m4DL

WELCOME

Schedule Day One

The Schedule for the workshop is as follows:

Thursday 21st April

12:00-13:00	Arrival and lunch
13:00-13:30	Opening and intro to the Maths4DL programme
13:30-14:00	Talk: Andrew Fitzgibbon (Graphcore)
14:00-14:20	Talk: Kweku Abraham (Cambridge)
14:20-14:40	Talk: Riccardo Barbano (UCL)
14:40-15:00	Talk: Teo Deveney (Bath)
15:00-15:30	Coffee
15:30-16:00	Talk: Gitta Kutyniok (LMU)
16:00-16:20	Talk: Subhadip Mukherjee (Cambridge)
16:20-16:50	Talk: Kwinten van Weverberg (Met Office)
16:50-17:10	Talk: Lisa Kreusser (Bath)
17:15	Talk: Reception and posters

Dinner 7.30pm Aqua Restaurant

Opening the Black Box

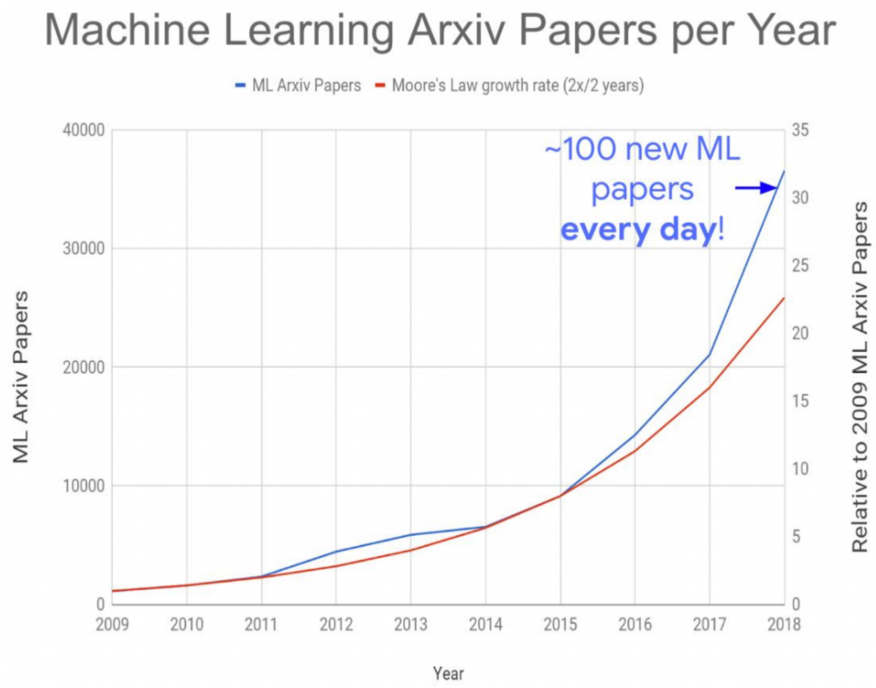
An EPSRC programme grant project on the

Mathematics for Deep Learning

Chris Budd (Bath)

Motivation

- Recent years have seen a huge explosion of deep learning
- Machine Learning (ML), in particular Deep Learning, has had a transformative effect on all areas of our life
- Applications in all disciplines, e.g.
    - (Bio-) Medical Sciences
    - Computer Vision
    - Physical sciences
    - Environmental sciences
    - Scientific computing
    - Speech Recognition
    - Gaming
    - Finance



But there are currently problems associated with the rapid growth

- Success of deep learning not understood
- Results are mysterious => black box lacks explainability
- Problems with unstructured data
- Problems with small data
- Much larger scale problematic

There currently is no complete rigorous mathematical theory for the setup, training and application performance of deep neural networks.

There is an urgent need to address this to make progress

Aim and Objectives of Maths4DL

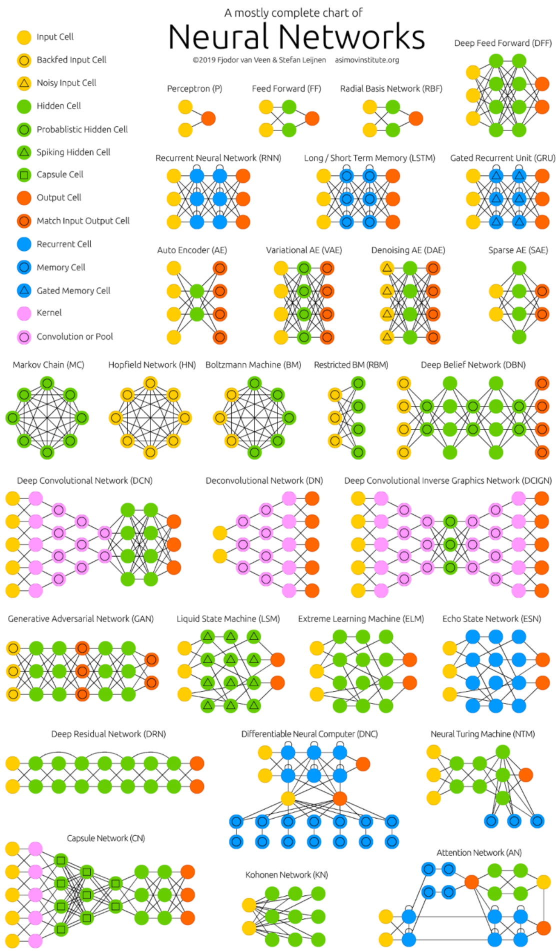
Aim: **Put deep learning onto a firm mathematical basis**

