Lec 10: Inverse problems

- Introduction
- Inverse problems (I): Structural topology optimzation
- Inverse problems (II): Bayesian nonlinear regression
- Techniques for time-dependent inverse problems

Introduction

PDE-constrained optimization

Consider a temperature control problem where we want to achieve a desired temperature field $u^d(\underline{x})$ by changing the heat source $f(\underline{x})$.

Mathematically, this is an optimization problem constrained by a PDE:

$$\min_{f,u} \ \underbrace{\frac{1}{2}\int_\Omega (u-u^d)^2}_{\text{data mismatch}} + \underbrace{\frac{\alpha}{2}\int_\Omega f^2}_{\text{regularization}}$$

$$\left.\begin{aligned} \text{s.t.} \quad -k\Delta u &= f \quad \forall \underline{x} \in \Omega \\ u &= g \quad \text{on } \Gamma_D \\ -k\nabla u\cdot\underline{n} &= h \quad \text{on } \Gamma_N \end{aligned}\right\} \rightarrow \text{PDE constraint}$$

We will adopt the discretize-then-optimize approach (the alternative is optimize-then-discretize). Let $\underline{\theta} \in \mathbb{R}^M$ be the discretized $f(\underline{x})$ and $\underline{U} \in \mathbb{R}^N$ be the discretized $u(\underline{x})$. The discretized optimization problem is:

$$\min_{\underline{\theta},\underline{U}} J(\underline{U},\underline{\theta}) \rightarrow \text{objective function}$$

$$\text{s.t. } \underline{R}(\underline{U},\underline{\theta}) = \underline{0} \rightarrow \text{discretized residual function}$$

where $J: \mathbb{R}^N \times \mathbb{R}^M \to \mathbb{R}$ is the objective function, and $\underline{R}: \mathbb{R}^N \times \mathbb{R}^M \to \mathbb{R}^N$ is the discretized residual function from FEM weak form.

Adjoint method - Direct formulation

Observation: For any given $\underline{\theta}$, we can find $\underline{U}$ by solving $\underline{R}(\underline{U}, \underline{\theta}) = \underline{0}$.

Then, $\underline{U}$ is a function of $\underline{\theta}$ implicitly, i.e., $\underline{U}(\underline{\theta})$. Then we can consider the equivalent (unconstrained) optimization:

$$\min_{\underline{\theta}} \hat{J}(\underline{\theta}) = J(\underline{U}(\underline{\theta}), \underline{\theta})$$

Since $\underline{\theta} \in \mathbb{R}^M$ can be a high-dimensional vector, the optimization problem can only be efficiently solved if the gradient $\dfrac{d\hat{J}}{d\underline{\theta}}$ is available.

Goal: Find $\dfrac{d\hat{J}}{d\underline{\theta}}$ using the adjoint method.

By the chain rule, we have

$$\underset{1\times M}{\frac{d\hat{J}}{d\underline{\theta}}} = \underset{1\times N}{\frac{\partial J}{\partial \underline{U}}} \underset{N\times M}{\frac{d\underline{U}}{d\underline{\theta}}} + \underset{1\times M}{\frac{\partial J}{\partial \underline{\theta}}} \qquad (*)$$

To find $\dfrac{d\underline{U}}{d\underline{\theta}}$, differentiate $\underline{R}(\underline{U}, \underline{\theta}) = \underline{0}$ w.r.t. $\underline{\theta}$ for both sides:

$$\underset{N\times N}{\frac{\partial \underline{R}}{\partial \underline{U}}} \underset{N\times M}{\frac{d\underline{U}}{d\underline{\theta}}} + \underset{N\times M}{\frac{\partial \underline{R}}{\partial \underline{\theta}}} = 0 \Rightarrow \underset{N\times M}{\frac{d\underline{U}}{d\underline{\theta}}} = -\underset{N\times N}{\left(\frac{\partial \underline{R}}{\partial \underline{U}}\right)^{-1}} \underset{N\times M}{\frac{\partial \underline{R}}{\partial \underline{\theta}}}$$ (tangent linear equation)

