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

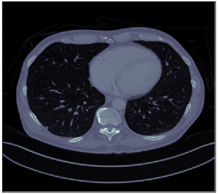
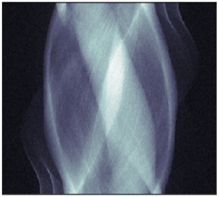
Lecture plan

- 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

Guiding example

Computed tomography: compute image $u \in X$ from $Y \ni y = Tu + n$, X, Y normed vector spaces. In computed tomography (CT) T is (sub-sampled) Radon transform.

T  $+\, n \; =$ 

Inverse problem: measuring (noisy & incomplete) data y compute image $u \Rightarrow$ typically ill-posed $\rightarrow$ well-posedness through regularisation.

CT image from Mayo Clinic open CT dataset[1]

[1] Chen, Baiyu, et al. "An open library of CT patient projection data." Medical Imaging 2016: Physics of Medical Imaging. Vol. 9783. International Society for Optics and Photonics, 2016.

Variational regularisation

General task: **restore** $\mathbf{u}$ from an **observed datum** $\mathbf{y}$ where

$$y = \underbrace{Tu}_{\text{forward model}} + \underbrace{n}_{\text{noise}},$$

and T is bounded and linear.

Variational approach: Compute u as a minimizer of

$$\mathcal{J}(u) = \alpha \underbrace{R(u)}_{\text{regularisation}} + \underbrace{D(Tu, y)}_{\text{data fidelity}} \to \min_{u \in X},$$

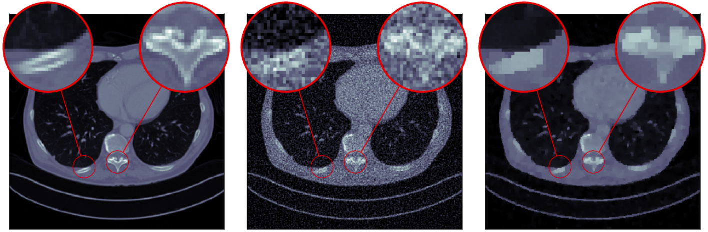
where

- $D(\cdot, \cdot)$ is a distance function, e.g. $D(Tu, y) = \|Tu - y\|_2^2$.
- $R(u)$ is a prior/regulariser that models a-priori information on u weighted by positive α, e.g., $R(u) = TV(u) = \|\nabla u\|_1$ (in infinite dimensions $|Du|(\Omega)$) and X suitable Banach space, e.g., $X = BV(\Omega)$.

Engl, Hanke, Neubauer '96; Rudin, Osher, Fatemi, Physica D '92; Natterer, Wübbeling '01; Candes, Romberg, Tao, IEEE Trans Inf Theory '06; Kaltenbacher, Neubauer, Scherzer '08; Schuster, Kaltenbacher, Hofmann, Kazimierski '12

Limits of knowledge driven models

Example: CT reconstruction:



Ground-truth | FBP: 21.63 dB, 0.24 | TV: 29.25 dB, 0.79

We only have a limited amount of modelling capacity ...

Photo courtesy of Subho Mukherjee

Can data driven models help?

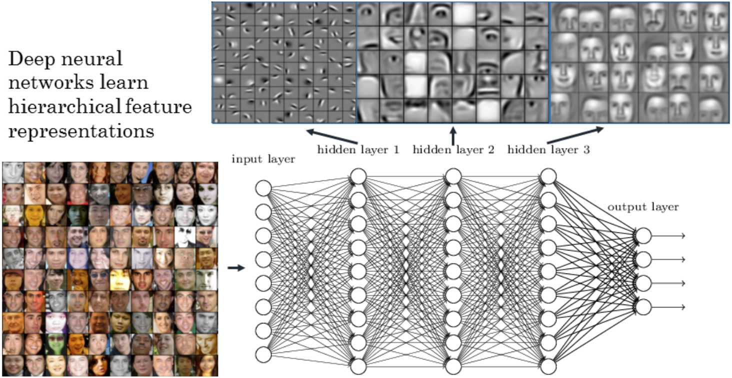
(a) $\|\nabla u\|_1$

(b) noisy

Why can our brain denoise this image so much better?

Can data driven models help?

Key idea: use highly overparametrised models, e.g. neural networks, and train model on data, agnostic to mathematical (physical or geometrical) modelling



LeCun, Y., Bengio, Y., Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444.

Knowledge vs data driven

Example: CT reconstruction: top row is based on mathematical/handcrafted models; bottom row is using novel deep learning based models.

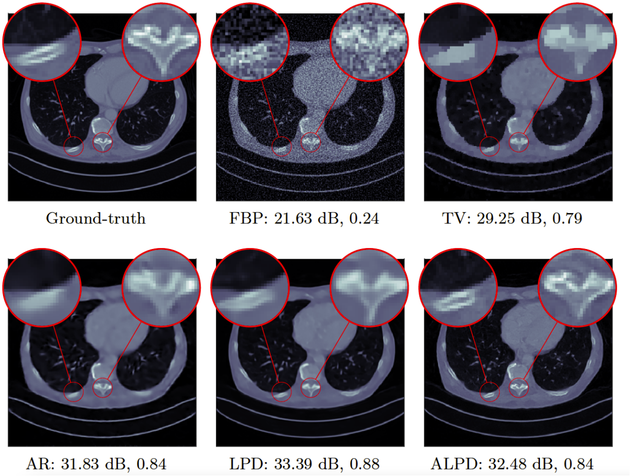


Photo courtesy of Subho Mukherjee

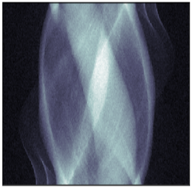
Knowledge vs data driven regularisation

UNIVERSITY OF CAMBRIDGE

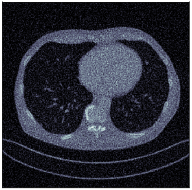
Example: Sparse-angle CT reconstruction

Knowledge driven models are limited by our ability to model images!

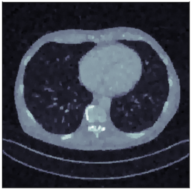
From left to right: we see the data, and then two knowledge driven solutions (FBP and TV) and one data driven solution (AR).



