THE UNIVERSITY OF UTAH

SCHOOL OF COMPUTING

AutoIP: A United Framework to Integrate Physics into Gaussian Processes

Da Long[1], Zheng Wang[1], Aditi S. Krishnapriyan[2 3], Robert M. Kirby[1], Shandian Zhe[1], Michael W. Mahoney[2 3 4]

[1] University of Utah [2] University of California, Berkeley [3] Lawrence Berkeley

[4] National Laboratory International Computer Science Institute

Presenter: Da Long

dl932@cs.utah.edu

Background

- Physical modeling:
  - Diffusion, heat, sound, fluid dynamics, etc.
- ML and data science method:
  - Learn target functions from observed data
  - Optimize data-dependent loss
- Physics-informed neural networks (PINN):
  - Fit the boundary/initial conditions and minimize a residual term to conform to the equation
  - Differentiation operators on the NN itself
  - Challenges:
    - Need massive collocation points
    - Equations need to be fully specified, e.g., unable to handle incomplete equations with latent sources
    - challenging in optimization, robustness, and uncertainty quantification

THE UNIVERSITY OF UTAH

SCHOOL OF COMPUTING

Main contributions

- Incorporating arbitrary differential equations into Gaussian processes
  - Flexible
    - Complete and incomplete equations.
    - Linear, nonlinear, spatial, temporal, spatio-temporal, etc.
  - Robust:
    - Works well with a small number of collocation points, even when the observations are noisy
    - Due to the nonparametric nature of GPs, our model is flexible enough to learn complex functions from data
    - Convenient to quantify the uncertainty

Model: Key idea

- Kernel differentiation
  - Differential equation is a combination of derivatives and functions
  - If f follows a GP, all of f’s derivatives follow GPs.
  - We can use kernel differentiation to construct a joint Gaussian prior distribution for f and all of f’s derivatives’ values

THE UNIVERSITY OF UTAH

Model

- Joint probability

$$f = [u; \hat{u}; \hat{u}_t; \hat{u}_{xx}; g]$$

at training inputs (u); at collocation inputs ($\hat{u}; \hat{u}_t; \hat{u}_{xx}$)

$$p(f, y, 0) = \underbrace{\mathcal{N}(f|0, \Sigma)}_{\text{prior over f}} \cdot \underbrace{\mathcal{N}(y|u, \beta^{-1}\mathbb{I})}_{\text{observation likelihood}} \cdot \underbrace{\mathcal{N}(0\,|\hat{u}_t - \nu\hat{u}_{xx} + \gamma\hat{u}\circ(\hat{u}\circ\hat{u} - 1) + g, v\mathbb{I})}_{\text{virtual likelihood to conform to ODEs/PDEs}}$$

g: latent source (if exists)

$$u \sim \mathcal{GP}(0, \mathcal{K}_u(\cdot,\cdot))$$
$$g \sim \mathcal{GP}(0, \mathcal{K}_g(\cdot,\cdot))$$

$$\text{cov}\left(u(\mathbf{z}_1), u(\mathbf{z}_2)\right) = \kappa_u(\mathbf{z}_1, \mathbf{z}_2),$$
$$\text{cov}\left(\partial_t u(\mathbf{z}_1), \partial_t u(\mathbf{z}_2)\right) = \frac{\partial^2 \kappa_u(\mathbf{z}_1, \mathbf{z}_2)}{\partial t_1 \partial t_2},$$
$$\text{cov}\left(\partial_x^2 u(\mathbf{z}_1), \partial_x^2 u(\mathbf{z}_2)\right) = \frac{\partial^4 \kappa_u(\mathbf{z}_1, \mathbf{z}_2)}{\partial x_1^2 \partial x_2^2},$$
$$\text{cov}\left(\partial_t u(\mathbf{z}_1), \partial_x^2 u(\mathbf{z}_2)\right) = \frac{\partial^3 \kappa_u(\mathbf{z}_1, \mathbf{z}_2)}{\partial t_1 \partial x_2^2},$$
$$\text{cov}\left(\partial_t u(\mathbf{z}_1), u(\mathbf{z}_2)\right) = \frac{\partial \kappa_u(\mathbf{z}_1, \mathbf{z}_2)}{\partial t_1},$$
$$\text{cov}\left(\partial_x^2 u(\mathbf{z}_1), u(\mathbf{z}_2)\right) = \frac{\partial^2 \kappa_u(\mathbf{z}_1, \mathbf{z}_2)}{\partial x_1^2},$$

