When Equations meet Data:
Inverse Problems and Hybrid-Cooperative Learning

Roberto Morales Ponce
University of Deusto, Spain

Supported by the CoDeFeL ERC project
Grant Number 101096251

June 4, 2026

European Commission

European Research Council
Established by the European Commission

Deusto
Universidad de Deusto
Deustuko Unibertsitatea
University of Deusto

Outline

Models based on Differential Equations

Models based on Differential Equations

General form

A mathematical model can often be written as

$$\mathcal{A}(u, \psi) = f,$$

where u is the state, ψ denotes a set of parameters, and f represents external inputs or sources.

A First Example: The SIR Model

A simple question

How can we describe the time evolution of an epidemic using a few interpretable variables?

- $S(t)$: susceptible individuals.
- $I(t)$: infected individuals.
- $R(t)$: recovered or removed individuals.

$$S \longrightarrow I \longrightarrow R$$

The population moves between compartments.

Key idea

These transitions can be encoded by a system of Differential Equations.

The SIR Equations

W. Kermack and A. McKendrick[1] proposed the following model:

$$\begin{cases} S'(t) = -\beta S(t)I(t), \\ I'(t) = \beta S(t)I(t) - \gamma I(t), \\ R'(t) = \gamma I(t), \\ S(0) = S_0,\ I(0) = I_0,\ R(0) = R_0, \end{cases}$$

where

- $N = S_0 + I_0 + R_0$ is the number of individuals.
- β is the transmission rate: it measures effective contacts.
- γ is the recovery rate: it measures how fast infected individuals leave I.
- The term βSI couples the state variables and creates nonlinear dynamics.

[1] Kermack, W. O., and McKendrick, A. G. (1927) A contribution to the Mathematical Theory of Epidemics. Proceedings of the Royal Society A.

Why is the advantage to use Differential Equations?

Notice that

- For each time $t > 0$, we have

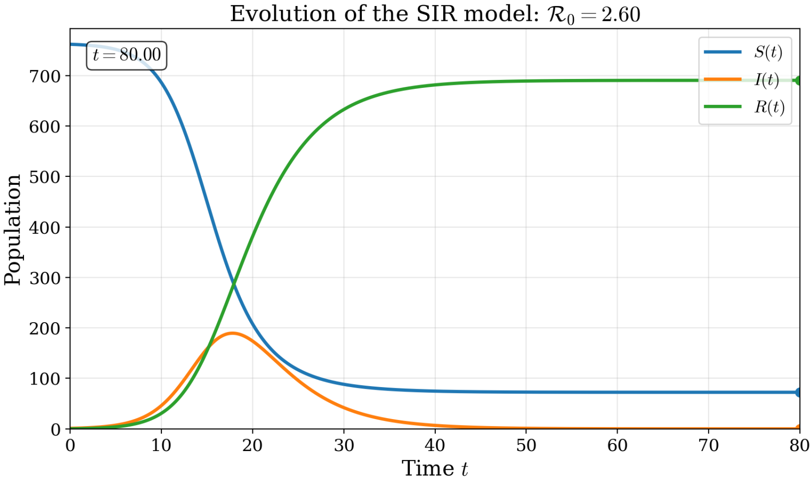
$$S(t) + I(t) + R(t) = N, \quad \forall t \geq 0.$$

- If $S_0 > 0$, $I_0 > 0$ and $R_0 \geq 0$, then

$$S(t) > 0, \quad I(t) > 0, \quad R(t) \geq 0, \quad \forall\, t \geq 0.$$

- Consider $\mathcal{R}_0 := \dfrac{\beta S_0}{\gamma}$.
  - If $\mathcal{R}_0 > 1$, then the infection initially increases,
  - If $\mathcal{R}_0 < 1$, then the infection immediately decreases.
- The peak of $I(t)$ occurs when $S(t) = \frac{\gamma}{\beta}$.