Substitute into (*):

$$\frac{d\hat{J}}{d\underline{\theta}} = \underbrace{\overbrace{\frac{\partial J}{\partial \underline{U}}\left(-\frac{\partial \underline{R}}{\partial \underline{U}}\right)^{-1}}^{\text{adjoint}}\frac{\partial \underline{R}}{\partial \underline{\theta}}}_{\text{tangent linear}} + \frac{\partial J}{\partial \underline{\theta}}$$

$1\times M \qquad 1\times N \qquad N\times N \qquad N\times M \qquad 1\times M$

Two strategies :

① Solve the tangent linear equation first and get $\frac{d\underline{U}}{d\underline{\theta}}$ so that

$$\frac{d\hat{J}}{d\underline{\theta}} = \frac{\partial J}{\partial \underline{U}}\frac{d\underline{U}}{d\underline{\theta}} + \frac{\partial J}{\partial \underline{\theta}} \quad ✓$$

② Solve the adjoint equation first and get $\underline{\lambda}$:

$$\frac{\partial J}{\partial \underline{U}}\left(-\frac{\partial \underline{R}}{\partial \underline{U}}\right)^{-1} = \underline{\lambda}^T \Rightarrow \left(\frac{\partial \underline{R}}{\partial \underline{U}}\right)^T \underline{\lambda} = -\left(\frac{\partial J}{\partial \underline{U}}\right)^T \quad \text{(adjoint equation)}$$

$1\times N \qquad N\times N \qquad 1\times N \qquad\qquad N\times N \qquad N \qquad N$

$$\Rightarrow \frac{d\hat{J}}{d\underline{\theta}} = \underline{\lambda}^T\frac{\partial \underline{R}}{\partial \underline{\theta}} + \frac{\partial J}{\partial \underline{\theta}} \quad ✓ .$$

Question : Which strategy is more efficient in what conditions ?

Adjoint method - Formulation through Lagrange function

A popular (equivalent) formulation of the adjoint method is through the Lagrange function method. Define the Lagrange function :

$$\mathcal{L}(\underline{U}, \underline{\theta}, \underline{\lambda}) := J(\underline{U}, \underline{\theta}) + \underline{\lambda}^T \underline{R}(\underline{U}, \underline{\theta})$$

Consider
$$\frac{\partial \mathcal{L}}{\partial \underline{\lambda}} = \underline{0} \iff \underline{R}(\underline{U}, \underline{\theta}) = \underline{0} \quad (\text{constraint equation})$$

$$\frac{\partial \mathcal{L}}{\partial \underline{U}} = \underline{0} \iff \frac{\partial J}{\partial \underline{U}} + \underline{\lambda}^T \frac{\partial \underline{R}}{\partial \underline{U}} = \underline{0} \quad (\text{adjoint equation})$$

$$\frac{\partial \mathcal{L}}{\partial \underline{\theta}} = \frac{\partial J}{\partial \underline{\theta}} + \underline{\lambda}^T \frac{\partial \underline{R}}{\partial \underline{\theta}} = \frac{d \hat{J}}{d \underline{\theta}} \quad (\text{gradient})$$

Remarks: Lagrange function formulation can be convenient.

Inverse problems (I): Structural topology optimization

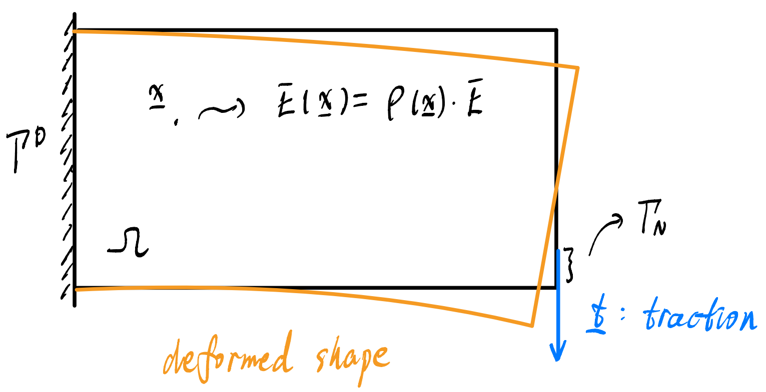
Problem statement

Structural topology optimization aims to optimize the structural performance by adjusting the material layout within a given design space.

