Data-driven approaches to inverse problems

Carola-Bibiane Schönlieb

Department for Applied Mathematics and Theoretical Physics
Cantab Capital Institute for the Mathematics of Information
EPSRC Centre for Mathematical Imaging in Healthcare
University of Cambridge, UK

Solving Inverse Problems with Deep Learning, September 2023

Lecture plan

UNIVERSITY OF CAMBRIDGE

- Lecture 1: Variational models & PDEs for inverse imaging problems
- Lecture 2: Learned variational regularisers & plug-and-play denoising
- Lecture 3: Learned iterative reconstruction & perspectives.

Based on: Arridge, Maass, Öktem, CBS, Acta Numerica '19

Deep learning for inverse imaging

Main paradigms:

- Fully Learned Models
- Learned Post Processing
- Learned Iterative Schemes
- Learning the Regulariser

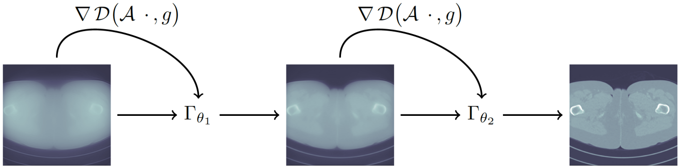
Reviews: McCann, Jin, Unser, IEEE Signal Processing Magazine ’17; Arridge, Maass, Öktem, CBS, Acta Numerica ’19

Learned iterative schemes

Reconstruct via

$$u_0 = T^\dagger y$$

$$u_{n+1} = \Psi_\Theta(u_n,\ \nabla_{u_n}\|Tu_n - y\|_2^2)$$



Adler, Öktem, Inverse Problems ’17.

Learned iterative schemes

Given an initial guess $u^0 \in X$, the learned reconstruction is given by

$$u := (\Lambda_{\Theta_N} \circ \cdots \circ \Lambda_{\Theta_1})(u^0).$$

Here,

- Λ_{Θ_k} can be interpreted as residual layers in neural network $\Psi_\Theta(y)$ with N layers which reconstructs u from y.
- Learned iterative schemes are usually inspired by iterative algorithms for solving variational regularisation problems.

Gregor and LeCun, ICML '10; Yang, Sun, Li, Xu, NeurIPS '16; Meinhardt, Moeller Hazirbas, Cremers, ICCV '17; Putzky, Welling, arXiv:1706.04008; Adler, Öktem, Inverse Problems '17; Adler, Öktem, IEEE TMI '18; Hammernik et al. MRM '18; Adler, Lunz, Verdier, CBS, Öktem, NeuIPS '18; Hauptmann et al., IEEE TMI '19; de Hoop, Lassas, Wong, arXiv:1912.11090; Gilton, Ongie, Willett, IEEE TCI '19; Mukherjee, Öktem, CBS, SSVM '21; Bubba et al., SIIMS '21.

UNIVERSITY OF CAMBRIDGE

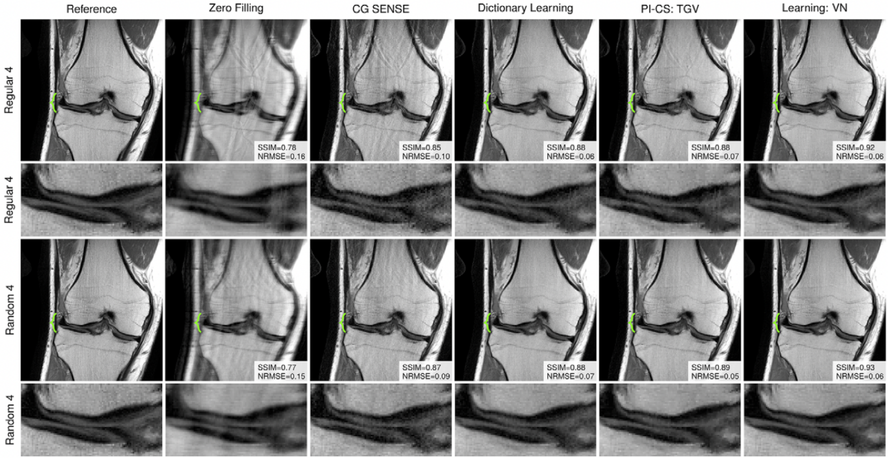
Variational networks

- Inspired by the conditional shrinkage fields (CSF) [Schmidt, Roth '14], we allow to change the parameters during the iterations:

$$\begin{cases} u^0 = f \\ u^{t+1} = u^t - \lambda^t \left(\sum_{k=1}^{q} (K_k^t)^\top (\rho_k^t)'(K_k^t u^t) + (u^t - f) \right), \quad t = 0 \ldots T-1 \end{cases}$$