Evolution of the SIR model



A PDE model

We consider a forced passive scalar transport equation in a prescribed two-dimensional cellular flow:

$$\begin{cases} \partial_t u + \vec{v} \cdot \nabla u = f(x,y), & (x,y,t) \in \Omega \times (0,T), \\ u(x,y,0) = u_0, & (x,y) \in \Omega, \\ u \text{ is periodic on } \partial\Omega. \end{cases}$$

Here,

- scalar quantity transported by the flow,
- $\vec{v}$ is fixed in advance,
- $f(x,y)$ is a spatially distributed source,
- u_0 is the initial state of the passive scalar.

A numerical example

Here, we consider $\Omega = (0,1)^2$, $T = 5$, and

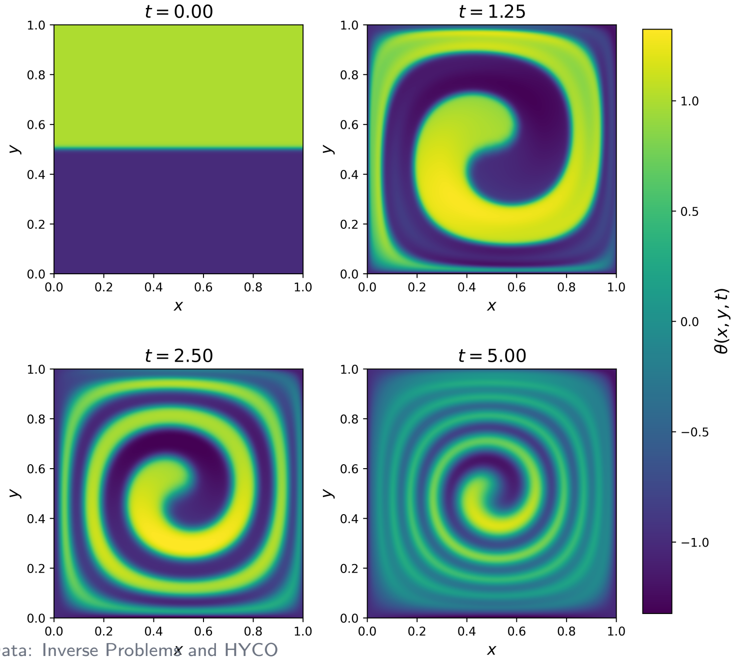
$$\begin{cases} \partial_t u + \vec{v} \cdot \nabla u = f(x,y), & (x,y,t) \in \Omega \times (0,T), \\ u(x,y,0) = u_0, & (x,y) \in \Omega, \\ u(0,y,t) = u(1,y,t),\ u(x,0,t) = u(x,1,t), & \forall t \in (0,T). \end{cases}$$

with

$$u_0(x,y) = \tanh\left(\frac{y-0.5}{0.01}\right), \quad \vec{v}(x,y) = (-\pi\sin(\pi x)\cos(\pi x), \pi\cos(\pi x)\sin(\pi y)),$$

and

$$f(x,y) = \sin(2\pi x)\cos(2\pi y).$$


$t = 0.00$
$t = 1.25$
$t = 2.50$
$t = 5.00$
x
y
$\theta(x, y, t)$
0.0
0.2
0.4
0.6
0.8
1.0
−1.0
−0.5
0.0
0.5
1.0

Inverse Problems and their applications

From Direct to Inverse Problems

Direct problem. Given the model and its parameters, predict the state of the system.

$$A(u, \psi) = f \quad \Longrightarrow \quad u.$$

Inverse problem. Given partial observations of the state, recover hidden information about the system.

$$\text{observations of } u \quad \Longrightarrow \quad \psi.$$

Main idea

Inverse problems ask whether we can infer causes from effects.

A Motivating Question

Inverse problems appear when the quantity of interest cannot be measured directly.

Can we determine what is inside a system by observing only its response?

Examples:

- Can we recover the interior conductivity of a body from boundary measurements?
- Can we identify a source term from temperature observations?
- Can we estimate epidemiological parameters from infection data?
- Can we reconstruct an image from indirect measurements?

Can One Hear the Shape of a Drum?