Example (Compliance minimization): Find the material density field $\rho(\underline{x})$ so that the work done by the external load is minimized.

$$\min_{\rho,\, u} \int_{\Gamma^N} \underline{u} \cdot \underline{t} \rightsquigarrow \text{compliance}$$

$$\text{s.t.} \int_{\Omega} \underline{\underline{\sigma}} : \nabla \underline{v} - \int_{\Gamma^N} \underline{t} \cdot \underline{v} = 0 \quad \forall \underline{v}$$



$$\boxed{\begin{aligned} &\min_{\underline{\theta},\, \underline{U}} \ \underline{U}^T \cdot \underline{F} \\ &\text{s.t.} \ \underline{\underline{K}}(\underline{\theta}) \underline{U} = \underline{F} \end{aligned}}$$

$\underline{\theta} \in \mathbb{R}^M$: discretized $\rho(\underline{x})$

$\underline{U} \in \mathbb{R}^N$: discretized $\underline{u}(\underline{x})$

$\rightsquigarrow$ (discretized) PDE-constrained optimization

$$\min_{\underline{\theta}} \hat{J}(\underline{\theta}) = \underline{U}^T(\underline{\theta}) \cdot \underline{F}$$

e.g., $\theta_c = 0.5$

$$\text{s.t.} \quad g(\underline{\theta}) = \frac{1}{M}\sum_{m=1}^{M} \theta_m - \theta_c \leq 0$$ (some additional volume constraint)

Optimization algorithm

$\underline{\theta} = \underline{\theta}_0$, $\Delta = 1.0$ (initialization)

while $\Delta > \varepsilon$: (ε is tolerance , e.g., $\varepsilon = 10^{-6}$)

  Find $\hat{J}(\underline{\theta})$, $\dfrac{d\hat{J}}{d\underline{\theta}}$ (adjoint method)

  Find $g(\underline{\theta})$, $\dfrac{dg}{d\underline{\theta}}$ (easy)

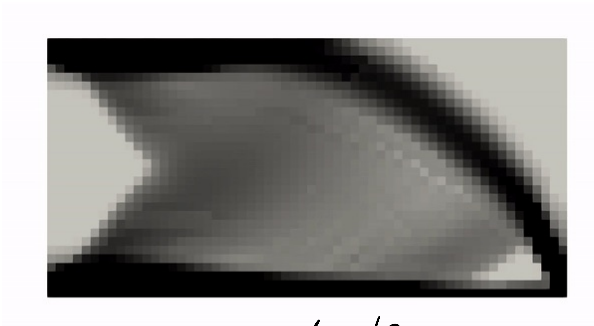
  $\underline{\theta} = \text{MMA}\left(\hat{J}, \dfrac{d\hat{J}}{d\underline{\theta}}, g, \dfrac{dg}{d\underline{\theta}}\right)$ (optimizer set to be Method of Moving Asymtotes)

  $\Delta = \| \underline{\theta} - \underline{\theta}^{prev} \|$ (update Δ for stopping condition)

return $\underline{\theta}$

Numerical example from JAX-FEM

step 0



step 10

step 20

step 30

Remarks:

* You may need some numerical treatments for better solutions:

  ① Sensitivity / density filtering to avoid "checker board patterns"

  ② Projection to binarize filtered densities

  ③ Penalization ($E(\underline{x}) = E_{min} + \rho^p(\underline{x})\,(E_{max} - E_{min})$) to avoid intermediate values.