For convenience, $\bar{u} = [u; \hat{u}]$

$$\Sigma = \begin{bmatrix} cov(\bar{u}, \bar{u}) & cov(\bar{u}, \hat{u}_t) & cov(\bar{u}, \hat{u}_{xx}) & \mathbf{0} \\ cov(\hat{u}_t, \bar{u}) & cov(\hat{u}_t, \hat{u}_t) & cov(\hat{u}_t, \hat{u}_{xx}) & \mathbf{0} \\ cov(\hat{u}_{xx}, \bar{u}) & cov(\hat{u}_{xx}, \hat{u}_t) & cov(\hat{u}_{xx}, \hat{u}_{xx}) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & cov(g, g) \end{bmatrix}$$

Inference

- Variational Inference
  - Whitening method to parameterize f: $f = \mathrm{A}\eta,\ \eta \sim \mathcal{N}(0, \mathbb{I}), \Sigma = \mathrm{A}\mathrm{A}^T$
    - Address the strong coupling of f and the kernel parameters

$$p(f, y, 0) = \mathcal{N}(f|0, \Sigma) \cdot \mathcal{N}(y|u, \beta^{-1}\mathbb{I}) \cdot \mathcal{N}(0|\hat{u}_t - \nu\hat{u}_{xx} + \gamma\hat{u} \circ (\hat{u} \circ \hat{u} - 1) + g, v\mathbb{I})$$

$$\downarrow$$

$$p(f, y, 0) = \mathcal{N}(\eta|0, \mathbb{I}) \cdot \mathcal{N}(y|A\eta) \cdot \mathcal{N}(0|A\eta)$$

  - A multivariate normal $q(\eta)$ as the variational posterior for η, $q(\eta) = \mathcal{N}(\eta|u, LL^T)$

Lower bound: $\mathcal{L} = -KL(q(\eta)||\mathcal{N}(0, \mathbb{I})) + \mathbb{E}_q[\log(p(y|A\eta))] + \mathbb{E}_q[log(p(0|A\eta))]$

- KL : Kullback-Leibler divergence
- Maximize the lower bound to estimate the $q(\eta)$ and other parameters, e.g., kernel parameters, noise variances

Model

- Insight
  - With Gaussian variational approximation, the posterior process for f is still a GP, and maintains the link between the function and its derivatives in terms of kernel differentiation.

1. predictive distribution is still a GP: $p(h|D)=\int p(h|f)\,p(f|D)\,df \approx \int p(h|f)\,q(f|m_f,V_f)\,df = N(h|m_h, cov(h,h)-cov(h,f)\cdot B\cdot cov(f,h)) \quad h=(u,u_t,u_{xx},...)$

$cov(\ )\ is\ obtained\ from\ the\ kernel\ K_u\ and\ its\ differentiation,\ see\ the\ last\ slide$

$m_h = cov(h,f)\Sigma^{-1}m_f \qquad B=\Sigma^{-1}-\Sigma^{-1}V_f\Sigma^{-1}$

2. new GP defines a new kernel for u (function): $cov_{new}(u_1,u_2)=k_u(z_1,z_2)-cov(u_1,f)\cdot B\cdot cov(u_2,f),\ u_1=u(z_1),\ u_2=u(z_2),\ z_1=(x_1,t_1), z_2=(x_2,t_2)$

3. kernel differentiation on the new kernel gives the same covariance as in the predictive distribution:

$$\frac{\partial cov_{new}(u_1,u_2)}{\partial t_2}=\frac{\partial k_u(z_1,z_2)}{\partial t_2}-cov(u_1,f)\cdot B\cdot\frac{\partial cov(u_2,f)}{\partial t_2}=cov(u_1,\partial_t u_2)-cov(u_1,f)\cdot B\cdot cov(f,\partial_t u_2)$$

$$=cov(u_1,\partial_t u_2)-cov(u_1,f)\cdot B\cdot cov(f,\partial_t,u_2),\ from\ predictive\ distribution,\ link\ preserves$$