Another famous inverse problem was popularized by Mark Kac [2]:

Can one hear the shape of a drum?

Mathematically, the vibration of a drum is modeled by the wave equation or by the eigenvalue problem

$$\begin{cases} -\Delta \varphi_k = \lambda_k \varphi_k, & x \in \Omega, \\ \varphi_k = 0, & x \in \partial\Omega. \end{cases}$$

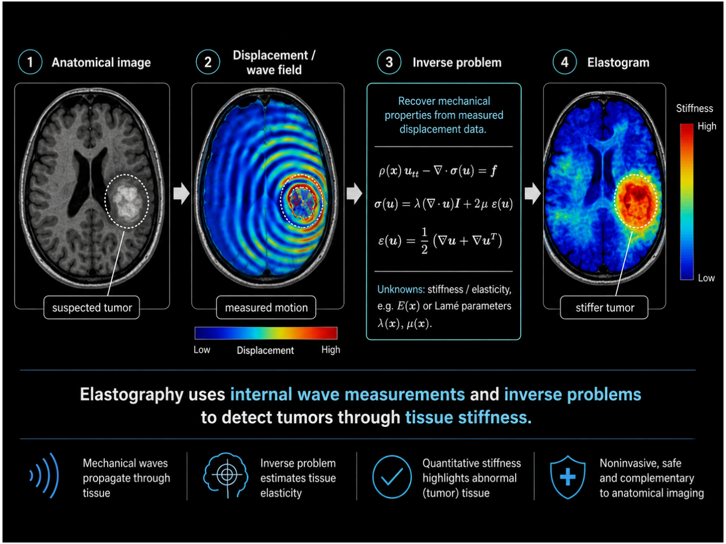
The numbers λ_k determine the frequencies of vibration.

Inverse question

Does the spectrum $\{\lambda_k\}_{k \geq 1}$ determine the geometry of Ω?

[2] Kac, M. (1966). Can one hear the shape of a drum?. The american mathematical monthly, 73(4P2), 1-23.

An example in Elastography



Why Are Inverse Problems Difficult?

Inverse problems are usually challenging because they are often **ill-posed**.

A problem is well-posed if:

- a solution exists,
- the solution is unique,
- the solution depends continuously on the data.

In inverse problems, one or more of these properties may fail.

Main difficulty

Small errors in the observations may produce large errors in the reconstruction.

A Variational Formulation

Given noisy observations y_{obs}, we seek an unknown set of parameters ψ by solving

$$\min_{\psi} \frac{1}{2}\|Cu(\psi) - y_{\text{obs}}\|^2 + \alpha R(\psi),$$

subject to the model equation

$$A(u, \psi) = f.$$

- C is the observation operator.
- $R(\psi)$ is a regularization term.
- $\alpha > 0$ balances data fitting and stability.

Interpretation

We search for a parameter that explains the data while avoiding unstable reconstructions.

Data, Neural Networks and Universal Approximation

When the equation is unknown

A different situation

Sometimes we observe a phenomenon, but we do not know the differential equation that governs it.

- We have measurements, experiments, or simulations.
- The underlying mechanisms may be too complex or partially unknown.
- Instead of starting from an equation, we start from data.

Main question

Can we learn the input–output relation directly from observations?

Supervised learning

Formulation

Assume that we are given data of the form

$$\{(x_i, y_i)\}_{i=1}^{N},$$

where x_i represents an input and y_i the corresponding observed output.
The goal is **to find a function** F such that

$$F(x_i) \approx y_i.$$

- In classification, y_i is a label.
- In regression, y_i is a continuous quantity.
- Training means adjusting F to fit the data.

Neural networks

Neural networks are flexible families of functions depending on parameters.
A simple feedforward neural network can be written as

$$F_\theta(x) = W_L\sigma\big(W_{L-1}\sigma(\cdots\sigma(W_1 x + b_1)\cdots) + b_{L-1}\big) + b_L.$$

- The matrices W_ℓ and vectors b_ℓ are trainable parameters.
- The activation function σ introduces nonlinearity.
- The architecture defines the class of functions we can approximate.

