The Report committee for David Michael Van Veen
Certifies that this is the approved version of the following report:

Compressed Sensing Recovery

With Unlearned Neural Networks

SUPERVISING COMMITTEE:

Sriram Vishwanath, Supervisor

Georgios-Alex Dimakis, Co-supervisor

Compressed Sensing Recovery with Unlearned Neural Networks

by

David Michael Van Veen

REPORT

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Master of Science in Engineering

THE UNIVERSITY OF TEXAS AT AUSTIN

May 2019

Compressed Sensing Recovery with

Unlearned Neural Networks

David Michael Van Veen, M.S.E.

The University of Texas at Austin, 2019

Supervisor: Sriram Vishwanath

Co-supervisor: Georgios-Alex Dimakis

This report investigates methods for solving the problem of compressed sensing, in which the goal is to recover a signal from noisy, linear measurements. Compressed sensing techniques enable signal recovery with far fewer measurements than required by traditional methods such as Nyquist sampling. Signal recovery is an incredibly important area in application domains such as consumer electronics, medical imaging, and many others.

While classical methods for compressed sensing recovery are well established, recent developments in machine learning have created wide opportunity for improvement. In this report I first discuss pre-existing approaches, both classical and modern. I then present my own contribution to this field: creating a method using untrained machine learning models. This approach has several advantages which enable its use in complex domains such as medical imaging.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

We consider the well-studied compressed sensing problem of recovering an unknown signal $x^* \in \mathbb{R}^n$ by observing a set of noisy measurements $y \in \mathbb{R}^m$ of the form

$$y = Ax^* + \eta. \tag{1.1}$$

Here $A \in \mathbb{R}^{m \times n}$ is a known measurement matrix, typically generated with random independent Gaussian entries. Since the number of measurements m is smaller than the dimension n of the unknown vector x^*, this is an under-determined system of noisy linear equations and hence ill-posed. There are many solutions, and some structure must be assumed on x^* to have any hope of recovery. Pioneering research [26, 13, 15] established that if x^* is assumed to be sparse in a known basis, a small number of measurements will be provably sufficient to recover the unknown vector in polynomial time using methods such as Lasso [83].

Sparsity in a known basis has proven successful for multiple signals of interest, but more complex models with additional structure have been recently proposed such as model-based compressive sensing [6] and manifold models [42, 41, 30]. Recently Bora et al. [9] showed that deep generative models can be used as excellent priors for images. They also showed that backpropagation can be used to solve the signal recovery problem by performing gradient descent in the generative latent space. Bora et al. [9] were able to reconstruct images with significantly fewer measurements compared to Lasso for a given reconstruction error. Compressed sensing using deep generative models was further improved in very recent work [86, 36, 47, 81, 34, 3]. Additionally a theoretical analysis of the nonconvex gradient descent algorithm [9] was proposed by Hand et al. [39] under some assumptions on the generative model.

Inspired by these impressive benefits of deep generative models, we chose to investigate the application of such methods for medical imaging, a canonical application of compressive sensing. A significant problem, however, is that all these previous methods require the existence of *pre-trained* models. While this has been achieved for various types of images, e.g. human faces of CelebA [56] via DCGAN [76], it remains significantly more challenging for medical images [96, 79, 72]. Instead of addressing this important problem in generative models, we found an easier way to circumvent it.

Surprising recent work by Ulyanov et al. [89] proposed Deep Image Prior (DIP), which uses *untrained* convolutional neural networks to perform inpainting and denoising. In DIP a convolutional neural network generator (e.g. DCGAN) is initialized with random weights; these weights are subsequently optimized to make the network produce an output as close to the target image as possible. This procedure is image-agnostic, using no prior information from other images. The prior is enforced only by the fixed convolutional structure of the generator network.

Our Contributions: Our novel contribution in this report is DIP for compressed sensing (CS-DIP). The basic method is as follows: initialize a DCGAN generator with random weights and optimize them using gradient descent to make the network produce an output which *agrees with the observed measurements* as much as possible. This method includes a novel *learned regularization* technique which regularizes the DCGAN weights throughout the optimization process.

Our results show that we require significantly fewer measurements to obtain similar reconstruction error compared to classical Lasso methods and even outperform unlearned BM3D-AMP and TVAL3 when the number of measurements is small. However, for a high number of measurements, BM3D-AMP tends to outperform our method.

We note that our reconstruction quality is not as high as the gains achieved by Bora et al. [9], but we have the advantage of not requiring a generative model pre-trained over a large dataset. We only require access to measurements from a small number of images for hyperparameter tuning and learned regularization. This is significantly easier than training a generative model on medical imaging tasks [96, 79, 72].

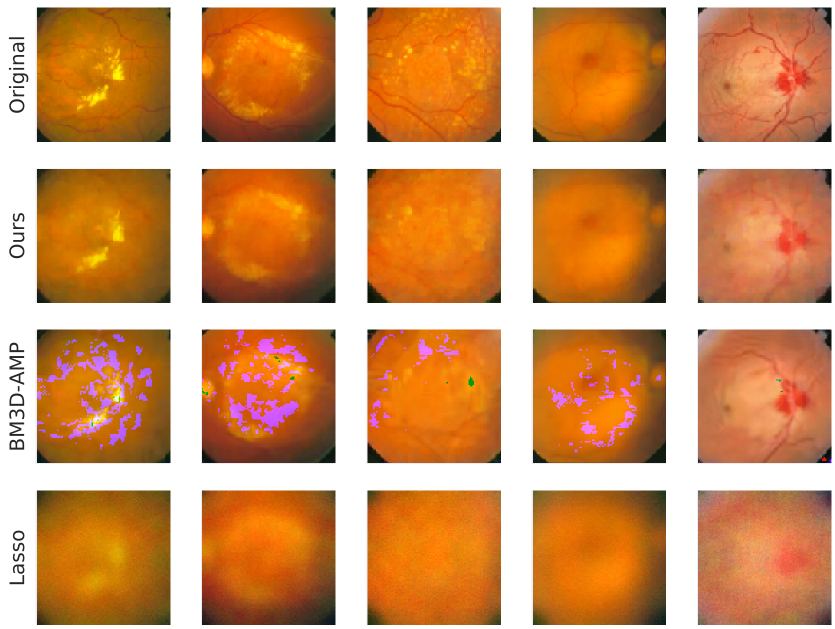


Figure 1.1: Reconstruction results on retinopathy images for m = 2000 measurements (of n = 49152 pixels). From top to bottom row: original image, reconstructions by our algorithm, then reconstructions by baselines BM3D-AMP and Lasso. In this case the number of measurements is much smaller than the number of pixels (roughly 4% ratio) and our algorithm successfully reconstructs the whole image. BM3D-AMP produces sharp reconstructions but fails to converge on some parts of the image, as demonstrated by erroneous green and purple pixels. We recommend viewing in color.