Combine theory, modelling, data, computation
to unlock the **next generation of deep learning**

Objectives

1. **Explainable AI** Develop a Fundamental Theory of Deep Learning
2. **Trustworthy AI** Determine the Limits of Deep Learning Technology
3. **New avenues** Bring Deep Learning to New Horizons

Explainable AI ?



Deep learning zoo - what works best and why?

- A myriad of proposed network architectures
- Various proposed loss functions

How to make an informed choice?

P. Liu, C. Li, C.-B Schönlieb, GANReDL: Medical Image enhancement using a generative adversarial network with real-order derivative induced loss functions, MICCAI 2019

Trustworthy AI?

Vulnerability of deep learning

Identified as a 45mph speed sign

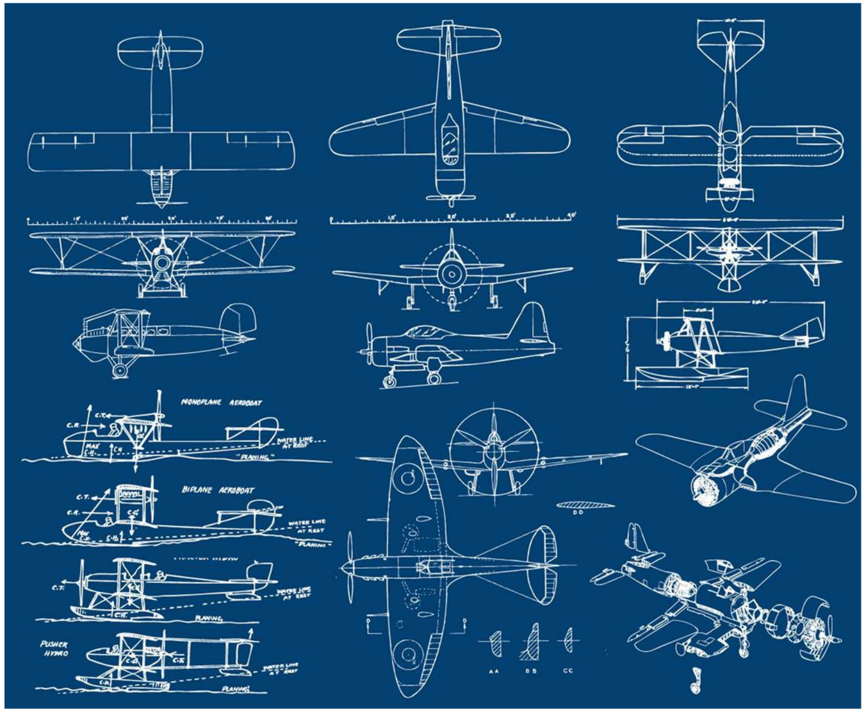
Safety danger: visually indistinguishable perturbed examples can break the network performance

How to render provably stable neural networks while keeping the performance?

Etmann, C., Lunz, S., Maass, P., & Schönlieb, C. B. On the Connection Between Adversarial Robustness and Saliency Map Interpretability, ICML 2019.

New Avenues

Does deep learning respect the physics?



Would you fly on an airplane which was built by looking at a lot of examples of existing airplanes?

Would you believe in weather predictions that are solely built on empirical observations?

How to introduce physics into deep learning solutions?

McRae, A. T., Cotter, C. J., & Budd, C. J. (2018). Optimal-transport--based mesh adaptivity on the plane and sphere using finite elements. *SIAM Journal on Scientific Computing*, *40*(2), A1121-A1148.

Hauptmann, A., Lucka, F., Betcke, M., Huynh, N., Adler, J., Cox, B., ... & Arridge, S. (2018). Model-based learning for accelerated, limited-view 3-d photoacoustic tomography. *IEEE transactions on medical imaging*, *37*(6), 1382-1393.