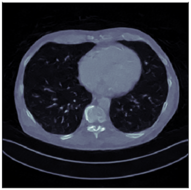
Data



FBP



TV



AR

In our case data driven means learned adversarial regularisers (AR) which I will describe a bit later . . .

Two paradigms in mathematical imaging

Knowledge driven imaging, based on mathematical models & physical principles, e.g. non-linear PDEs

$$u_t = \mathrm{div}\left(\frac{\nabla u}{|\nabla u|}\right),$$

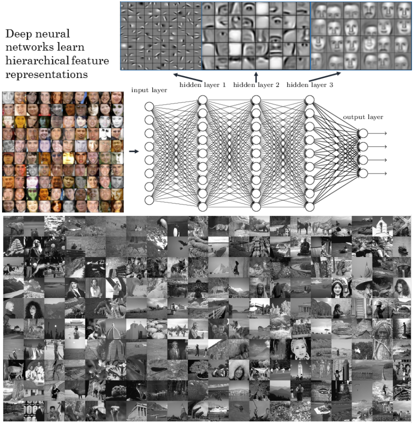
variational principles

$$\min_{\xi}\{\int |D\xi| + \int \xi(c_1 - f)^2 + \int (1-\xi)(c_2 - f)^2\},$$

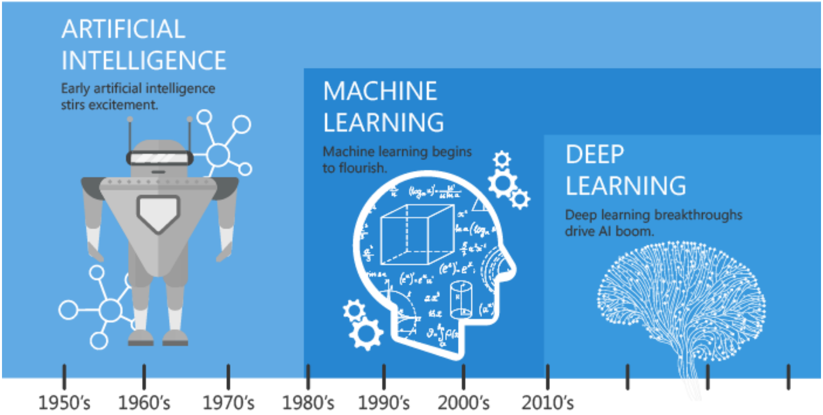
multi-resolution/sparsity

$$u \approx Su_\Lambda = \sum_{\lambda\in\Lambda} u_\lambda \psi_\lambda.$$

Data driven imaging, based on overparametrised models, e.g. support vector machines or neural networks.



Deep learning



Picture from `https://towardsdatascience.com`

UNIVERSITY OF CAMBRIDGE

Computing power

Big data

Picture from `https://towardsdatascience.com`

Deep learning based imaging

So lets have a closer look at this idea

Neural network in a nutshell:

$$\Psi : X \times \Theta \to Y$$
$$(x, \theta) \mapsto z^K.$$

where

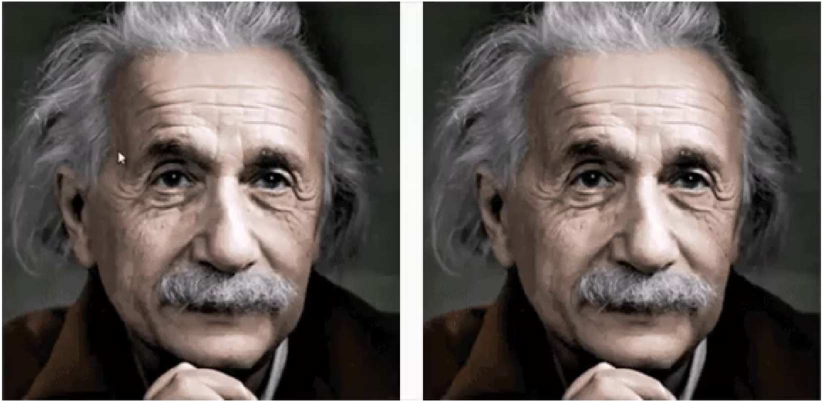
$$z^0 = x \in \mathcal{X}$$
$$z^{k+1} = f^k(z^k, \theta^k), \quad k = 0, \ldots, K-1,$$

with $f^k : X^k \times \Theta^k \to X^{k+1}$, X^k feature (vector) space, Θ^k parameter (vector) space.
Typical choices: $f^k(z) = \sigma(A^k z + b^k)$, where σ is an element-wise nonlinearity (ReLU, tanh etc.) and A is often replaced by a convolutional operator.
Training goal:

$$\min_{\theta \in \Theta} \frac{1}{N} \sum_{n=1}^{N} L_n(\Psi(x_n, \theta), y_n) + R(\theta) \tag{1}$$

for dataset $\{(x_n, y_n)\}_n$. How to use this in mathematical imaging?

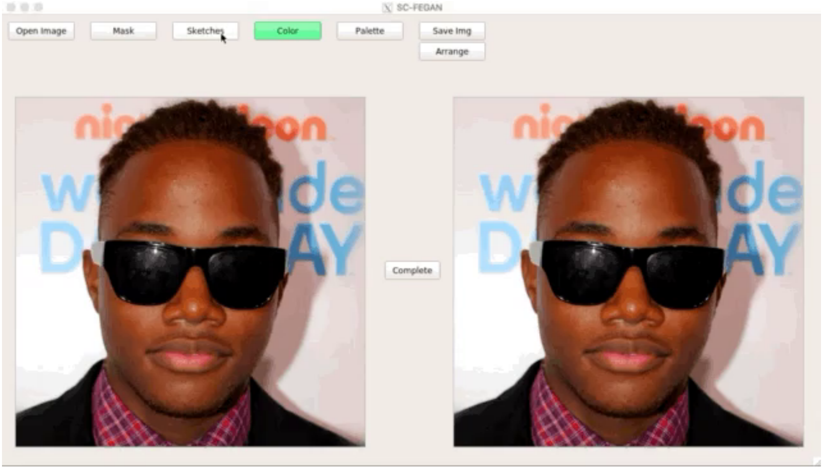
Deep learning for inpainting



Highly accurate and computationally fast solutions

G. Liu et al. *Image inpainting for irregular holes using partial convolutions*. ECCV 2018

Deep learning for editing



Highly accurate and computationally fast solutions

Y. Jo, J. Park. *SC-FEGAN: Face Editing Generative Adversarial Network with User's Sketch and Color*. ICCV 2019

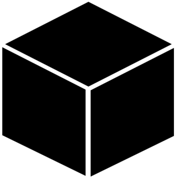
Deep learning challenges

State of the art deep neural networks have ‘too many’ degrees of freedom:

- millions of free parameters, i.e., Θ is super high-dimensional;
- complex concatenation of diverse mathematical constructs (convolutions, activations, skip connections, normalization, dropout, ...);
- high-dimensional and non-convex optimisation problems.

The deep learning result is influenced by all this, as well quality of the training set, optimisation approach used ....

Blackbox



Deep learning is a black box

In short: this is super complex and messy!

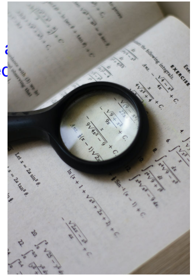
Deep learning challenges

Why is black box a problem?

No systematic way to design them

Identified as a
45mph speed
sign

Safety danger



Lack of interpretation

Good recent reference: Berner, Grohs, Kutyniok, Petersen, The Modern Mathematics of Deep Learning, arXiv:2105.04026

Data driven inverse problems

UNIVERSITY OF CAMBRIDGE

- Machine learning and in particular deep learning offers interesting opportunities for inverse problems – high quality solutions through data-adaptivity.
- They only work in practice when combined with mathematical and statistical modelling.
- In general, there is a lack of understanding of these approaches and a need for more mathematical scrutiny. This lecture series discusses what we know and what we do not know.
- TODAY: Learned variational models & Plug-and-play denoisers as a direct development from what we have seen so far.

Examples for learning regularisation

- Sparse coding & dictionary learning '06–: Aharon, Allard, Binev, Bruckstein, Bruna, Chandrasekaran, DeVore, Elad, Fadili, Mallat, Mayral, Papyan, Peyre, Ponce, Sulam, . . . :

$$\min_{\gamma,\phi} \|T(\sum_i \gamma_i \phi_i) - y\|_2^2 + \|\gamma\|_1$$

- Black-box denoiser: Plug-and-Play Prior (P^3) method Venkatakrishnan, Bouman, Wohlberg, GlobalSIP '13; Wei, Aviles-Rivero, Liang, Fu, Huang, CBS, ICML '20 best paper award, Regularisation by Denoising (RED) Romano, Elad, Milanfar, SIIMS '17; Terris, Repetti, Pesquet, Wiaux, ICASSP '20,

$$\min_u D(T(u), y) + \alpha R(u), \quad \text{with } R(u) = \langle u, u - \Lambda(u)\rangle, \quad \Lambda : X \to X \text{ denoiser.}$$

- Bilevel optimisation '03–: Chung, De Los Reyes, Fonseca, Haber, Hintermüller, Horesh, Kunisch, Langer, Liu, Pock, Tappen, Tenorio, CBS, . . .

$$\min_\lambda F(u_\lambda) \quad \text{s.t. } u_\lambda = \operatorname{argmin}_v R(\lambda, u) + D(T(u), y)$$

- Deep neural networks as regularisers . . .

Deep learning for inverse imaging

Main paradigms:

- Fully Learned Models [2]
- Learned Post Processing [3]
- Learned Iterative Schemes [4]
- Learning the Regulariser [5] and Plug & Play Approaches [6]

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

[2] Zhu, Bo, Liu, Cauley, Rosen, Rosen, Nature ’18.

[3] Jin, McCann, Froustey, Unser, IEEE TIP, ’17; Kang, Min, Ye, Medical Physics ’17.

[4] Yang et al. NeurIPS ’16; Meinhardt et al. ICCV ’17; Putzky, Welling ’17; Adler, Öktem, Inverse Problems ’17; Hammernik et al. MRM ’18; Hauptmann et al., IEEE TMI ’19; de Hoop, Lassas, Wong, Neural Networks ’20; Gilton, Ongie, Willett, IEEE TCI ’21; Mukherjee, Öktem, CBS, SSVM ’21; Pesquet, Repetti, Terris, Wiaux, SIIMS ’21; Bubba et al., SIIMS ’21.

[5] Lunz, Öktem, CBS, NeurIPS ’18; Ye, Ravishankar, Long, Fessler, IEEE TMI ’18; Schwab et al. Inverse Problems ’20; Kobler et al. CVPR ’20; Pinetz et al., SIIMS ’21; Kabri et al. ’22

[6] Venkatakrishnan, Bouman, Wohlberg, IEEE SIP ’13, Romano, Elad, Milanfar. SIIMS ’17; Sun, Wu, Xu, Wohlberg, Kamilov IEEE TCI ’21; Hauptmann, Mukherjee, CBS, Sherry ’23

Deep Learning for Inverse Problems

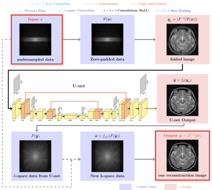
Let Ψ_Θ be a neural network with parameters Θ.

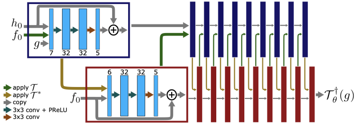
- Postprocessing. Reconstruct via

$$u = \Psi_\Theta T^\dagger y$$

- Learned Iterative Reconstruction.

$$u_0 = T^\dagger y$$
$$u_{n+1} = \Psi_\Theta(u_n,\ \nabla_{u_n}\|Tu_n - y\|_2^2)$$





Advantage: can perform very well; disadvantage: mostly a black box! [7]

[7] Only few theoretical results, e.g. Hertrich, Neumayer, Steidl, Linear Algebra and its Applications ’21; Gilton, Ongie, Willett, IEEE TCI ’21.