- In each step we perform one gradient descent on a learned variational energy
- Can be interpreted as one cycle of a block incremental gradient descent
- Can also be interpreted as learned non-linear diffusion, trying to "invert" the convolution $\int p(f|u)p(u)\mathrm{d}u$
- And it can be interpreted as a convolutional neural network with T layers

Variational networks for MRI



Hammernick et al. Magnetic Resonance in Medicine '18; Effland, Kobler, Kunisch, Pock, JMIV '20. See here `https://fastmri.org` for open source MRI data.

Learned gradient descent

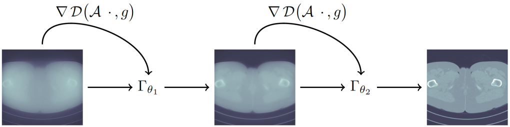
Again, the reconstruction is given by

$$u := (\Lambda_{\Theta_N} \circ \cdots \circ \Lambda_{\Theta_1})(u^0).$$

but here, the variational network parametrisation is generalised to

$$\Lambda_{\Theta_k} := \mathrm{Id} + \Gamma_{\Theta_k} \circ \nabla D\big(T\cdot, y\big).$$

where $\Gamma_{\Theta_k} : X \to X$ is learned from supervised data.



Adler, Öktem, Inverse Problems '17

Learned proximal descent

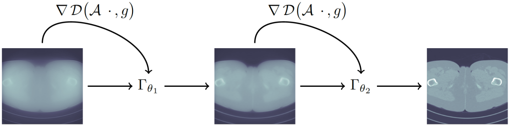
Again, the reconstruction is given by

$$u := (\Lambda_{\Theta_N} \circ \cdots \circ \Lambda_{\Theta_1})(u^0).$$

with

$$\Lambda_{\Theta_k} := \Gamma_{\Theta_k}\big(\mathrm{Id} + \nabla D(T\cdot, y)\big),$$

where $\Gamma_{\Theta_k} : X \to X$ is learned from supervised data.



Adler, Öktem, Inverse Problems ’17

Generalised learned gradient descent

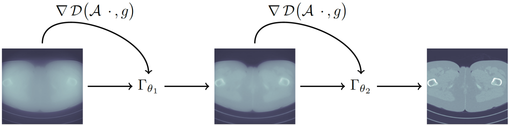
Again, the reconstruction is given by

$$u := (P_X \circ \Lambda_{\Theta_N} \circ \ldots \circ \Lambda_{\Theta_1})(u^0, y)$$

with a more general parametrisation of the gradient steps

$$\Lambda_{\Theta_k} := \Gamma_{\Theta_k}\Big(u, y, T^* y, Tu, \nabla \mathrm{R}_\alpha(u)\Big) \quad \text{for } (u, \tilde{y}) \in X \times Y,$$

where $\Gamma_{\Theta_k} : X \to X$ is learned from supervised data.



Adler, Öktem, Inverse Problems ’17

Learned primal-dual

Idea: introduce an explicit learned iteration also in the measurement (dual) space.

Inspired by the primal-dual hybrid gradient approach, we consider

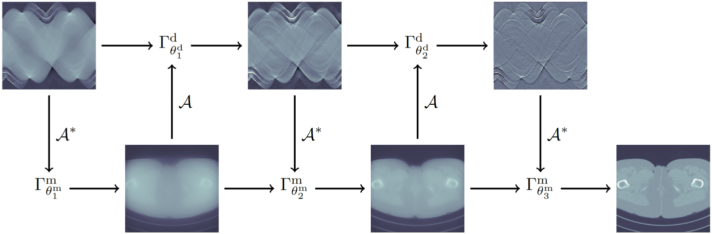
$$\begin{cases} \tilde{y}^0 = y \text{ and } u^0 \in X \text{ given} \\ \tilde{y}^{k+1} = \Gamma^{\mathsf{d}}_{\Theta^{\mathsf{d}}_k}\left(\tilde{y}^k, Tu^k, y\right) \\ u^{k+1} = \Gamma^{\mathsf{m}}_{\Theta^{\mathsf{m}}_k}\left(u^k, T^*\tilde{y}^{k+1}\right) \end{cases} \qquad \text{for } k = 0, \ldots, N-1,$$