The first Aim of Maths4DL

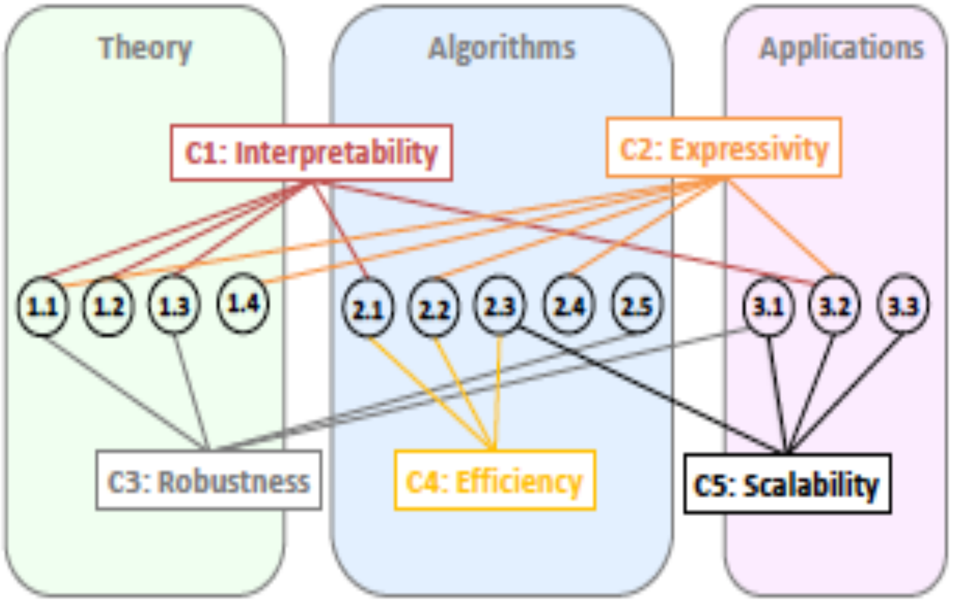
To put DL onto a firm mathematical basis, by looking at DL through the magnifying glass of continuum models.

The second Aim of Maths4DL

To provide a new physics-driven DL paradigm. We believe this to be crucial to unlock the next generation of DL that realises its full potentials within physical domains.

Work Packages

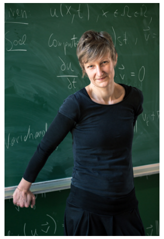
WP1.1	DL as Optimal Control
WP1.2	Physics Based Neural Networks
WP1.3	Uncertainty Quantification for DL
WP1.4	DL on Manifolds
WP2.1	Training of Continuum Models
WP2.2	Novel Metric Based Training Methods
WP2.3	Large Nonconvex Optimisation
WP2.4	Imaging Off-the-Grid
WP2.5	Learning Approximate Models
WP3.1	Environmental Problems
WP3.2	DL in Computational Physics
WP3.3	Inverse Problems in Imaging

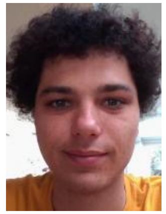


The Team so far .....

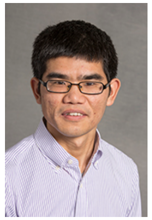


- **Bath**: Chris Budd, Matthias Ehrhardt, Teo Deveney, Lisa Kreusser, Tatiana Bubba, Simone Appella, Margaret Duff





- **Cambridge**: Carola-Bibiane Schönlieb, Richard Nickl, Kweku Abraham, Yury Korolev, GeorgiosBatzolis, Sören Dittmer,Tamara Grossmann, Subhadip Mukherjee, Jan Stanczuk, Junqi Tang.


- **UCL**: Simon Arridge, Bangti Jin, Rob Tovey, Andreas Hauptmann
- **Project Manager**: Helena Lake

Industrial Partners

BT

Unilever

THE ALAN TURING INSTITUTE

GE Healthcare

Microsoft Research

AstraZeneca

gsk do more feel better live longer

NPL National Physical Laboratory

Canon
CANON MEDICAL RESEARCH EUROPE LTD.

AVIVA

Met Office

DASSAULT SYSTEMES

SIEMENS Healthineers

Opportunities to get involved

Planned activities include:

- Deep learning **reading group** every Thursday (Tatiana Bubba & Yury Korolev)
- July 2022: Sponsored workshop with Matt Thorpe and the ICMS:
  Analytic and Geometric Approaches to Machine Learning
- September 2022: Workshop on **Deep learning and environmental problems**
- Late Autumn 2022: Workshop on **Deep learning for industry**
- June 2023: International conference on **Deep learning in computational physics**
- Partnership with the Millennium Mathematics Project on Public engagement

Physics Inspired Deep Learning

Simon Arridge, Bangti Jin (University College London)

Goal of Math4DL: To establish a mathematical, computational and statistical framework

Combining knowledge / model with deep learning

- Classical techniques: specialized, with provable guarantees

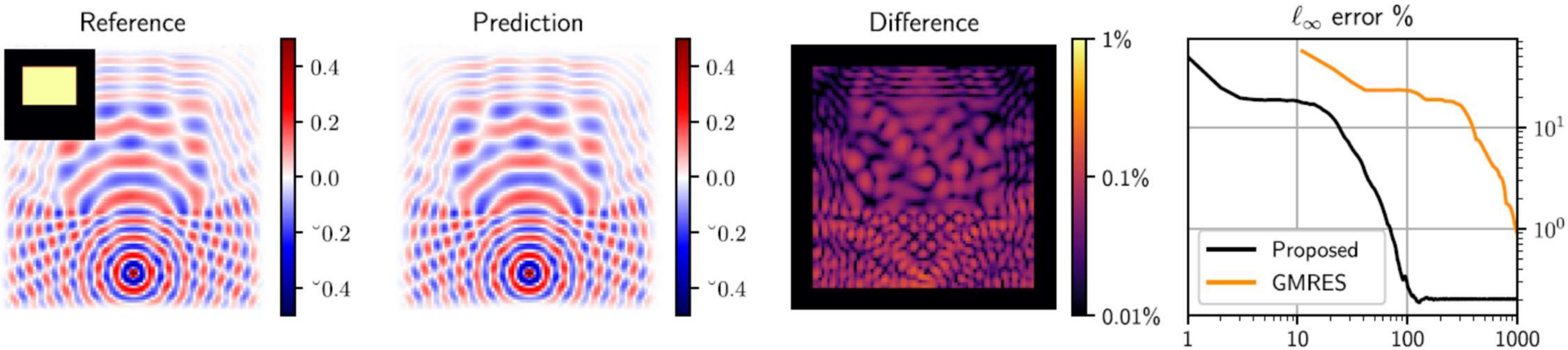
  fast iterative solvers, physical model, variational regularization, optimal transport …

