How intelligent is artificial intelligence? – On the surprising and mysterious secrets of deep learning

Vegard Antun (UiO)
Anders C. Hansen (Cambridge, UiO)

Joint work with:

B. Adcock (SFU), M. Colbrook (Cambridge)
N. Gottschling (Cambridge) C. Poon (Bath),
F. Renna (Porto)

22 May 2019

The 'weird events' that make machines hallucinate

© Getty Images

Computers can be made to see a sea turtle as a gun or hear a concerto as someone's voice, which is raising concerns about using artificial intelligence in the real world.

What could possibly go wrong?

AI replacing standard algorithms

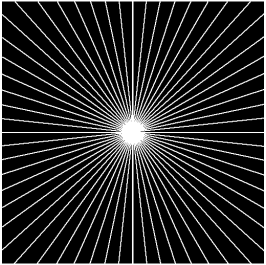
Mathematical setup

Image reconstruction in medical imaging

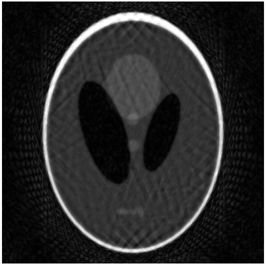
- $x \in \mathbb{C}^N$ the true image (interpreted as a vector).
- $A \in \mathbb{C}^{m \times N}$ measurement matrix ($m < N$).
- $y = Ax$ the measurements.

True image x

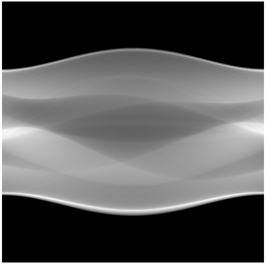
Sampling pattern Ω



$|\tilde{x}| = |A^* y|$



Sinogram $y = Ax$



Back proj. $\hat{x}_1 = A^* y$

FBP: $\hat{x}_2 = By$

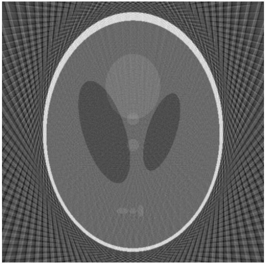


Image reconstruction methods

- Deep learning approach: For a given a set $\{x_1, \ldots, x_n\}$, train a neural network $f\colon \mathbb{C}^m \to \mathbb{C}^N$ such that

$$\|f(Ax_i) - x_i\| \ll \|A^* x_i - x_i\|$$

Image reconstruction methods

- Deep learning approach: For a given a set $\{x_1, \ldots, x_n\}$, train a neural network $f\colon \mathbb{C}^m \to \mathbb{C}^N$ such that

$$\|f(Ax_i) - x_i\| \ll \|A^* x_i - x_i\|$$

- Sparse regularization

$$\text{minimize}_{z \in \mathbb{C}^N} \|Wz\|_{\ell_1} \quad \text{subject to} \quad \|Az - y\|_{\ell_2} \leq \eta$$

Image x

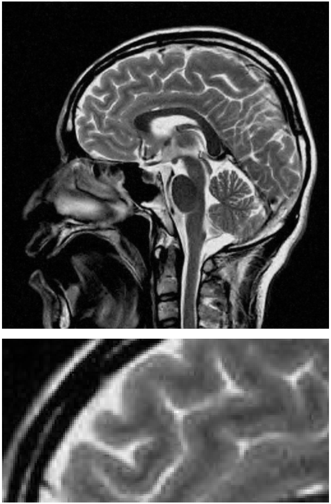
Wx

W = Wavelets

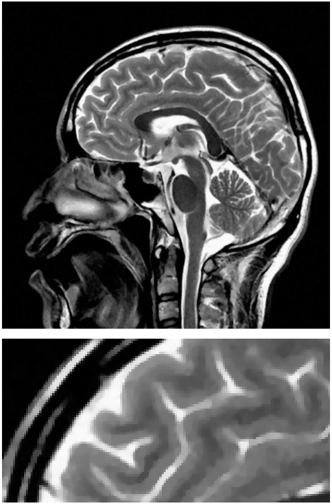
Wx

$W = \nabla$

Sparse regularization reconstruction



DB4 wavelets



TV

Typical sparse regularization result

Let $A \in \mathbb{C}^{m \times N}$ with $m < N$ and $y = Ax + e$, with $\|e\|_2 \leq \eta$. Let $W \in \mathbb{C}^{N \times N}$ be unitary, and suppose that AW^{-1} satisfies the *restricted isometry property in levels* (RIPL). Then any minimizer $\hat{x}$ of

$$\text{minimize}_{z \in \mathbb{C}^N} \|Wz\|_1 \quad \text{subject to} \quad \|Az - y\| \leq \eta$$

satisfies

$$||\hat{x} - x||_2 \lesssim \frac{\sigma_{\mathbf{s},\mathbf{M}}(Wx)_1}{\sqrt{s}} + \eta$$

where

$$\sigma_{\mathbf{s},\mathbf{M}}(Wx)_1 = \inf\{||Wx - z||_1 : z \text{ is } (\mathbf{s}, \mathbf{M})\text{-sparse}\}$$

Neural network image reconstruction approaches

- **Pure denoisers**. Train a neural network ϕ to learn the noise.

$$f(y) = A^*y - \phi(A^*y)$$

Neural network image reconstruction approaches

- **Pure denoisers**. Train a neural network ϕ to learn the noise.

$$f(y) = A^* y - \phi(A^* y)$$

- **Data consistent denoisers**. Train n networks ϕ_i, $i = 1, \ldots, n$ and ensure that the final image is consistent with your data

  1: Pick $\alpha \in [0, 1]$.
  2: Set $\tilde{y}_1 = y$.
  3: **for** $i = 1, \ldots n$ **do**
  4: $\quad \tilde{x}_i = A^* \tilde{y}_i - \phi_i(A^* \tilde{y}_i)$
  5: $\quad \hat{y} = A\tilde{x}_i$
  6: $\quad \tilde{y}_{i+1} = \alpha \hat{y} + (1 - \alpha) y$, (Enforce data consistency)
  7: **Return:** $\tilde{x}_n$.

Neural network image reconstruction approaches

- **Learn the physics**. Do **not** warm start your network with A^*.

  Rather learn $\quad f(y_i) = x_i, \quad i = 1, \ldots, n \quad$ directly

Neural network image reconstruction approaches

- **Learn the physics**. Do **not** warm start your network with A^*.

  Rather learn $\quad f(y_i) = x_i, \quad i = 1, \ldots, n \quad$ directly

- **Unravel n steps with sparse regularization solver**. Learn λ_i, K_i, and Ψ_i for $i = 1, \ldots, n$.

  1: $x_1 = A^* y$
  2: **for** $i = 1, \ldots n$ **do**
  3: $\quad \tilde{x}_{i+1} = \tilde{x}_i - (K_i)^T \Psi_i (K_i \tilde{x}_i) + \lambda_i A^*(A\tilde{x}_i - y)$
  4: **Return:** $\tilde{x}_{n+1}$.