where $\Gamma^{\mathsf{m}}_{\Theta^{\mathsf{m}}_k} : X \times X \to X$ and $\Gamma^{\mathsf{d}}_{\Theta^{\mathsf{d}}_k} : Y \times Y \times Y \to Y$ are update operators in image and measurement space, respectively.
Then the reconstruction is given by $u := (P_X \circ \Lambda_{\Theta_N} \circ \ldots \circ \Lambda_{\Theta_1})(u^0, y)$ with

$$\Lambda_{\Theta_k}(u, y) := \left(\Gamma^{\mathsf{m}}_{\Theta^{\mathsf{m}}_k}\left(u, T^*(\Gamma^{\mathsf{d}}_{\Theta^{\mathsf{d}}_k}(\tilde{y}, Tu, y))\right), \Gamma^{\mathsf{d}}_{\Theta^{\mathsf{d}}_k}(\tilde{y}, Tu, y)\right).$$

Adler, Öktem, Inverse Problems '17

Learned primal-dual



Adler, Öktem, Inverse Problems ’17

Learned iterative schemes

Some comments:

- Performance: For mildly ill-posed inverse problems the quality of the reconstruction is excellent and these schemes are extremely versatile, e.g. used for task-adapted reconstruction Adler, Lunz, Verdier, CBS, Öktem, NeuIPS '18 & Inverse Problems '22. Brief diversion . . .
- Computational bottleneck: Memory! Evaluation of the forward operator and its adjoint in every network layer of the neural network updates. Hence, scalability is challenging. Some works in this direction use invertible neural networks Rudzusika, Bajic, Öktem, CBS, Etmann, NeurIPS '21 Workshop on DL and Inverse Problems, stochastic subsampling of the forward operator Tang, Mukherjee, CBS '21 or greedy training Hauptmann et al., IEEE TMI '18.
- Theory?: Learned iterative schemes approximate conditional mean Adler, Öktem, Inverse Problems '17 . . . also, learned iterative schemes are trained for a fixed number of iterations (typically ≤ 20) due to computational constraints, and the reconstruction deteriorates if more iterations are performed at test time . . .

Multi-tasking: more together than alone

- Starting point: inverse problems, i.e. given data y the task is to recover physical quantity u with $y = Tu + n$, where T is the forward operator and n noise.
- Prototypical example: inverse problems in imaging.

Background: Inverse problems in practice rarely constitute a single problem but rather consist of a sequence of problems (tasks) that need to be solved.

- Tasks: image data acquisition, image reconstruction (from indirect measurements, e.g. computed tomography (CT) in medical imaging), image segmentation, classification, ...

Hypothesis: these tasks can positively influence each other.

Biomedical imaging pathway - current

The path from imaging data acquisition to prediction / diagnosis / treatment planning features several processing and analysis steps which usually are performed **sequentially**.

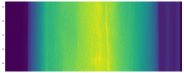
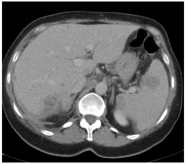
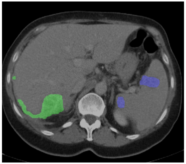


image recon. $\Longrightarrow$ $\sim T^\dagger f$



lesion segment. $\Longrightarrow$



statistical anal. $\Longrightarrow$

prediction $\Longrightarrow$ **treatment response**

CT data and segmentation are courtesy of Evis Sala and Ramona Woitek.

Multi-tasking: more together than alone

UNIVERSITY OF CAMBRIDGE

Background: Inverse problems in practice rarely constitute a single problem but rather consist of a sequence of problems (tasks) that need to be solved.

- Tasks: image data acquisition, image reconstruction (from indirect measurements, e.g. computed tomography (CT) in medical imaging), image segmentation, classification, ...

Hypothesis: these tasks can positively influence each other.

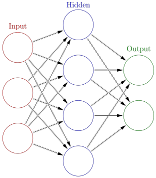
Example in this talk: task-adapted tomographic reconstruction.