- **Learned iterative solver for physical problems**
- $\boldsymbol{u_{k+1} = u_k + f_\theta(u_k, e_k)},$ $\boldsymbol{e_k = F(u_k) - g},$ / $\boldsymbol{e_k = F'(u_k)^*(F(u_k) - g)}$
- Very effective for challenging tasks in physical simulation, imaging, uncertainty quantification ….

2D wave propagation

$$\left[\nabla^2 + \left(\frac{\omega}{c(r)}\right)^2\right] u(r) = \rho(r)$$

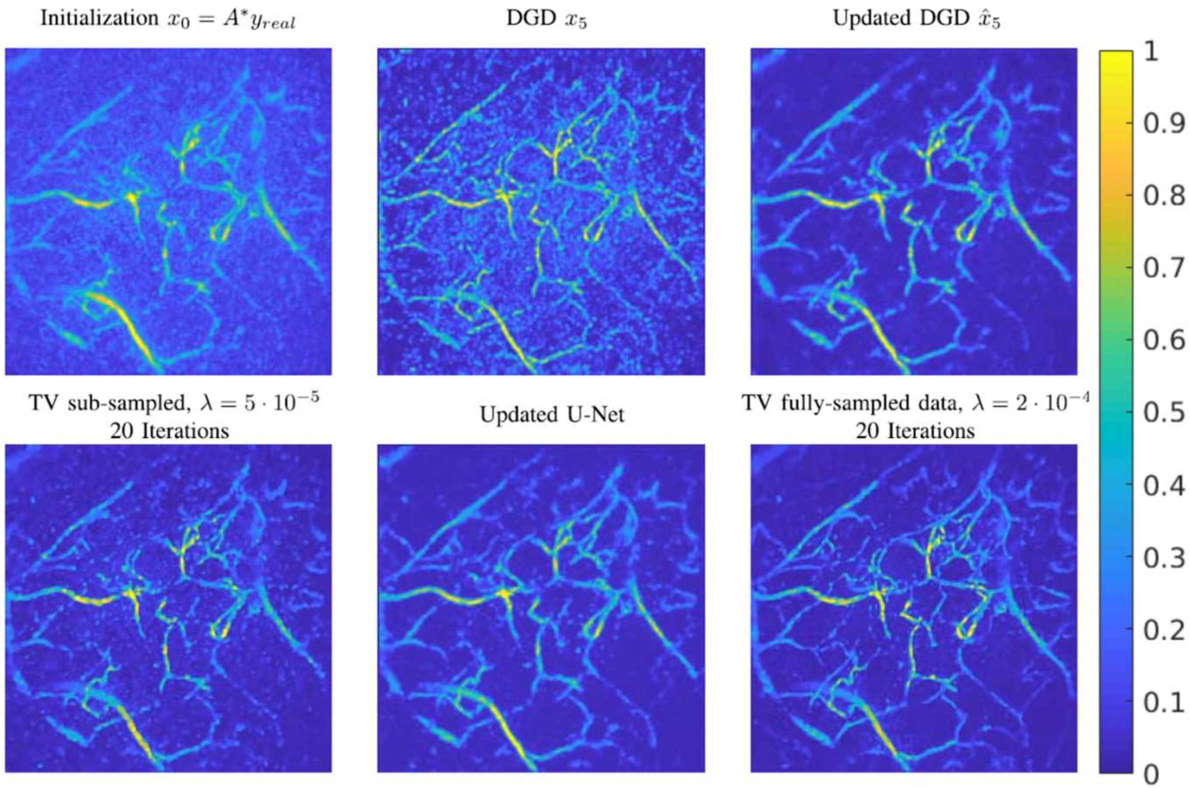
$$\text{s.t.} \lim_{|r|\to\infty} |r|^{\frac{n-1}{2}} \left(\frac{\partial}{\partial |r|} - i\frac{\omega}{c_0}\right) u(r) = 0 \ .$$



A. Stanziola, S. Arridge, B. Cox, B. Treeby. J. Comput. Phys. 2021

Unsupervised Helmoltz solvers beat GMRES !

