PDE Analysis via Inference: Treating Dynamical Systems as Data-Generating Processes

Dalton A R Sakthivadivel

Stony Brook University

28th October 2021

Outline

1. Problem and Motivation
2. Previous Results
3. Max Ent and Stationary Heat
4. Concluding Remarks

Slides and other material can be found later at darsakthi.github.io

Inference

Inference is fundamentally dynamical in character

- Given the state space of some data, D, and an unknown process $f : \mathbb{R} \to D$, we seek to infer a likely f for all $\mathbb{R}$ explaining current observations of D
- Non-trivial in the presence of noise—f is no longer uniquely specified under $f(t) + \eta(t)$
  - Allows several different *particular realisations* of noise, i.e., sample paths
- Given probabilities over sample paths $f(t) + \eta(t)$, what can we say about the states (data) generated by f?

Inference

Inference is fundamentally dynamical in character

- Given the state space of some data, D, and an unknown process $f : \mathbb{R} \to D$, we seek to infer a likely f for all $\mathbb{R}$ explaining current observations of D
- Non-trivial in the presence of noise—f is no longer uniquely specified under $f(t) + \eta(t)$
  - Allows several different *particular realisations* of noise, i.e., sample paths
- Given probabilities over sample paths $f(t) + \eta(t)$, what can we say about the states (data) generated by f?

Inference

Inference is fundamentally dynamical in character

- Given the state space of some data, D, and an unknown process $f : \mathbb{R} \to D$, we seek to infer a likely f for all $\mathbb{R}$ explaining current observations of D
- Non-trivial in the presence of noise—f is no longer uniquely specified under $f(t) + \eta(t)$
  - Allows several different *particular realisations* of noise, i.e., sample paths
- Given probabilities over sample paths $f(t) + \eta(t)$, what can we say about the states (data) generated by f?

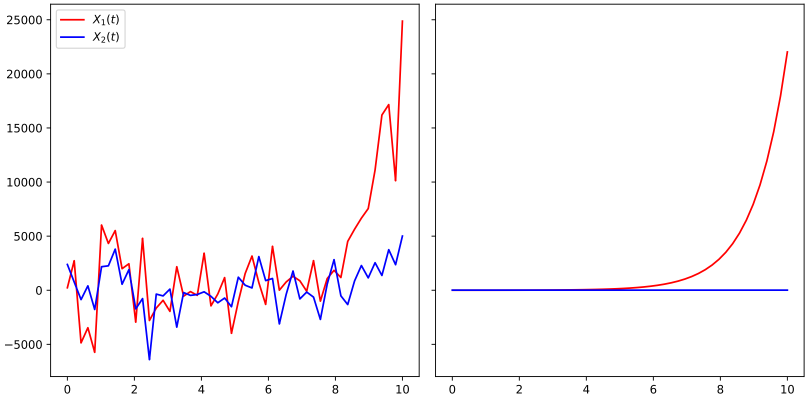
Inference

Inference is fundamentally dynamical in character

- Given the state space of some data, D, and an unknown process $f : \mathbb{R} \to D$, we seek to infer a likely f for all $\mathbb{R}$ explaining current observations of D
- Non-trivial in the presence of noise—f is no longer uniquely specified under $f(t) + \eta(t)$
  - Allows several different *particular realisations* of noise, i.e., sample paths
- Given probabilities over sample paths $f(t) + \eta(t)$, what can we say about the states (data) generated by f?

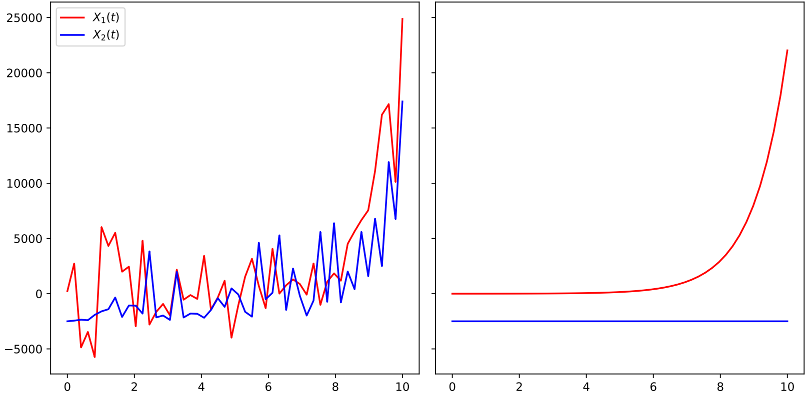
Example one

Consider the following samples of two D-valued random variables, $X_1(t)$ and $X_2(t)$:



Here, $X_1(t) = \exp\{t\} + \mathcal{N}(0, \sigma)$ whilst $X_2(t) = \mu + \mathcal{N}(0, \sigma)$

Example two



Here, $X_1(t) = a\exp\{t\} + \mathcal{N}(0,\sigma)$ whilst $X_2(t) = \mu + b\mathcal{N}_s(0,t)$

Problem

Problem: given an unknown stochastic dynamical system, how can we find the corresponding probability density, in general?

- Inferring $f \implies$ finding the probability of given samples $x \in D$, $p(x)$, by observing sample paths
- This probability density is given by the Fokker-Planck equation in the case of η being a Wiener process

Problem

Problem: given an unknown stochastic dynamical system, how can we find the corresponding probability density, in general?

- Inferring $f \implies$ finding the probability of given samples $x \in D$, $p(x)$, by observing sample paths
- This probability density is given by the Fokker-Planck equation in the case of η being a Wiener process

Problem

Problem: given an unknown stochastic dynamical system, how can we find the corresponding probability density, in general?

- Inferring $f \implies$ finding the probability of given samples $x \in D$, $p(x)$, by observing sample paths
- This probability density is given by the Fokker-Planck equation in the case of η being a Wiener process

The Fokker-Planck equation

The Fokker-Planck equation for states of a Gaussian process

$$f(t) + \mathcal{N}(0, t)$$

is

$$\frac{\partial p(x,t)}{\partial t} = -\frac{\partial f(t) p(x,t)}{\partial x} + \frac{\partial^2 p(x,t)}{\partial x^2}.$$

Can we find the solution to this equation by using inference to derive $p(x, t)$?

Objective

We seek to solve arbitrary diffusion processes, ideally in a closed form, inspired by inference

The Fokker-Planck equation

The Fokker-Planck equation for states of a Gaussian process

$$f(t) + \mathcal{N}(0, t)$$

is

$$\frac{\partial p(x,t)}{\partial t} = -\frac{\partial f(t) p(x,t)}{\partial x} + \frac{\partial^2 p(x,t)}{\partial x^2}.$$

Can we find the solution to this equation by using inference to derive $p(x, t)$?

Objective

We seek to solve arbitrary diffusion processes, ideally in a closed form, inspired by inference

Maximum entropy inference

Let S be the entropy

$$-\int_D \ln\{p(x)\}p(x)\,\mathrm{d}x - \lambda\int_D p(x)\,\mathrm{d}x + 1$$

maximised by some optimal $p_*(x)$. The Euler-Lagrange equation

$$\frac{\partial}{\partial p(x)}\big(-\ln\{p(x)\}p(x) - \lambda p(x)\big) = 0$$

maximises S, so we have $p_*(x)$ such that

$$-\ln\{p(x)\} - 1 - \lambda = 0.$$

Thus

$$p_*(x) = \exp\{-1-\lambda\}.$$

Maximum entropy inference with constraints

Let S be the constrained entropy

$$-\int_D \ln\{p(x)\}p(x)\,\mathrm{d}x - \left(\lambda_1 \int_D p(x)\,\mathrm{d}x - 1\right) - \left(\lambda_2 \int_D x\,p(x)\,\mathrm{d}x - \mathbb{E}[x]\right)$$

The Euler-Lagrange equation is now

$$-\ln\{p(x)\} - 1 - \lambda_1 - \lambda_2 x = 0,$$

yielding

$$p_*(x) = Z^{-1}\exp\{-\lambda_2 x\}.$$

This is the Boltzmann distribution.

Villani and Jaynes

Two interesting, but unconnected, statements:

- All physical processes ought to maximise entropy at equilibrium [Jaynes, 1957]
- The Fokker-Planck equation maximises entropy at equilibrium [Markowich and Villani, 2000]

Villani and Jaynes

Two interesting, but unconnected, statements:

- All physical processes ought to maximise entropy at equilibrium [Jaynes, 1957]
- The Fokker-Planck equation maximises entropy at equilibrium [Markowich and Villani, 2000]

Villani and Jaynes

Two interesting, but unconnected, statements:

- All physical processes ought to maximise entropy at equilibrium [Jaynes, 1957]
- The Fokker-Planck equation maximises entropy at equilibrium [Markowich and Villani, 2000]

The heat equation

In the restricted case of the heat equation and maximum entropy inference, we have analytic results.

