On the existence of stable and accurate neural networks for image reconstruction

Matthew Colbrook*, Vegard Antun, Anders Hansen***

*DAMTP, University of Cambridge **Department of Mathematics, University of Oslo

(E-mail: m.colbrook@damtp.cam.ac.uk)

Introduction

▪ Existence of stable, accurate and fast methods for image reconstruction from incomplete noisy measurements is a crucial problem in applications.

▪ Over the last decade compressed sensing and sparse regularisation have become standard tools in imaging, providing reduced scanning time and enhanced image resolution.

▪ Deep learning has emerged as a competitive new tool in image reconstruction, yet many questions remain open regarding stability and robustness to noise (a serious safety concern).

Model: recover image $x \in \mathbb{C}^N$ from noisy measurements with modality $A \in \mathbb{C}^{m\times N}, m \ll N$:

$$y = Ax + e$$

A Stability Test

▪ Suppose we have an algorithm (e.g. neural network) ϕ which seeks to recover images x. We use the stability test of [1] which searches for a perturbation r so that $\|r\|_2$ is small yet $\|\phi(y+Ar)-\phi(y)\|_2$ is large.

▪ This is done through a search for local maxima of

$$\|\phi(y+Ar)-\phi(y)\|_2^2 - \lambda\|r\|_2^2$$

via gradient ascent with momentum.

Example: ϕ taken as AUTOMAP network [2] used for MRI reconstruction with 60% subsampling (considered state-of-the-art). The results shown in Figure 1 demonstrate severe instability to adversarial (tiny) noise.

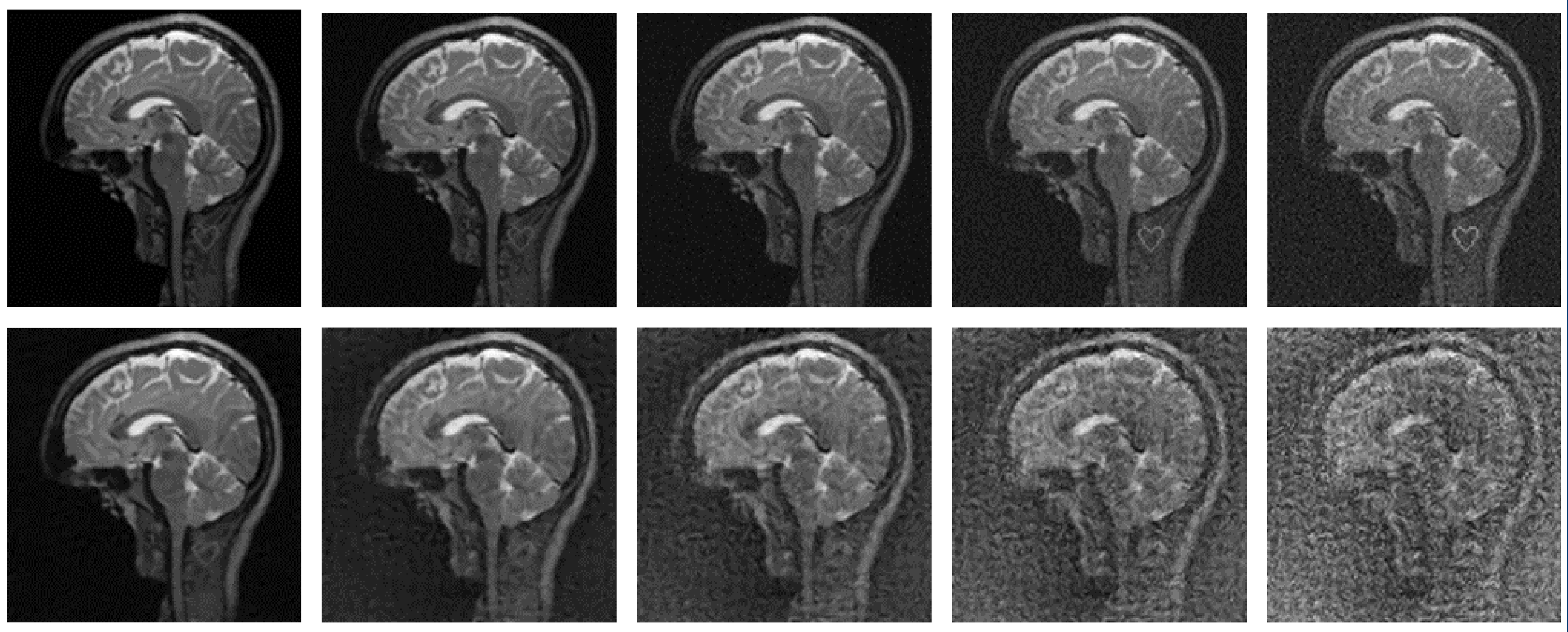

Figure 1: Stability test for AUTOMAP. Top: Image plus adversarial perturbation w.r.t network (original image on left). Bottom: Output of neural network.