Chapter 2

Background

2.1 Compressed Sensing: Classical Sparsity Approaches

Recall Eqn. 1.1, $y = Ax^* + \eta$, where the goal is to solve for $x^* \in \mathbb{R}^n$ given measurements $y \in \mathbb{R}^m$ such that $m \ll n$. Compressed sensing leverages the convenient fact that many natural signals have concise representations in some basis. Consider, for example, the wavelet transform of a natural image. Most wavelet coefficients are small, and the few large coefficients capture most of the information.

Thus a classical assumption made in compressed sensing is that the vector x^* is k-sparse in some basis such as wavelet or discrete cosine transform (DCT). The problem then becomes finding the sparsest solution to the underdetermined linear system of equations, i.e.

$$\begin{aligned} x^* = \arg\min_x \quad & \|\Phi x\|_0 \\ \text{s.t.} \quad & y = Ax, \end{aligned} \tag{2.1}$$

where Φ is the basis transformation. Solving this optimization problem is NP-hard in general; however, this objective can be relaxed to the ℓ_1-norm. Minimizing ℓ_1 with respect to linear constraints can be recast as a linear program, leveraging a rich history of convex programming methods. Candes et al. [14] prove that if Φx is sufficiently sparse, then recovery via ℓ_1-minimization is exact given sufficient number of measurements m.

Another fundamental result in compressed sensing is a condition on the measurement matrix, $A \in \mathbb{R}^{m \times n}$, called the Restricted Isometry Property (RIP) [15]. Matrix A satisfies RIP if

$$(1 - \delta_k)\|v\|_2^2 \leq \|Av\|_2^2 \leq (1 + \delta_k)\|v\|_2^2 \tag{2.2}$$

for all k-sparse vectors v, where the isometry constant $\delta_k \in [0, 1)$ is not too close to one. This implies that k-sparse vectors cannot be in the null space because A preserves the length of k-sparse vectors. Similar conditions on the measurement matrix exist, such as the Restricted Eigenvalue Condition (REC) [26, 83].

While initial work in compressed sensing utilized convex programs, these methods are computationally prohibitive for recovering large signals such as images. Thus less expensive iterative methods were developed, such as matching pursuit [61], orthogonal matching pursuit [87], compressive sampling matching pursuit [70], approximate message passing [27], iterative hard-thresholding [8], and iterative soft-thresholding [33], among others. These methods have been compared in different application domains, e.g. face recognition [97]. Below we provide a brief discussion of these methods that are relevant in current state-of-the-art recovery algorithms.

With these developments x^* can be provably recovered in polynomial time via convex relaxations [88] or iterative methods. Another name for this problem is high-dimensional sparse linear regression, for which there is extensive literature regarding assumptions on A, numerous recovery algorithms, and variations of RIP and REC [7, 71, 1, 5, 57].

We will now discuss recent state-of-the-art unlearned methods for compressed sensing recovery, BM3D-AMP [68, 67] and TVAL3 [99, 54], as these methods are baselines against which we later compare our proposed method, compressed sensing with deep image prior (CS-DIP).

2.2 Compressed Sensing: Modern, Unlearned Approaches

While sparsity has proven successful for compressed sensing recovery of many signals, this assumption is not as effective for imaging applications because natural images do not have sparse representations in any known basis. For example given a natural image, the majority of wavelet coefficients are non-zero; many of these non-zero coefficients have large magnitude. Thus sparsity-based algorithms are not equipped to recover this signal of natural images exactly.

2.2.1 BM3D-AMP

Recent work has used other priors to solve linear inverse problems. Plug-and-play priors [91, 18] and Regularization by Denoising [77] have shown how image denoisers can be used to solve general linear inverse problems. A key example of this is BM3D-AMP, which applies a Block-Matching and 3D filtering (BM3D) denoiser to an Approximate Message Passing (D-AMP) algorithm [68, 67]. We will begin our discussion of BM3D-AMP by providing relevant background as to how Approximate Message Passing (AMP) connects to iterative soft-thresholding (IST) before introducing the BM3D denoiser.

With the goal of recovering x_o, IST algorithms take the form

$$\begin{aligned} x^{t+1} &= \eta_\tau(\mathbf{A}^* z^t + x^t), \\ z^t &= y - \mathbf{A}x^t, \end{aligned} \tag{2.3}$$

where $\eta_\tau(y) = (|y| - \tau)_+ \mathrm{sign}(y)$ is a thresholding non-linearity, x^t is the estimate of x_o at iteration t, and z^t denotes the residual estimate $y - Ax_o$ at iteration t.

AMP extends IST by adding an Onsager correction term to the residual:

$$\begin{aligned} x^{t+1} &= \eta_\tau(\mathbf{A}^* z^t + x^t), \\ z^t &= y - \mathbf{A}x^t + \tfrac{n}{m} z^{t-1} \langle \eta_\tau'(\mathbf{A}^* z^{t-1} + x^{t-1}) \rangle. \end{aligned} \tag{2.4}$$

where $\langle \cdot \rangle$ represents the average of a vector, η_τ' denotes the derivative of η_τ, and $\frac{n}{m}\langle \eta_\tau'(\mathbf{A}^* z^{t-1} + x^{t-1}) \rangle$ is the Onsager correction term. The inclusion of this term imposes the effective noise at each iteration of AMP to be approximately Gaussian. This feature enables linear convergence of x^t [60], accurate algorithm analysis [27, 27], and optimal parameter tuning [69].

