Random Neural Networks with applications to Image Recovery

Paul Hand

Northeastern University

Collaborators: Vladislav Voroninski, Wen Huang, Oscar Leong, Reinhard Heckel
Examples of inverse problem

- Compressed sensing
- Super-res
- Inpainting
- Denoising
- Phase Retrieval

\[ x_0 \xrightarrow{\Phi} \Phi(x_0) \]
A common prior: sparsity

Original image

Sparse approximation
Recovery guarantee for sparse signals

Fix \( k \)-sparse vector \( x_0 \in \mathbb{R}^n \).

Let \( A \in \mathbb{R}^{m \times n} \) be a random gaussian matrix with \( m = \Omega(k \log n) \).

\[
\begin{align*}
\min & \quad \|x\|_1 \\
\text{s.t.} & \quad Ax = Ax_0
\end{align*}
\] (L1)


The global minimizer of (L1) is \( x_0 \) with high probability.
Deep generative models

Given a set of natural images in $\mathbb{R}^n$ sampled from distribution $\pi_0$, produce a generator neural network $G : \mathbb{R}^k \mapsto \mathbb{R}^n$.

Consider a random variable $X \sim \mathcal{N}(0, I_{k \times k})$ on $\mathbb{R}^k$.

The goal is to have $k << n$ and $G(X) \equiv^d \pi_0$. 
An artificial celebrity
Generative models learn to impressively sample from complex signal classes

- Faces
- Bedrooms
- MRI scans
- Cells
- Fingerprints
Deep Compressive Sensing

\[
\min_{z \in \mathbb{R}^k} \left\| AG(z) - Ax_0 \right\|^2
\]

Bora, Jalal, Price, Dimakis
Random generative priors allow rigorous recovery guarantees

Let: \[ G : \mathbb{R}^k \rightarrow \mathbb{R}^n \]
\[ G(z) = \text{relu}(W_d \ldots \text{relu}(W_2\text{relu}(W_1z)) \ldots) \]

Given: \[ W_i \in \mathbb{R}^{n_i \times n_{i-1}}, A \in \mathbb{R}^{m \times n}, y := AG(z_0) \in \mathbb{R}^m \]

Find: \[ x_0 \]

- **Expansivity**: Let \( n_i > cn_{i-1} \log n_{i-1} \)

- **Gaussianicity**: Let \( W_i \) and \( A \) have iid Gaussian entries.

- **Biasless**: No bias terms in \( G \).
Compressive sensing with random generative prior has favorable geometry for optimization

Theorem (Hand and Voroninski, 2017)

Let $d \geq 2$. If

1. $G$ is gaussian and sufficiently expansive,

2. $m = \Omega(kd \log(\prod_{i=1}^{d} n_i))$,

then w.h.p. there exists $v_z$ such that $D_{-v_z} f < 0$ for all $z$ outside a neighborhood of $z_0$ and $-\rho_d z_0$. Additionally, 0 is a local max.
Theorem (Hand and Voroninski, 2017)

Fix $\epsilon < 1 / \text{poly}(d)$ and $d \geq 2$.

If $n_{i+1} \geq cn_i \log n_i$ and $m > cdk \log(\prod_{i=1}^{d} n_i)$, then there exists a $v_x$ s.t.

$$D_{-v_x} f(x) < 0, \quad \forall x \notin B(x_0, d\epsilon\|x_0\|_2) \cup B(-\rho_d x_0, \text{poly}(d)\epsilon\|x_0\|_2) \cup \{0\} \quad D_y f(0) < 0, \quad \forall y \neq 0,$$

simultaneously for all $x_0 \neq 0$ with probability at least $1 - C \sum n_i e^{-\gamma n_i - 1} - Ce^{-\gamma m}$. Here, $\rho_d$ is a $d$-dependent positive scalar. All constants depend polynomially on $\epsilon$. 
Proof Outline

- Explicit formula for $v_z = \nabla \text{objective}(z)$
- Explicit formula for $g_z = \mathbb{E} \nabla \text{objective}(z)$
- Show $v_z \approx g_z$ uniformly in $z$
- Show $g_z \neq 0$ away from $z_0, 0, -\rho d z_0$
Proof Requires Concentration of Discontinuous Matrix-Valued Random Functions