Take the heat equation as

$$\frac{\partial}{\partial t} p(x,t) = \frac{\partial^2}{\partial x^2} p(x,t)$$

The stationary solution to the heat equation $p(x,t) = 0$ is some constant, the value of which depends on boundary conditions [Haberman 1987, 2.4.1].

Maximising entropy

Maximising the unconstrained entropy yields the constant function

$$\exp\{-\lambda - 1\},$$

so with the degree of freedom in λ being determined by the data, we can converge to the stationary heat equation.

Can we *prove* that this solves the heat equation?

Theorem (maximum entropy solves heat)

Theorem

For any finite volume and stationary $p(x)$,

$$\arg\max_{p(x)}(S[p(x)]) = p(x) = \exp\{-\lambda - 1\}$$

such that maximising the entropy computes the stationary solution to the heat equation.

A useful lemma

Lemma

The gradient flow of S is $\partial_{xx} p(x, t)$.

Proof.

First, set $\partial_t p(x, t) = \nabla S[p(x, t)]$ as a governing equation.

It is a fundamental result in *harmonic analysis* that $\frac{\partial^2}{\partial x^2} f$ can be deduced as the function for the dynamics of the gradient flow (descent) of certain functionals, like $S[f]$.

Therefore, $\frac{\partial^2}{\partial x^2} = \nabla S$, and,

$$\partial_t p(x, t) = \nabla S[p(x, t)] \iff \partial_t p(x, t) = \frac{\partial^2}{\partial x^2} p(x, t). \quad \square$$

Another useful lemma

Lemma

The gradient flow of $\partial_t S$ is $\partial_{xx} \ln\{p(x,t)\}$.

Proof.

Integrating by parts, we can write $\partial_t S$ as $\int_D \partial_x \ln\{p(x,t)\}\partial_x p(x,t)\,\mathrm{d}x$. The gradient flow of this expression is

$$\partial_x \partial_{\partial_x p(x,t)} \big(\partial_x \ln\{p(x,t)\}\partial_x p(x,t)\big),$$

which equals $\partial_{xx} \ln\{p(x,t)\}$. □

The integration by parts identity in this calculation was noted by Terence Tao in ‘Some notes on Bakry-Émery theory,’ 2013.

Proof of Theorem 1

Proof.

$\partial_{xx} p(x) \mapsto \partial_{xx} \ln\{p(x)\} \iff \nabla S[p(x)] \mapsto \nabla \partial_t S[p(x)]$ (Lem 2).

Set $\nabla \partial_t S[p(x)] = \partial_{xx} \ln\{p(x)\} = \partial_t \ln\{p(x)\}$.

Assume gradient descent, such that $\nabla \partial_t S[p(x)] = 0$. Then $\ln\{p(x)\}$ is stationary, i.e., $\partial_t \ln\{p(x)\} = 0$.

Furthermore, $\int_0^\tau \nabla \partial_t S[p(x)] = \nabla S[p(x)]\Big|_0^\tau = 0$ (Leibniz' rule).

Integrating the other side of Step 2, we get $\int_0^\tau \partial_t \ln\{p(x)\} = \ln\{p(x)\}\Big|_0^\tau$ such that

$$\ln\{p(x)\} = \nabla S[p(x)] - \lambda - 1.$$

Exponentiating both sides, Theorem 1 follows from $\nabla S = 0$. □

Why is this interesting?

For the heat equation, a gradient descent on entropy hides a time integral behind some algebra.

Deep learning as an inspiration

Where is this going? Will it become a tool for PDE analysis? A general theory of PDEs?

- Deep learning predicts dynamics of hopelessly complex SDE problems; solves their corresponding PDEs over states, from observing outputs
  - Is a very powerful technique
- Can we use deep learning to understand physical processes by inference?
  - In a sense, we already do
  - However—we don't understand DL
- Very little hope for analytic use of these techniques right now

Deep learning as an inspiration

Where is this going? Will it become a tool for PDE analysis? A general theory of PDEs?

- Deep learning predicts dynamics of hopelessly complex SDE problems; solves their corresponding PDEs over states, from observing outputs
  - Is a very powerful technique
- Can we use deep learning to understand physical processes by inference?
  - In a sense, we already do
  - However—we don't understand DL
- Very little hope for analytic use of these techniques right now

Deep learning as an inspiration

Where is this going? Will it become a tool for PDE analysis? A general theory of PDEs?