* The topology optimization method discussed here is known as the SIMP method, while other methods exist (e.g., level set, BESO, MMC).

Inverse problems (II): Bayesian nonlinear problems

In the context of scientific machine learning, we will show a connection between supervised learning and FEM-based inverse problems.

Bayesian regression for supervised learning

Problem setup: We want to train a model to predict from $\underline{x}$ to $\underline{y}$.

$\underline{x}$ → **Neural networks parametrized by $\underline{w}$** → $\underline{y}$

$\underline{x}$: Input (e.g., features of a house), $\underline{x} \in \mathbb{R}^L$

$\underline{y}$: Output (e.g., price of the house), $\underline{y} \in \mathbb{R}^N$

$\underline{w}$: Trainable parameters, $\underline{w} \in \mathbb{R}^M$

We have a dataset $\{(\underline{x}_i, \underline{y}_i)\}_{i=1}^{I}$ that follows

$$\underbrace{\underline{y}_i}_{\text{observation}} = \underbrace{\underline{f}_{\underline{w}}(\underline{x}_i)}_{\text{neural network prediction}} + \underbrace{\underline{\varepsilon}_i}_{\text{noise}}, \quad \underline{\varepsilon}_i \sim \mathcal{N}(\underline{0}, \sigma^2 \underline{\underline{I}})$$

Assume a Gaussian prior to parameters $\underline{w}$:

$$p(\underline{w}) = \mathcal{N}(\underline{w} \mid \underline{0}, \alpha^{-1}\underline{\underline{I}}) \quad (\alpha > 0 \text{ controls the precision})$$

The likelihood of observing $\underline{y}_i$:

$$p(\underline{y}_i \mid \underline{x}_i, \underline{w}) = \mathcal{N}(\underline{y}_i \mid \underline{f}_{\underline{w}}(\underline{x}_i), \sigma^2 \underline{\underline{I}})$$

The posterior distribution is:

$$p(\underline{w} \mid \underbrace{\underline{x}_1, \underline{x}_2, \cdots, \underline{x}_I}_{\underline{X}}, \underbrace{\underline{y}_1, \underline{y}_2, \cdots, \underline{y}_I}_{\underline{Y}}) = \frac{p(\underline{w} \mid \underline{x}_1, \cdots, \underline{x}_I)\, p(\underline{y}_1, \cdots, \underline{y}_I \mid \underline{w}, \underline{x}_1, \cdots, \underline{x}_I)}{p(\underline{y}_1, \cdots, \underline{y}_I \mid \underline{x}_1, \cdots, \underline{x}_I)}$$

$$= \frac{p(\underline{w}) \cdot p(\underline{y}_1 \mid \underline{w}, \underline{x}_1) \cdot p(\underline{y}_2 \mid \underline{w}, \underline{x}_2) \cdots p(\underline{y}_I \mid \underline{w}, \underline{x}_I)}{p(\underline{y}_1, \cdots, \underline{y}_I \mid \underline{x}_1, \cdots, \underline{x}_I)}$$

The maximum a posteriori (MAP) estimate states that

$$\underline{w}_{MAP} = \arg\max_{\underline{w}} p(\underline{w} \mid \underline{X}, \underline{Y})$$

Equivalent to:

$$\underline{w}_{MAP} = \arg\min_{\underline{w}} -\log p(\underline{w} \mid \underline{X}, \underline{Y}) = \arg\min_{\underline{w}} \left(-\log p(\underline{w}) - \sum_{i=1}^{I} p(\underline{y}_i \mid \underline{w}, \underline{x}_i)\right)$$

$$= \arg\min_{\underline{w}} \underbrace{\frac{\alpha}{2}\|\underline{w}\|^2}_{\text{regularization}} + \underbrace{\frac{1}{2\sigma^2}\sum_{i=1}^{I} \|\underline{y}_i - \underline{f}_{\underline{w}}(\underline{x}_i)\|^2}_{\text{data mismatch}}$$

→ loss function $L(\underline{w})$

Training: Use gradient descent ($\underline{w} \leftarrow \underline{w} - \eta \frac{d\mathcal{L}}{d\underline{w}}$) to find a good $\underline{w}$.

