches and solutions.
- Data is everywhere, but needs to be correctly "constrained" by the models.
- Intelligent "blending" of model- and data-space inversion techniques can provide excellent results.
- But, when we have lots of parameters thanks to sophisticated instrumentation (there are many examples of this), a causality approach would seem to be indicated.

Thoughts and Conclusions

- Inverse problems are tough.
- UQ is even tougher.
- Tough problems require creative approaches and solutions.
- Data is everywhere, but needs to be correctly "constrained" by the models.
- Intelligent "blending" of model- and data-space inversion techniques can provide excellent results.
- But, when we have lots of parameters thanks to sophisticated instrumentation (there are many examples of this), a causality approach would seem to be indicated.

Thoughts and Conclusions

- Inverse problems are tough.
- UQ is even tougher.
- Tough problems require creative approaches and solutions.
- Data is everywhere, but needs to be correctly "constrained" by the models.
- Intelligent "blending" of model- and data-space inversion techniques can provide excellent results.
- But, when we have lots of parameters thanks to sophisticated instrumentation (there are many examples of this), a causality approach would seem to be indicated.

Thoughts and Conclusions

- Inverse problems are tough.
- UQ is even tougher.
- Tough problems require creative approaches and solutions.
- Data is everywhere, but needs to be correctly “constrained” by the models.
- Intelligent “blending” of model- and data-space inversion techniques can provide excellent results.
- But, when we have lots of parameters thanks to sophisticated instrumentation (there are many examples of this), a causality approach would seem to be indicated.

Thoughts and Conclusions

- Inverse problems are tough.
- UQ is even tougher.
- Tough problems require creative approaches and solutions.
- Data is everywhere, but needs to be correctly "constrained" by the models.
- Intelligent "blending" of model- and data-space inversion techniques can provide excellent results.
- But, when we have lots of parameters thanks to sophisticated instrumentation (there are many examples of this), a causality approach would seem to be indicated.

Thoughts and Conclusions

- Inverse problems are tough.
- UQ is even tougher.
- Tough problems require creative approaches and solutions.
- Data is everywhere, but needs to be correctly "constrained" by the models.
- Intelligent "blending" of model- and data-space inversion techniques can provide excellent results.
- But, when we have lots of parameters thanks to sophisticated instrumentation (there are many examples of this), a causality approach would seem to be indicated.

Plan

References

- M. Asch, M. Bocquet, M. Nodet.
  Data Assimilation - Methods, Algorithms, Applications.
  SIAM, 2017.

- Caers Group, Dorlofsky Group, Tchelepi Group.
  https://profiles.stanford.edu/jef-caers,
  https://profiles.stanford.edu/louis-durlofsky,
  https://profiles.stanford.edu/hamdi-tchelepi

- Ghattas Group.
  http://users.ices.utexas.edu/~omar/

- A. Stuart (2010). Inverse problems: A Bayesian perspective. *Acta Numerica*, **19**, 451-559. doi:10.1017/S0962492910000061.

Backup: gPC & Stochastic Galerkin

The implementation of gPC involves the following seven steps.

- (I) Identify the sources of uncertainty in the model in question.
- (II) Choose independent random variables with appropriate PDFs to represent these sources of uncertainty
- (III) Construct a generalised Polynomial Basis.
- (IV) Use this basis to construct a gPC expansion of the sources of uncertainty - initial conditions, parameters etc.
- (V) Substitute the gPC expansions into the governing equations.
- (VI) Perform a Galerkin projection to transform the stochastic equation into a set of coupled deterministic equations.
- (VII) Solve the resulting system of equations with appropriate numerical methods.

Backup: Functional Data Analysis

- point 1
- point 2