BM3D-AMP [68, 67] now aims to leverage the rich history of denoising algorithms to enhance signal recovery. This assumes that any denoiser D_σ, when applied to a signal x_o plus Gaussian noise, will return an estimate closer to x_o than the original corrupted signal. This denoiser D_σ is henceforth treated as a black box, where knowledge of the algorithm's interworkings is not required for analysis. Thus D-AMP employs a denoiser from the previous AMP (Eqn. 2.4) in the following way:

$$\begin{aligned} x^{t+1} &= D_{\hat{\sigma}^t}(x^t + \mathbf{A}^* z^t), \\ z^t &= y - \mathbf{A}x^t + z^{t-1} \mathrm{div} D_{\hat{\sigma}^{t-1}}(x^{t-1} + \mathbf{A}^* z^{t-1})/m, \\ (\hat{\sigma}^t)^2 &= \frac{\|z^t\|_2^2}{m}. \end{aligned} \tag{2.5}$$

Here again we have that x^t and z^t are estimates of x_o at iteration t and of the residual, respectively. The Onsager correction term is $z^{t-1}\mathrm{div}D_{\hat{\sigma}^{t-1}}(x^{t-1} + \mathbf{A}^* z^{t-1})/m$, where $\mathrm{div}D_{\hat{\sigma}^{t-1}}$ is the divergence of the denoiser.

At a high level, D-AMP applies an existing denoising algorithm to obtain a better estimate of x_o at every iteration and eventually converge to the signal of interest. Empirically it has shown to outperform classical sparsity-based approaches while also being robust to measurement noise [68, 67].

2.2.2 TVAL3

Another related algorithm is TVAL3 [99, 54] which leverages augmented Lagrangian multipliers to achieve impressive performance on compressed sensing problems. Instead of the ℓ_1-norm, TVAL3 uses total variation (TV), which more accurately preserves high-frequency components such as edges or boundaries. Here the main assumption behind TV regularization is that natural images have sparse gradients. Over the past few decades, this approach has become very popular for tasks such as image denoising [78, 17], deconvolution [18, 92], and restoration [11, 98].

The model for total variation can be written as

$$\min_x TV(x) = \textstyle\sum_i \|D_i x\|_p \quad \text{s.t. } y = Ax, \tag{2.6}$$

where $D_i x$ is the discrete gradient of x at the i^{th} pixel. Note that the ℓ_p-norm could be either anisotropic if $p = 1$, or isotroptic if $p = 2$. Here we will focus on the isotropic case, thus $\|\cdot\|$ refers to the ℓ_2 norm.

To account for noise in the measurements, we place the constraint of Eqn. 2.6 into the objective, i.e.

$$\min_x TV(x) = \sum_i \|D_i x\| + \frac{\mu}{2}\|Ax - y\|^2. \tag{2.7}$$

Now our goal is to separate the non-differentiable TV term to achieve an easily solvable augmented Lagrangian function. This can be accomplished by introducing a splitting variable $q_i = D_i x$. Thus Eqn. 2.7 is equivalent to:

$$\begin{aligned} \min_{q_i, x} \quad & \sum_i \|q_i\| + \frac{\mu}{2}\|Ax - y\|^2 \\ \text{s.t.} \quad & q_i = D_i x \; \forall \; i. \end{aligned} \tag{2.8}$$

This framework presents an efficient Lagrangian method for total variation minimization. Empirically this has delivered improved image reconstruction compared to classical sparsity-based methods and also various iterative soft-thresholding algorithms [99].

2.3 Compressed Sensing: Learned Approaches

While sparsity in some chosen basis is well-established, recent work has shown better empirical performance when neural networks are used [9]. This success is attributed to the fact that neural networks are capable of learning image priors from very large datasets [35, 48]. There is significant recent work on solving linear inverse problems using various learned techniques. Mardani et al. [63] propose recurrent generative models while Dave et al. [24] apply auto-regressive models. Additionally approximate message passing (AMP) has been extended to a learned setting by Metzler et al. [66].

Bora et al. [9] is the closest set-up to our proposed algoirthms. In this work the authors assume that the unknown signal is in the range of a pre-trained differentiable generative model like a generative adversarial network (GAN) [35] or a variational autoencoder (VAE) [48]. The recovery of the unknown signal is then obtained via gradient descent in the latent space to search for a generated signal that satisfies the measurements. This can be directly applied for linear inverse problems and more generally to any differentiable measurement process. Chang et al. [20] solve a problem similar to Bora et al. but with a different optimization technique. Very recent work has built upon the method of Bora et al. using amortized variational compressed sensing [36], modelling sparse deviations [25], and task-aware generator training [47].

The key point is that all this prior work requires pre-trained generative models. In contrast, as we discussed, our proposed algorithm applies DIP [89] which uses an untrained model and optimizes the network weights for linear measurements taken from an individual image. We further leverage access to measurements from only a few (roughly 5 – 10) similar images to learn the prior distribution of the weights of network layers. This results in an informative prior using much less data than would be required to train a GAN or a VAE over a large image dataset. As such, we compare our algorithm to all these unlearned methods: BM3D-AMP, TVAL3, and Lasso in various bases. We perform this comparison with different datasets, different measurement processes, and various levels of measurements.

2.4 Compressed Sensing: Applications

Compressed sensing methods have many applications such as data compression, channel coding, and inverse problems. It has led to profoud developments in imaging, for example the single-pixel camera (SPC) [29] where micro-mirrors create measurements from a single light sensor that subsequently are used to reconstruct images using compressed sensing reconstruction algorithms. Bell Labs leveraged this technique to take still photographs using repeated snapshots of randomly chosen apertures from a grid. Medical tomographic applications include x-ray radiography, microwave imaging, magnetic resonance imaging (MRI), and computed tomography (see e.g. [95, 21, 58] and references therein). The overarching goal in developing new compressed sensing recovery methods is to reduce the number of measurements while maintaining good reconstruction quality. Obtaining measurements can often be costly, time-consuming, and in some cases dangerous, e.g. exposing a patient to harmful x-ray radiation [75].

Our method is unlearned, requiring little or no measurements to tune the network; as such we choose to place emphasis on the field of medical imaging for which learned models are currently infeasible. This lack of feasibility can be attributed to two different characteristics of learned models: (1) they require a large amount of training data (2) they are not able to reconstruct complex signals such as chest x-rays or retinopathy images. We circumvent these undesirable characteristics by using a model that is unlearned and also often outperforms many of the unlearned methods proposed in Section 2.2.

Chapter 3

Methods

Let $x^* \in \mathbb{R}^n$ be the signal that we are trying to reconstruct, $A \in \mathbb{R}^{m \times n}$ be the measurement matrix, and $\eta \in \mathbb{R}^m$ be independent noise. Given the measurement matrix A and the observations $y = Ax^* + \eta$, we wish to reconstruct an $\hat{x}$ that is close to x^*.