Training principle

Choose θ by minimizing a loss function, for example

$$\min_\theta \frac{1}{N}\sum_{i=1}^{N} |F_\theta(x_i) - y_i|^2.$$

Universal approximation

Universal approximation theorems provide a theoretical explanation for the expressive power of neural networks.

Informal statement

Under suitable assumptions on the activation function, neural networks can approximate any continuous function on a compact set with arbitrary accuracy.

That is, for a target function f and any tolerance $\varepsilon > 0$, one can find a neural network F_θ such that

$$\|f - F_\theta\| < \varepsilon.$$

- This gives a theoretical guarantee of expressivity.
- It does not automatically guarantee good training.
- It does not replace physical understanding.

From data-driven models to hybrid models

Purely data-driven models are powerful, but in scientific problems we often know something about the system.

- We may know conservation laws.
- We may know a partial differential equation.
- We may know constraints, symmetries, or qualitative properties.

Transition

This motivates hybrid approaches: use neural networks to learn from data, but guide the learning process with mathematical models.

Hybrid-Cooperative Learning

Why Hybrid-Cooperative Learning?

Motivation

Differential equations provide structure and interpretability, while data-driven models provide flexibility and adaptivity.

- Model-based approaches may depend on unknown parameters or incomplete mechanisms.
- Data-driven approaches may require large datasets and may ignore physical laws.
- Hybrid-Cooperative Learning[3] combines both sources of information.

Key idea

Equations and data should cooperate, not compete.

[3] Liverani, L., Steynberg, M., & Zuazua, E. (2025). Hyco: Hybrid-cooperative learning for data-driven pde modeling. arXiv preprint arXiv:2509.14123.

Two Complementary Agents

Physical agent

- Encodes the differential equation.
- Preserves constraints and mechanisms.
- Provides interpretable quantities.

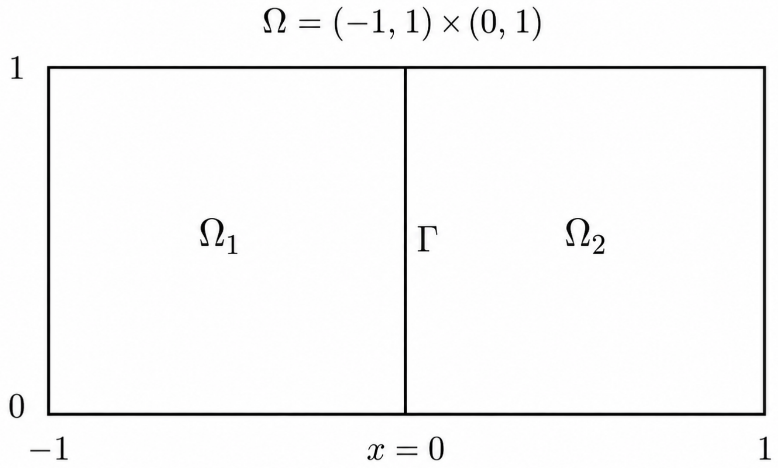
Data-driven agent

- Learns from observations.
- Approximates hidden states or corrections.
- Adapts to noisy and partial data.

Cooperation

The two agents exchange information through consistency and training losses.

An example of an Elliptic Transmission equation



An example of an Elliptic Transmission equation

We consider the elliptic boundary value problem[4]

$$\begin{cases} -\text{div}(a(x)\nabla u(x,y)) = \sin(\pi x)\sin(\pi y) & \text{in } \Omega, \\ u = 0, & \text{on } \partial\Omega, \end{cases}$$