**Lemma:** Fix $\epsilon$. Let $W \in \mathbb{R}^{n \times k}$ have i.i.d. $\mathcal{N}(0, 1/n)$ entries. If $n > ck \log k$, then with probability at least $1 - 8ne^{-\gamma k}$, we have for all $x, y \neq 0 \in \mathbb{R}^k$,

$$\left\| \sum_{i=1}^{n} 1_{w_i \cdot x > 0} 1_{w_i \cdot y > 0} \cdot w_i w_i^T - \mathbb{E}[\cdots] \right\| \leq \epsilon$$

The constants depend polynomially on $\epsilon$. 

Lemma (Baraniuk et al. 2008)

Let $A \in \mathbb{R}^{m \times n}$ have iid $\mathcal{N}(0, 1/m)$ entries. Fix $\epsilon$. Fix a subspace $T \subset \mathbb{R}^n$ of dimension $2k < m$. With probability at least $1 - (c_1/\epsilon)^{2k} e^{-\gamma_1 \epsilon^{m}}$, \[
|\langle Ax, Ay \rangle - \langle x, y \rangle| \leq \epsilon \|x\|_2 \|y\|_2, \quad \forall x, y \in T
\]
RIP-like property and control of subspaces

Lemma (Baraniuk et al. 2008)

Let \( A \in \mathbb{R}^{m \times n} \) have iid \( \mathcal{N}(0, 1/m) \) entries. Fix \( \epsilon \). Fix a subspace \( T \subset \mathbb{R}^n \) of dimension \( 2k < m \). With probability at least

\[
1 - (c_1/\epsilon)^{2k} e^{-\gamma_1 \epsilon m},
\]

\[
|\langle Ax, Ay \rangle - \langle x, y \rangle| \leq \epsilon \|x\|_2 \|y\|_2, \quad \forall x, y \in T
\]

- Apply lemma to all subspaces given by ReLU patterns
- \( \|W_{+,x}^t A^t A W_{+,y} - W_{+,x}^t W_{+,y}\| \leq \epsilon \) \( \forall x, y \) whp if \( m \gtrsim k \log n \)
Guarantees for compressive sensing under generative priors have been extended to convolutional architectures

Invertibility of Convolutional Generative Networks from Partial Measurements

Fangchang Ma*
MIT
fcma@mit.edu
Ulas Ayaz*
MIT
uayaz@mit.edu
uayaz@lyft.com
Sertac Karaman
MIT
sertac@mit.edu

Abstract

The problem of inverting generative neural networks \textit{(i.e., to recover the input latent code given partial network output)}, motivated by image inpainting, has recently been studied. Prior work focused on fully-connected networks for mathematical simplicity. In this work, we present new results on convolutional networks, which are more widely used. The network inversion problem is highly non-convex, and hence is typically computationally intractable and without optimality guarantees. However, we rigorously prove that, for a 2-layer convolutional generative network with ReLU and Gaussian-distributed random weights, the input latent code can be deduced from the network output efficiently using simple gradient descent. This new theoretical finding implies that the mapping from the low-dimensional latent space to the high-dimensional image space is one-to-one, under our assumptions. In addition, the same conclusion holds even when the network output is only partially observed \textit{(i.e., with missing pixels)}. We further demonstrate, empirically, that the same conclusion extends to networks with multiple layers, other activation functions (leaky ReLU, sigmoid and tanh), and weights trained on real datasets.
Random Neural Networks as Unlearned Priors

Untrained neural nets as a model of natural images

- No training data
- Nonlinear signal model
The **deep decoder** is an image generating network that is

- untrained
- yields compression and image restoration performance on par with state-of-the-art
- underparameterized
- is simple
- comes with theoretical guarantees
The deep decoder is a non-learned deep image model:

$$\text{img} = \text{DD(}C\text{')}$$

Best representation of an image:

$$\hat{C} = \arg \min_C \|\text{img} - \text{DD}(C)\|_2$$
Compression

compression by factor 30.8
Image compression

PSNR's for random images from ImageNet

The deep decoder is an extremely concise image representation
The deep decoder

- Linear combinations, linear upsampling, ReLU
  \[ B_{i+1} = \text{relu}(U_i B_i C_i) \]