A generative model is a deterministic function $G(z; w)$: $\mathbb{R}^k \to \mathbb{R}^n$ which takes as input a seed $z \in \mathbb{R}^k$ and a set of parameters (or "weights") $w \in \mathbb{R}^d$, producing an output $G(z; w) \in \mathbb{R}^n$. These models have shown excellent performance generating real-life signals such as images [35, 48] and audio [90]. We investigate *deep convolutional* generative models, a special case in which the model architecture has multiple cascaded layers of convolutional filters [49]. In this paper we apply a DCGAN [76] model and restrict the signals to be images.

3.1 Compressed Sensing with Deep Image Prior (CS-DIP)

Our approach is to find a set of weights for the convolutional network such that the measurement matrix applied to the network output, i.e. $AG(z; w)$, matches the measurements y we are given. Hence we initialize an *untrained* network $G(z; w)$ with some fixed z and solve the following optimization problem:

$$w^* = \arg\min_w \|y - AG(z; w)\|^2. \tag{3.1}$$

This is, of course, a non-convex problem because $G(z; w)$ is a complex feed-forward neural network. Still we can use gradient-based optimizers for any generative model and measurement process that is differentiable. Ulyanov et al. observed that generator networks such as DCGAN are biased toward smooth, natural images due to their convolutional structure; thus the network structure alone provides a good prior for reconstructing images in problems such as inpainting and denoising [89]. Our finding is that this applies to general linear measurement processes. We restrict our solution to lie in the span of a convolutional neural network, and if a sufficient number of measurements m is given, we obtain an output such that $x^* \approx G(z; w^*)$.

Note that this method uses an untrained generative model and optimizes over the network weights w. In contrast previous methods, such as that of Bora et al. [9], use a trained model and optimize over the latent z-space, solving $z^* = \arg\min_z \|y - AG(z; w)\|^2$. We instead initialize a random z with Gaussian iid entries and keep this fixed throughout the optimization process.

Note that DIP must be tuned to avoid overfitting; we rely on early stopping and also on two different regularizaton terms: $LR(w)$, a novel learned regularization technique and also $TV(G(z; w))$, the well-established total variation norm [78, 92].

Thus the final optimization problem becomes:

$$w^* = \arg\min_w \|y - A\, G(z; w)\|^2 + R(w; \lambda_L, \lambda_T). \tag{3.2}$$

Where the regularization term contains hyperparameters λ_L and λ_T for learned regularization and total variation: $R(w; \lambda_L, \lambda_T) = \lambda_L LR(w) + \lambda_T TV(G(z; w))$. We discuss our regularization techniques below.

3.2 Learned Regularization

Without regularization CS-DIP relies only on linear measurements taken from one unknown image. We now introduce a novel method which leverages a small amount of training data to optimize regularization. In this case training data refers to measurements from additional ground truth of a similar type, for example other x-ray images.

To leverage this additional information, we pose Eqn. (3.2) as a Maximum a Posteriori (MAP) estimation problem and propose a novel prior on the weights of the generative model. This prior then acts as a regularization term, penalizing the model toward an optimal set of weights w^*.

For a set of weights $w \in \mathbb{R}^d$, we model the *likelihood* of the measurements $y = Ax, y \in \mathbb{R}^m$, as a Gaussian distribution given by

$$p(y|w) = \frac{1}{\sqrt{(2\pi)^m \lambda}} \exp\left(-\frac{\|y - AG(z;w)\|^2}{2\lambda}\right), \tag{3.3}$$

and the prior on the weights w as a Gaussian given by

$$p(w) = \frac{1}{\sqrt{(2\pi)^d |\Sigma|}} \exp\left(-\frac{1}{2}(w-\mu)^T \Sigma^{-1} (w-\mu)\right), \tag{3.4}$$

where $\mu \in \mathbb{R}^d$ and $\Sigma \in \mathbb{R}^{d \times d}$.

In this setting we want to find a set of weights w^* that maximize the posterior on w given y, i.e.,

$$\begin{aligned} w^* &= \arg\max_w p(w|y), \\ &= \arg\max_w \frac{p(y|w)p(w)}{p(y)}, \\ &\equiv \arg\min_w \|y - AG(z;w)\|^2 + \lambda_L (w-\mu)^T \Sigma^{-1} (w-\mu). \end{aligned} \tag{3.5}$$

This gives us the learned regularization term

$$LR(w) = (w - \mu)^T \Sigma^{-1} (w - \mu)\,, \tag{3.6}$$

where the coefficient λ_L in Eqn. (3.5) controls the strength of the prior.

Notice that when $\mu = 0$ and $\Sigma = I_{d\times d}$, this regularization term is equivalent to ℓ_2-regularization. Thus this method can be thought of as a more strategic version of standard weight decay.

3.2.1 Learning the Prior Parameters

In the previous section we introduced the learned regularization term:

$$LR(w) = (w - \mu)^T \Sigma^{-1} (w - \mu)\,.$$

However we do not yet know good values for (μ, Σ) that will give high quality reconstructions. For a fixed set of measurements m and a measurement

Algorithm 1 Estimate (μ, Σ) for a distribution of network weights W^*

input Set of optimal weights $W^* = \{w_1^*, w_2^*, \cdots, w_K^*\}$ obtained from L-layer DCGAN run over K images; number of samples S; number of iterations T.
output mean vector $\mu \in \mathbb{R}^L$; covariance matrix $\Sigma \in \mathbb{R}^{L\times L}$.

for $t = 1$ to T **do**
  Sample k uniformly from $\{1, ..., K\}$
  for $l = 1$ to L {for each layer} **do**
    Get $v \in \mathbb{R}^S$, a vector of S uniformly sampled weights from the l^{th} layer of w_k^*
    $M_t[l, :] \leftarrow v^T$ where $M_t[l, :]$ is the l^{th} row of matrix $M_t \in \mathbb{R}^{L\times S}$
    $\mu_t[l] \leftarrow \frac{1}{S}\sum_{i=1}^{S} v_i$
  end for
  $\Sigma_t \leftarrow \frac{1}{S} M_t M_t^T - \mu_t \mu_t^T$
end for
$\mu \leftarrow \frac{1}{T}\sum_{t=1}^{T} \mu_t$
$\Sigma \leftarrow \frac{1}{T}\sum_{t=1}^{T} \Sigma_t$

matrix A, we now propose a way to estimate (μ, Σ) such that prior knowledge of the network weights can be incorporated.