Methods from Compressed Sensing

▪ Two standard optimisation problems used in compressed sensing are LASSO (L) and Basis Pursuit (BP) defined respectively as

$$\min \ \lambda\|\tilde{x}\|_1 + \|A\tilde{x}-y\|_2^2 \qquad (L)$$

$$\min \ \|\tilde{x}\|_1 \ s.t. \ \|A\tilde{x}-y\|_2 \leq \varepsilon \quad (BP)$$

▪ Perhaps surprisingly, these are susceptible to instabilities too!

▪ Figure 2 shows the instability of using FISTA [3] to solve (L). The instability of using Chambolle and Pock's primal-dual algorithm [4] to solve (BP) is similar.

▪ Let φ_A denote solution map for (L) or (BP). Given a finite set $M = \{y_j\}_{j=1}^r$, there exists a neural network Φ such that $\Phi(y_j) = \varphi_A(y_j)$ for $j = 1, \ldots, r$.

▪ Suppose as training data we can access $A, \varphi_A(y_j), y_j$ to n digits:

$$\mathcal{T} = \{A_n, \varphi_{j,n}, y_{j,n}\}_{n\in\mathbb{N}}.$$

This models computer storage and a form of numerical stability.

THEOREM: Let $K > 2$, $L \in \mathbb{N}$ and d be a norm on $\mathbb{C}^N$ with $N > 3$. Then there exists a well-conditioned class (A, M) such that:

1. No algorithm can use $\mathcal{T}$ to reconstruct Φ to K correct digits (measured in d) on M.
2. There exists a recursive algorithm that uses $\mathcal{T}$ to reconstruct a neural network approximating Φ to $K-1$ correct digits (measured in d) on M, but any algorithm producing such a network will need arbitrary many samples of elements from $\mathcal{T}$.
3. There exists a recursive algorithm that uses $\mathcal{T}$ to reconstruct a neural network approximating Φ to $K-2$ correct digits (measured in d) on M, using L samples of elements from $\mathcal{T}$.

QUESTION: Is there an algorithm computing stable reconstructions?