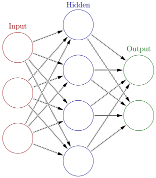
NN-based task-adapted reconstruction

$$(\theta^*, \vartheta^*) \in \arg\min_{(\theta,\vartheta)\in\Theta\times\Xi} \left\{ \frac{1}{m} \sum_{i=1}^{m} \ell_{\text{joint}}\Big(\big(x_i, \tau(z_i)\big), \big(\mathcal{A}_\theta^\dagger(\mathsf{y}_i), \mathcal{T}_\vartheta \circ \mathcal{A}_\theta^\dagger(\mathsf{y}_i)\big) \Big) \right\}$$

$$\ell_{\text{joint}}\big((x,d),(x',d')\big) := (1-C)\ell_X(x,x') + C\ell_D(d,d') \quad \text{for fixed } C \in [0,1].$$

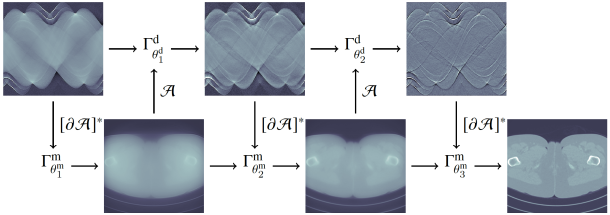
Reconstruction X | Task D





Adler, Lunz, Verdier, CBS, Öktem, In NIPS 2018 meets medical imaging. Also available as arXiv e-print: https://arxiv.org/abs/1809.00948. Some related works: Wu, Kim, Dong, Li '17; Liu, Wen, Liu, Wang, Huang '17.

Physics-constrained NN-based reconstruction

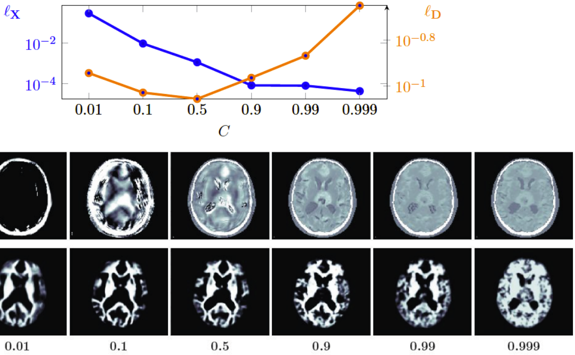


$$\begin{cases} v^0 = g \text{ and } f^0 \in X \text{ given} \\ v^{k+1} = \Gamma^{\mathrm{d}}_{\theta^{\mathrm{d}}_k}(v^k, \mathcal{A}(f^k), g)) \\ f^{k+1} = \Gamma^{\mathrm{m}}_{\theta^{\mathrm{m}}_k}(f^k, [\partial\,\mathcal{A}(f)]^*(v^{k+1})) \end{cases} \qquad \text{for } k = 0, \ldots, N-1.$$

Acta Numerica 2019 review paper on Data-driven methods for solving inverse problems with S. Arridge, P. Maass, and O. Öktem. Related work Adler, Öktem, IEEE TMI '18; Hammernik, Klatzer, Kobler, Recht, Sodickson, Pock, Knoll, Magnetic Resonance in Medicine '18.

Task-adapted reconstruction

CNN-based MRI reconstruction (X) and CNN-based MRI segmentation (D). Both are trained with combined loss $C\,\ell_X + (1-C)\,\ell_D$ for $C \in [0,1]$.



First row: CNN-based reconstructions; second row: CNN-based segmentations. Both trained with combined loss for varying C.

J. Adler, S. Lunz, O. Verdier, CBS, O. Öktem '18

Learned iterative schemes

Some comments:

- Performance: For mildly ill-posed inverse problems the quality of the reconstruction is excellent and these schemes are extremely versatile, e.g. used for task-adapted reconstruction Adler, Lunz, Verdier, CBS, Öktem, NeuIPS '18 & Inverse Problems '22.
- Computational bottleneck: Memory! Evaluation of the forward operator and its adjoint in every network layer of the neural network updates. Hence, scalability is challenging. Some works in this direction use invertible neural networks Rudzusika, Bajic, Öktem, CBS, Etmann, NeurIPS '21 Workshop on DL and Inverse Problems, stochastic subsampling of the forward operator Tang, Mukherjee, CBS '21 or greedy training Hauptmann et al., IEEE TMI '18.
- Theory?: Learned iterative schemes approximate conditional mean Adler, Öktem, Inverse Problems '17 . . . also, learned iterative schemes are trained for a fixed number of iterations (typically ≤ 20) due to computational constraints, and the reconstruction deteriorates if more iterations are performed at test time . . .