Assume we have a set of measurements $S_Y = \{y_1, y_2, \cdots, y_K\}$ from K different images $S_X = \{x_1, x_2, \cdots, x_K\}$, each obtained with a different measurement matrix A. For each measurement $y_i, i \in \{1, 2, ..., K\}$, we run CS-DIP to solve the optimization problem in Eqn. (3.2) and obtain an optimal set of weights $W^* = \{w_1^*, w_2^*, \cdots, w_K^*\}$. Note that when optimizing for the weights W^*, we only have access to the measurements S_Y, not the ground truth S_X.

The number of weights d in deep networks tends to be very large. As such, learning a distribution over each weight, i.e. estimating $\mu \in \mathbb{R}^d$ and $\Sigma \in \mathbb{R}^{d \times d}$, becomes intractable. We instead use a layer-wise approach: with L network layers, we have $\mu \in \mathbb{R}^L$ and $\Sigma \in \mathbb{R}^{L \times L}$. Thus each weight within layer $l \in \{1, 2, ..., L\}$ is modeled according to the same $\mathcal{N}(\mu_l, \Sigma_{ll})$ distribution. For simplicity we assume $\Sigma_{ij} = 0 \; \forall i \neq j$, i.e. that network weights are independent across layers. The process of estimating statistics (μ, Σ) from W^* is described in Algorithm (1), where we find different (μ, Σ) for each measurement number m.

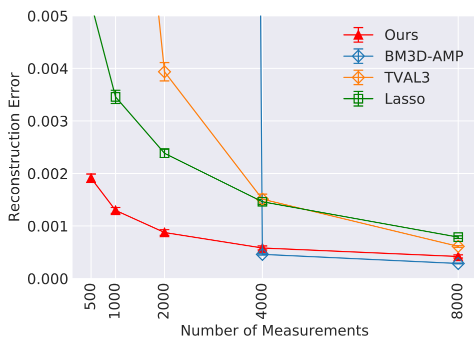
We use this learned (μ, Σ) in the regularization term $LR(w)$ from Eqn. (3.6) for reconstructing measurements of images. We refer to this technique as *learned regularization*. While this technique may seem analogous to batch normalization [44], note that we only use (μ, Σ) to penalize the ℓ_2-norm of the weights and do not normalize the layer outputs themselves.

3.2.2 Discussion

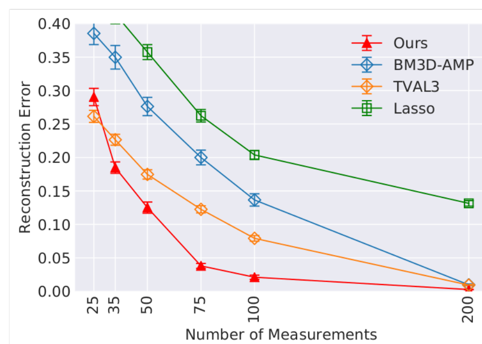
The proposed CS-DIP algorithm is data-agnostic if no learned regularization is used. That is, given measurements for any single unknown image $x^* \in \mathbb{R}^n$, we can search for good weights w^* such that the generator network produces an output which approximately satisfies these measurements. Learned regularization utilizes a small amount of prior information, as it only requires access to measurements from a small number of images (roughly $5 - 10$). In contrast, other pre-trained models such as that of Bora et al. [9] require access to ground truth from a large number of similar images (tens of thousands for CelebA). If such a large dataset is available and if a good generative model can be trained on that dataset, we expect that methods which use pre-trained models [9, 36, 47, 63] would outperform our method. Our approach is instead more suitable for reconstructing problems where large amounts of data or good generative models are not readily available.

3.3 Total Variation Regularization

In addition to our novel learned regularization method, we also leverage total variation (TV) norm [78, 92, 55] regularization in our objective function. TV loss penalizes the sum of absolute difference for neighboring pixel values. This makes reconstructions smoother and reduces high frequency noise in the reconstructed image. Note also that in parallel to our work, total variation regularization was proposed as a method to improve DIP very recently by Liu et al. [55].

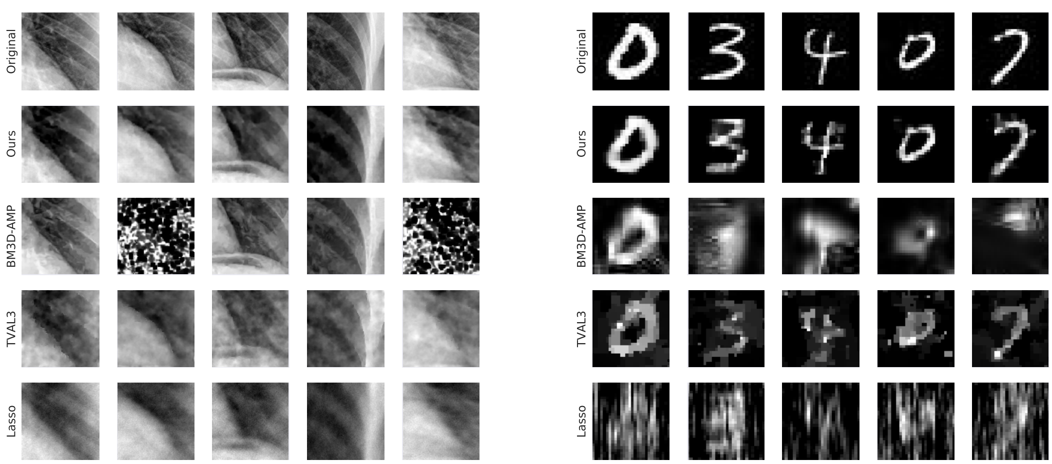


(a) MSE - Chest X-ray (65536 pixels)



(b) MSE - MNIST (784 pixels)

Figure 3.1: We compare the performance of our algorithm with baselines on the x-ray and MNIST datasets, plotting per-pixel reconstruction error (MSE) vs. number of measurements, where vertical bars indicate 95% confidence intervals. Notice that when the number of measurements is below 4000, BM3D-AMP frequently fails to converge. This is demonstrated in the graph since its reconstruction error values are large and hence far above our vertical axis.



(a) Reconstructions - Chest X-ray

(b) Reconstructions - MNIST

Figure 3.2: Reconstruction results on x-ray images for m = 2000 measurements (of n = 65536 pixels) and MNIST for m = 75 measurements (of n = 784 pixels). From top to bottom row: original image, reconstructions by our algorithm, then reconstructions by baselines BM3D-AMP, TVAL3, and Lasso. For x-ray images the number of measurements obtained are 3% the number of pixels (i.e. $\frac{m}{n} = .03$), for which BM3D-AMP often fails to converge.