Disadvantages of most DL solutions

- Almost no theoretical underpinnings[8]; well-posed regularised solution? controversy around deep learning for inverse problems[9]
- Interpretation of reconstruction model is missing (or known only asymptotically, e.g. learned iterative schemes with L^2 loss converge to conditional mean in the infinite data limit[10])
- Data consistency is in general not guaranteed[11]
- Usually requires lots of supervision which can be problematic in real-life problems.

[8] Some exceptions, e.g. Hertrich, Neumayer, Steidl, Lin Alg Appl '21; Sun, Wu, Xu, Wohlberg, Kamilov IEEE TCI '21; Gilton, Ongie, Willett, IEEE TCI '21.

[9] Genzel, Macdonald, März, IEEE PAMI '22

[10] Adler, Öktem, Inverse Problems '17

[11] Some exceptions, e.g. Null-space networks, Schwab, Antholzer, Haltmeier, IP '19

Deep learning for inverse imaging

Main paradigms:

- Fully Learned Models [12]
- Learned Post Processing [13]
- Learned Iterative Schemes [14]
- Learning the Regulariser [15] and Plug & Play Approaches [16]

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

[12] Zhu, Bo, Liu, Cauley, Rosen, Rosen, Nature ’18.

[13] Jin, McCann, Froustey, Unser, IEEE TIP, ’17; Kang, Min, Ye, Medical Physics ’17.

[14] Yang et al. NeurIPS ’16; Meinhardt et al. ICCV ’17; Putzky, Welling ’17; Adler, Öktem, Inverse Problems ’17; Hammernik et al. MRM ’18; Hauptmann et al., IEEE TMI ’19; de Hoop, Lassas, Wong, Neural Networks ’20; Gilton, Ongie, Willett, IEEE TCI ’21; Mukherjee, Öktem, CBS, SSVM ’21; Pesquet, Repetti, Terris, Wiaux, SIIMS ’21; Bubba et al., SIIMS ’21.

[15] Lunz, Öktem, CBS, NeurIPS ’18; Ye, Ravishankar, Long, Fessler, IEEE TMI ’18; Schwab et al. Inverse Problems ’20; Kobler et al. CVPR ’20; Pinetz et al., SIIMS ’21; Kabri et al. ’22

[16] Venkatakrishnan, Bouman, Wohlberg, IEEE SIP ’13, Romano, Elad, Milanfar. SIIMS ’17; Sun, Wu, Xu, Wohlberg, Kamilov IEEE TCI ’21; Hauptmann, Mukherjee, CBS, Sherry ’23

Learning the regulariser

Variational Problem: $y \in Y$ and $T : X \to Y$ linear and bounded, X, Y Banach spaces. Then, consider

$$\arg\min_u \|Tu - y\|_2^2 + \alpha R(u)$$

- Interpretability: explicit prior.
- Stability and convergence results.
- Regularisation parameter to adopt to noise level.
- Incorporates forward model explicitly.
- Incorporates noise statistics.

Do not learn how to reconstruct,
learn an ‘image prior’ that is data-adaptive and amenable to mathematical analysis.

Joint work with



Marcello Carioni Sören Dittmer Sebastian Lunz



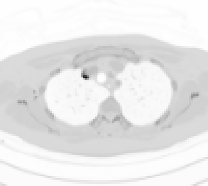
Subhadip Mukherjee Ozan Öktem Zak Shumaylov

S. Lunz, O. Öktem, CBS, Adversarial Regularizers in Inverse Problems, in NeurIPS '18; S. Mukherjee, S. Dittmer, Z. Shumaylov, S. Lunz, O. Öktem, CBS, Learned convex regularizers for inverse problems, arXiv:2008.02839; Carioni, Mukherjee, CBS, End-to-end reconstruction meets data-driven regularization for inverse problems, NeurIPS '21

Train regulariser adversarially

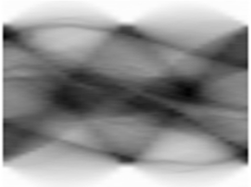
How to learn a regulariser for $\min_v \alpha\mathcal{R}(v) + \|Tv - y\|_2^2$?

'Good guys'



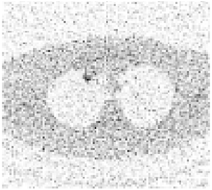
$U^* \sim \mathbb{P}_U$

Data



$Y \sim \mathbb{P}_Y$

'Bad guys'



$U \sim \mathbb{P}_n := T^\dagger_\sharp \mathbb{P}_Y$

Train regulariser in adversarial manner to be small for samples from $\mathbb{P}_U$ and large for samples from $\mathbb{P}_n$.

Lunz, Öktem, CBS, NeurIPS '18; Arridge, Maass, Öktem, CBS, Acta Numerica '19; Lunz, PhD thesis University of Cambridge '21

Train regulariser adversarially

- Idea from Wasserstein GANs [17]: use 1-Wasserstein distance as (unsupervised) loss for regulariser

$$\mathrm{Wass}(\mathbb{P}_n, \mathbb{P}_U) = \sup_{R \in 1-Lip} \mathcal{E}_{U \sim \mathbb{P}_n} R(U) - \mathcal{E}_{U \sim \mathbb{P}_U} R(U)$$

- Train the regulariser to learn image statistics – good and bad guys DO NOT need to be paired :)
- Parametrization for R^*: $R_\Theta(u) = \Psi_\Theta(u) + \rho_0 \|u\|_2^2$, where $\Psi_\Theta(u)$ is a [convex[18]] CNN.
- Train the Network with the loss [19]

$$\min_\Theta \mathbb{E}_{U \sim \mathbb{P}_U} [\Psi_\Theta(U)] - \mathbb{E}_{U \sim \mathbb{P}_n} [\Psi_\Theta(U)] + \mu \cdot \mathbb{E}\left[(\|\nabla_u \Psi_\Theta(U)\|_* - 1)_+^2 \right].$$

[17] Arjovsky, Chintala & Bottou, '17

[18] Amos, Xu, Kolter ICML '17; S. Mukherjee, S. Dittmer, Z. Shumaylov, S. Lunz, O. Öktem, CBS, arXiv:2008.02839