3D Photoacoustic tomography



A. Hauptmann, F. Lucka, …, S. Arridge. IEEE Trans. Med. Imag. 2018

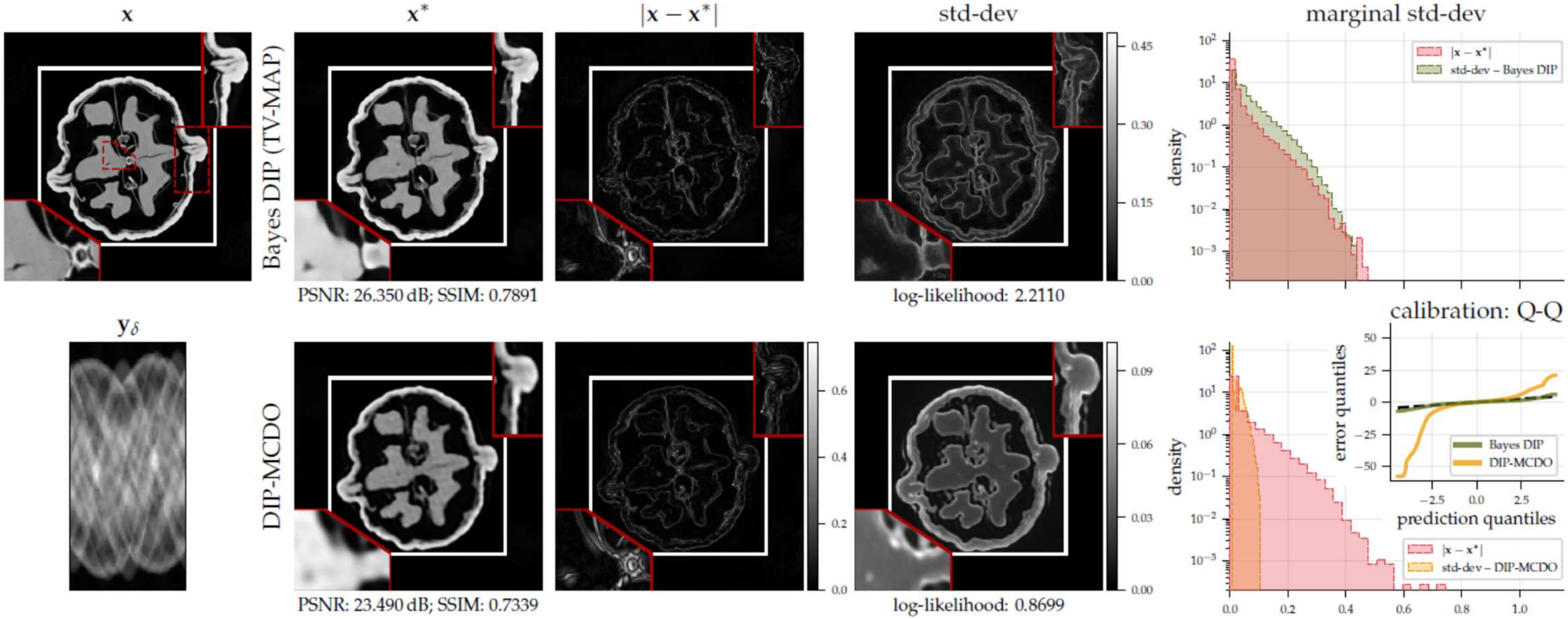
$$(\partial_{tt} - c_0^2 \Delta) p(r,t) = 0$$

$$p\Big|_{t=0} = x, \qquad \partial_t p\Big|_{t=0} = 0$$

$$y = Mp\Big|_{\partial\Omega \times (0,T)}$$

- Supervised training with ellipses
- Test with real palm data
- **beat standard TV with full data**

Uncertainty quantification



J Antorán, R Barbano, J Leuschner, JM Hernández-Lobato, B Jin. arXiv:2203.00479

A probabilistic formulation of **deep image prior** on Walnut data, with **calibrated uncertainty**

What is next ?

- New physics inspired approaches for scientific problems with better interpretability
- Theoretical studies, especially convergence issues
    - Algorithmic convergence
    - Error estimates via stability analysis
    - Generalization errors
    - Statistical estimation
    - …

Bayesian inverse problems, UQ and DL

Richard Nickl

University of Cambridge (UK)

Bath, April 21 2022

UNIVERSITY OF CAMBRIDGE

European Research Council

Established by the European Commission

Statistical inverse regression models

Consider statistical observations arising as random vectors $(Y_i, X_i)_{i=1}^N$, where the X_i represent a discretisation of a d-dimensional manifold $\mathcal{X}$, and with

$$Y_i = \mathscr{G}_\theta(X_i) + \varepsilon_i,\ \varepsilon_i \sim^{i.i.d.} \mathcal{N}(0,1),$$

where a collection of 'regression' fields $\{\mathscr{G}_\theta : \theta \in \Theta\}$ is indexed by the high-dimensional parameter $\theta \in \Theta$.