  (omitting some details here)

Networks considered

- *AUTOMAP*
  - Low resolution images, 60% subsampling, single coil MRI.
  - B. Zhu, J. Z. Liu, S. F. Cauley, B. R. Rosen and M. S. Rosen, '*Image reconstruction by domain-transform manifold learning*', Nature, vol. 555, no. 7697, p. 487, Mar. 2018.
- *DAGAN*
  - Medium resolution, 20% subsampling, single coil MRI.
  - G. Yang, S. Yu, H. Dong, G. Slabaugh, P. L. Dragotti, X. Ye, F. Liu, S. Arridge, J. Keegan, Y. Guo et al., *DAGAN: Deep de-aliasing generative adversarial networks for fast compressed sensing MRI reconstruction*, IEEE Transactions on Medical Imaging, 2017.
- *Deep MRI*
  - Medium resolution, 33% subsampling, single coil MRI.
  - J. Schlemper, J. Caballero, J. V. Hajnal, A. Price and D. Rueckert, *A deep cascade of convolutional neural networks for MR image reconstruction*, in International conference on information processing in medical imaging, Springer, 2017, pp. 647–658.

Networks considered

- *Ell 50 and Med 50* (FBPConvNet)
  - CT or any Radon transform based inverse problem, with 50 uniformly spaced lines.
  - K. H. Jin, M. T. McCann, E. Froustey and M. Unser, '*Deep convolutional neural network for inverse problems in imaging*', IEEE Transactions on Image Processing, vol. 26, no. 9, pp. 4509–4522, 2017.
- *MRI-VN*
  - Medium to high resolution, parallel MRI with 15 coil elements and 15% subsampling.
  - K. Hammernik, T. Klatzer, E. Kobler, M. P. Recht, D. K. Sodickson, T. Pock and F. Knoll, '*Learning a variational network for reconstruction of accelerated MRI data*', Magnetic resonance in medicine, vol. 79, no. 6, pp. 3055–3071, 2018.

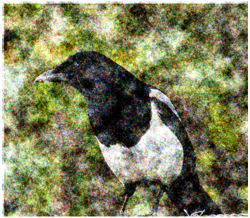
How to measure image quality?

(1) Base image

(2) Translated

(3) Add $\epsilon = 0.32$ to pixels

(4) Nosiy



(5) Another bird

(6) Different image

Image	(1)	(2)	(3)	(4)	(5)	(6)
$\ell_2 - \mathit{distance}$	0	215.04	204.80	167.26	216.44	193.15

Figure from: M. Lohne, *Parseval Reconstruction Networks*, Master thesis, UiO, 2019

Three types of instabilities

(1) Instabilities with respect to tiny perturbations. That is $\tilde{y} = A(x + r)$ with $\|r\|$ very small.

(2) Instabilities with respect to small structural changes, for example a tumour, may not be captured in the reconstructed image

(3) Instabilities with respect to changes in the number of samples. Having more information should increase performance.

V. Antun, F. Renna, C. Poon, B.Adcock, A. Hansen. *On instabilities of deep learning in image reconstruction - Does AI come at a cost?* (arXiv 2019)

Finding tiny perturbations

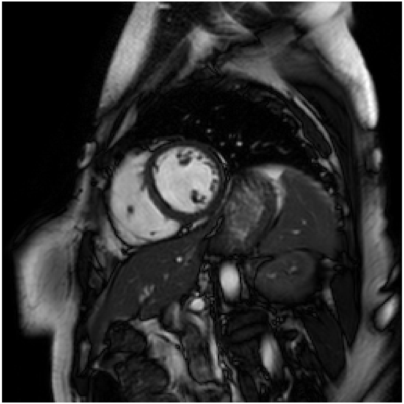
Try to maximize

$$Q_x(r) = \frac{1}{2}\|f(A(x+r)) - f(Ax)\|_{\ell_2}^2 - \frac{\lambda}{2}\|r\|_{\ell_2}^2, \quad \lambda > 0$$

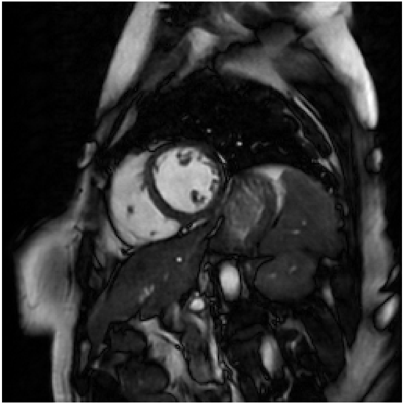
using a gradient ascent proceedure.