Chapter 4

Experiments

To replicate these experiments or run new experiments using our method, please see our GitHub repository at `https://github.com/davevanveen/compsensing_dip`.

4.1 Experimental Setup

Measurements: We evaluate our algorithm using two different measurements processes, i.e. matrices $A \in \mathbb{R}^{m \times n}$. First we set the entries of A to be Gaussian iid, such that $A_{i,j} \sim \mathcal{N}(0, \frac{1}{m})$. Recall m is the number of measurements, and n is the number of pixels in the ground truth image. This measurement process is standard practice in compressed sensing literature, and hence we use it on each dataset. Additionally in Section 4.2 we use a Fourier measurement process common in MRI applications [64, 62, 37, 51, 59] and evaluate it on the chest x-ray dataset. In that case measurements obtained are Fourier coefficients sampled according to a radial pattern shown in Figure 4.2 of the appendix.

Datasets: We use our algorithm to reconstruct both grayscale and RGB images. For grayscale we use the first 100 images in the test set of MNIST [50] and also 60 random images from the Shenzhen Chest X-Ray Dataset [46], selecting a 512x512 crop and then downsampling to 256x256 pixels. For RGB images we use the Structured Analysis of the Retina (STARE) dataset [43] with 512x512 crops downsized to 128x128 pixels.

Baselines: We compare our algorithm to state-of-the-art unlearned methods such as BM3D-AMP [68, 67], TVAL3 [52, 54, 99], and Lasso in a DCT basis [2]. We also evaluated the performance of Lasso in a Daubechies wavelet basis [22, 94] but found this performed worse than Lasso - DCT on all datasets. Thus for simplicity we refer to Lasso - DCT as "Lasso" and do not include results of Lasso - Wavelet. To reconstruct RGB retinopathy images, we must use the colored version CBM3D-AMP. Unfortunately an RGB version of TVAL3 does not currently exist, although similar TV algorithms such as FTVd can performs similar tasks such as denoising RGB images [92].

Metrics: To quantitatively evaluate the performance of our algorithm, we use per-pixel mean-squared error (MSE) between the reconstruction $\hat{x}$ and true image x^*, i.e. $\frac{\|\hat{x}-x^*\|^2}{n}$. Note that because these pixels are over the range $[-1, 1]$, it's possible for the MSE to be greater than 1.

Implementation: To find a set of weights w^* that minimize Eqn. (3.2), we use PyTorch [73] with a DCGAN architecture. For baselines BM3D-AMP and TVAL3, we use the repositories provided by the authors Metzler et al. [65] and Li et al. [53], respectively. For baseline reconstructions Lasso, we use a scikit-learn [74] implementation. Experimental details are discussed in the paragraphs below to provide heuristic intuition for those wishing to implement our work.

Our algorithm CS-DIP is implemented in PyTorch using the RMSProp optimizer [85] with learning rate 10^{-3} and momentum 0.9. We take 1000 update steps for every set of measurements. On larger images such as xray ($n = 65536$) and retinopathy ($n = 49152$), we found no difference using random restarts of the initial seed z. However for smaller vectors such as MNIST ($n = 784$), restarts did provide some benefit. As such our experiments utilize 5 random restarts for MNIST and one initial seed (no restarts) for x-ray and retinopathy images.

The convergence, i.e. Error vs. Iterations, of CS-DIP with RMSProp could be unstable for some learning rates, even though error gradually decreased. As such we implemented a stopping condition which chooses the reconstruction with least error over the last 20 iterations. Note we choose this reconstruction based off measurement loss and do not look at the ground truth image.

4.2 Results

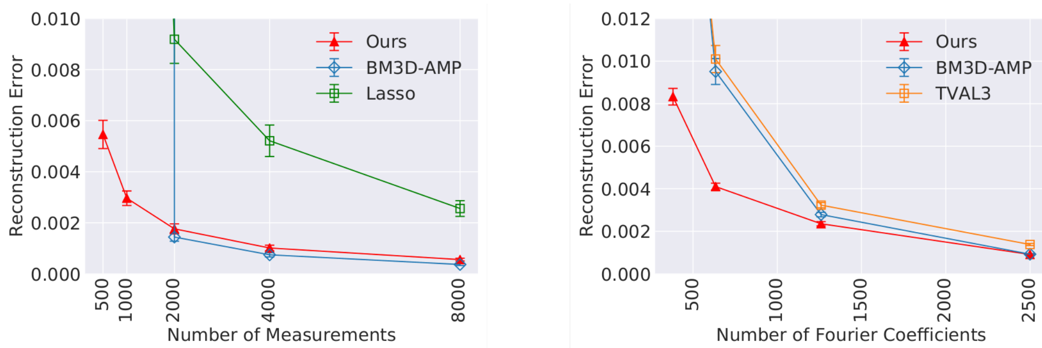
We first compare our algorithm (CS-DIP) to baselines on all three datasets using a measurement matrix with Gaussian iid entries. Then we also demonstrate CS-DIP with a Fourier measurement process on the x-ray dataset.

MNIST

In Figure 3.1b we plot reconstruction error with varying number of measurements m of n = 784. This demonstrates that our algorithm outperforms baselines in almost all cases. Figure 3.2b shows reconstructions for 75 measurements, while remaining reconstructions are in the appendix.

Chest X-Rays

In Figure 3.1a we plot reconstruction error with varying number of measurements m of n = 65536. Figure 3.2a shows reconstructions for 2000 measurements, while the remaining reconstructions are in the appendix. On this dataset we outperform all baselines except BM3D-AMP for higher m, which produces sharp reconstructions. However for lower m, e.g. when the ratio $\frac{m}{n} \leq 3\%$, BM3D-AMP



(a) MSE - Retinopathy with Gaussian measurements

(b) MSE - Chest X-ray with Fourier measurements

Figure 4.1: Per-pixel reconstruction error (MSE) vs. number of measurements, where vertical bars indicate 95% confidence intervals.

often doesn't converge. This finding seems to support the work of Metzler et al. [67]: BM3D-AMP performs impressively on higher m, e.g. $\frac{m}{n} \geq 10\%$, but recovery at lower sampling rates are not demonstrated.