Statistical inverse regression models

Consider statistical observations arising as random vectors $(Y_i, X_i)_{i=1}^N$, where the X_i represent a discretisation of a d-dimensional manifold $\mathcal{X}$, and with

$$Y_i = \mathscr{G}_\theta(X_i) + \varepsilon_i,\ \varepsilon_i \sim^{i.i.d.} \mathcal{N}(0,1),$$

where a collection of 'regression' fields $\{\mathscr{G}_\theta : \theta \in \Theta\}$ is indexed by the high-dimensional parameter $\theta \in \Theta$. In inverse problems

$$\mathscr{G} : \Theta \to L^2(\mathcal{X}),\quad \theta \mapsto \mathscr{G}_\theta,$$

is the possibly non-linear solution 'forward' map of a PDE.

Statistical inverse regression models

Consider statistical observations arising as random vectors $(Y_i, X_i)_{i=1}^N$, where the X_i represent a discretisation of a d-dimensional manifold $\mathcal{X}$, and with

$$Y_i = \mathscr{G}_\theta(X_i) + \varepsilon_i, \ \varepsilon_i \sim^{i.i.d.} \mathcal{N}(0,1),$$

where a collection of 'regression' fields $\{\mathscr{G}_\theta : \theta \in \Theta\}$ is indexed by the high-dimensional parameter $\theta \in \Theta$. In inverse problems

$$\mathscr{G} : \Theta \to L^2(\mathcal{X}), \quad \theta \mapsto \mathscr{G}_\theta,$$

is the possibly non-linear solution 'forward' map of a PDE.

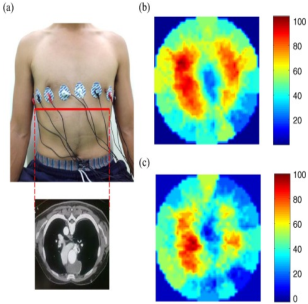
Example: Suppose $\mathscr{G}_\theta = u_\theta$ is the solution to

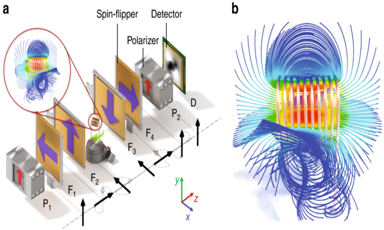
$$\begin{aligned} \mathscr{D}u + \theta u &= 0 \ \text{ on } \mathcal{X}, \\ u &= g \ \text{ on } \partial\mathcal{X}, \end{aligned}$$

with $\mathscr{D}$ a given linear partial differential op. (e.g., Laplacian, or geodesic vector field).

Examples for $\mathscr{G}$ from partial differential equations (PDEs)

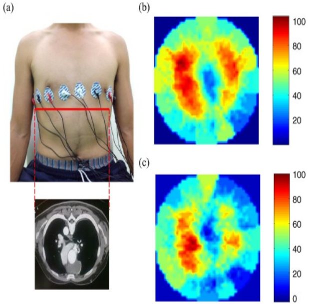
- Elliptic PDEs: Electric impedance tomography (Caldéron problem), groundwater flow & oil reservoir analysis, photoacoustics, etc..

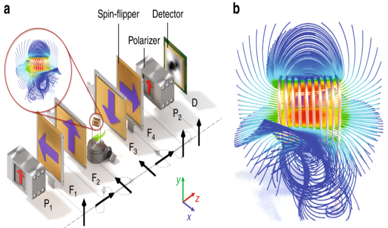




Examples for $\mathscr{G}$ from partial differential equations (PDEs)

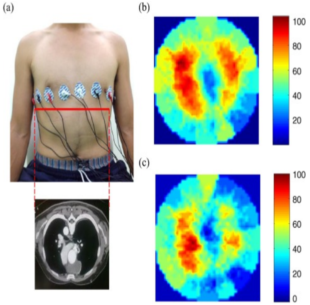
- Elliptic PDEs: Electric impedance tomography (Caldéron problem), groundwater flow & oil reservoir analysis, photoacoustics, etc..
- Transport PDEs: X-ray transforms in (non-linear) tomography

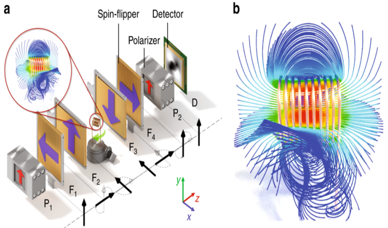




Examples for $\mathscr{G}$ from partial differential equations (PDEs)

- Elliptic PDEs: Electric impedance tomography (Caldéron problem), groundwater flow & oil reservoir analysis, photoacoustics, etc..
- Transport PDEs: X-ray transforms in (non-linear) tomography
- Time evolution equations: wave, heat, Euler and Navier-Stokes equations





Bayesian Inverse Problems (Stuart (2010))

The ‘log-likelihood’ or ‘least squares fit’ function is

$$\ell_N(\theta) = -\frac{1}{2}\sum_{i=1}^{N} \| Y_i - \mathscr{G}_\theta(X_i)\|_V^2, \quad \theta \in \Theta.$$

Note $-\ell_N$ is non-convex when $\mathscr{G}$ is non-linear.