Tiny perturbation – Deep MRI net

$|x|$

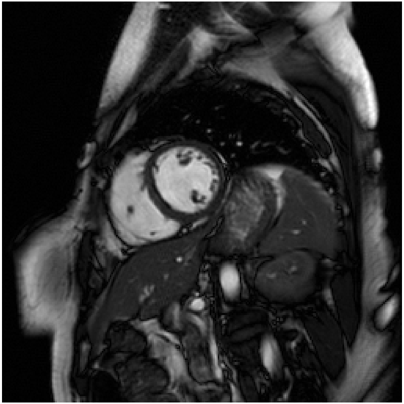


$|f(Ax)|$

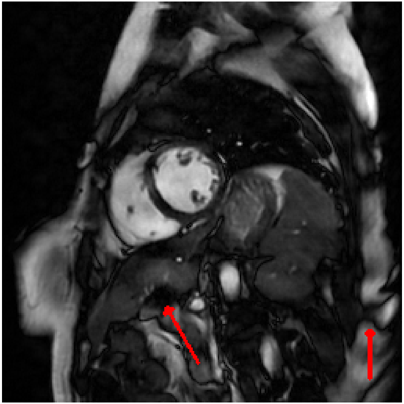


Tiny perturbation – Deep MRI net

$|x + r_1|$

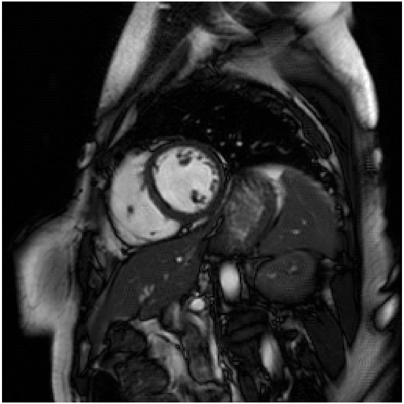


$|f(A(x + r_1))|$

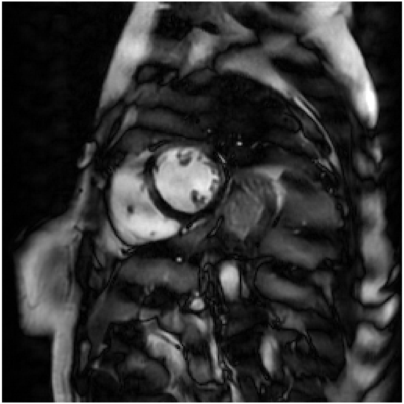


Tiny perturbation – Deep MRI net

$|x + r_2|$

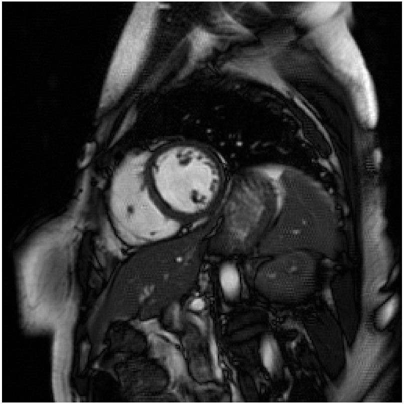


$|f(A(x + r_2))|$

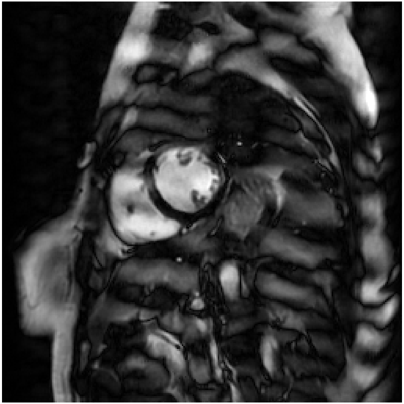


Tiny perturbation – Deep MRI net

$|x + r_3|$

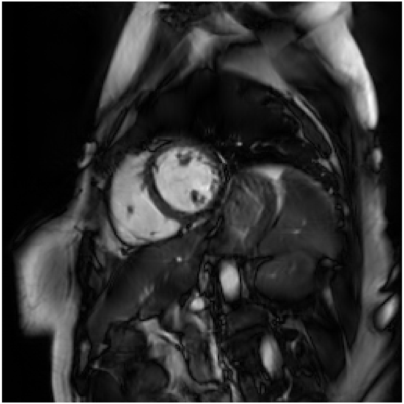


$|f(A(x + r_3))|$

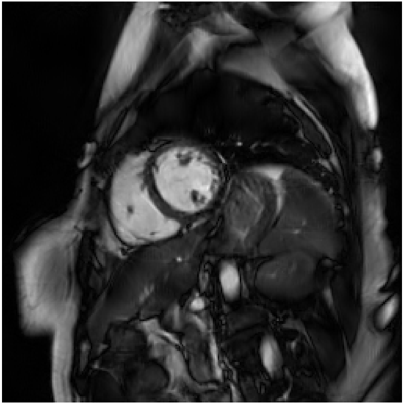


Tiny perturbation – Deep MRI net

SoA from Ax

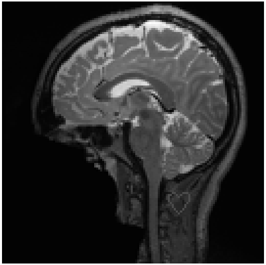


SoA from $A(x + r_3)$

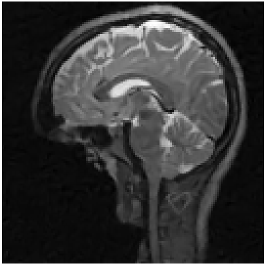


Tiny perturbation – AUTOMAP

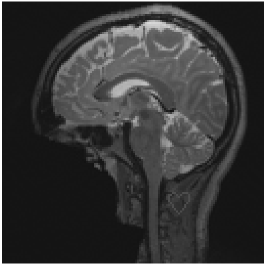
$|x|$



$|f(Ax)|$

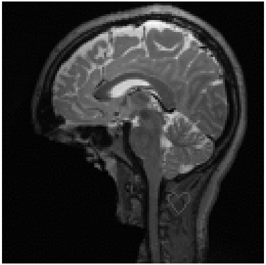


SoA from Ax

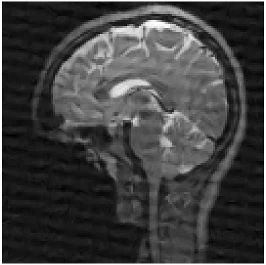


Tiny perturbation – AUTOMAP

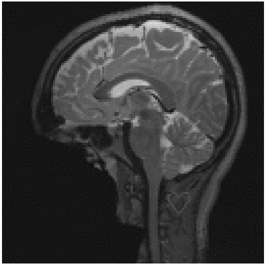
$|x + r_1|$



$|f(A(x + r_1))|$

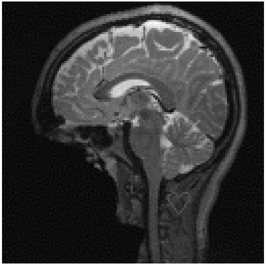


SoA from $A(x + r_1)$

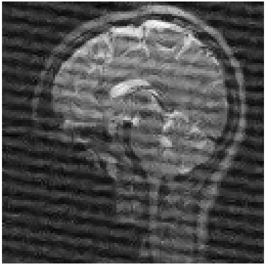


Tiny perturbation – AUTOMAP

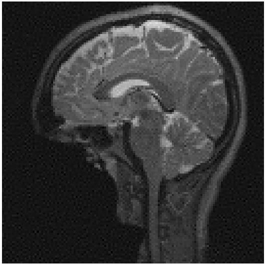
$|x + r_2|$



$|f(A(x + r_2))|$

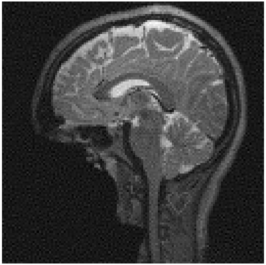


SoA from $A(x + r_2)$

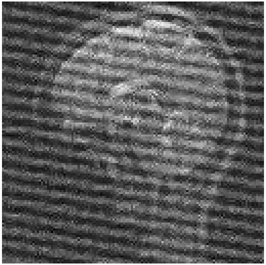


Tiny perturbation – AUTOMAP

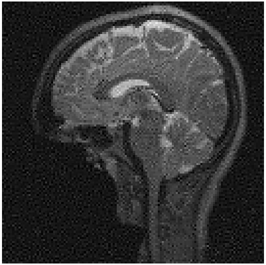
$|x + r_3|$



$|f(A(x + r_3))|$

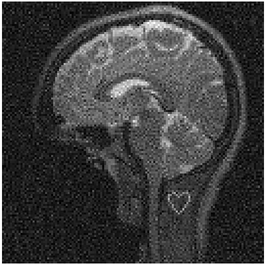


SoA from $A(x + r_3)$

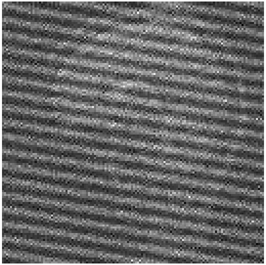


Tiny perturbation – AUTOMAP

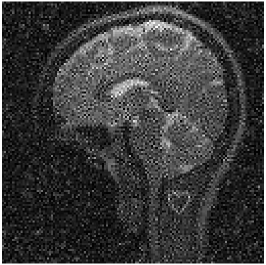
$|x + r_4|$



$|f(A(x + r_4))|$



SoA from $A(x + r_4)$



Finding tiny perturbations

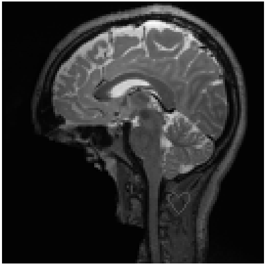
What if we tried to maximize

$$Q_x(r) = \frac{1}{2}\|f(A(x+r)) - x\|_{\ell_2}^2 - \frac{\lambda}{2}\|r\|_{\ell_2}^2, \quad \lambda > 0$$