- Linear combinations, sigmoid:
  \[ DD(C) = \text{sigm}(B_d C_d) \]

- \( B_1 \) has full rank and is fixed

- Parameters: \( C_1, \ldots, C_d \)
Solving inverse problems with the deep decoder
Image recovery with models

\[ \text{img} = \text{DD}(\mathbf{C}) \]

Recovery:

1. Estimate model parameter

\[ \hat{\mathbf{C}} = \arg \min_{\mathbf{C}} \| \text{measurement} - f(\text{DD}(\mathbf{C})) \|_2 \]

2. Recover image as \( \text{DD}(\hat{\mathbf{C}}) \)
Deep decoder is on par with state of the art for denoising.
Inpainting

original image  

noisy image  

12.9dB  

DD  

33.6dB  

DIP  

34dB
Compressive sensing in MRI

on facebook/NYU fastMRI dataset

full rec. mask

LS 25.82dB
L1-Wav 29.04dB
DD 30.08dB
Why does the deep decoder work?

With many parameters, deep decoder can represent any image.

With few parameters, the deep decoder can approximate natural images well.

With few parameters, the deep decoder cannot represent noise well.
Why does the deep decoder work?

With many parameters, deep decoder can represent any image.

With few parameters, the deep decoder can approximate natural images well.

With few parameters, the deep decoder cannot represent noise well.
Why does the deep decoder denoise so well?

... because the deep decoder can not fit noise!
Theory: Deep Decoder can only fit so much noise

Theorem (Heckel and Hand, 2018)

- DD: one layer deep decoder with \( m \) parameters
- \( n \): output image dimension
- \( z \): be standard Gaussian noise

With high probability

\[
\min_C \|z - DD(C)\|_2^2 \geq \|z\|_2^2 \left(1 - c \frac{m}{n}\right)
\]

Typically \( m \ll n \), so deep decoder provably filters out noise
Denoising rates

Denoising rate determined by number of model parameters, $m$, and model complexity

Noise energy reduced to fraction:

- Subspace in $\mathbb{R}^n$ of dimension $m$:
  \[
  \frac{m}{n}
  \]

- Deep decoder:
  \[
  \frac{m}{n} d \log(n)
  \]
Proof Strategy

$G(C)$ lies in union of $n^{k^2}$ many $k^2$-dimensional subspaces
Proof

Consider network with one layer:

\[ G(C) = \text{relu}(UB_C)c'. \]

Let \( W_i = \text{diag}(1, 0, 1, \ldots, 0) \) such that:

\[
G(C) = \begin{bmatrix}
W_1A, & \ldots, & W_kA
\end{bmatrix}_{n \times k^2}
\begin{bmatrix}
c_1c_1' \\
\vdots \\
c_kc_k'
\end{bmatrix}
\]

Image \( G(C) \) lies in union of \( k^2 \)-dimensional subspaces

How many subspaces?
Proof

Consider network with one layer:

\[ G(C) = \text{relu}(U B C) c'. \]

Let \( W_i = \text{diag}(1, 0, 1, \ldots, 0) \) such that:

\[ G(C) = [W_1 A, \ldots, W_k A] n \times k^2 \begin{bmatrix} c_1 c_1' \\ \vdots \\ c_k c_k' \end{bmatrix} \]

Image \( G(C) \) lies in union of \( k^2 \)-dimensional subspaces

How many subspaces? Number of sets \( \{W_1, \ldots, W_k\} \).

Lemma

For any \( A \in \mathbb{R}^{n \times k} \), \( |\{\text{sign}(A c) \mid c \in \mathbb{R}^k\}| \leq n^k \).
How can linear upsampling, ReLUs, and linear combinations synthesize images efficiently?
Random Neural Networks

- Random Neural Networks as a model for trained networks can allow rigorous recovery guarantees for inverse problems with generative priors.
- Random Neural Networks can be fit to a single image in order to perform restoration.