Inverse problems: Integrating Data with PDE-based Models Under Uncertainty

Noémi Petra[1]

[1] Department of Applied Mathematics
University of California, Merced

Power of Diversity in Uncertainty Quantification
SISSA (International School for Advanced Studies), Italy
February 26, 2024

Acknowledgements

- The **US National Science Foundation (NSF)**, Office of Advanced Cyberinfrastructure (OAC), for supporting the development and integration of the MUQ/hIPPYlib software and tutorial materials through grants ACI-1550487, ACI-1550547, and ACI-1550593. *(PIs: Omar Ghattas (UT Austin), Youssef Marzouk (MIT), Matthew Parno (Dartmouth), Noemi Petra (UC Merced), Umberto Villa (UT Austin))*
- The 2018 **Gene Golub SIAM Summer School** on Inverse Problems (www.g2s3.com): "Systematic Integration of Data with Models under Uncertainty" Organizers: Omar Ghattas (UT Austin), Youssef Marzouk (MIT), Matthew Parno (Dartmouth), Noemi Petra (UC Merced), Georg Stadler (NYU), Umberto Villa (UT Austin). This talk is based on material developed for the g2s3 summer school.
- The US National Science Foundation (NSF), Division of Mathematical Sciences (DMS), **CAREER award** #1654311

Outline

Motivation

- Recent years have seen tremendous growth in the volumes of **observational and experimental data** that are being collected, stored, processed, and analyzed.
- The central question that has emerged in the rapidly-growing field of **big data science** is:

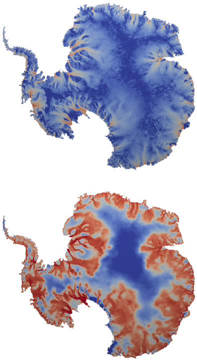
  How do we extract knowledge and insight from all of this data?

- When the data correspond to observations of (natural or engineered) systems, and these systems can be represented by mathematical models this **knowledge-from-data problem** is fundamentally a **mathematical inverse problem**.
- This **inverse problem** can we formulated as:

  Given (possibly noisy) data and (a possibly uncertain) model, infer parameters that characterize the model.

Examples of inverse problems governed by PDEs

- Inverse problems abound in all areas of science, engineering, technology, and medicine.
- For example we may infer:
  - ice sheet basal friction from satellite observations of surface flow
  - inertia of the generators for power grid
  - earth structure from reflected seismic waves
  - 3D bone structure from X-ray CT measurements
  - subsurface contaminant plume spread from crosswell electromagnetic measurements
  - ocean state from surface temperature observations.

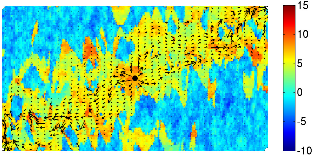


In general, any endeavor to infer cause from effect **to extract knowledge from data** can be viewed as an inverse problem.

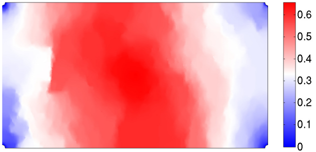
Example I: Coefficient inversion in an elliptic PDE

Model for single phase flow, ground water filtration

$$-\nabla \cdot (m \nabla u) = f \quad \text{in } \Omega \subset \mathbb{R}^d$$
$$+ \text{ boundary conditions}$$



- f: given source function
- m: permeability, hydraulic permittivity
- u: unknown pressure
- $d = 1, 2,$ or 3.

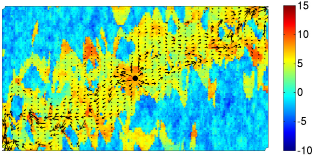


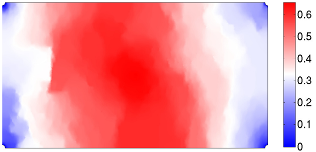
- forward problem: the parameter-to-solution mapping $m \to u$ (i.e., solve the PDE)
- data/measurements: pressure u at points/parts of Ω
- parameter field/image: $m = m(x)$
- inverse problem: given (measurements of) u, find m

Example I: Coefficient inversion in an elliptic PDE

Model for single phase flow, ground water filtration

$$-\nabla \cdot (m \nabla u) = f \quad \text{in } \Omega \subset \mathbb{R}^d$$
$$+ \text{ boundary conditions}$$





- f: given source function
- m: permeability, hydraulic permittivity
- u: unknown pressure
- $d = 1, 2,$ or 3.

Details in:

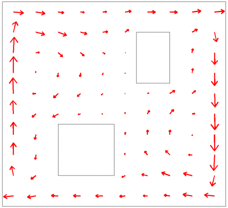
N. Petra and G. Stadler. "Model Variational Inverse Problems Governed by Partial Differential Equations", ICES REPORT 11-05, 2011.

A. Alexanderian, N. Petra, G. Stadler, O. Ghattas. A Fast and Scalable Method for A-Optimal Design of Experiments for Infinite-dimensional Bayesian Nonlinear Inverse Problems, SISC, 38(1), 2016.

Example II: Inversion for the initial condition

in an advection-diffusion equation

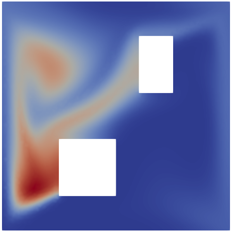
$$
\begin{aligned}
u_t - \kappa \Delta u + \mathbf{v} \cdot \nabla u &= 0 \quad \text{in } \Omega \times [0, T], \\
\kappa \nabla u \cdot \mathbf{n} &= 0 \quad \text{on } \partial\Omega \times [0, T], \\
u(0, x) &= u_0 \quad \text{in } \Omega.
\end{aligned}
$$



$\mathbf{v}$: given velocity field

u : concentration field

u_0 : unknown initial condition



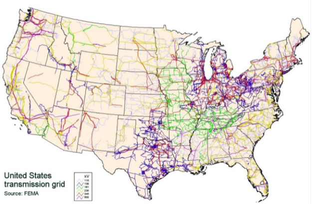
- data/measurements: concentration u at points/parts of $\partial\Omega$
- parameter field: $u_0(x)$
- inverse problem: given (measurements of) u, find u_0

Details in: U. Villa, N. Petra, and O. Ghattas, *hIPPYlib: An Extensible Software Framework for Large-Scale Deterministic and Linearized Bayesian Inverse Problems*. Journal of Open Source Software, 30(3), 2018.