with

$$a(x,y) = \begin{cases} a_1, & x < 0, \\ a_2, & x > 0. \end{cases}$$

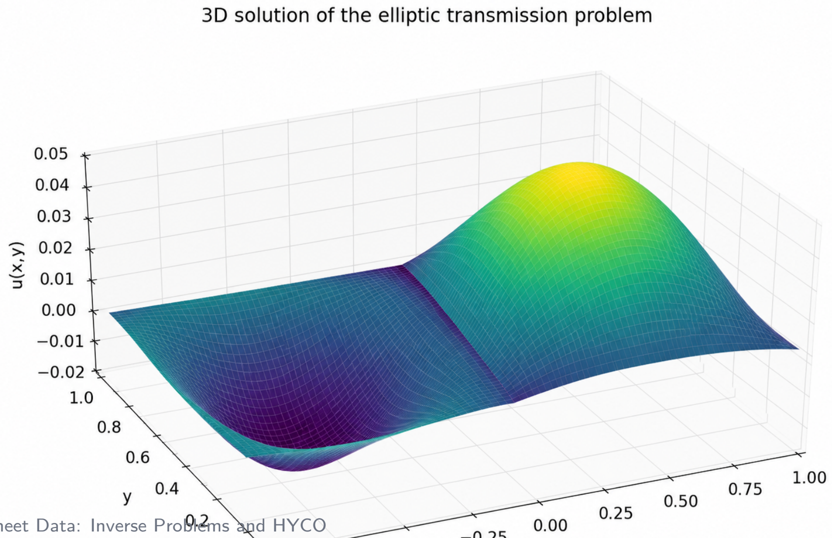
Inverse problem

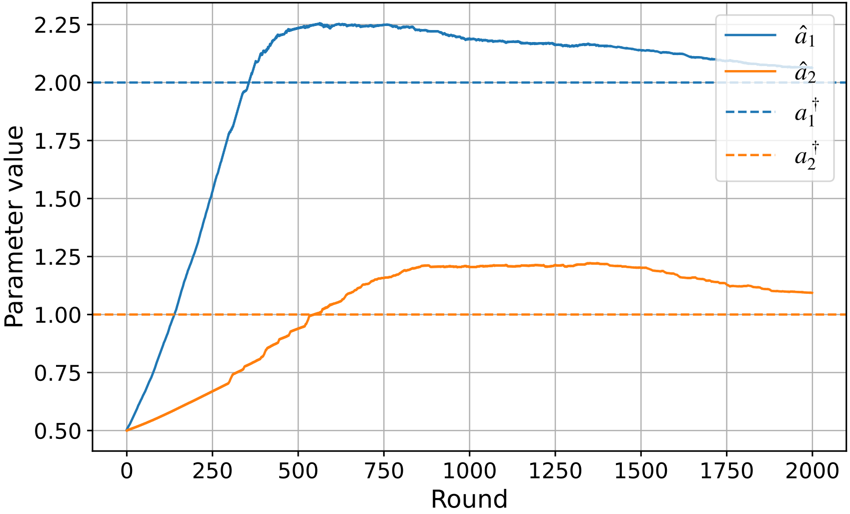
To recover the parameters (a_1, a_2) from fragmented partial observations.

[4] Biccari, U., Chen, J., Morales, R., & Zuazua, E. A Bi-objective HYCO framework for PDE-constrained inverse problems. In preparation.

The target

In our case, the true parameters are given by $a_1 = 2$ and $a_2 = 1$.





2.25
2.00
1.75
1.50
1.25
1.00
0.75
0.50
Parameter value
0
250
500
750
1000
1250
1500
1750
2000
Round
$\hat{a}_1$
$\hat{a}_2$
$a_1^\dagger$
$a_2^\dagger$

Robustness

Metric	η				
	0.00	0.05	0.10	0.15	0.20
$[u]_{\mathrm{syn}}$	2.56e-03	2.51e-03	2.60e-03	2.11e-03	2.27e-03
$[u]_{\mathrm{phy}}$	2.64e-03	2.51e-03	2.61e-03	2.69e-03	2.57e-03
$[a\nabla u \cdot n]_{\mathrm{phy}}$	2.47e-06	2.41e-06	2.50e-06	2.40e-06	2.46e-06
$[a\nabla u \cdot n]_{\mathrm{syn}}$	0.0711	0.0720	0.0711	0.0677	0.0640

Thank you!

Questions?