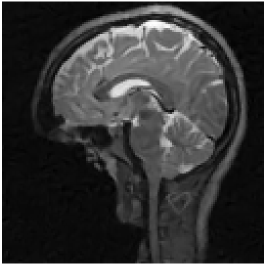
instead

Tiny perturbation – AUTOMAP

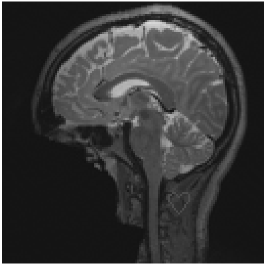
$|x|$



$|f(Ax)|$

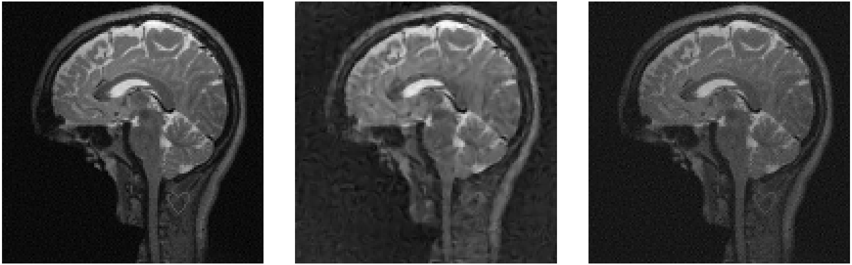


SoA from Ax



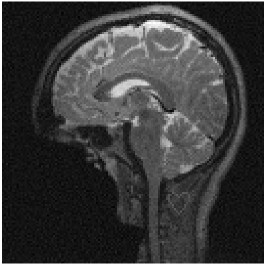
Tiny perturbation – AUTOMAP

$|x + r_1|$ $\qquad$ $|f(A(x + r_1))|$ $\qquad$ SoA from $A(x + r_1)$



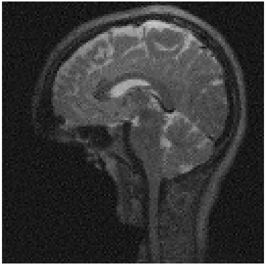
Tiny perturbation – AUTOMAP

$|x + r_2|$



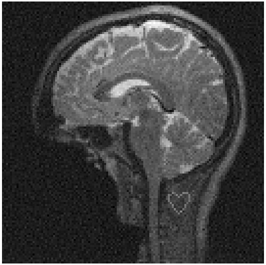
$|f(A(x + r_2))|$

SoA from $A(x + r_2)$



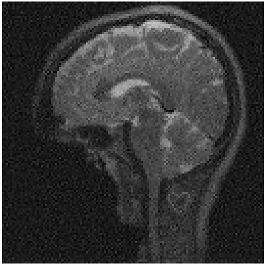
Tiny perturbation – AUTOMAP

$|x + r_3|$



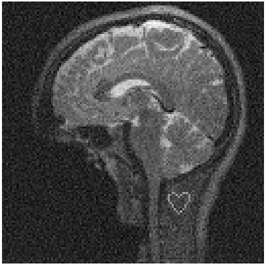
$|f(A(x + r_3))|$

SoA from $A(x + r_3)$



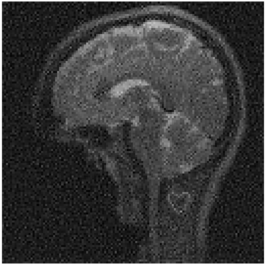
Tiny perturbation – AUTOMAP

$|x + r_4|$



$|f(A(x + r_4))|$

SoA from $A(x + r_4)$



Tiny perturbation – MRI-VN

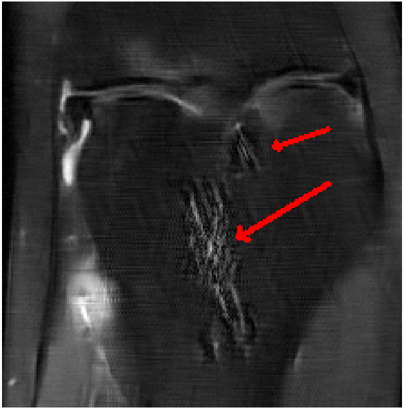
Original x

$x + r_1$

Tiny perturbation – MRI-VN

$f(Ax)$ (zoomed)

$f(A(x + r_1))$ (zoomed)

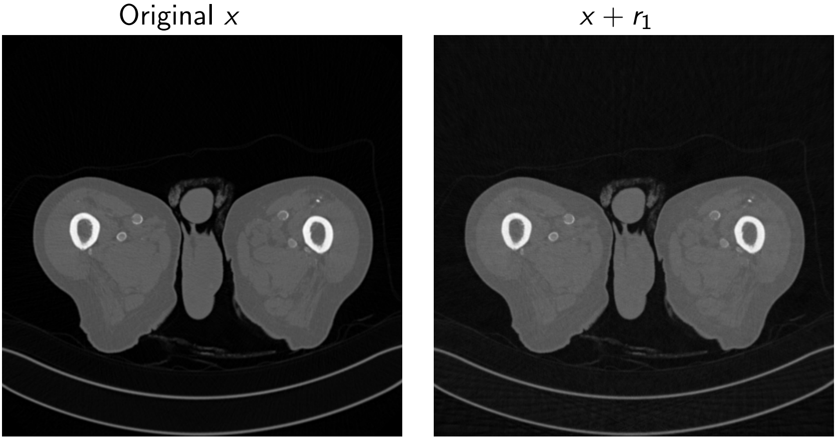


Tiny perturbation – MRI-VN

SoA from Ax (zoomed)

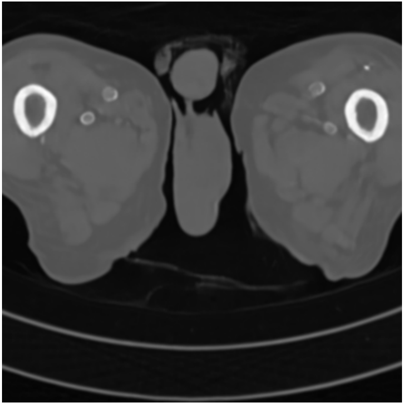
SoA from $A(x + r_1)$ (zoomed)

Tiny perturbation – Med 50

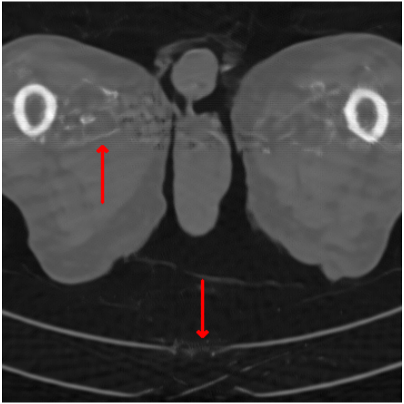


Tiny perturbation – Med 50

$f(Ax)$ (zoomed)