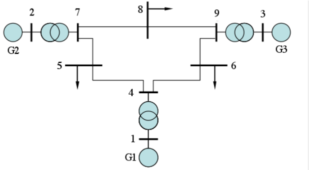
Example III: Parameter estimation for power grid

Differential-algebraic equation (DAE) driven inverse problem

$$\dot{x} = h(x, y, m, t)$$
$$0 = g(x, y, t)$$
$$x(0) = x_0$$



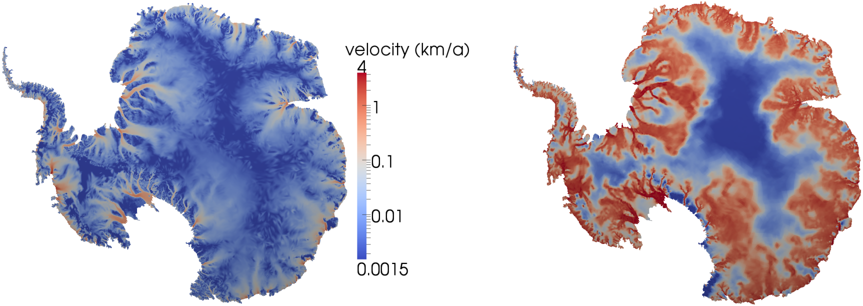
- Estimate/infer model parameters m (e.g., inertia of the generators)
- Observational data (e.g., bus voltages)



Goal: Enable real-time power grid operational tasks such as monitoring fault-detection, dynamic stability assessment, etc.

Details in: Petra, N.; Petra, C.; Zhang, Z.; Constantinescu, E. and Anitescu, M. A Bayesian Approach for Parameter Estimation with Uncertainty for Dynamic Power Systems. IEEE Transactions on Power Systems, 32(4), (2017).

Example IV: Inferring the basal boundary condition in ice dynamics



- data/measurements: surface ice flow velocity $\boldsymbol{u}$
- parameter field: basal sliding coefficient field $\beta(x)$
- inverse problem: given (measurements of) $\boldsymbol{u}$, find β

Details in: T. Isaac, N. Petra, G. Stadler, and O. Ghattas. "Scalable and efficient algorithms for the propagation of uncertainty from data through inference to prediction for large-scale problems, with application to flow of the Antarctic ice sheet", Journal of Computational Physics, 296, 348-368 (2015).

Outline

- Given some input parameter field (coefficient, right hand side, initial condition, boundary condition, etc.) m solve

$$r(u, m) = 0,$$

  for the *observable* u, where
  - $r : \mathcal{V} \times \mathcal{M} \to \mathcal{V}^*$ represents the PDE problem,
  - $\mathcal{V}$ and $\mathcal{M}$ are suitable spaces of functions.
- Tipically we assume that the foward problem is well-posed (i.e., it has a unique solution, is stable with respect to perturbation).

Anatomy of inverse problems governed by PDEs

The inverse problem

- Consists of using available observational data $\boldsymbol{d}$ to infer the values of the unknown parameter field m that characterize a physical process modeled by PDEs.
- Mathematically this inverse relationship is expressed as

$$\boldsymbol{d} = \mathcal{F}(m) + \boldsymbol{\eta}.$$

  - The map $\mathcal{F} : \mathcal{M} \to \mathbb{R}^q$ is the so-called *parameter-to-observable* map.
  - Evaluations of $\mathcal{F}$ involve the solution of a PDE given m, followed by the application of an observation operator $\mathcal{B} : \mathcal{V} \to \mathbb{R}^q$ to extract the observations from the state.
  - $\boldsymbol{\eta}$ accounts for noisy measurements and model errors and is modeled as $\boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Gamma}_{\text{noise}})$, i.e., a centered Gaussian at $\mathbf{0}$ with covariance $\boldsymbol{\Gamma}_{\text{noise}}$.

Anatomy of inverse problems governed by PDEs

Deterministic inversion

- Ill-posed: observations are usually sparse and the forward operator smoothing; many different parameter values may be consistent with the observations/data.
- Occam's approach to ill-posedness: employ regularization to penalize unwanted solution features to guarantee unique solution:

$$\min_{m \in \mathcal{M}} \mathcal{J}(m) := \frac{1}{2} \|\mathcal{F}(m) - \boldsymbol{d}\|_W^2 + \frac{\alpha}{2} \|m - m_{\text{ref}}\|_R^2$$

  - The first term in the cost functional, $\mathcal{J}(m)$, represents the misfit between the observations, $\boldsymbol{d}$, and that predicted by the parameter-to-observable map $\mathcal{F}(m)$ (weighted by W).
  - The second term is a regularization term which imposes regularity on the inversion field m, such as smoothness; α is a regularization parameter.
  - For instance, one can use Tikhonov regularization, which penalizes oscillatory components of the parameter m.

Anatomy of inverse problems governed by PDEs

Deterministic inversion

- Ill-posed: observations are usually sparse and the forward operator smoothing; many different parameter values may be consistent with the observations/data.
- Occam's approach to ill-posedness: employ regularization to penalize unwanted solution features to guarantee unique solution:

$$\min_{m \in \mathcal{M}} \mathcal{J}(m) := \frac{1}{2} \|\mathcal{F}(m) - \boldsymbol{d}\|_W^2 + \frac{\alpha}{2} \|m - m_{\text{ref}}\|_R^2$$

Remarks:

- deterministic problem
- large-scale (PDE-constrained) optimization
- no quantification of uncertainty
- influence of $R(\cdot)$ on m is unclear

Anatomy of inverse problems governed by PDEs

Deterministic inversion

- Ill-posed: observations are usually sparse and the forward operator smoothing; many different parameter values may be consistent with the observations/data.
- Bayesian approach to ill-posedness: describe probability of all models that are consistent with the observations/data and any prior knowledge about the parameters:

$$d\mu_{\text{post}} \propto \exp\left\{ -\frac{1}{2}\|\mathcal{F}(m) - \boldsymbol{d}\|^2_{\boldsymbol{\Gamma}^{-1}_{\text{noise}}} - \frac{1}{2}\|m - m_{\text{pr}}\|^2_{\mathcal{C}^{-1}_{\text{prior}}}\right\}.$$

  - The first term in the exponential ($\pi_{\text{like}}(\boldsymbol{d}|m)$) is the negative log-likelihood (the likelihood represents the probability that a given set of parameters might give rise to the observed data)
  - The second term represent the negative log-prior, e.g., a Gaussian prior, i.e., $m \sim \mathcal{N}(m_{\text{pr}}, \mathcal{C}_{\text{prior}})$.