FEM-based inverse problems

Problem setup: For a heat problem with inhomogeneous conductivity $k(\underline{x})$, a given heat source $f(\underline{x})$ can lead to a temperature field $u(\underline{x})$. Suppose we have observed $\{(\underline{F}_i, \underline{U}_i)\}_{i=1}^{I}$, and the goal is to infer $\underline{\theta}$.

$$\underline{F} \rightarrow \boxed{\begin{array}{l}\text{FEM solver} \\ \text{parametrized by } \underline{\theta}\end{array}} \rightarrow \underline{U}$$

$\underline{F}$: Input (discretized $f(\underline{x})$), $\underline{F} \in \mathbb{R}^L$

$\underline{U}$: Output (discretized $u(\underline{x})$), $\underline{U} \in \mathbb{R}^N$

$\underline{\theta}$: Optimizable parameters (discretized $k(\underline{x})$), $\underline{\theta} \in \mathbb{R}^M$

Suppose due to measurement noise,

$$\underbrace{\underline{U}_i}_{\text{observation}} = \underbrace{\underline{U}_\theta(\underline{F}_i)}_{\text{FEM prediction}} + \underbrace{\underline{\varepsilon}_i}_{\text{noise}}, \quad \underline{\varepsilon}_i \sim \mathcal{N}(\underline{0}, \sigma^2 \underline{\underline{I}})$$

Following exactly the same procedure of Bayesian regression:

$$\underline{\theta}_{MAP} = \underset{\underline{\theta}}{\operatorname{argmin}} \underbrace{\overbrace{\frac{\alpha}{2}\|\underline{\theta}\|^2}^{\text{regularization}} + \overbrace{\frac{1}{2\sigma^2}\sum_{i=1}^{I}\|\underline{U}_i - \underline{U}_\theta(\underline{F}_i)\|^2}^{\text{data mismatch}}}_{\text{loss function } \mathcal{L}(\underline{\theta})}$$

Training requires $\frac{d\mathcal{L}}{d\underline{\theta}}$ (involving $\frac{d\underline{U}_\theta}{d\underline{\theta}}$) to be found by adjoint method.

Techniques for time-dependent inverse problems

From previous discussions, we understand the importance of finding the gradient $\frac{dL}{d\underline{\theta}}$ (or $\frac{d\hat{J}}{d\underline{\theta}}$). Here, we present the workflow of finding the gradient for time-dependent problems.

Problem setup

Consider a time-dependent problems with both time and space discretized:

$$\underline{U}^n = \underline{U}^n(\underline{U}^{n-1}, \underline{\theta}) = \begin{cases} \underline{F}(\underline{U}^{n-1}, \underline{\theta}) & \text{explicit} \\ \underline{U}^n \text{ by solving } \underline{R}(\underline{U}^n, \underline{U}^{n-1}, \underline{\theta}) = \underline{0} & \\ \text{given } \underline{U}^{n-1} \text{ and } \underline{\theta} & \text{implicit} \end{cases}$$

$$\underline{U}^0 = \underline{\beta}$$

$\underline{\theta} \in \mathbb{R}^M$

$\underline{U}^n \in \mathbb{R}^N$

$n = 0, 1, 2, \cdots, n_t$

Assume that the objective function $L(\underline{\theta}, \underline{\beta}) = h(\underline{U}^{n_t})$, the goal is to find $\frac{dL}{d\underline{\theta}}$ and $\frac{dL}{d\underline{\beta}}$.