[19] Gulrajani et al., '17

Algorithm summary

UNIVERSITY OF CAMBRIDGE

- Create training data $\mathbb{P}_U$ (good guys) and $\mathbb{P}_n$ (bad guys) – do not need to be paired!
- Train regularizer in adversarial and unsupervised fashion by minimizing

$$\mathbb{E}_{U\sim\mathbb{P}_U}\left[\Psi_\Theta(U)\right] - \mathbb{E}_{U\sim\mathbb{P}_n}\left[\Psi_\Theta(U)\right] + \mu\,\mathbb{E}\left[\left(\|\nabla_u\Psi_\Theta(U)\|_* - 1\right)_+^2\right].$$

- Deploy the learned adversarial [convex] regularizer (A[C]R) into variational problem

$$\arg\min_u \|Tu - y\|_2^2 + \alpha\left(\Psi_\Theta(u) + \rho_0\,\|u\|_2^2\right)$$

- Solve the above via (sub)gradient descent.

This [convex] variational model is amenable to analysis! And we get well-posedness :)

Code available at `https://github.com/Subhadip-1`

Unrolling AR: end-to-end meets variational regularisation

UNIVERSITY OF CAMBRIDGE

- Data model: $U \sim \mathbb{P}_U$, $Y = TU + N \sim \mathbb{P}_Y$
- Learn two networks G_ϕ and R_θ adversarially:

$$\min_{\phi} \max_{\theta : R_\theta \in \mathbb{L}_1} \mathbb{E}_{\mathbb{P}_Y} \|Y - TG_\phi(Y)\|_2^2 + \lambda \left(\mathbb{E}_{\mathbb{P}_Y} \left[R_\theta(G_\phi(Y)) \right] - \mathbb{E}_{\mathbb{P}_U} \left[R_\theta(U) \right] \right)$$

- Equivalent to: $\min_{\phi} \mathbb{E}_{\mathbb{P}_Y} \|Y - TG_\phi(Y)\|_2^2 + \lambda \ \mathrm{Wass}(\mathbb{P}_U, (G_\phi)_{\#}\mathbb{P}_Y)$
- Alt. min.: init. $\theta_0 \leftarrow \underset{\theta : R_\theta \in \mathbb{L}_1}{\arg\min} \mathbb{E}_{\mathbb{P}_U} \left[R_\theta(U) \right] - \mathbb{E}_{\mathbb{P}_Y} \left[R_\theta(T^\dagger Y) \right]$:
  1. $\phi_n \leftarrow \arg\min_{\phi} \mathbb{E}_{\mathbb{P}_Y} \left[\|Y - TG_\phi(Y)\|_2^2 + \lambda R_{\theta_n}(G_\phi(Y)) \right]$
  2. $\theta_{n+1} \leftarrow \underset{\theta : R_\theta \in \mathbb{L}_1}{\arg\min} \mathbb{E} \left[R_\theta(U) \right] - \mathbb{E} \left[R_\theta(G_{\phi_n}(Y)) \right]$.
- Returns a regularizer R_{θ^*} and a network G_{ϕ^*} that minimizes the corresponding variational loss
- Fits within the variational framework, yet offers a fast and efficient reconstruction.

Carioni, Mukherjee, CBS, End-to-end reconstruction meets data-driven regularization for inverse problems, NeurIPS '21

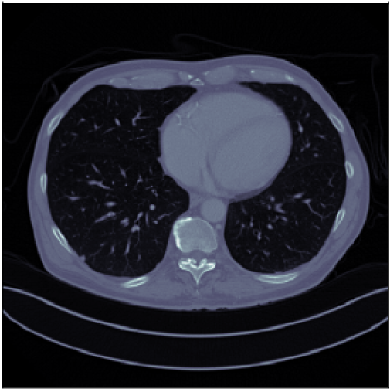
Computed tomography example

Dataset: Mayo clinic open CT data[20]

- Training data: 2250 slices extracted from 9 patients; test data: 128 slices from 1 patient
- Sparse-view geometry: parallel beam, 200 angles, 400 lines/angle, additive Gaussian noise with $\sigma = 2.0$
- Acronyms:
  - FBP: filtered back projection
  - TV: total variation regularisation Rudin, Osher, Fatemi, Physics D '92
  - AR: AR with unrolled updating G of $\mathbb{P}_n$ Carioni, Mukherjee, CBS, End-to-end reconstruction meets data-driven regularization for inverse problems, NeurIPS '21
  - ACR: adversarial convex regularizer Mukherjee, Dittmer, Shumaylov, Lunz, Öktem, CBS, arXiv:2008.02839
  - LPD: learned primal-dual (supervised) Adler, Öktem, IEEE TMI '18

[20] Chen, Baiyu, et al. "An open library of CT patient projection data." Medical Imaging 2016: Physics of Medical Imaging. Vol. 9783. International Society for Optics and Photonics, 2016.