Anatomy of inverse problems governed by PDEs

Bayesian inversion

- Describe probability of all models that are consistent with the observations/data and any prior knowledge about the parameters:

$$d\mu_{\text{post}} \propto \exp\left\{ -\frac{1}{2}\|\mathcal{F}(m) - \boldsymbol{d}\|^2_{\boldsymbol{\Gamma}^{-1}_{\text{noise}}} - \frac{1}{2}\|m - m_{\text{pr}}\|^2_{\mathcal{C}^{-1}_{\text{prior}}} \right\}.$$

  - The first term in the exponential $(\pi_{\text{like}}(\boldsymbol{d}|m))$ is the negative log-likelihood.
  - The second term represent the negative log-prior, e.g., a Gaussian prior, i.e., $m \sim \mathcal{N}(m_{\text{pr}}, \mathcal{C}_{\text{prior}})$.

Remarks:

- systematic method to incorporate measurement and model errors and prior knowledge
- allows quantification of uncertainty and probabilistic predictions
- related to regularization approach
- high-dimensional probability density

Anatomy of inverse problems governed by PDEs

Bayesian inversion

- Describe probability of all models that are consistent with the observations/data and any prior knowledge about the parameters:

$$d\mu_{\text{post}} \propto \exp\left\{ -\frac{1}{2}\|\mathcal{F}(m) - \boldsymbol{d}\|^2_{\boldsymbol{\Gamma}^{-1}_{\text{noise}}} - \frac{1}{2}\|m - m_{\text{pr}}\|^2_{\mathcal{C}^{-1}_{\text{prior}}} \right\}.$$

  - The first term in the exponential ($\pi_{\text{like}}(\boldsymbol{d}|m)$) is the negative log-likelihood.
  - The second term represent the negative log-prior, e.g., a Gaussian prior, i.e., $m \sim \mathcal{N}(m_{\text{pr}}, \mathcal{C}_{\text{prior}})$.

Target:

- characterize $d\mu_{\text{post}}$
- for functions m (large vectors after discretization)
- for expensive $\mathcal{F}$
- use connection to optimization

- The *maximum a posteriori* (MAP) point m_{MAP} is defined as the parameter field that maximizes the posterior distribution:

$$m_{\text{MAP}} := \underset{m \in \mathcal{M}}{\operatorname{argmin}} (-\log d\mu_{\text{post}}(m)) = \underset{m \in \mathcal{M}}{\operatorname{argmin}} \frac{1}{2} \|\mathcal{F}(m) - \boldsymbol{d}\|^2_{\boldsymbol{\Gamma}^{-1}_{\text{noise}}} + \frac{1}{2} \|m - m_{\text{pr}}\|^2_{\mathcal{C}^{-1}_{\text{prior}}}.$$

- When $\mathcal{F}$ is linear, due to the particular choice of prior and noise model, the posterior measure is Gaussian, $\mathcal{N}(m_{\text{MAP}}, \mathcal{C}_{\text{post}})$

$$\mathcal{C}_{\text{post}} = \mathcal{H}^{-1} = (\mathcal{F}^* \boldsymbol{\Gamma}^{-1}_{\text{noise}} \mathcal{F} + \mathcal{C}^{-1}_{\text{prior}})^{-1}, \qquad m_{\text{MAP}} = \mathcal{C}_{\text{post}} (\mathcal{F}^* \boldsymbol{\Gamma}^{-1}_{\text{noise}} \boldsymbol{d} + \mathcal{C}^{-1}_{\text{prior}} m_{\text{pr}}),$$

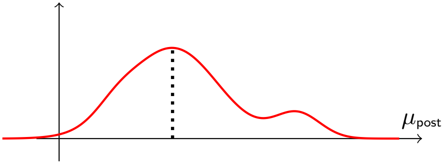
  where $\mathcal{F}^* : \mathbb{R}^q \to \mathcal{M}$ is the adjoint of $\mathcal{F}$.

- In the general case of nonlinear parameter-to-observable map $\mathcal{F}$ the posterior distribution is not Gaussian. In this case one would need to use sampling to characterize the posterior.

Approximation of the posterior distribution

Despite the explicit form of the posterior, its exploration is difficult due to:

- the high/infinite dimension of m
- the expensive PDE-based parameter-to-observable map $\mathcal{F}$



Approximation I: MAP estimation

Finding the maximum a posteriori (MAP) point

- requires solution of PDE-constrained optimization problem $\sim$ deterministic inversion;
- computation of derivatives using adjoint methods: 2(+) PDE solves per gradient.

Outline

- Suppose we are interested in finding solutions $u^\star$ to the minimization problem:

$$\min_{u \in H_0^1(\Omega)} \mathcal{J}(u)$$

where

$$\mathcal{J}(u) := \frac{1}{2} \int_\Omega k(u) \nabla u \cdot \nabla u \, dx - \int_\Omega f u \, dx$$

  - $u(\boldsymbol{x}) \in H_0^1(\Omega)$: e.g., the transverse deflection of a membrane at a point $\boldsymbol{x} \in \Omega$
  - $H_0^1(\Omega)$: is the Sobolev space of functions in $L^2(\mathcal{D})$ with square integrable derivatives.
  - $f \in L^2(\Omega)$: given (input) function
  - $k(u) := k_1 + k_2 u^2$, $k_1 > 0$ and $k_2 \geq 0$: e.g., the stiffness of the membrane with square integrable derivatives that satisfy homogenous boundary conditions
  - $\Omega \subset \mathbb{R}^2$: an open bounded domain with sufficiently smooth boundary Γ
  - $u(\boldsymbol{x}) \in \mathcal{V}_0$: e.g., the transverse deflection of a membrane at a point $\boldsymbol{x} \in \Omega$