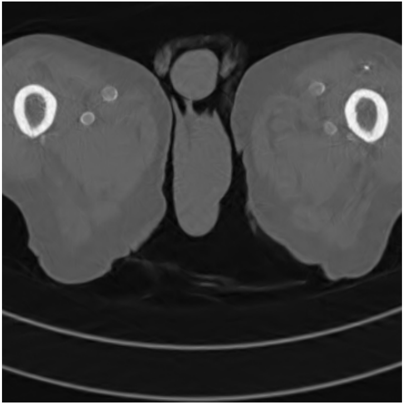


$f(A(x + r_1))$ (zoomed)



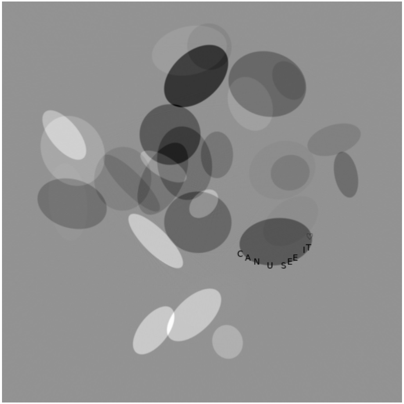
Tiny perturbation – Med 50

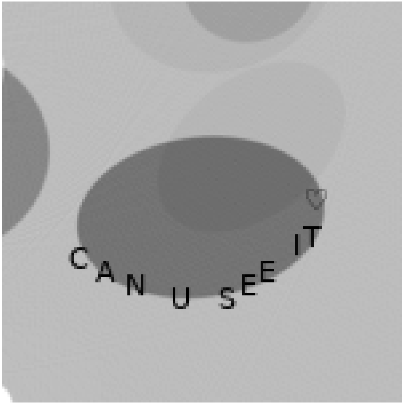
SoA from Ax (zoomed)



SoA from $A(x + r_1)$ (zoomed)

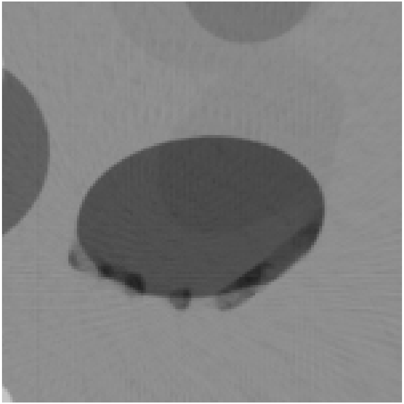
Small structural change – Ell 50





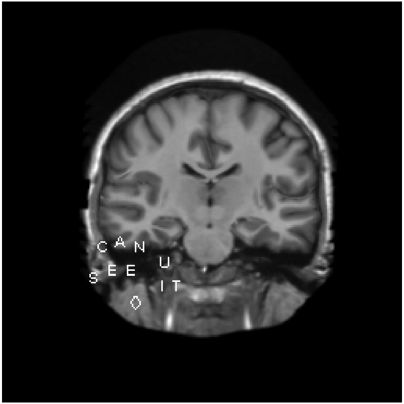
Small structural change – Ell 50

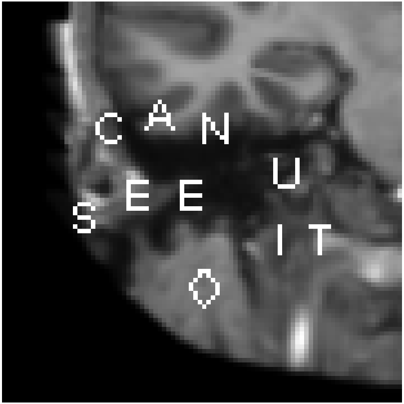
$f(Ax)$



SoA from Ax

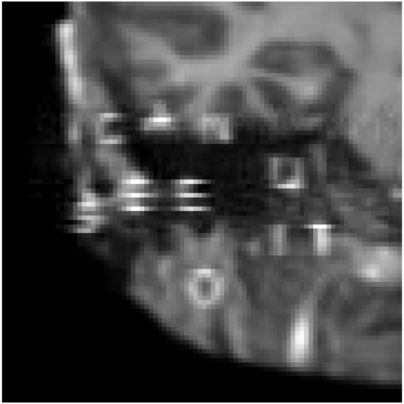
Small structural change – DAGAN



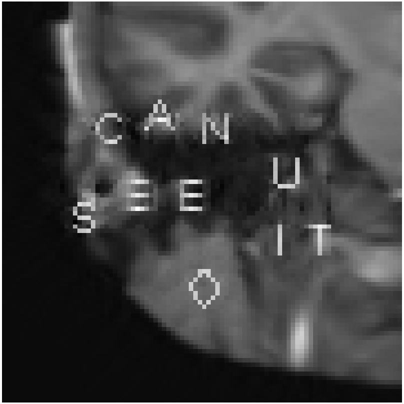


Small structural change – DAGAN

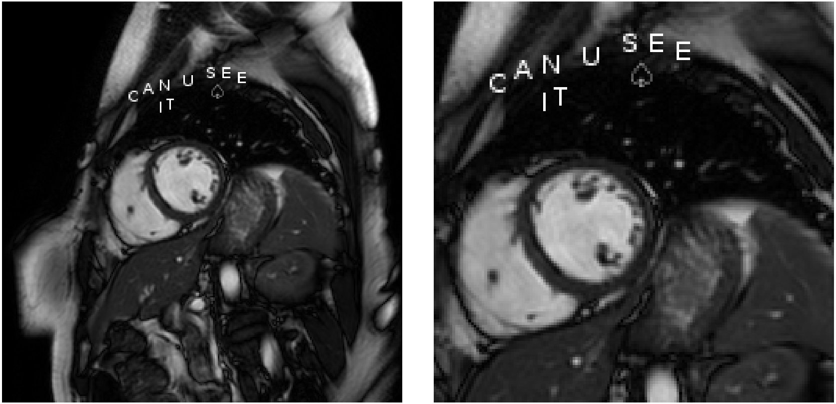
$f(Ax)$



SoA from Ax



Small structural change – Deep MRI

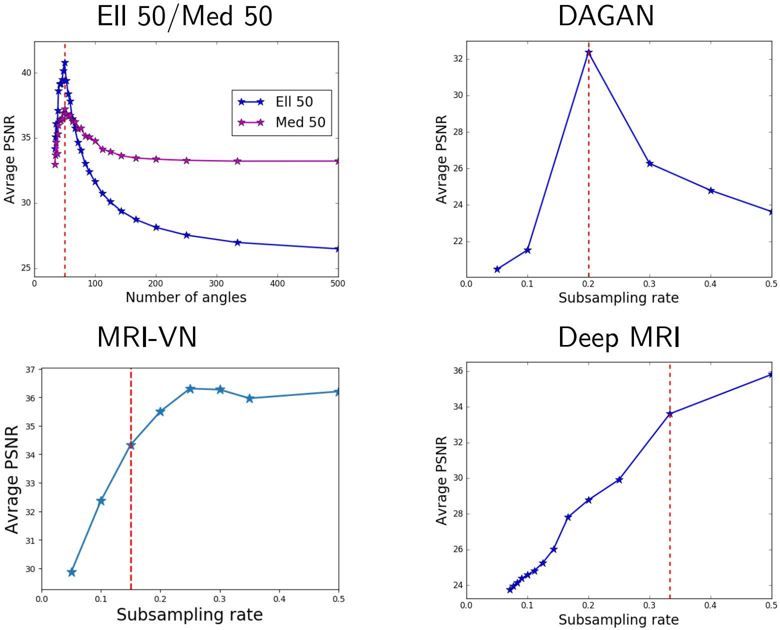


Small structural change – Deep MRI

$f(Ax)$

SoA from Ax

Adding more samples



Summary of so far...