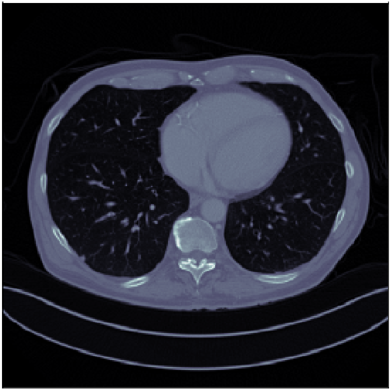
Sparse-view CT



ground-truth

FBP: 21.6262 dB, 0.2435

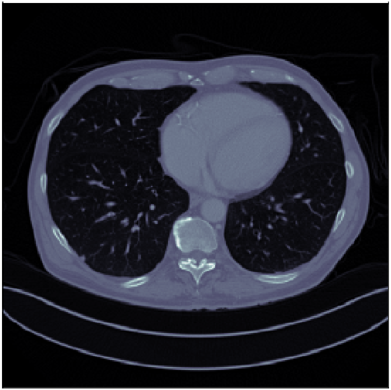
Sparse-view CT



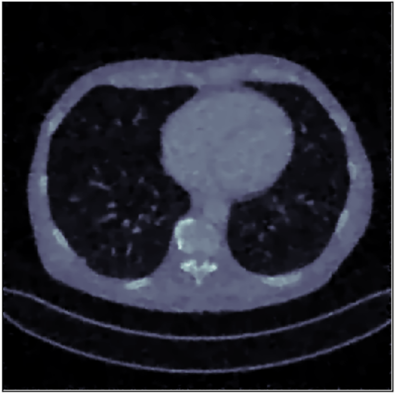
ground-truth

TV: 29.2506 dB, 0.7905

Sparse-view CT

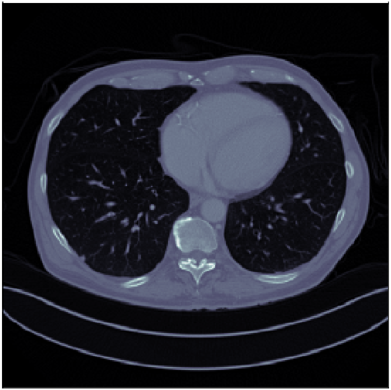


ground-truth

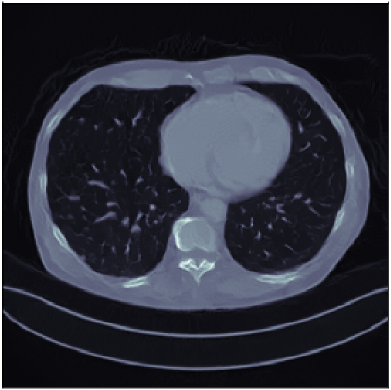


ACR: 30.0016 dB, 0.8246

Sparse-view CT

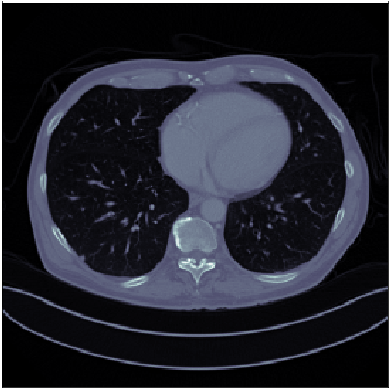


ground-truth



AR: 32.91 dB, 0.8942

Sparse-view CT



ground-truth

LPD: 33.6218 dB, 0.8871

Decreasing the Wasserstein Distance

Doing gradient descent over R^*:

- Definitions: consider u sampled from $\mathbb{P}_n$ and let

$$g_\eta(u) := u - \eta \cdot \nabla_u R^*(u).$$

$$\mathbb{P}_\eta := (g_\eta)_\# \mathbb{P}_n$$

- Assume that $\eta \mapsto \mathrm{Wass}(\mathbb{P}_u, \mathbb{P}_\eta)$ admits a left and a right derivative at $\eta = 0$, and that they are equal. Then,

$$\frac{\mathrm{d}}{\mathrm{d}\eta} \mathrm{Wass}(\mathbb{P}_u, \mathbb{P}_\eta)|_{\eta=0} = -\mathcal{E}_{U \sim \mathbb{P}_n} \left[\|\nabla_u \Psi_\Theta(U)\|^2 \right] = -1.$$

- This is the fastest decrease in Wasserstein distance for any regularization functional with normalized gradients

Existence of optimal R

- Data Manifold Assumption (DMA): The measure $\mathbb{P}_u$ is supported on the weakly compact set $\mathcal{M}$, i.e. $\mathbb{P}_u(\mathcal{M}^c) = 0$
- Denote by $P_{\mathcal{M}} : D \to \mathcal{M}$, $u \mapsto \arg\min_{v \in \mathcal{M}} \|u - v\|$ the projection onto the data manifold
- Projection Assumption: $(P_{\mathcal{M}})_{\#}(\mathbb{P}_n) = \mathbb{P}_r$
- Corresponds to a low-noise assumption - noise level low in comparison to manifold curvature

Existence of optimal R

Theorem

Assume DMA and low-noise assumption. Then, the distance function to the data manifold

$$u \mapsto \min_{v \in \mathcal{M}} \|u - v\|_2$$

is a maximizer to the Wasserstein Loss

$$\sup_{R \in 1-Lip} \mathbb{E}_{U \sim \mathbb{P}_n} R(U) - \mathbb{E}_{U \sim \mathbb{P}_r} R(U).$$

The above served us as an intuition why training the regulariser in this way could be good... unfortunately, it does not hold in practice, see Stanczuk, Etmann, Kreusser, CBS: Wasserstein GANs work because they fail (to approximate the Wasserstein distance), arXiv:2103.01678.

Well-posedness of A[C]R problem

AR

1. Existence: by Lipschitz continuity and coercivity
2. Stability: Let $y_n \to y$ in Y and

$$x_n \in \arg\min_{x \in X} \|Tx - y_n\|^2 + \alpha R_\Theta(x),$$

then x_n has weakly convergent subsequence with limit

$$\hat{x} \in \arg\min_{x \in X} \|Tx - y\|^2 + \alpha R_\Theta(x).$$

ACR

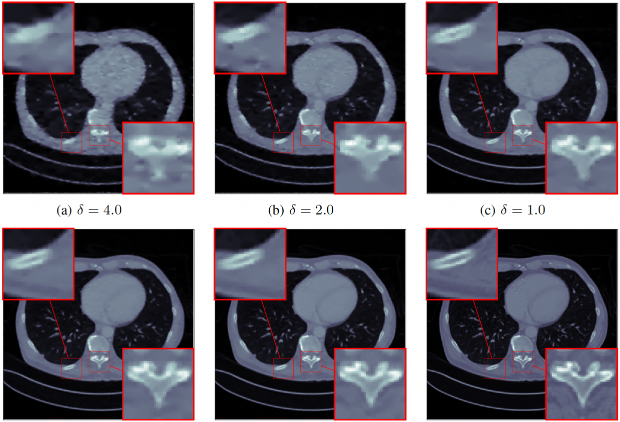
1. Existence and uniqueness: by strong-convexity
2. Stability: $\left\|\hat{x}_\alpha\left(y^{\delta_1}\right) - \hat{x}_\alpha\left(y\right)\right\|_2 \leq \frac{4\beta_1\delta_1}{\alpha\rho_0}$, if $\left\|y^{\delta_1} - y\right\|_2 \leq \delta_1$
3. Convergence: $\hat{x}_\alpha(y) \to x^\dagger$ if $\alpha \to 0$ and $\frac{\delta}{\alpha} \to 0$ when $\delta = \|\boldsymbol{e}\|_2 \to 0$, where

$$x^\dagger = \arg\min_{x \in X} R_\Theta(x) \text{ s.t. } Tx = y_0,$$

where y_0 is noise-free data.
4. Convergent sub-gradient algorithm

Lunz, CBS, Öktem, NeurIPS '18; Mukherjee, Dittmer, Shumaylov, Lunz, Öktem, CBS, arXiv:2008.02839