- Deep learning predicts dynamics of hopelessly complex SDE problems; solves their corresponding PDEs over states, from observing outputs
  - Is a very powerful technique
- Can we use deep learning to understand physical processes by inference?
  - In a sense, we already do
  - However—we don't understand DL
- Very little hope for analytic use of these techniques right now

Deep learning as an inspiration

Where is this going? Will it become a tool for PDE analysis? A general theory of PDEs?

- Deep learning predicts dynamics of hopelessly complex SDE problems; solves their corresponding PDEs over states, from observing outputs
  - Is a very powerful technique
- Can we use deep learning to understand physical processes by inference?
  - In a sense, we already do
  - However—we don't understand DL
- Very little hope for analytic use of these techniques right now

Computational techniques

If not analytic, what about numerical?

- Maximum entropy for the heat equation was just a special case. We can do the same in general, using numerical techniques rather than solving by hand
- Physics inspired neural networks
  - Lagrangian neural networks
  - Wave equations and fluid mechanics [Cai et al, 2021]
    - Karniadakis group
  - Fourier neural operator [Li et al, 2021]
    - Anandkumar group
- Many algorithms easily handle non-linearity, extreme dimensionality [Han, Jentzen, and E, 2018]

Computational techniques

If not analytic, what about numerical?

- Maximum entropy for the heat equation was just a special case. We can do the same in general, using numerical techniques rather than solving by hand
- Physics inspired neural networks
  - Lagrangian neural networks
  - Wave equations and fluid mechanics [Cai et al, 2021]
    - Karniadakis group
  - Fourier neural operator [Li et al, 2021]
    - Anandkumar group
- Many algorithms easily handle non-linearity, extreme dimensionality [Han, Jentzen, and E, 2018]

Computational techniques

If not analytic, what about numerical?

- Maximum entropy for the heat equation was just a special case. We can do the same in general, using numerical techniques rather than solving by hand
- Physics inspired neural networks
  - Lagrangian neural networks
  - Wave equations and fluid mechanics [Cai et al, 2021]
    - Karniadakis group
  - Fourier neural operator [Li et al, 2021]
    - Anandkumar group
- Many algorithms easily handle non-linearity, extreme dimensionality [Han, Jentzen, and E, 2018]

Computational techniques

If not analytic, what about numerical?

- Maximum entropy for the heat equation was just a special case. We can do the same in general, using numerical techniques rather than solving by hand
- Physics inspired neural networks
  - Lagrangian neural networks
  - Wave equations and fluid mechanics [Cai et al, 2021]
    - Karniadakis group
  - Fourier neural operator [Li et al, 2021]
    - Anandkumar group
- Many algorithms easily handle non-linearity, extreme dimensionality [Han, Jentzen, and E, 2018]

A look to the future

Lots of interesting questions in pure and applied maths, data and computer science, etc

ARTIFICIAL INTELLIGENCE

Latest Neural Nets Solve World's Hardest Equations Faster Than Ever Before

20

Two new approaches allow deep neural networks to solve entire families of partial differential equations, making it easier to model complicated systems and to do so orders of magnitude faster.

From *Quanta Magazine*, https://www.quantamagazine.org/latest-neural-nets-solve-worlds-hardest-equations-faster-than-ever-before-20210419/, by Anil Ananthaswamy.

References

1. Edwin T Jaynes. Information theory and statistical mechanics. *Physical Review*, 106(4):620–630, 1957.
2. Peter A Markowich and Cédric Villani. On the trend to equilibrium for the Fokker-Planck equation: An interplay between physics and functional analysis. *Matemática Contemporânea (SBM)*, 19:1–29, 2000 (preprinted 1999).
3. Richard Haberman. *Elementary Applied Partial Differential Equations With Fourier Series And Boundary Value Problems*, 1987.
4. Shengze Cai et al. Physics-informed neural networks (PINNs) for fluid mechanics: A review. 2021. Preprint arXiv:2105.09506.
5. Zongyi Li et al. Fourier neural operator for parametric partial differential equations. 2021. Preprint arXiv:2010.08895.
6. Jiequn Han, Arnulf Jentzen, and Weinan E. Solving high-dimensional partial differential equations using deep learning. *PNAS*, 115(34):8505–8510, 2018.