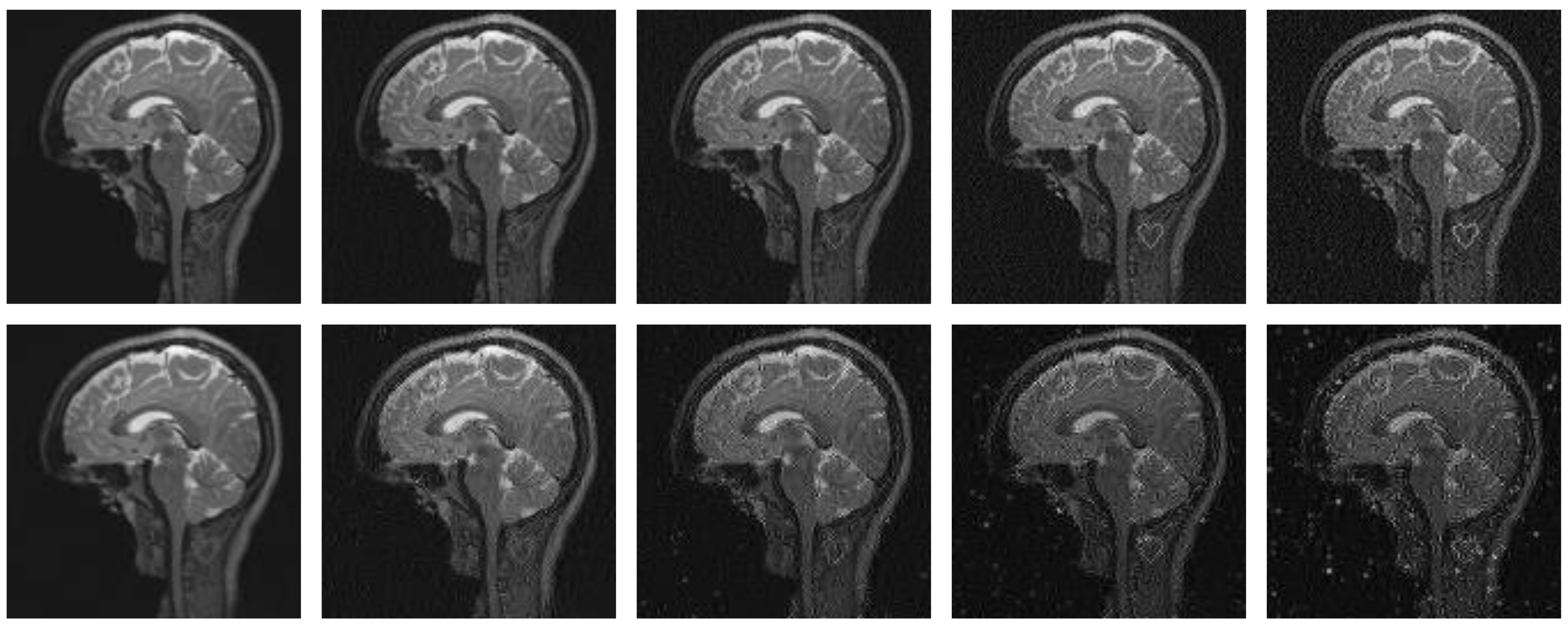

Figure 2: Stability test for FISTA solving (L) Top: Image plus adversarial perturbation w.r.t iterative solver FISTA (original image on left). Bottom: Reconstruction of algorithm.

A Stability Theorem

▪ Use the framework of sparsity in levels [5]: $\Sigma_{\mathbf{s}}$ set of vectors with s_k non-zero entries in kth wavelet level.

$$\sigma_{\mathbf{s}}(x)_{l_w^1} = \min\{\|x-v\|_{l_w^1} : v \in \Sigma_{\mathbf{s}}\}, \qquad \|q\|_{l_w^1} = \sum_{j=1}^N w_j|q_j|,$$

Where weights are constant and equal to $w_{(j)}$ in jth level, and $s = \sum s_j$.

▪ Let Wx be wavelet coefficients of x, then we expect $\sigma_{\mathbf{s}}(Wx)_{l_w^1}$ to be small.

▪ Consider the case where A is a subsampled discrete FT in d dimensions (modelling MRI), $N = 2^{r\cdot d}$ with r levels. Subsample $m_{\mathbf{k}}$ frequencies uniformly at random from tensor product of dyadic band indexed by $\mathbf{k} = (k_1, k_2, \ldots, k_d)$, so that $m = \sum m_{\mathbf{k}}$.

▪ Define the quantities $\alpha = \frac{\sum s_j w_{(j)}^2}{\min s_j w_{(j)}^2}$, $\beta = \sum s_j w_{(j)}^2$,

$$\mathcal{M}(\mathbf{s},\mathbf{k}) = \sum_{l=1}^{\|\mathbf{k}\|_\infty} s_l \prod_{i=1}^d 2^{-|k_i-l|} + \sum_{l=\|\mathbf{k}\|_\infty+1}^r s_l 2^{-2(l-\|\mathbf{k}\|_\infty)} \prod_{i=1}^d 2^{-|k_i-l|}.$$

THEOREM: Let $\varepsilon_{\mathbb{P}} > 0$ and suppose that

$$m_{\mathbf{k}} \gtrsim \alpha \cdot \mathcal{M}(\mathbf{s},\mathbf{k}) \cdot \left(\log(m)\cdot r^2 \cdot \log^2(\alpha s) + \log(\varepsilon_{\mathbb{P}}^{-1})\right).$$

Then for each $n \in \mathbb{N}$, we construct, using $\mathcal{T}$, an explicit neural network ϕ_n with $3n$ layers such that the following stable uniform recovery guarantee holds with probability at least $1-\varepsilon_{\mathbb{P}}$. For any input $y \in \mathbb{C}^m$ and image $x \in \mathbb{C}^N$ (assumed to be in some bounded ball):

$$\|\phi_n(y)-x\|_2 \lesssim \frac{\alpha^{\frac{1}{4}}}{\sqrt{\beta}}\sigma_{\mathbf{s}}(Wx)_{l_w^1} + \frac{\alpha^{\frac{1}{4}}\|A\|}{n} + \alpha^{\frac{1}{4}}\|y-Ax\|_2.$$

▪ Hence for large n, we obtain stable reconstruction near the manifold of sparse vectors.