- Tiny perturbations lead to a myriad of different artefacts.
- Variety in failure of recovering structural changes.
- Networks must be retrained on any subsampling pattern?
- Universality – Instabilities regardless of architecture?
- Rare events? – Empirical tests are needed.

Can we fix it?

- Computational power is increasing. We can train, test and at a substantial higher rate than just a few years ago.
- The datasets are growing.
- Increased knowledge about good learning techniques

Winner's Curse?
On Pace, Progress, and Empirical Rigor

D. Sculley, Jasper Snoek, Ali Rahimi, Alex Wiltschko
{dsculley, jsnoek, arahimi, alexbw}@google.com
Google AI

Abstract

The field of ML is distinguished both by rapid innovation and rapid dissemination of results. While the pace of progress has been extraordinary by any measure, in this paper we explore potential issues that we believe to be arising as a result. In particular, we observe that the rate of empirical advancement may not have been

Troubling Trends in Machine Learning Scholarship

Zachary C. Lipton* & Jacob Steinhardt*
Carnegie Mellon University, Stanford University
zlipton@cmu.edu, jsteinhardt@cs.stanford.edu

July 27, 2018

1 Introduction

Collectively, machine learning (ML) researchers are engaged in the creation and dissemination of knowledge about data-driven algorithms. In a given paper, researchers might aspire to any subset of the following goals, among others: to theoretically characterize what is learnable, to obtain

Theorem 1

Let $A : \mathbb{C}^N \to \mathbb{C}^m$ be a linear sampling map and let $f : \mathbb{C}^m \to \mathbb{C}^N$. Suppose that there are $x, \eta, \xi_\eta, \xi_x \in \mathbb{C}^N$ with $\|\xi_\eta\|, \|\xi_x\| \leq \delta \in (0, 1/2)$ such that

$$f(Ax) = x + \xi_x, \quad f(A(x+\eta)) = x + \eta + \xi_\eta, \tag{1}$$

where $\|\eta\| = 1$ and $\|A\eta\| = \delta > 0$. Then we have the following.

(i) *(Instabilities) Then the local Lipschitz constant of f at Ax, defined for $\epsilon \geq \delta > 0$, satisfies*

$$\begin{aligned} L^{\epsilon}_{Ax} &= \sup_{0<\|Az\|\leq\epsilon} \frac{\|f(Ax + Az) - f(Ax)\|}{\|Az\|} \\ &\geq \frac{1 - 2\delta}{\epsilon} \end{aligned}$$

Theorem 2

Let $A : \mathbb{C}^N \to \mathbb{C}^m$ *be a linear sampling map and let* $f : \mathbb{C}^m \to \mathbb{C}^N$*. Suppose that there are* $x, \eta, \xi_\eta, \xi_x \in \mathbb{C}^N$ *with* $\|\xi_\eta\|, \|\xi_x\| \leq \delta \in (0, 1/2)$ *such that*

$$f(Ax) = x + \xi_x, \quad f(A(x + \eta)) = x + \eta + \xi_\eta, \tag{2}$$

where $\|\eta\| = 1$ *and* $\|A\eta\| = \delta > 0$*. Then we have the following.*

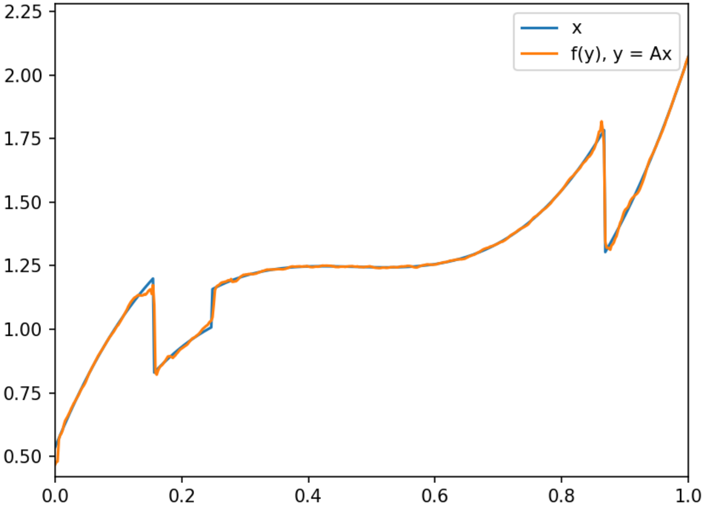
(ii) *(False positives) Moreover, there exists a perturbation* $r \in \mathbb{C}^m$ *with* $\|r\| = \delta$ *such that*

$$\|f(r + Ax) - (x + \eta)\| \leq \delta.$$

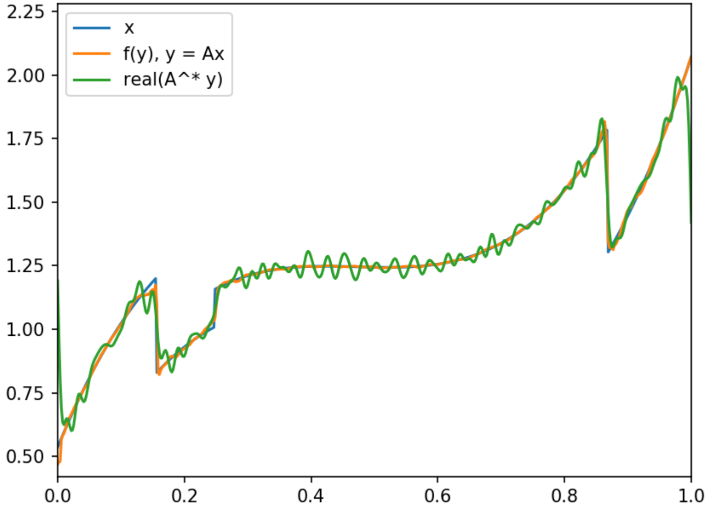
(iii) *(False negatives) and there exists a perturbation* $r \in \mathbb{C}^m$ *with* $\|r\| = \delta$ *such that*