Bayesian Inverse Problems (Stuart (2010))

The 'log-likelihood' or 'least squares fit' function is

$$\ell_N(\theta) = -\frac{1}{2}\sum_{i=1}^{N} \|Y_i - \mathscr{G}_\theta(X_i)\|_V^2, \quad \theta \in \Theta.$$

Note $-\ell_N$ is non-convex when $\mathscr{G}$ is non-linear. There is need for regularisation of the complexity of θ, which can be modelled by a Bayesian approach.

Let Π be prior probability on Θ. The posterior distribution $\Pi(\cdot|(Y_i, X_i)_{i=1}^N)$ on Θ then equals

$$d\Pi(\theta|(Y_i, X_i)_{i=1}^N) \propto e^{\ell_N(\theta)} d\Pi(\theta) \quad \theta \in \Theta.$$

Bayesian Inverse Problems (Stuart (2010))

The 'log-likelihood' or 'least squares fit' function is

$$\ell_N(\theta) = -\frac{1}{2}\sum_{i=1}^{N} \|Y_i - \mathscr{G}_\theta(X_i)\|_V^2, \quad \theta \in \Theta.$$

Note $-\ell_N$ is non-convex when $\mathscr{G}$ is non-linear. There is need for regularisation of the complexity of θ, which can be modelled by a Bayesian approach.

Let Π be prior probability on Θ. The posterior distribution $\Pi(\cdot|(Y_i, X_i)_{i=1}^N)$ on Θ then equals

$$d\Pi(\theta|(Y_i, X_i)_{i=1}^N) \propto e^{\ell_N(\theta)} d\Pi(\theta) \quad \theta \in \Theta.$$

Avoiding optimisation methods one can estimate θ by the posterior mean

$$E^\Pi[\theta|(Y_i, X_i)_{i=1}^N)] = \int_\Theta \theta d\Pi(\theta|(Y_i, X_i)_{i=1}^N).$$

Can we consider priors that arise with neural network architectures?

Computation: gradient based MCMC

Let us single out one popular MCMC method for illustration:

Unadjusted discretised Langevin algorithm

The log-posterior density on a high-dimensional discretisation space $\Theta \supset \mathbb{R}^D$ is

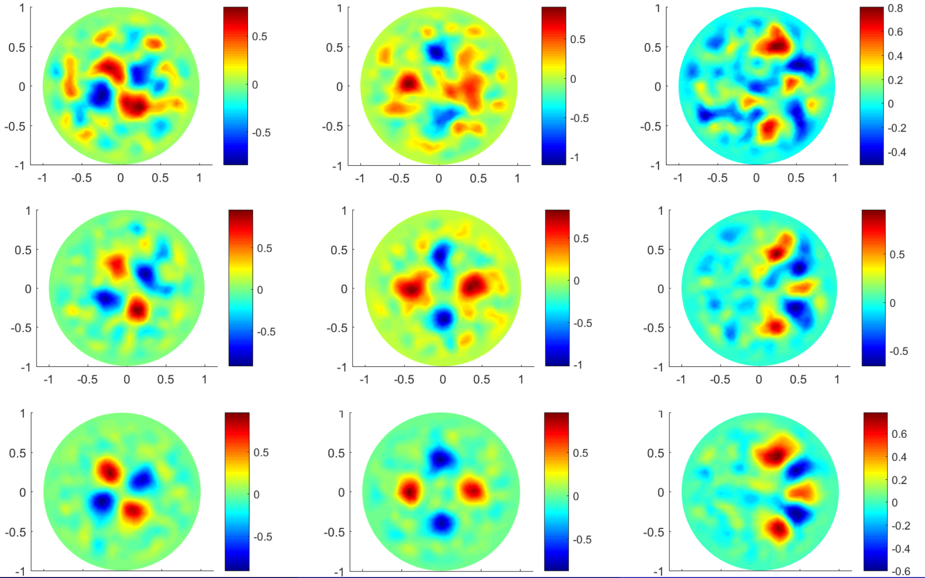
$$p_N(\theta) := \log \pi_N(\theta|(Y_i, X_i)_{i=1}^N), \quad \theta \in \mathbb{R}^D.$$

Then fix $\delta > 0$ and initialise at ϑ_0. For $k \geq 0$ and $\xi_k \sim^{iid} \mathcal{N}(0, I)$ in $\mathbb{R}^D$, do:

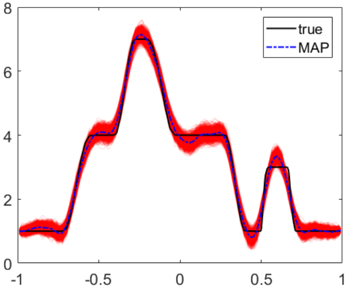
$$\vartheta_{k+1} = \vartheta_k + \delta \nabla p_N(\vartheta_k) + \sqrt{2\delta}\xi_k.$$

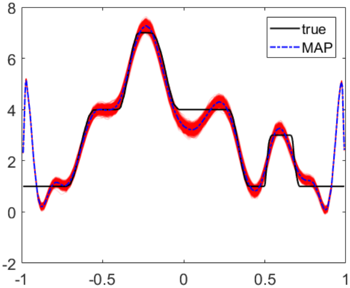
Bayesian inversion with Gaussian process priors in action