▪ Up to log-factors, equivalent to oracle estimator (as $n \to \infty$).

▪ Given instability results for (L) and (BP), and the impossibility theorem, more subtle than unravelling your favourite optimisation solver.

▪ This stability is demonstrated in Figure 3, for the same instability test (search for network dependent adversarial perturbations) is shown.

▪ Can be extended to other modalities such as binary measurements.

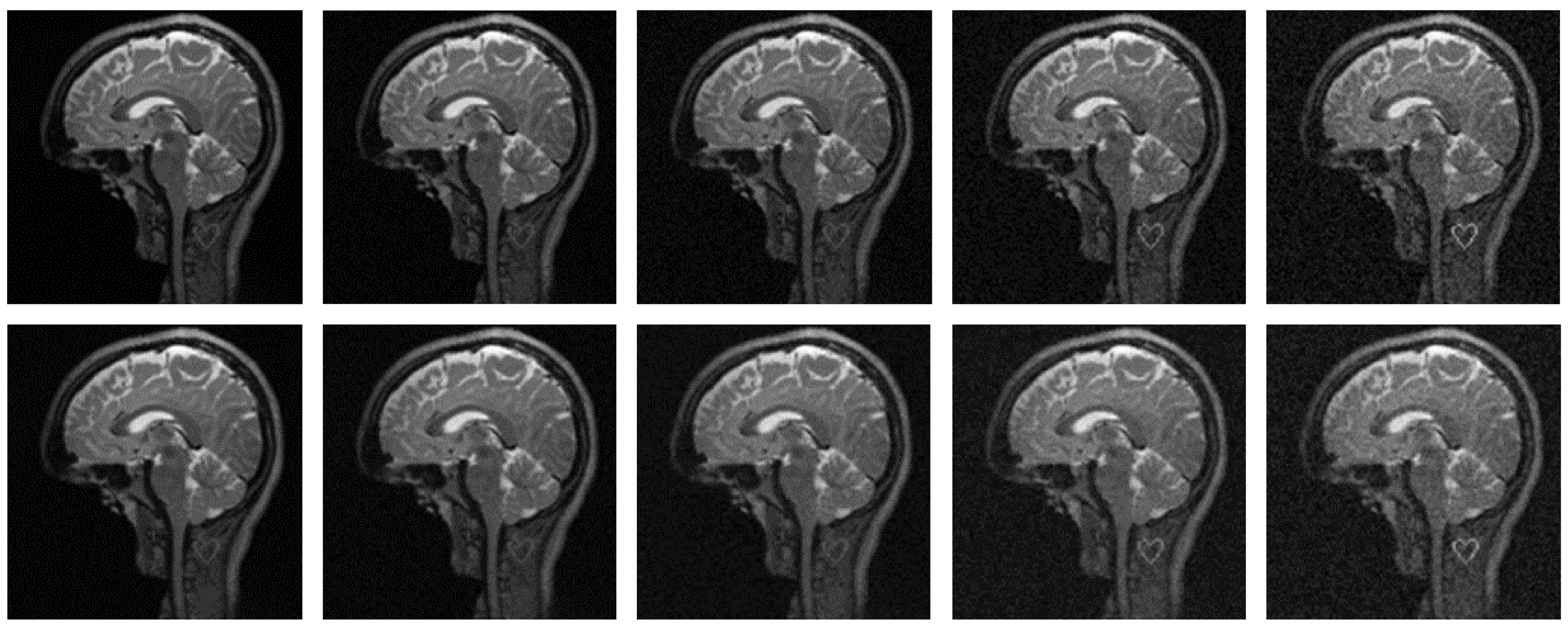

Figure 3: Stability test for new neural networks. Top: Image plus adversarial perturbation w.r.t network (original image on left). Bottom: Reconstruction of algorithm.

References:

[1] V. Antun, F. Renna, C. Poon, B. Adcock, and A. C. Hansen. On instabilities of deep learning in image reconstruction - Does AI come at a cost? Submitted, 2019.

[2] B. Zhu, J. Z. Liu, S. F. Cauley, B. R. Rosen, and M. S. Rosen. Image reconstruction by domain-transform manifold learning. Nature, 2018.

[3] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM journal on imaging sciences, 2009.

[4] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. Journal of mathematical imaging and vision, 2011.

[5] B. Adcock, A. Hansen, C. Poon, and B. Roman. Breaking the coherence barrier: A new theory for compressed sensing. In Forum of Mathematics, Sigma, 2017.