ACR is provably convergent regulariser



(a) $\delta = 4.0$ (b) $\delta = 2.0$ (c) $\delta = 1.0$

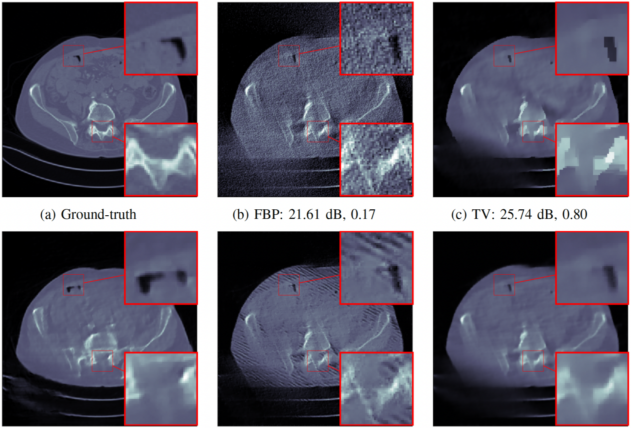
(d) $\delta = 0.50$ (e) $\delta = 0.25$ (f) Ground-truth

ACR is a convergent regulariser, that is the reconstructed image converges to the ground-truth as $\delta \to 0$, subject to an appropriate parameter choice rule for the scalar regularization parameter: $\delta \mapsto \lambda(\delta) > 0$.

Mukherjee, Hauptmann, Öktem, Pereyra, CBS, IEEE SPM '23

Heavily ill-posed problem - no magic

Limited-angle tomography: Deep learning cannot do magic and also hits boundaries of what is mathematically possible.



(a) Ground-truth (b) FBP: 21.61 dB, 0.17 (c) TV: 25.74 dB, 0.80

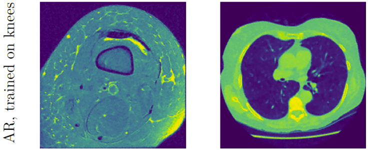
(d) LPD: 29.51 dB, 0.85 (e) AR: 26.83 dB, 0.71 (f) ACR: 27.98 dB, 0.84

Mukherjee, Hauptmann, Öktem, Pereyra, CBS, IEEE Signal Processing Magazine '22

Extensions

- Generalisability of machine learned regularisers – only empirical investigations so far, see PhD thesis of Sebastian Lunz
- Stronger constraints on learned regulariser (e.g., source condition Mukherjee, CBS, Burger, NeurIPS workshop '21)
- Learned regularisers to fulfill certain qualitative properties (e.g. equivariant NNs for regularisers that are invariant to affine transformations Sherry, Celledoni, Ehrhardt, Etmann, Owren, CBS, Inverse Problems '21
- Choice of optimality criteria (e.g., task-adapted inversion joint with Mukherjee, Etmann, Sala and Öktem)[21]
- Training regularisers for large- and high-dimensional inverse problems invertible networks Etmann, Ke, CBS, MLSP '20
- Uncertainty quantification with learned convex regulariser – e.g. via proximal MCMC Pereyra, Statistics and Computing '16

Cross dataset experiments: left knee, right lung



[21] http://all-in-one.maths.cam.ac.uk

Plug & Play Prior (P^3) method

Plug-and-Play Prior (P^3) is based on the idea of operator splitting methods (such as ADMM, see e.g. Setzer, IJCV '11) for optimisation, i.e. the observation that

$$\min_u \{D(Tu, y) + \alpha R(u)\}$$

is equivalent to

$$\min_{u,v} \{D(Tu, y) + \alpha R(v)\} \quad \text{s.t. } u = v.$$

Then, the associated Lagrangian reads

$$L_\lambda(u, v, h) = D(Tu, y) + \alpha R(v) + \frac{\lambda}{2}\|u - v + h\|_2^2 - \frac{\lambda}{2}\|h\|_2^2,$$

where h is the Lagrange dual variable and $\lambda > 0$ a penalty parameter.
Venkatakrishnan, Bouman, Wohlberg, IEEE SIP '13

Then ADMM consists of approximating a solution to the saddle point problem for L_λ by iterating

$$\begin{aligned} u^{k+1} &= \arg\min_u L_\lambda(u, v^k, u^k) \\ v^{k+1} &= \arg\min_v L_\lambda(u^{k+1}, v, u^k) \\ u^{k+1} &= u^k + (u^{k+1} - v^{k+1}). \end{aligned}$$

In particular, the update in u and v read

$$u^{k+1} = \arg\min_u D(Tu, y) + \frac{\lambda}{2}\|u - v^k + u^k\|_2^2 \tag{2}$$

$$v^{k+1} = \operatorname{prox}_{\tau\alpha R}\left(v^k - u + u\right), \tag{3}$$

This decouples the reconstruction problem, reconstructing u from measurements y, from regularising the reconstruction by denoising v.

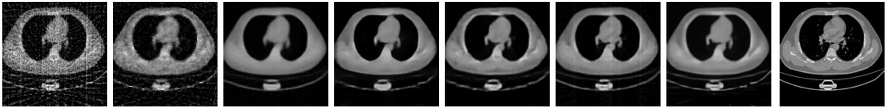
This decouples the reconstruction problem, reconstructing u from measurements y, from regularising the reconstruction by denoising v.

$$v^{k+1} = \underbrace{\mathrm{prox}_{\tau\alpha R}}_{=:D}\left(v^k - u + u\right)$$

This opens the door to switching in any demonstrably successful denoising algorithms for D such as BM3D Dabov et al., SPARS'09, nonlocal means Buades, Coll, Morel, CVPR'05, or deep learning based-denoisers, without redesigning the reconstruction algorithm. Hence the name 'Plug-and-Play'.