$$\|f(r + A(x + \eta)) - x\| \leq \delta.$$

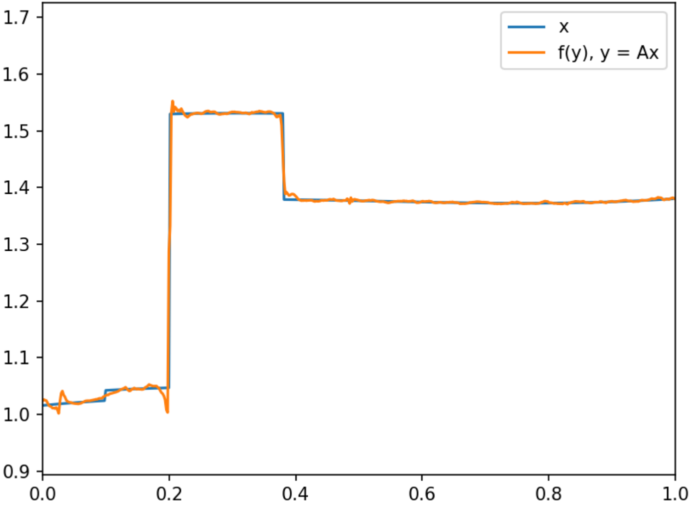
Neural network reconstruction



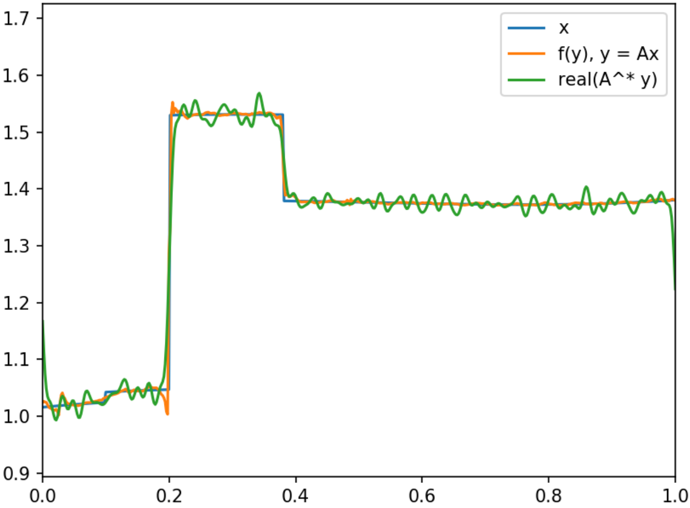
Neural network reconstruction



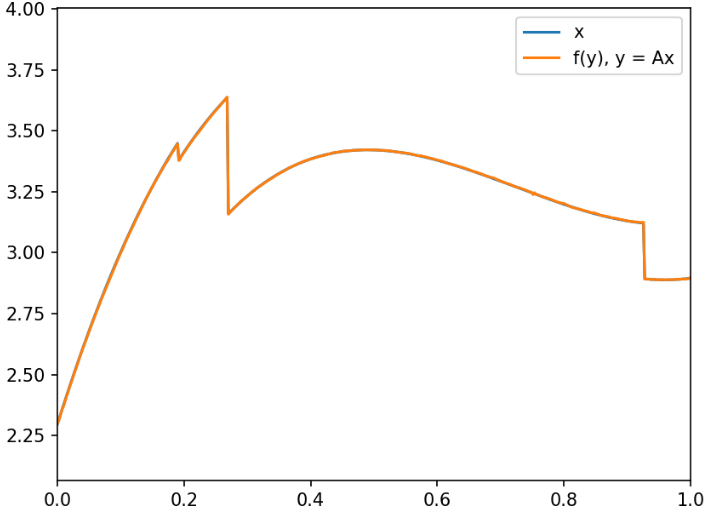
Neural network reconstruction



Neural network reconstruction

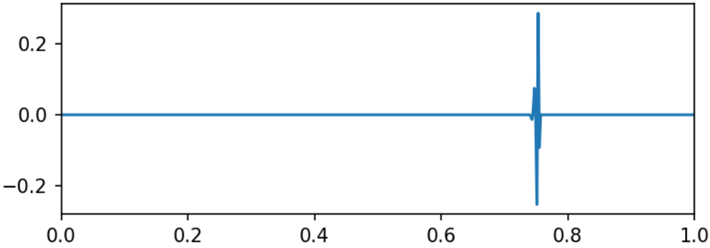


Neural network reconstruction

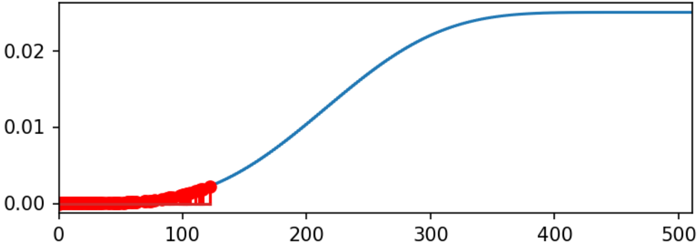


Neural network reconstruction

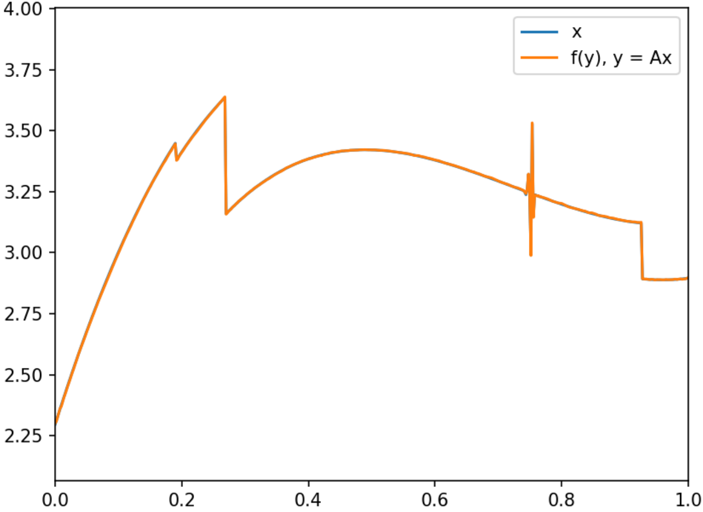
η



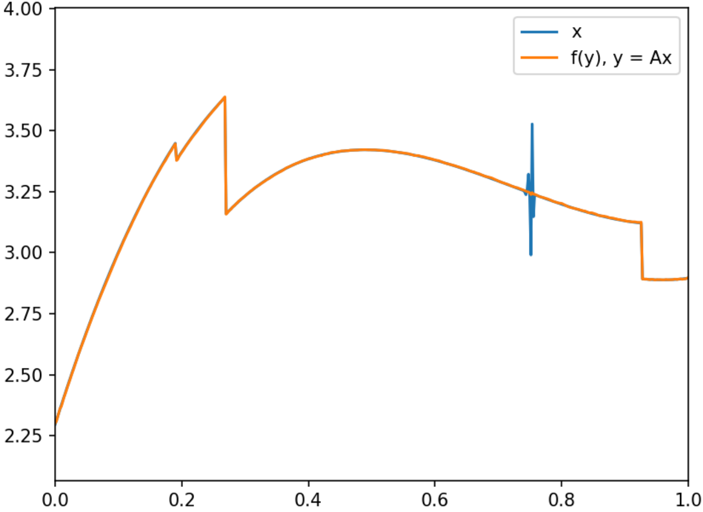
$|F\eta|$, F- Fourier matrix



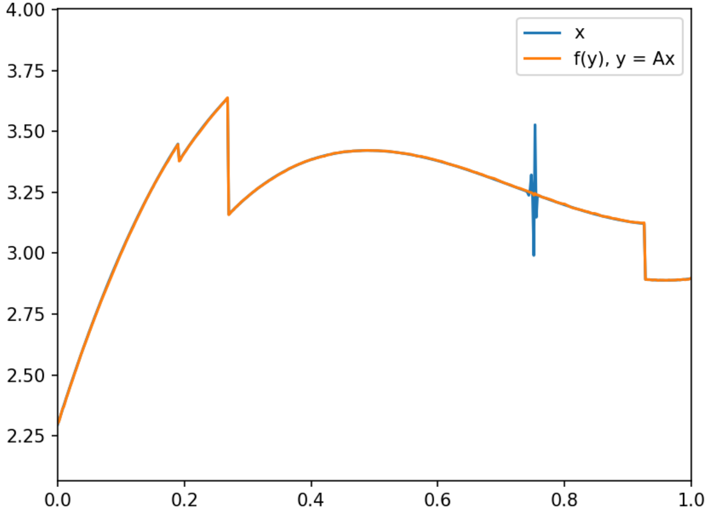
Neural network reconstruction



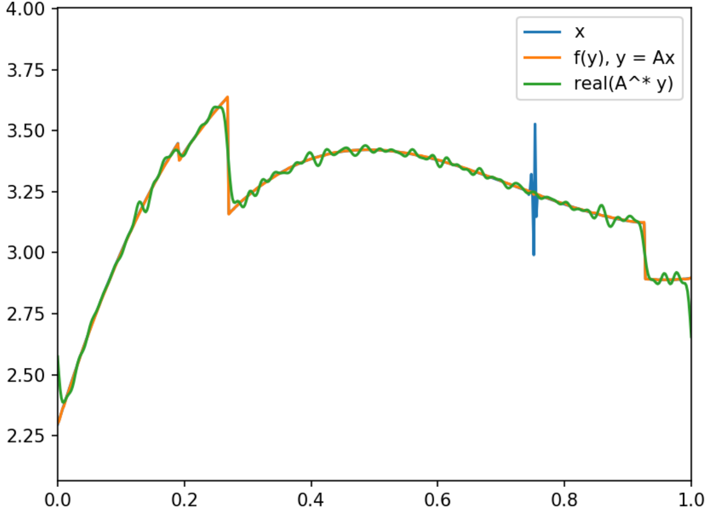
Neural network reconstruction



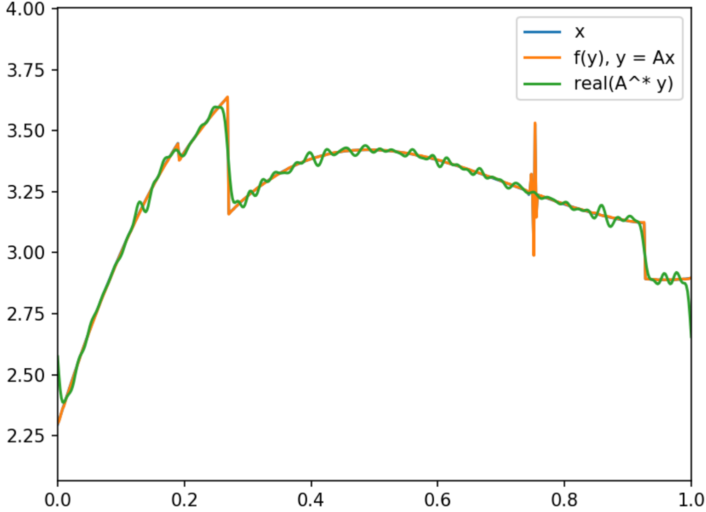
Neural network reconstruction



Neural network reconstruction