Experiments

- Nonlinear pendulum and diffusion-reaction system, with complete and incomplete(with unknown latent source) equations
- Real-world datasets
  - Metal pollution in Swiss Jura
  - Motion capture

Experiment - Nonlinear pendulum

First row: no damping force

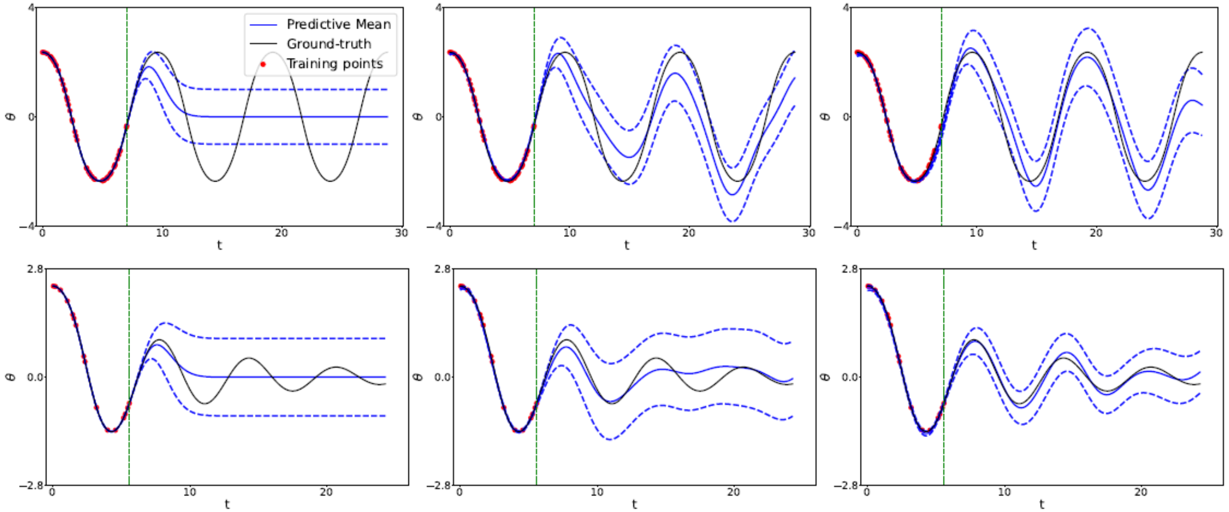
Complete equation $\frac{\mathrm{d}^2\theta}{\mathrm{d}t^2} + \sin(\theta) = 0,$ AutoIP-C

Incomplete equation $\frac{\mathrm{d}^2\theta}{\mathrm{d}t^2} + g(t) = 0.$ AutoIP-I