Learned fixed point iterations

UNIVERSITY OF CAMBRIDGE

Example: Deep equilibrium networks Assumption: the reconstruction is a fixed point of a learned fixed point operator Γ_Θ, i.e.

$$u = \Gamma_\Theta(u; y).$$

> *Idea: Design learned iterative schemes which provably converge to a fixed point as the number of iterations goes to infinity.*

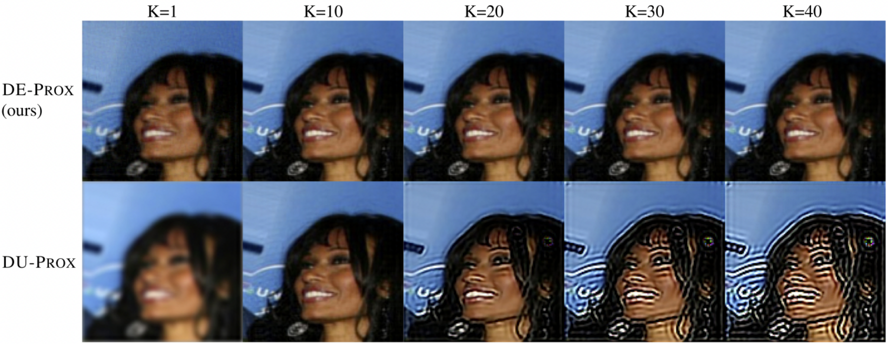
Example: deep equilibrium gradient descent where

$$\Gamma_\Theta(u; y) = u + \eta T^*(y - Tu) - \eta \, \mathrm{R}_\Theta(u),$$

and R_Θ is a trainable neural network.

Gilton, Ongie, Willett, IEEE Trans Comp Imaging '21; Hertrich, Neumayer, Steidl, Linear Algebra and its Applications '21.

Deep equilibrium networks



Example taken from Gilton, Ongie, Willett, IEEE Trans Comp Imaging '21

Deep equilibrium networks

In order to guarantee convergence of iterations to a fixed point, most natural way is to restrict Γ_Θ to be a contraction, i.e., we have the following theorem for deep equilibrium gradient descent

Theorem

Assume $\mathrm{R}_\Theta - \mathrm{Id}$ *is* ϵ*-Lipschitz continuous and let* $L = \lambda_{max}(T^*T)$ *and* $\mu = \lambda_{min}(T^*T)$*. If* $0 < \eta < 1/(L+1)$ *then* Γ_Θ *fulfills*

$$\|\Gamma_\Theta(u;y) - \Gamma_\Theta(\tilde{u};y)\| \leq \underbrace{(1 - \eta(1+\mu) + \eta\epsilon)}_{=\gamma} \|u - \tilde{u}\|, \quad \textit{for all } u, \tilde{u} \in X.$$

Then Γ_Θ *is a contraction, i.e.* $\gamma < 1$ *if* $\epsilon < 1 + \mu$ *and hence the iterates converge.*

Gilton, Ongie, Willett, IEEE Trans Comp Imaging '21

Comments:

- An interesting additional avenue with convergent learned iterative schemes is the ability to accelerate their convergence by increasing the memory of the iterations, i.e. the dependency of each iteration from just the previous iterate only to a couple of previous iterates, for instance via Anderson acceleration as presented in Gilton, Ongie, Willett, IEEE Trans Comp Imaging '21.
- Convergence to a fixed point is an interesting property. Still outstanding: what does this fixed point actually fulfil?

Take Away Messages

UNIVERSITY OF CAMBRIDGE

- The role of deep learning as a data-adaptive image prior / regulariser for inverse problems
- The role of deep learning for computationally more efficient solutions for inverse problems
- Many open questions wrt generalisability, interpretability, trade-off between theoretical & computational performance. More concretely, some are:
  - Richer parametrisations of convex neural networks or gradients of convex neural networks;
  - More qualitative properties on learned regularisation, e.g. invariances wrt certain transformations;
  - Learned iterative schemes that encode a provable regularisation;
  - Uncertainty quantification for inverse problems with deep learning.

Arridge, Maass, Öktem, CBS, Acta Numerica 2019.

UNIVERSITY OF CAMBRIDGE

And alongside the theory we need to work on convincing use-cases.

Headline: DL & maths turn CT/MRI . . . into a clinical screening tool!

Lecture plan

- Lecture 1: Variational models & PDEs for inverse imaging problems
- Lecture 2: Learned variational regularisers & plug-and-play denoising
- Lecture 3: Learned iterative reconstruction & perspectives.

Based on: Arridge, Maass, Öktem, CBS, Acta Numerica '19

Thank you very much for your attention!

More information see:

http://www.ccimi.maths.cam.ac.uk

http://www.cmih.maths.cam.ac.uk

http://www.damtp.cam.ac.uk/research/cia/

Email: cbs31@cam.ac.uk