Retinopathy

In Figure 4.1a we plot the reconstruction error with varying number of measurements m of $n = 49152$. On this RGB dataset we quantitatively outperform all baselines except BM3D-AMP on higher m; however, even at these higher m, patches of green and purple pixels corrupt the image reconstructions as seen in Figure 1.1. Similar to x-ray for lower m, BM3D-AMP fails to produce anything sensible as demonstrated by additional reconstructions located in the appendix.

Fourier Measurement Process

All previous experiments in this section used a measurement matrix A containing Gaussian iid entries. We now consider the case where the measurement

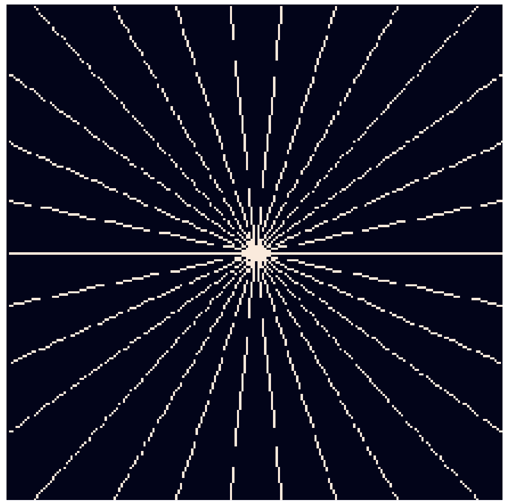


Figure 4.2: This figure shows a radial sampling pattern of coefficients Ω in the Fourier domain. The measurements are obtained by sampling Fourier coefficients along these radial lines.

matrix is a subsampled Fourier matrix. That is, for a 2D image x and a set of indices Ω, the measurements we receive are given by $y_{(i,j)} = [\mathcal{F}(x)]_{(i,j)}, (i, j) \in \Omega$, where $\mathcal{F}$ is the 2D Fourier transform. In our experiments we choose Ω to be indices along radial lines, as shown in Figure 4.2. This choice of Ω is common in literature [13] and has also been used in MRI applications [62, 59, 32]. We run our algorithm along with BM3D-AMP and TVAL3 baselines on the chest X-ray dataset for $\{3, 5, 10, 20\}$ radial lines in the Fourier domain, which corresponds to $\{381, 634, 1260, 2500\}$ Fourier coefficients, respectively.

In Figure 4.1b we plot the reconstruction error with varying number of Fourier coefficients. In the appendix we show reconstructions obtained by our algorithm versus BM3D-AMP and TVAL3.

Algorithm	1000	2000	4000	8000
CS-DIP	15.7	**17.2**	**20.6**	**30.1**
BM3D-AMP	51.1	54.0	67.8	71.2
TVAL3	**13.8**	22.1	31.9	56.7
Lasso DCT	27.1	33.0	52.2	96.4

Table 4.1: Runtime (seconds) for each algorithm with varying number of measurements.

Runtime

In Table 4.1 we show runtime of CS-DIP on the x-ray dataset. Our algorithm has the capability of utilizing GPU, as we run experiments on an NVIDIA GTX 1080-Ti. The other baselines are implemented in MATLAB or sci-kit learn [74] and as such are restricted to CPU. Comparisons aside, this demonstrates that our method executes in a reasonable amount of time.

Chapter 5

Conclusion and Future Work

We first discuss the landscape of compressed sensing algorithms: classical sparsity-based approaches, modern unlearned approaches, and learned approaches using neural networks. We then demonstrate compressed sensing recovery using untrained, randomly initialized convolutional neural networks. Our method outperforms previous state of the art unlearned methods in several cases, especially when the number of obtained measurements is small.

There are several interesting directions for future work. We suspect improved performance from data-driven network initialization, e.g. initializing network weights according to the learned distribution for W^*. Another extension could be to apply our method multiple times over patches within an image, e.g. similar to PatchGAN proposed by Isola et al. [45]. These Deep Image Prior techniques may be applicable to other inverse problems e.g. phase retrieval, inspired by Hand et al. [38]. Further, very recent work showed that convolutions can be replaced by linear interpolation for Deep Image Prior [40]; it would be interesting to combine the deep decoder with our methods of regularization.

Overall we believe that the inductive benefits of convolutional neural networks will provide significant impact in the future of inverse problem reconstruction algorithms. We are now working to apply the proposed method to improve the performance of signal recovery for rapid MRI imaging. As discussed in Section 2.4, measurements for medical imaging (e.g. MRI, CT, x-ray), can be costly, time-consuming, and in some cases dangerous by exposing the patient to harmful radiation. Hence there is great incentive to obtain quality reconstructions with fewer measurements.

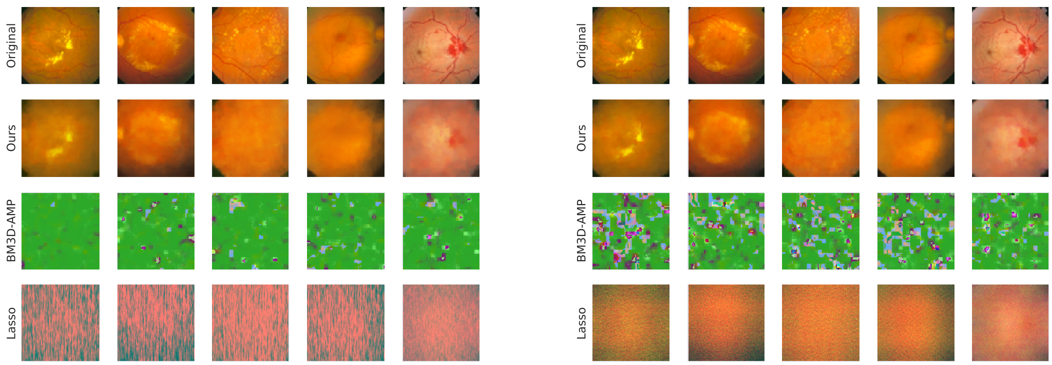
To this end we have recently begun a collaboration with Dr. Tom Yankeelov, a professor who is an expert in quantitative MRI for clinical breast cancer imaging. With his domain knowledge, we intend to make impact in medical imaging by utilizing generative neural networks. Specifically we plan to show that CS-DIP will allow for similar spatial and temporal resolution with fewer MRI measurements. Beyond MRI, this project will provide intuition for how these compelling algorithms can be applied to real world problems.