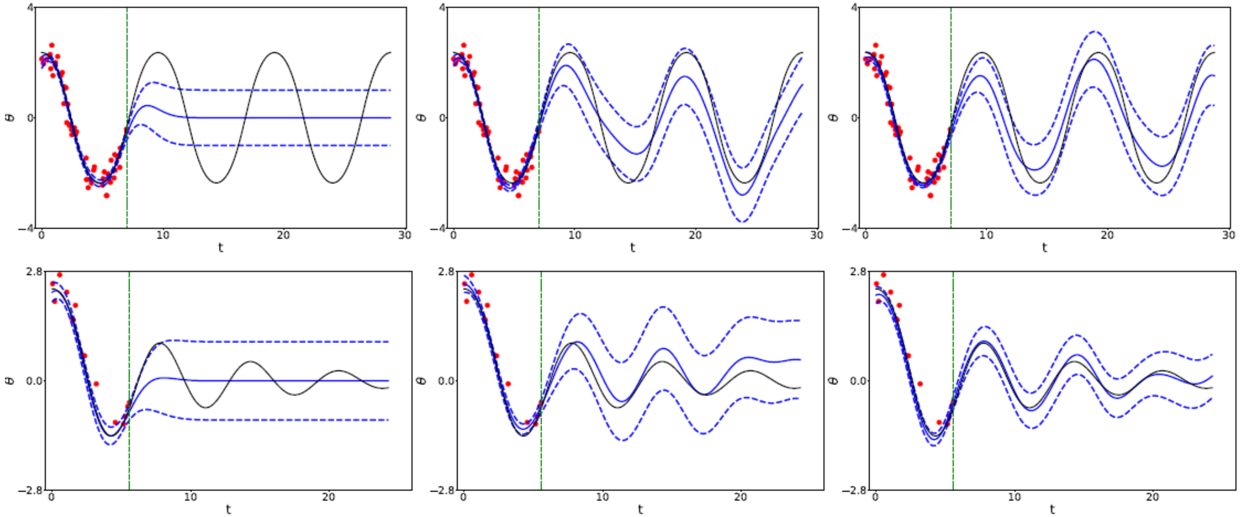
Second row: with damping force

Complete equation $\frac{\mathrm{d}^2\theta}{\mathrm{d}t^2} + \sin(\theta) + b\frac{\mathrm{d}\theta}{\mathrm{d}t} = 0,$ b is unknown

Incomplete equation $\frac{\mathrm{d}^2\theta}{\mathrm{d}t^2} + g(t) = 0.$



(a) GPR (b) AutoIP-I (c) AutoIP-C



(a) GPR (b) AutoIP-I (c) AutoIP-C

Experiment - Nonlinear pendulum

- Comparing with GPR, PINNs

AutoIP-I: incomplete equation, AutoIP-C: complete equation

Our model uses 20 collocation points for both settings

Method	No damping/Exact training	No damping/Noisy training	Damping/Exact training	Damping/Noisy training
PINN-5 (20)	1.955 ± 0.214	1.895 ± 0.261	0.310 ± 0.019	0.310 ± 0.050
PINN-10 (20)	2.122 ± 0.179	1.824 ± 0.231	0.290 ± 0.018	0.342 ± 0.020
PINN-50 (20)	2.238 ± 0.541	1.927 ± 0.250	0.297 ± 0.044	0.361 ± 0.017
PINN-100 (20)	2.042 ± 0.273	2.407 ± 0.353	0.320 ± 0.074	0.384 ± 0.066
PINN-5 (10K)	1.479 ± 0.115	1.783 ± 0.297	0.110 ± 0.015	0.248 ± 0.037
PINN-10 (10k)	1.852 ± 0.320	1.548 ± 0.141	0.049 ± 0.023	0.194 ± 0.044
PINN-50 (10k)	1.367 ± 0.575	1.658 ± 0.074	$\mathbf{0.00007 \pm 0.00001}$	0.157 ± 0.051
PINN-100 (10k)	1.862 ± 0.584	1.993 ± 0.357	$\mathbf{0.00007 \pm 0.00002}$	0.186 ± 0.045
AutoIP-I	0.585 ± 0.017	0.691 ± 0.030	0.212 ± 0.014	0.268 ± 0.013
AutoIP-C	$\mathbf{0.416 \pm 0.050}$	$\mathbf{0.488 \pm 0.036}$	0.096 ± 0.004	$\mathbf{0.133 \pm 0.010}$

No damping	RMSE	MNLL
GPR	1.354 ± 0.005	1.97 ± 0.015
AutoIP-I	0.585 ± 0.017	1.02 ± 0.013
AutoIP-C	$\mathbf{0.416 \pm 0.050}$	$\mathbf{0.892 \pm 0.032}$
With damping		
GPR	0.262 ± 0.0003	0.744 ± 0.008
AutoIP-I	0.212 ± 0.014	0.678 ± 0.02
AutoIP-C	$\mathbf{0.096 \pm 0.0035}$	$\mathbf{0.155 \pm 0.01}$

(a) Exact training data

No damping	RMSE	MNLL
GPR	1.44 ± 0.017	2.242 ± 0.055
AutoIP-I	0.691 ± 0.030	1.206 ± 0.024
AutoIP-C	$\mathbf{0.488 \pm 0.036}$	$\mathbf{1.061 \pm 0.028}$
With damping		
GPR	0.381 ± 0.018	1.07 ± 0.029
AutoIP-I	0.268 ± 0.013	0.937 ± 0.011
AutoIP-C	$\mathbf{0.133 \pm 0.010}$	$\mathbf{0.428 \pm 0.017}$