Calculus of variations and weak forms

Energy minimization problem

- A minimum $u^\star$ is characterized by

$$\mathcal{J}(u^\star + \varepsilon\tilde{u}) \geq \mathcal{J}(u^\star), \forall \tilde{u} \in \mathcal{V}_0, \text{ and } \varepsilon > 0 \text{ with } u^\star + \varepsilon\tilde{u} \in \mathcal{V}_0.$$

- Thus, a minimum $u^\star$ must satisfy the *Euler-Lagrange* conditions for stationarity, namely

$$J_u(u^*, \tilde{u}) := \left.\frac{d\mathcal{J}(u^\star + \varepsilon\tilde{u})}{d\varepsilon}\right|_{\varepsilon=0} = 0 \text{ for all } \tilde{u} \in H_0^1(\Omega).$$

- Details in:
  - Jorge Nocedal and Stephen J. Wright, "Numerical Optimization", Springer-Verlag, 1999
  - I. M. Gelfand and S. V. Fomin, "Calculus of variations", Dover, 2000
  - Peter J. Olver, "The Calculus of Variations", Lecture notes, 2021

Calculus of variations and weak forms

Energy minimization problem

- The optimization problem (assume $k_2 = 0$ for now):

$$\min_{u \in H_0^1(\Omega)} \mathcal{J}(u) := \frac{1}{2} \int_\Omega k_1 \nabla u \cdot \nabla u \, dx - \int_\Omega f u \, dx$$

- The Euler-Lagrange equation:

$$J_u(u^*)(\tilde{u}) := \left. \frac{d\mathcal{J}(u^\star + \varepsilon \tilde{u})}{d\varepsilon} \right|_{\varepsilon=0} = 0 \text{ for all } \tilde{u} \in H_0^1(\Omega)$$

- Differentiating with respect to ε we obtain

$$\frac{d\mathcal{J}(u^\star + \varepsilon \tilde{u})}{d\varepsilon} = \int_\Omega k_1 \nabla (u^\star + \varepsilon \tilde{u}) \cdot \nabla \tilde{u} \, dx - \int_\Omega f \tilde{u} \, dx$$

- Setting $\varepsilon = 0$, we obtain the *weak* (or *variational*) form:

$$\text{Find } u^\star \in H_0^1(\Omega) \text{ s.t. } \int_\Omega k_1 \nabla u^\star \cdot \nabla \tilde{u} \, dx = \int_\Omega f \tilde{u} \, dx, \text{ for all } \tilde{u} \in H_0^1(\Omega).$$

Calculus of variations and weak forms

Linear elliptic model problem: weak-to-strong form

- Use Green's identity, which is a multidimensional version of integration-by-parts. The identity states that for all $u, v \in H_0^1(\Omega)$ holds

$$\int_\Omega k_1 \nabla u \cdot \nabla v \, dx = -\int_\Omega u \nabla \cdot k_1 \nabla v) \, dx + \int_{\partial\Omega} (k_1 \nabla v \cdot \boldsymbol{n}) u \, ds$$

- Using this identity for the first term in the weak form we obtain

$$\begin{aligned} 0 &= -\int_\Omega \nabla \cdot (k_1 \nabla u^\star) \tilde{u} \, dx - \int_\Omega f \tilde{u} \, dx \\ &= \int_\Omega - \left[f + \nabla \cdot (k_1 \nabla u^\star) \right] \tilde{u} \, dx, \text{ for all } \tilde{u} \in H_0^1(\Omega). \end{aligned}$$

- Since $\tilde{u}$ is arbitrary, this implies that the factors multiplying $\tilde{u}$ must vanish. (This is a very common argument in variational calculus.)

- Since $\tilde{u}$ is arbitrary in Ω

$$-\nabla \cdot (k_1 \nabla u^\star) = f \text{ on } \Omega.$$

- Since $u^\star$ is in $H_0^1(\Omega)$, it satisfies the Dirichlet boundary condition

$$u^\star = 0 \text{ on } \partial\Omega.$$

- These equations are the *strong form* for the variational problem.

- Note: the optimality conditions for this problem are nonlinear when $k_2 > 0$, hence in general an iterative "outer' optimization procedure (e.g., gradient descent or Newton's method) must be applied to determine the minimizer u^*.

- The *second variation*:

$$J_{uu}(u^\star)(\tilde{u}, \hat{u}) := \left. \frac{d\mathcal{J}_u(u^\star + \varepsilon\hat{u})(\tilde{u})}{d\varepsilon} \right|_{\varepsilon=0}$$

- Newton's method at iteration k:

$$\begin{aligned} J_{uu}(u_k)(\tilde{u}, \hat{u}) &= -J_u(u_k)(\tilde{u}) \\ u_{k+1} &= u_k + \alpha\hat{u}, \end{aligned}$$

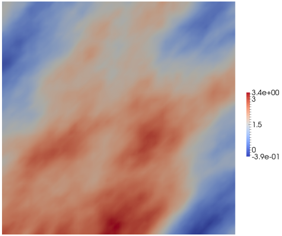
  where α is the step size (found via backtracking line search).

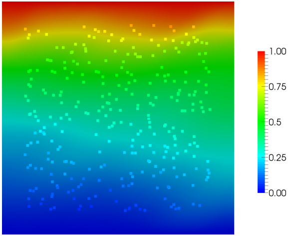
Coefficient field inversion in an elliptic PDE

The elliptic equation in strong form

$$\begin{aligned} -\nabla \cdot (e^m \nabla u) &= f \quad \text{in } \mathcal{D} \\ u &= g \quad \text{on } \Gamma_D \\ e^m \nabla u \cdot \boldsymbol{n} &= h \quad \text{on } \Gamma_N \end{aligned}$$

- u is the state variable (e.g., pressure field)
- $m \in \mathcal{M}$ is the inversion parameter (e.g., the log permeability field)
- $f \in L^2(\mathcal{D})$ is a source term
- $g \in H^{1/2}(\Gamma_D)$ and $h \in L^2(\Gamma_N)$ are Dirichlet and Neumann boundary data
- $\mathcal{D} \subset \mathbb{R}^2$ is an open bounded domain with sufficiently smooth boundary $\Gamma = \Gamma_D \cup \Gamma_N$, $\Gamma_D \cap \Gamma_N = \emptyset$