Venkatakrishnan, Bouman, Wohlberg, IEEE SIP '13

P^3 properties



- Convergence: In general P^3 not provably convergent. Typically, convergence results for PnP methods either assume D to satisfy hard Lipschitz bound, or require data-fidelity term to be strongly-convex in u. Some new variants of P^3 such as gradient-step (GS) denoisers alleviate the need for such restrictive assumptions, see e.g. Hurault et al. '21.
- Regularisation: In general, there is a lack of an explicit representation of the regulariser, hence no Bayesian interpretation as a prior probability distribution in MAP. Variant of P^3 with explicit regulariser 'Regularisation by Denoising' (RED) Romano, Elad, Milanfar, SIIMS '17 and GS.

Numerical example from Wei, Aviles-Rivero, Liang, Fu, Huang, CBS, ICML '20. For theoretical results see also Chan, Wang, Elgendy, IEEE Comput. Imaging '16; Ryu et al., ICML '19.

GS denoisers and RED

- Gradient-step (GS) denoisers Hurault, Leclaire, Papadakis, arxiv '21: Model the denoiser as $D_\theta = Id - \nabla R_\theta$ where $R_\theta(u) = \frac{1}{2}\|u - \Psi_\theta(u)\|_2^2$ with Ψ_θ being any differentiable deep network. This also provides an explicit representation of the regulariser R_θ.
- Regularisation by Denoising (RED) Romano, Elad, Milanfar, SIIMS '17: Construct explicit variational regulariser using a denoiser $R(u) = \langle u, u - D(u)\rangle$. Similar to above, if $\nabla R(u) = u - D(u)$ then RED recovers stationary point of

$$\frac{1}{2}\|Tu - y\|^2 + \alpha R(u)$$

  Weak point: gradient condition only holds for denoisers with symmetric Jacobian Reehorst, Schniter, IEEE Comp. Imaging '19.

Convergent Plug & Play

Joint work with

Andreas Hauptmann

Subhadip Mukherjee

Ferdia Sherry

Mukherjee, Hauptmann, Öktem, Pereyra, CBS, IEEE Signal Processing Magazine '22; Hauptmann, Mukherjee, CBS, Sherry '23.

Linear denoiser Plug & Play

- Under certain conditions on the learned denoiser Ψ_Θ these iterative schemes can be shown to be convergent to a fixed point.
- Learned denoisers are typically trained at fixed noise level: it is not clear how we can adjust regularisation strength to achieve convergent regularisation.
- Empirically: tune regularisation strength by denoiser scaling Xu, Liu, Sun, Wohlberg, Kamilov, IEEE TCI '21.

Hauptmann, Mukherjee, CBS, Sherry '23

Controlling the regularisation strength of a linear denoiser

- Assume a linear denoiser $D_\sigma : X \to X$: by the characterisation of proximal operators Moreau, Bulletin de la Societe mathematique de France '65; Gribonval, Nikolova, JMIV '20, D_σ is symmetric and p.s.d. iff $D_\sigma = \operatorname{prox}_J$ for some $J : X \to \mathbb{R} \cup \{\infty\}$.
- Up to additive constants, J is also determined by D_σ. Control the regularisation strength by scaling J, but how to implement this when we only have D_σ?
- Spectral filtering of D_σ is the answer:

$$\begin{tikzcd} J \arrow[r, "\text{prox}"] \arrow[d, "\text{scale by } \tau"'] & D_\sigma \arrow[d, "\text{spectral filtering by } g_\tau"] \\ \tau J \arrow[r, "\text{prox}"] & g_\tau(D_\sigma) \end{tikzcd}$$

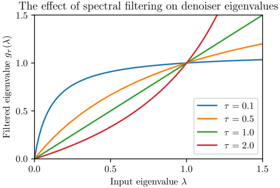


Figure: Spectral filtering of the linear denoiser by $g_\tau(\lambda) = \lambda/(\tau - \lambda(\tau - 1))$ enacts scaling of the underlying regularisation functional

Hauptmann, Mukherjee, CBS, Sherry '23

Spectral filtering of linear denoisers

- Contrast this with traditional spectral regularisation Engl, Hanke, Neubauer '96: we modify the denoiser, whereas the traditional approach modifies the SVD of the forward operator.
- We can generalise to other spectral filters satisfying technical conditions: $(1 - g_\tau(\lambda))/(\tau g_\tau(\lambda))$ is bounded above and below by positive values and converges as $\tau \to 0$.
- We get convergent PnP regularisation: if $y = T^* x$, $\|y^\delta - y\| \leq \delta$, $\tau^\delta \sim \delta$ and $\hat{x}(y^\delta, \tau^\delta) = \lim_{n\to\infty} x^n$ with $x^{n+1} = g_\tau(D_\sigma)(x^n - T^*(Tx^n - y^\delta))$, then $\hat{x}(y^\delta, \tau^\delta)$ converges to a J^*-minimising least-squares solution $x^\dagger$ of $Tx = y$ as $\delta \to 0$:

Hauptmann, Mukherjee, CBS, Sherry '23

Spectral filtering of linear denoisers

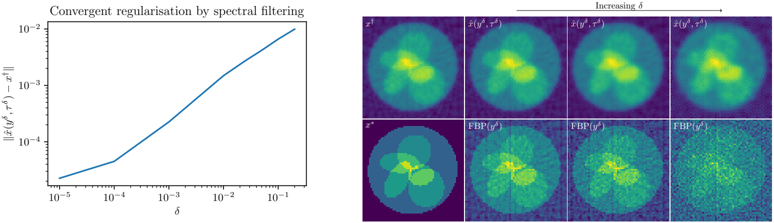


Figure: Using spectral filtering to control the regularisation strength to achieve convergent regularisation in the inverse problem of CT reconstruction

Hauptmann, Mukherjee, CBS, Sherry '23

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

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