(b) Noisy training data

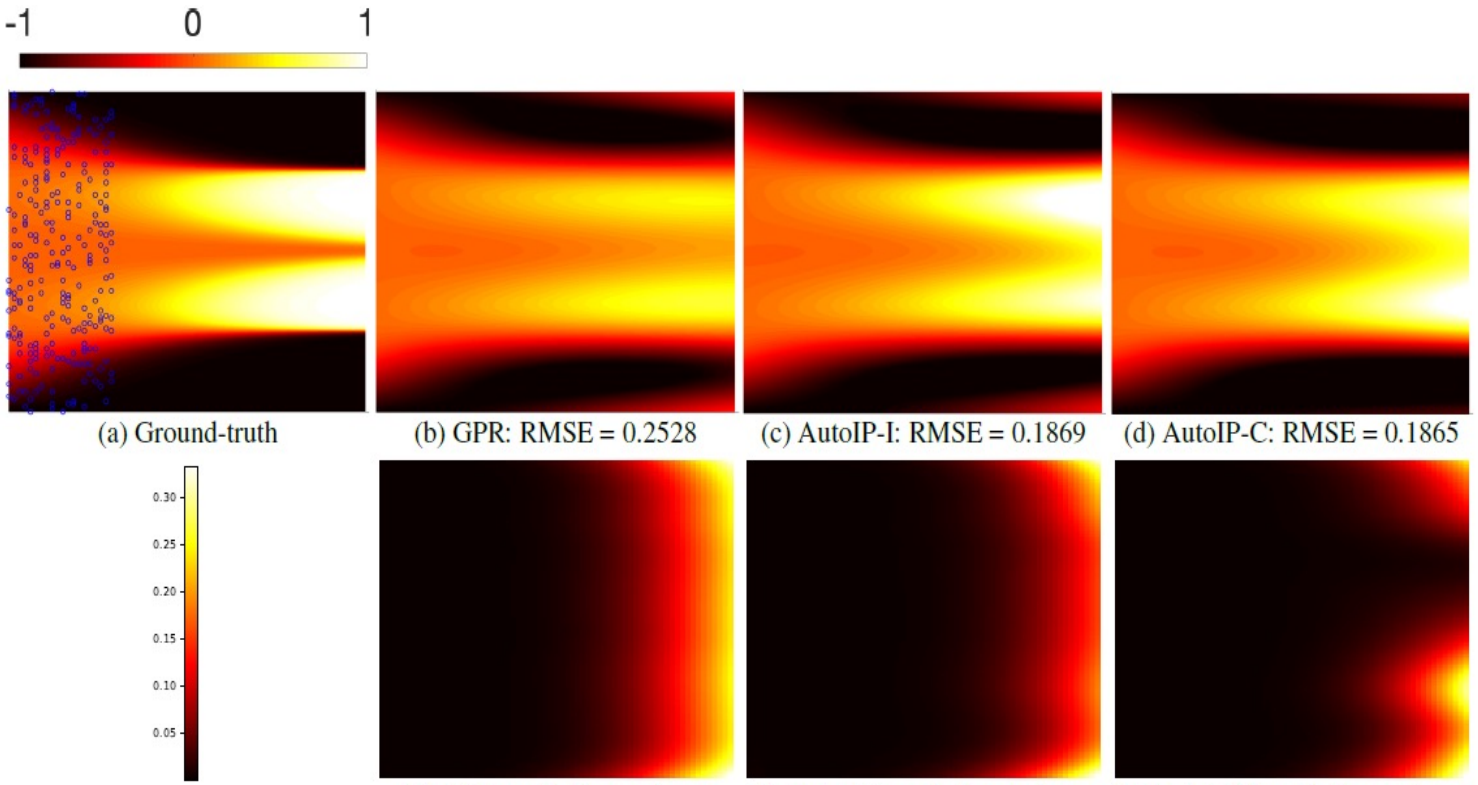
Experiment – diffusion-reaction system

Complete equation, AutoIP-I $$\frac{\partial u}{\partial t} - 0.0001\frac{\partial^2 u}{\partial x^2} + 5u^3 - 5u = 0,$$