- Define the spaces

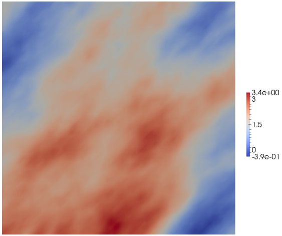
$$\mathcal{V}_g = \{v \in H^1(\mathcal{D}) : v\big|_{\Gamma_D} = g\}$$
$$\mathcal{V}_0 = \{v \in H^1(\mathcal{D}) : v\big|_{\Gamma_D} = 0\},$$

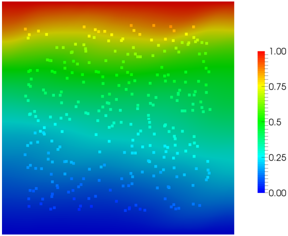
where $H^1(\mathcal{D})$ is the Sobolev space of functions in $L^2(\mathcal{D})$ with square integrable derivatives.

- The weak form: find $u \in \mathcal{V}_g$ such that

$$\langle e^m \nabla u, \nabla p\rangle = \langle f, p\rangle + \langle h, p\rangle_{\Gamma_N}, \quad \forall p \in \mathcal{V}_0,$$

where $\langle \cdot, \cdot \rangle$ and $\langle \cdot, \cdot \rangle_{\Gamma_N}$ denote the standard inner products in $L^2(\mathcal{D})$ and $L^2(\Gamma_N)$, respectively.





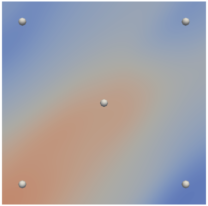
Coefficient field inversion in an elliptic PDE

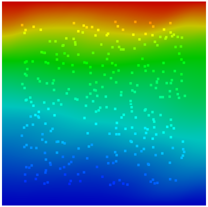
The inverse problem governed by an elliptic PDE

$$\min_{m} \mathcal{J}(m) := \frac{1}{2} \|\mathcal{B}u(m) - \boldsymbol{d}\|^2_{\boldsymbol{\Gamma}^{-1}_{\text{noise}}} + \frac{1}{2} \|m - m_{\text{pr}}\|^2_{\mathcal{C}^{-1}_{\text{prior}}}$$

where u solves the elliptic PDE for given parameter field m.

- $\mathcal{B}$ is a linear observation operator that extracts measurements from u
- m_{pr} is the mean of the log coefficient field
- $\boldsymbol{d} \in \mathbb{R}^q$ is a given data vector
- $\boldsymbol{\Gamma}^{-1}_{\text{noise}}$ and $\mathcal{C}^{-1}_{\text{prior}}$ appropriately chosen weights (to account for noise in measurements and properties of the parameter)
- We solve this optimization problem via Newton's method, therefore need to derive gradient and Hessian of $\mathcal{J}(m)$ with respect to the parameter m.





Gradient computation

using standard variational approach

- The Lagrangian functional:

$$\mathcal{L}(u,m,p) := \mathcal{J}(m) + \langle e^m \nabla u, \nabla p\rangle - \langle f, p\rangle - \langle p, h\rangle_{\Gamma_N},$$

  where $p \in \mathcal{V}_0$ is the Lagrange multiplier.

- Optimality system:

$$\begin{aligned}
\mathcal{L}_p(\tilde{p}) &= \langle e^m \nabla u, \nabla \tilde{p}\rangle - \langle f, \tilde{p}\rangle - \langle \tilde{p}, h\rangle_{\Gamma_N} &= 0 \quad \text{(state)}\\
\mathcal{L}_u(\tilde{u}) &= \langle e^m \nabla \tilde{u}, \nabla p\rangle + \left\langle \mathcal{B}^* \mathbf{\Gamma}_{\text{noise}}^{-1}(\mathcal{B}u - \boldsymbol{d}), \tilde{u}\right\rangle &= 0 \quad \text{(adjoint)}\\
\mathcal{L}_m(\tilde{m}) &= \langle m - m_{\text{pr}}, \tilde{m}\rangle_{\mathcal{C}_{\text{prior}}^{-1}} + \langle \tilde{m} e^m \nabla u, \nabla p\rangle &= 0 \quad \text{(gradient)}
\end{aligned}$$

  for all variations $(\tilde{u}, \tilde{p}, \tilde{m}) \in \mathcal{V}_0 \times \mathcal{V}_0 \times \mathcal{M}$.

- The gradient (also in weak form):

$$(J_u(m)(\tilde{m}) :=)\ \mathcal{G}(m)(\tilde{m}) = \langle m - m_{\text{pr}}, \tilde{m}\rangle_{\mathcal{C}_{\text{prior}}^{-1}} + \langle \tilde{m} e^m \nabla u, \nabla p\rangle,$$

  where u and p are solutions to the state and adjoint equations, respectively.

The action of the Hessian in a direction

- The *meta*-Lagrangian functional

$$\begin{aligned}
\mathcal{L}^{\mathcal{G}}(u, m, p; \hat{u}, \hat{m}, \hat{p}) := {} & \mathcal{G}(m)(\hat{m}) && \text{(gradient)} \\
& + \langle e^m \nabla u, \nabla \hat{p} \rangle - \langle f, \hat{p} \rangle - \langle \hat{p}, h \rangle_{\Gamma_N} && \text{(state)} \\
& + \langle e^m \nabla \hat{u}, \nabla p \rangle + \left\langle \mathcal{B}^* \mathbf{\Gamma}_{\text{noise}}^{-1} (\mathcal{B}u - \boldsymbol{d}), \hat{u} \right\rangle && \text{(adjoint)}
\end{aligned}$$