Neural network reconstruction



Typical sparse regularization result

Recall that if AW^{-1} satisfies the *restricted isometry property in levels* (RIPL).

$$\hat{x} \in \operatorname{argmin}_{z \in \mathbb{C}^N} \|Wz\|_1 \quad \text{subject to} \quad \|Az - y\| \leq \eta$$

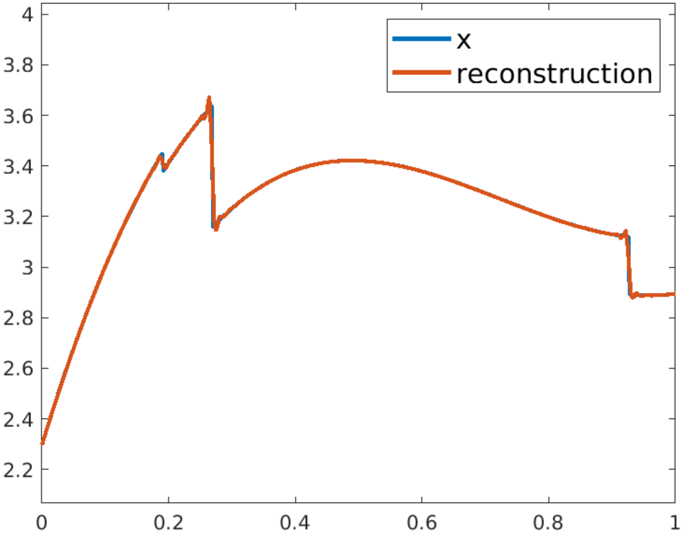
satisfies

$$||\hat{x} - x||_2 \lesssim \frac{\sigma_{\mathbf{s},\mathbf{M}}(Wx)_1}{\sqrt{s}} + \eta$$

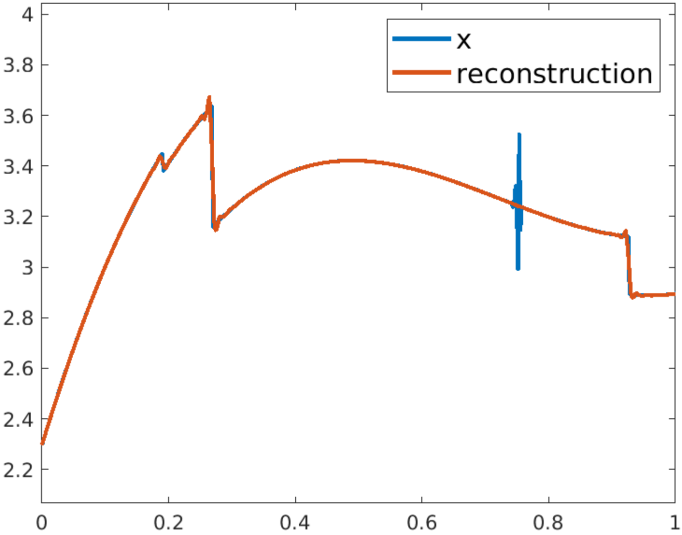
where

$$\sigma_{\mathbf{s},\mathbf{M}}(Wx)_1 = \inf\{||Wx - z||_1 : z \text{ is } (\mathbf{s}, \mathbf{M})\text{-sparse}\}$$

Wavelet reconstruction



Wavelet reconstruction



Summary

- Kernel awareness is important
- It seems hard to protect against to high preformance.
- Universality – Instabilities regardless of architecture