Posterior mean fields for $N = 200, 400, 800$ and $N_s = 100000$ MCMC iterations



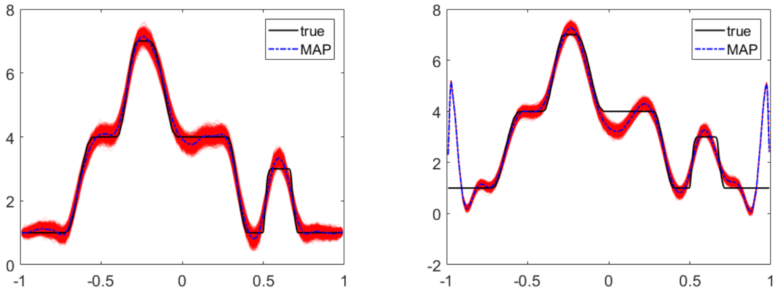
Algorithmic guarantees: Uncertainty quantification





Do Bayesian methods provide valid frequentist error bars?

Algorithmic guarantees: Uncertainty quantification



Do Bayesian methods provide valid frequentist error bars?

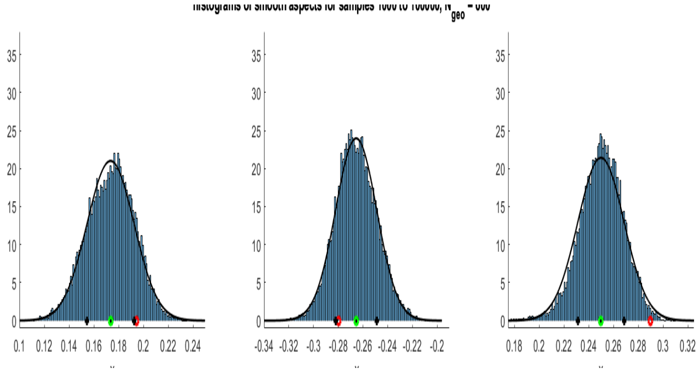
Confident credibility: If $R_N \equiv R_\alpha((Y_i, X_i)_{i=1}^N)$ are posterior quantiles for some norm, and $\bar{\theta}$ is the posterior mean or mode, do we have

$$P_{\theta_0}^N(\theta_0 \in \{\bar{\theta} - R_N, \bar{\theta} + R_N\}) \approx 1 - \alpha, \quad \text{as} \quad N \to \infty?$$

Recent progress in the field shows that this can be true: Monard, Nickl, Paternain (2021a, b). **Does this extend to Deep Learning architectures ?**

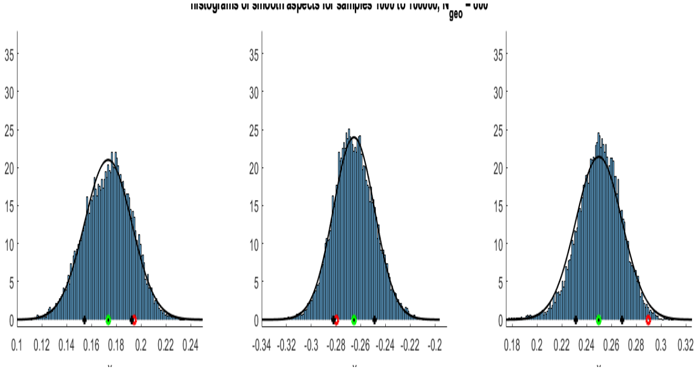
Numerical illustration

• Numerical MCMC plots of posterior draws of $\langle \theta, \psi \rangle_{L^2(\mathcal{X})} | (Y_i, X_i)_{i=1}^N$ around the posterior mean (green). The true value is marked in red, a Gaussian superimposed.



Numerical illustration

• Numerical MCMC plots of posterior draws of $\langle \theta, \psi \rangle_{L^2(\mathcal{X})} | (Y_i, X_i)_{i=1}^N$ around the posterior mean (green). The true value is marked in red, a Gaussian superimposed.



• These findings illustrate that the non-Gaussian posterior measures arising in non-linear inverse problems actually may produce Gaussian statistics for moderate $N = 600$. In particular this illustrates that uncertainty quantification based on posterior 'credible sets' often works in practice.

References

F. Monard, R. Nickl, G.P. Paternain, Consistent inversion of noisy non-Abelian X-ray transforms, *Comm. Pure Appl. Math.* (2021)

R. Nickl, Bernstein-von Mises theorems for statistical inverse problems I: Schrödinger equation, *J. Eur. Math. Soc. (JEMS)* 22 (2020)

F. Monard, R. Nickl, G.P. Paternain, Statistical guarantees for Bayesian uncertainty quantification in non-linear inverse problems with Gaussian process priors, *Ann. Statist.* (2021)

R. Nickl, S. Wang, On polynomial-time computation of high-dimensional posterior measures by Langevin-type algorithms, *J. Eur. Math. Soc. (JEMS)*, (2022)