Workflow

With chain rule, we have

$$\frac{dL}{d\underline{\theta}} = \frac{dh}{d\underline{U}^{n_t}} \cdot \left(\frac{\partial \underline{U}^{n_t}}{\partial \underline{\theta}} + \frac{\partial \underline{U}^{n_t}}{\partial \underline{U}^{n_t-1}} \cdot \frac{d\underline{U}^{n_t-1}}{d\underline{\theta}} \right)$$

$$= \frac{dh}{d\underline{U}^{n_t}} \cdot \left(\frac{\partial \underline{U}^{n_t}}{\partial \underline{\theta}} + \frac{\partial \underline{U}^{n_t}}{\partial \underline{U}^{n_t-1}} \cdot \left(\frac{\partial \underline{U}^{n_t-1}}{\partial \underline{\theta}} + \frac{\partial \underline{U}^{n_t-1}}{\partial \underline{U}^{n_t-2}} \cdot \frac{d\underline{U}^{n_t-2}}{d\underline{\theta}} \right) \right)$$

$$= \cdots$$

$$\frac{dL}{d\underline{\beta}} = \frac{dh}{d\underline{U}^{n_t}} \cdot \frac{d\underline{U}^{n_t}}{d\underline{\beta}}$$

$$= \frac{dh}{d\underline{U}^{n_t}} \cdot \frac{d\underline{U}^{n_t}}{d\underline{U}^{n_t-1}} \cdot \frac{d\underline{U}^{n_t-1}}{d\underline{\beta}}$$

$$= \cdots$$

Define auxiliary variables $\underline{V}_{\theta}^{n}$ associated with $\underline{\theta}$ and $\underline{V}_{\beta}^{n}$ with $\underline{\beta}$, where $\underline{V}_{\theta}^{n}, \underline{V}_{\beta}^{n} \in \mathbb{R}^{N}$, $n = n_t, n_{t-1}, \cdots, 0$ such that

$$\begin{cases} (\underline{V}_{\theta}^{n-1})^T = (\underline{V}_{\beta}^{n})^T \dfrac{\partial \underline{U}^{n}}{\partial \underline{\theta}} + (\underline{V}_{\theta}^{n})^T \\ (\underline{V}_{\beta}^{n-1})^T = (\underline{V}_{\beta}^{n})^T \dfrac{\partial \underline{U}^{n}}{\partial \underline{U}^{n-1}} \end{cases} \quad \text{with starting conditions} \quad \begin{cases} \underline{V}_{\theta}^{n_t} = \underline{0} \\ \underline{V}_{\beta}^{n_t} = \dfrac{dh}{d\underline{U}^{n_t}} \end{cases}$$

You can show that $\boxed{\dfrac{dL}{d\underline{\theta}} = \underline{V}_{\theta}^{0}}$ and $\boxed{\dfrac{dL}{d\underline{\beta}} = \underline{V}_{\beta}^{0}}$.

Explicit situation

Just replace $\dfrac{\partial \underline{U}^{n}}{\partial \underline{\theta}}$ with $\dfrac{\partial \underline{F}}{\partial \underline{\theta}}$ and replace $\dfrac{\partial \underline{U}^{n}}{\partial \underline{U}^{n-1}}$ with $\dfrac{\partial \underline{F}}{\partial \underline{U}^{n-1}}$.

Implicit situation

$$(\underline{V}_{\beta}^{n})^T \frac{\partial \underline{U}^{n}}{\partial \underline{\theta}} = (\underline{\lambda}^{n})^T \frac{\partial \underline{R}}{\partial \underline{\theta}}$$

$$(\underline{V}_{\beta}^{n})^T \frac{\partial \underline{U}^{n}}{\partial \underline{U}^{n-1}} = (\underline{\lambda}^{n})^T \frac{\partial \underline{R}}{\partial \underline{U}^{n-1}}$$

where $\left(\dfrac{\partial \underline{R}}{\partial \underline{U}^{n}}\right)^T \underline{\lambda}^{n} = -\underline{V}_{\beta}^{n}$ (adjoint equation)