Incomplete equation, AutoIP-C $$\frac{\partial u}{\partial t} - 0.0001\frac{\partial^2 u}{\partial x^2} + g(x,t) = 0,$$

Comparing with GPR, PINNs

Our model uses 100 collocation points



(a) Ground-truth (b) GPR: RMSE = 0.2528 (c) AutoIP-I: RMSE = 0.1869 (d) AutoIP-C: RMSE = 0.1865

(e) GPR predictive variance (f) AutoIP-I predictive variance (g) AutoIP-C predictive variance

Complete equation result

GPR	AutoIP-I	AutoIP-C	PINN (100)	PINN (10K)
0.2528	0.1869	0.1865	0.4388	**0.0169**

Experiments – real-world datasets

- Motion capture

$$\frac{\partial u}{\partial t} + b \cdot u(t) - c = g(t),$$ b and c are unknown; g(t) is the latent force

Method	Joint 1	Joint 50
GPR	1.727 ± 0.026	0.257 ± 0.007
LFM	1.671 ± 0.016	0.257 ± 0.006
AutoIP-T	1.511 ± 0.007	0.224 ± 0.006
AutoIP-H	1.489 ± 0.03	0.225 ± 0.005
AutoIP-W	$\mathbf{1.103 \pm 0.027}$	$\mathbf{0.215 \pm 0.009}$

(a) RMSE

Method	Joint 1	Joint 50
GPR	1.368 ± 0.020	3.431 ± 0.242
LFM	1.721 ± 0.020	N/A
AutoIP-T	1.138 ± 0.024	2.615 ± 0.149
AutoIP-H	1.208 ± 0.081	2.664 ± 0.154
AutoIP-W	$\mathbf{1.121 \pm 0.084}$	$\mathbf{2.495 \pm 0.111}$

(b) NMLL

AutoIP-T: uses the training inputs as the collocation points
AutoIP-H: uses 200 random collocation points in the training region only
AutoIP-W: uses 200 random collocation points across the whole-time span of the trajectory

THE UNIVERSITY OF UTAH

SCHOOL OF COMPUTING

Experiments – real-world datasets

- Metal pollution

$$\frac{\partial^2 u}{\partial {x_1}^2} + \frac{\partial^2 u}{\partial {x_2}^2} = g(x_1, x_2)$$

	GPR	LFM	AutoIP
Task 1	0.299 ± 0.009	0.384 ± 0.010	$\mathbf{0.284 \pm 0.011}$
Task 2	0.304 ± 0.012	0.381 ± 0.011	$\mathbf{0.284 \pm 0.008}$
Task 3	0.232 ± 0.009	0.358 ± 0.005	$\mathbf{0.224 \pm 0.006}$
Task 4	0.261 ± 0.005	0.296 ± 0.005	$\mathbf{0.247 \pm 0.004}$

(a) RMSE

	GPR	LFM	AutoIP
Task 1	1.16 ± 0.064	1.36 ± 0.058	$\mathbf{1.10 \pm 0.069}$
Task 2	1.274 ± 0.093	1.471 ± 0.157	$\mathbf{1.219 \pm 0.129}$
Task 3	0.979 ± 0.058	1.31 ± 0.044	$\mathbf{0.849 \pm 0.067}$
Task 4	1.383 ± 0.098	1.496 ± 0.097	$\mathbf{1.303 \pm 0.091}$

(b) NMLL

Our model uses 50 points for training; the collocation points are exactly the training inputs.

Task 1, predict Zn with the location, and Cd, Ni concentration
Task 2, predict Zn with the location, and Co, Ni, Cr concentration
Task 3, predict Ni with the location, and Cr concentration
Task 4, predict Cr with the location, and Co concentration

THE UNIVERSITY OF UTAH

SCHOOL OF COMPUTING

Welcome to our poster!

dl932@cs.utah.edu