- The Hessian in a direction $\hat{m}$ as the variation of $\mathcal{L}^{\mathcal{G}}$ with respect to m:

$$\begin{aligned}
(J_{uu}(m)(\tilde{m}, \hat{m}) :=)\ \mathcal{H}(m)(\tilde{m}, \hat{m}) = {} & \langle \tilde{m} e^m \nabla \hat{u}, \nabla p \rangle + \langle \hat{m}, \tilde{m} \rangle_{\mathcal{C}_{\text{prior}}^{-1}} \\
& + \langle \tilde{m} \hat{m} e^m \nabla u, \nabla p \rangle + \langle \tilde{m} e^m \nabla u, \nabla \hat{p} \rangle
\end{aligned}$$

- u and p are the solutions of the state and adjoint equations
- $\hat{u}$, $\hat{p}$ are the *incremental state and adjoint* variables

The action of the Hessian in a direction

The incremental ("second order") state and adjoint

- The Hessian in a direction $\hat{m}$ as the variation of $\mathcal{L}^{\mathcal{G}}$ with respect to m:

$$\mathcal{H}(m)(\tilde{m},\hat{m}) = \langle \tilde{m} e^m \nabla \hat{u}, \nabla p\rangle + \langle \hat{m}, \tilde{m}\rangle_{\mathcal{C}^{-1}_{\text{prior}}} + \langle \tilde{m}\hat{m} e^m \nabla u, \nabla p\rangle + \langle \tilde{m} e^m \nabla u, \nabla \hat{p}\rangle$$

- u and p are the solutions of the state and adjoint equations
- $\hat{u}$, $\hat{p}$ are the *incremental state and adjoint* variables, and are obtained by taking variations of $\mathcal{L}^{\mathcal{G}}$ with respect to p and u:

$$\langle e^m \nabla \hat{u}, \nabla \tilde{p}\rangle + \langle \hat{m} e^m \nabla u, \nabla \tilde{p}\rangle = 0,\ \forall \tilde{p} \in \mathcal{V}_0 \quad \text{(inc. state)}$$

$$\left\langle \mathcal{B}^* \mathbf{\Gamma}^{-1}_{\text{noise}} \mathcal{B}\hat{u}, \tilde{u}\right\rangle + \langle \hat{m} e^m \nabla \tilde{u}, \nabla p\rangle + \langle e^m \nabla \tilde{u}, \nabla \hat{p}\rangle = 0,\ \forall \tilde{u} \in \mathcal{V}_0 \quad \text{(inc. adjoint).}$$

The inexact Newton-conjugate gradient algorithm

To compute the maximum a posteriori (MAP) point

Algorithm 1 The inexact Newton-CG algorithm

$i \leftarrow 0$
Given $\boldsymbol{m}_0$ solve the forward/state problem to obtain $\boldsymbol{u}_0$
Given $\boldsymbol{m}_0$ and $\boldsymbol{u}_0$ compute the cost functional $\mathcal{J}_0$
while $i <$ max_iter **do**
  Given $\boldsymbol{m}_i$ and $\boldsymbol{u}_i$ solve the adjoint problem to obtain $\boldsymbol{p}_i$
  Given $\boldsymbol{m}_i$, $\boldsymbol{u}_i$ and $\boldsymbol{p}_i$ evaluate the gradient $\boldsymbol{g}_i$
  if $\|\boldsymbol{g}_i\| \leq \tau$ **then**
    break
  end if
  Given $\boldsymbol{m}_i$, $\boldsymbol{u}_i$ and $\boldsymbol{p}_i$ define a linear operator $\mathbf{H}_i$ that implements the Hessian-action
  Using conjugate gradients, find a search direction $\widehat{\boldsymbol{m}}_i$ such that

$$\|\mathbf{H}_i \widehat{\boldsymbol{m}}_i + \boldsymbol{g}_i\| \leq \eta_i \|\boldsymbol{g}_i\|, \text{ with } \eta_i = \left(\frac{\|\boldsymbol{g}_i\|}{\|\boldsymbol{g}_0\|} \right)^{\frac{1}{2}}$$

  $j \leftarrow 0$, $\alpha^{(0)} \leftarrow 1$
  while $j <$ max_backtracking_iter **do**
    Set $\boldsymbol{m}^{(j)} = \boldsymbol{m}_i + \alpha^{(j)} \widehat{\boldsymbol{m}}_i$
    Given $\boldsymbol{m}^{(j)}$ solve the forward problem to obtain $\boldsymbol{u}^{(j)}$
    Given $\boldsymbol{m}^{(j)}$ and $\boldsymbol{u}^{(j)}$ compute the cost $\mathcal{J}^{(j)}$
    if $\mathcal{J}^{(j)} < \mathcal{J}_i + \alpha^{(j)} c_{\text{armijo}} \boldsymbol{g}_i^T \widehat{\boldsymbol{m}}_i$ **then**
      $\boldsymbol{m}_{i+1} \leftarrow \boldsymbol{m}^{(j)}$, $\mathcal{J}_{i+1} \leftarrow \mathcal{J}^{(j)}$
      break
    end if
    $\alpha^{(j+1)} \leftarrow \alpha^{(j)}/2, \quad j \leftarrow j+1$
  end while
  $i \leftarrow i+1$
end while

Outline

hIPPYlib: **I**nverse **P**roblem **PY**thon **lib**rary

http://hippylib.github.io

- Developed at UT Austin and UC Merced.
- Authors: Umberto Villa (UT Austin), Noemi Petra (UC Merced) and Omar Ghattas (UT Austin)
- Supported by NSF-SSI2: *Integrating Data with Complex Predictive Models under Uncertainty: An Extensible Software Framework for Large-Scale Bayesian Inversion*
  - Joint with: Omar Ghattas (UT Austin), Umberto Villa (UT Austin), Youssef Marzouk (MIT) and Matthew Parno (US Army Cold Regions Research and Engineering Laboratory (CRREL))
- Implements state-of-the-art scalable adjoint-based algorithms for PDE-based deterministic and (linear) Bayesian inverse problems.
- Builds on FEniCS for the discretization of the underlying PDEs and on PETSc for scalable and efficient linear algebra operations and solvers.

hIPPYlib: **I**nverse **P**roblem **PY**thon **lib**rary

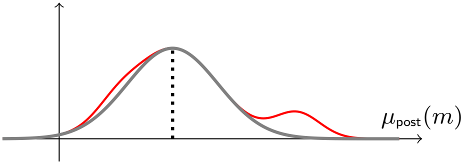
- Features:
  - Friendly, compact, near-mathematical FEniCS notation to express the PDE and likelihood in weak form
  - Automatic generation of efficient code for the discretization of weak forms using FEniCS
  - Symbolic differentiation of weak forms to generate derivatives and adjoint information
  - Globalized inexact Newton-CG method to solve the inverse problem
  - Randomized algorithm for the generalized eigenvalue problem
  - Low rank representation of the posterior covariance using randomized algorithms
  - Sampling from the prior and from the Gaussian approximation of the posterior
  - Extract pointwise variance of prior and posterior
- Release:
  - `hIPPYlib` 3.1 is now available at: http://hippylib.github.io

Details in: U. Villa, N. Petra, and O. Ghattas, *hIPPYlib: An Extensible Software Framework for Large-Scale Inverse Problems Governed by PDEs: Part I: Deterministic Inversion and Linearized Bayesian Inferences*. ACM Transactions on Mathematical Software (TOMS), 47(2), 2018.

Approximation of the posterior distribution

Despite the explicit form of $\mu_{\text{post}}(m)$, its exploration is difficult due to:

- the high/infinite dimension of m
- the expensive PDE-based parameter-to-observable map $\mathcal{F}$



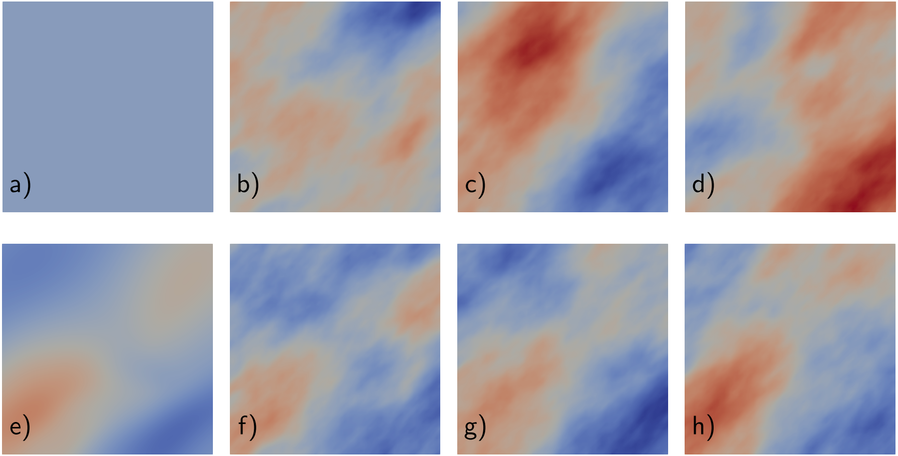
Approximation II: Gaussian around MAP point

Use a Gaussian approximation around the MAP point based on second derivatives (Hessians) of $\mathcal{J}$.

- This requires the Hessian matrix which is not directly available for PDE-constrained problems: but the Hessian can be applied to vectors by solving 2(+) PDEs.

Example 2: Coefficient field inversion in an elliptic PDE

Uncertainty quantification via Gaussian/Laplace approximation of the posterior

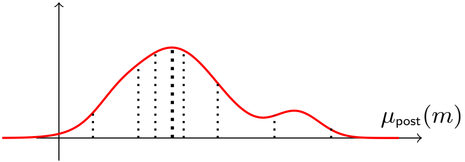


Top: Prior mean m_{pr} (a), and samples drawn from the prior distribution (b)–(d). Bottom: The MAP point (e) and samples drawn from the Laplace approximation of posterior distribution (f)–(h).

Approximation of the posterior distribution

Despite the explicit form of $\mu_{\text{post}}(m)$, its exploration is difficult due to:

- the high/infinite dimension of m
- the expensive PDE-based parameter-to-observable map $\mathcal{F}$



Approximation III: Sampling

Use sampling (Metropolis Hastings/Markov chain Monte Carlo) to approximate statistics.

- Sampling in high dimensions is challenging, requires many evaluations of $\boldsymbol{f}$ (and good proposal distributions).
- Exploit low rank properties of update of distribution; “feels” the curse of the effective dimensionality rather than the discretization dimension.

hIPPYlib-MUQ

- Features:
  - Friendly, compact, near-mathematical FEniCS notation to express the PDE and likelihood in weak form
  - Builds on the MUQ: MIT Uncertainty Quantification library. It allows:
    - exact sampling of non-Gaussian distributions (e.g., Markov chain Monte Carlo and importance sampling),
    - approximating computationally intensive forward models (e.g., polynomial chaos expansions and Gaussian process regression),
    - characterizing predictive uncertainties, etc.
    - **Details in:** M. Parno, A. Davis, and L. Seelinger, MUQ: MUQ: The MIT Uncertainty Quantification Library, Journal of Open Source Software, 6(68), 2021.
- Release:
  - hIPPYlib2MUQ is available on github: `https://github.com/hippylib/hippylib2muq.git`

Details in: K. Kim, U. Villa, M. Parno, Y. Marzouk, O. Ghattas, N. Petra, *hIPPYlib-MUQ: A Bayesian Inference Software Framework for Integration of Data with Complex Predictive Models under Uncertainty*. ACM Transactions on Mathematical Software (TOMS), 2023

Example: Coefficient field inversion in an elliptic PDE

Uncertainty quantification via MCMC sampling of the posterior

Method	**AR** (%)	**MPSRF**	**Min. ESS** (index)	**Max. ESS** (index)	**Avg. ESS**	**NPS/ES**
pCN (5.0E-3)	24	2.629	25 (24)	225 (8)	84	5,952
MALA (6.0E-6)	48	2.642	26 (22)	874 (5)	148	10,135
∞-**MALA** (1.0E-5)	57	2.943	25 (23)	1,102 (5)	160	9,375
H-pCN (4.0E-1)	27	1.192	64 (1)	3,598 (15)	2,314	216
H-MALA (6.0E-2)	60	1.014	545 (1)	8,868 (19)	6,459	232
H-∞**-MALA** (1.0E-1)	71	1.016	582 (1)	8,417 (18)	5,905	254
DR (H-pCN (1.0E0), **H-MALA** (6.0E-2))	(4, 61)	1.013	641 (1)	12,522 (17)	9,222	215
DR (H-pCN (1.0E0), **H-**∞**-MALA** (2.0E-1))	(4, 48)	1.011	613 (1)	12,812 (17)	9,141	213
DILI-PRIOR (0.8, 0.1)	(60, 33)	1.064	314 (1)	4,667 (13)	3,216	548
DILI-LA (0.8, 0.1)	(83, 36)	1.017	562 (1)	10,882 (17)	7,192	245
DILI-MAP (0.8, 0.1)	(77, 22)	1.006	1,675 (1)	10,271 (20)	8,692	202

MCMC convergence diagnostics

- Comparison of the performance of several MCMC methods: pCN, MALA, ∞-MALA, DR, DILI, and their Hessian-informed versions.
- **AR**: Acceptance rate, **MPSRF**: multivariate potential scale reduction factor, **ESS**: effective sample sample size
- Used 20 MCMC chains, 500,000 samples in total.

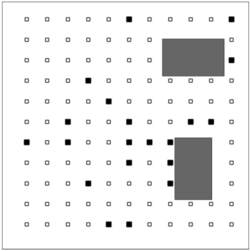
Outline

Infinite-dimensional parameters

- **Inverse problems governed by PDEs** are formulated as optimization problems **in function spaces** and the inversion parameters are typically functions.
  - **Large-scale inverse problems** governed by PDEs
  - **Optimize-then-discretize (OTD) versus discretize-then-optimize (DTO)**
  - Often **bound constraints** are needed, interior-point methods in infinite-dimensions
- **Fast and scalable optimization methods** require gradient and Hessian (of the cost functional) information
  - Variational calculus and **adjoint-based techniques** provide efficient means for deriving first and second order derivative information for PDE-constrained inverse problems.
  - **Mesh-independence** (i.e., the cost measured in terms of the number of required PDE solves does not scale with the dimension of the inversion parameter).
  - **Exploit low-dimensional structure** (global and/or off diagonal low rank, parameter and state dimension reduction, preconditioning for the Newton system, etc.)

Optimal experimental design (OED)

- In many inverse problem one has access to only sparse and noisy data, i.e., only a few measurements can be collected.
- This makes parameter estimation challenging.
- One needs to strike a balance between the competing goals of stabilizing the inverse problem via regularization and extracting as much information as possible from limited data.



- Optimal data aquisition ploblem: find optimal sensor locations that minimize the "posterior uncertainty".
- A-optimal design approach leads to the following optimization problem:

$$\min_{\boldsymbol{w} \in [0,1]^{ns}} \mathsf{tr}(\mathcal{C}_{\text{post}}(\boldsymbol{w}))$$

Details in: A. Alexanderian, N. Petra, G. Stadler, O. Ghattas. *A Fast and Scalable Method for A-Optimal Design of Experiments for Infinite-dimensional Bayesian Nonlinear Inverse Problems*, SISC 38 (1), 2016

Parameter estimation and optimal experimental design under (additional) uncertainty

- In practice, in addition to the inversion parameters, the PDEs governing an inverse problem contain other parameters that might be uncertain.
- Such parameters, which are referred to as **auxiliary parameters**, **nuisance parameters** or **secondary uncertain parameters** are not being estimated, but are needed to fully specify the governing model.
- It is important to understand the impact of these parameters on the inverse problem solution and to account for this additional model uncertainties.
- Hyper-differential sensitivity analysis (HDSA), can point to sources of modeling uncertainties that cannot be ignored.
- Once one has identifies the key modeling uncertainties, it is important to account for these uncertainties in both the parameter inference and experimental design (OED) stages.

Details in:

- J. Kaipio and E. Somersalo, *Statistical and Computational Inverse Problems*, Springer, 2005
- A. Alexanderian, N. Petra, R. Nicholson. *Optimal design of large-scale nonlinear Bayesian inverse problems under model uncertainty*, Under review. arXiv preprint arXiv:2211.03952

References

1. **Theory and computational methods for inverse problems:**
   - Heinz Engl, Michael Hanke, and Andreas Neubauer, *Regularization of Inverse Problems*, Dordrecht, 2nd ed., 1996.
   - Curtis R. Vogel, *Computational Methods for Inverse Problems*, SIAM, 2002.
   - Guy Chavent, *Nonlinear Least Squares for Inverse Problems*, Springer, 2009.
2. **Numerical optimization background:**
   - Jorge Nocedal and Stephen J. Wright, *Numerical Optimization*, Springer-Verlag, 1999.
   - C. Tim Kelley, *Iterative Methods of Optimization*, SIAM, 1999.
3. **Optimization of systems governed by PDEs:**
   - A. Borzi and V. Schulz, *Computational Optimization of Systems Governed by Partial Differential Equations*, SIAM, 2012
   - Fredi Tröltzsch, *Optimal Control of Partial Differential Equations: Theory, Methods and Applications*, Graduate Studies in Mathematics Vol. 112, AMS, 2010.
   - M. Hinze, R. Pinnau, M. Ulbich, and S. Ulbrich, *Optimization with PDE constraints*, Springer, 09.
4. **Finite element background and FEniCS:**
   - Eric B. Becker, Graham F. Carey, and J. Tinsley Oden, *Finite Elements: Volume I, An Introduction*, Prentice Hall, 1981.
   - Mark S. Gockenbach, *Understanding and Implementing the Finite Element Method*, SIAM, 2006.
   - Howard Elman, David Silvester, and Andrew Wathen, *Finite Elements and Fast Iterative Solvers*, Oxford University Press, 2005.
   - A. Logg, K. A. Mardal, and G. Wells, *Automated solution of differential equations by the finite element method: The FEniCS book*, vol. 84, Springer Science & Business Media, 2012.
5. **Probabilistic approach to inverse problems:**
   - Albert Tarantola, *Inverse Problem Theory and Methods for Model Parameter Estimation*, SIAM, 2005.
   - Jari Kaipio and Erkki Somersalo, *Statistical and Computational Inverse Problems*, Springer, 2005.
   - Andrew Stuart, Inverse problems, *A Bayesian approach*, Acta Numerica, 2010.
   - Daniela Calvetti and Erkki Somersalo, *Bayesian Scientific Computing*, Springer, 2023.