Apart from methods developed with my collaborators at the University of Texas, I am further exploring how machine learning models can improve medical imaging. This fall 2019 I will be working as a Research Scientist Intern at Subtle Medical. Subtle is a start-up that develops deep learning models to reduce the number of measurements required to reconstruct a particular signal. MRI is a particularly interesting modality as there are opportunities to reduce both the temporal and three-dimensional spatial resolutions at which measurements must be acquired. Subtle is the first company that has obtained FDA clearance to implement their algorithms in a clinical setting; it will be exciting to work with a company who is making real impact in that space. Skills I gain in this work experience will be useful for academic research upon my return to UT-Austin.

Another research goal upon returning to UT is to theoretically analyze various descent algorithms for this family of non-convex optimization problems under assumptions of the deep generative model G(z). The proposed signal reconstruction requires minimizing a non-convex loss $\|y - AG(z; w)\|^2$. Empirically gradient descent delivers excellent performance minimizing this loss; however, we have little theoretical understanding of how or why it actually works. Investigating this would be very useful for improving the algorithm and also knowing when we can trust the reconstructions. I intend to investigate conditions on both the generative model and measurement process that allow us to establish performance bounds for solving this non-convex problem. Exciting recent work [39] has developed results for this problem with random independent weights. I plan on extending these results to models with more structure or different assumptions.

Overall I believe this proposed method has great potential, and I look forward to building upon this research in the future.

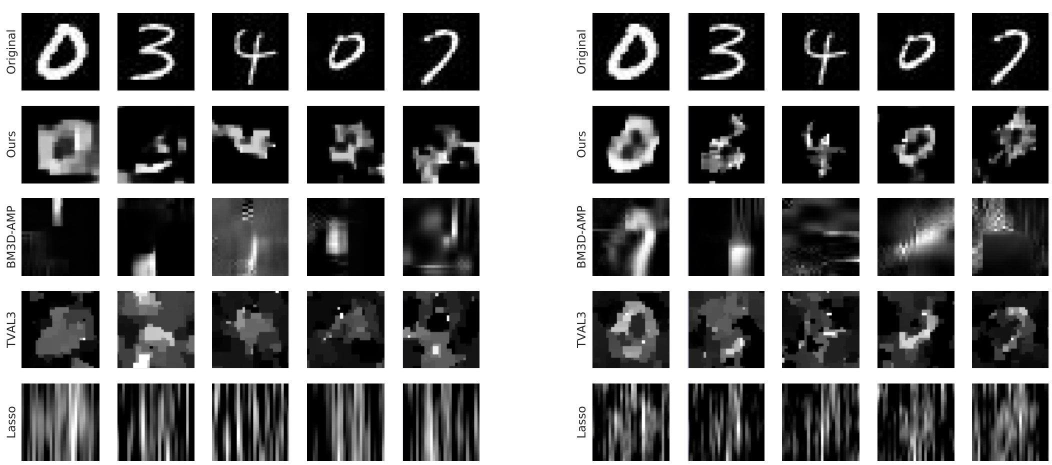
Appendix



(a) 500 measurements

(b) 1000 measurements

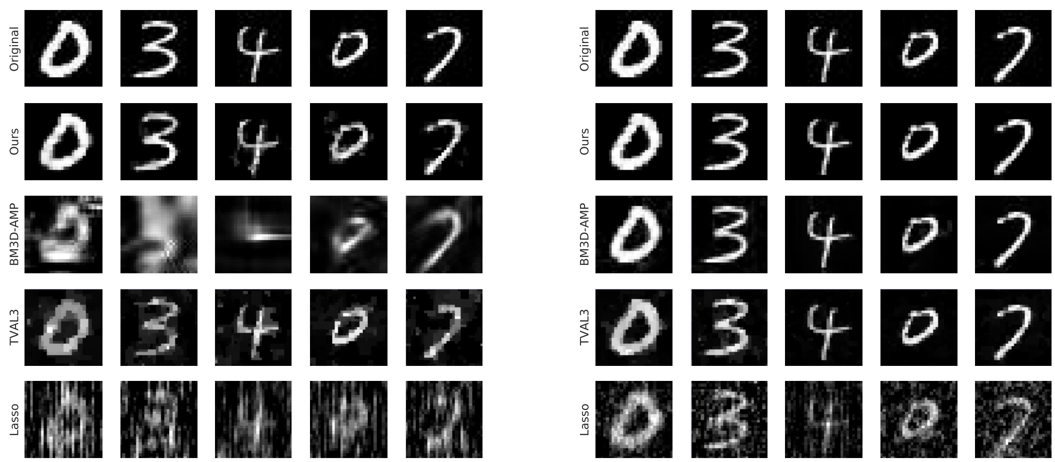
Figure A.1: Reconstruction results on retinopathy images for m = 500, 1000 measurements respectively (of n = 49152 dimensional vector). From top to bottom row: original image, reconstructions by our algorithm, then reconstructions by baselines BM3D-AMP and Lasso.



(a) 25 measurements

(b) 50 measurements

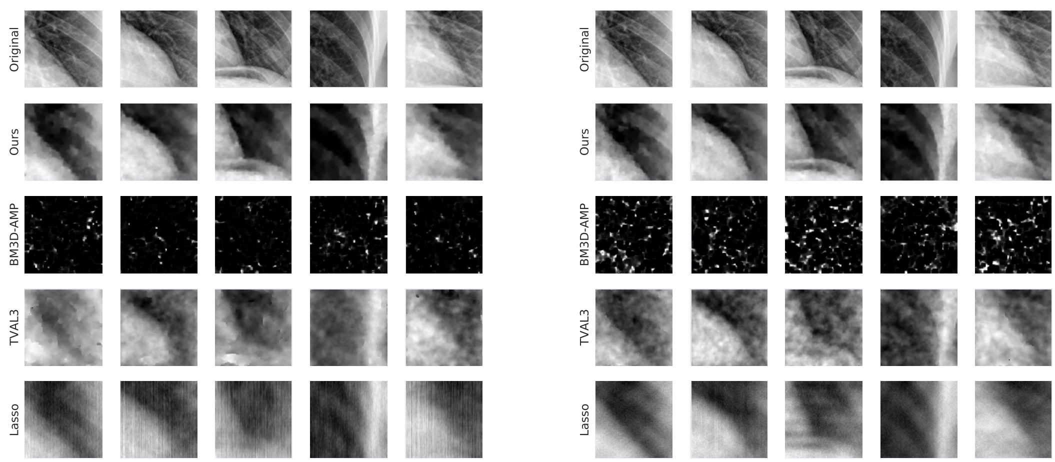
Figure A.2: Reconstruction results on MNIST for m = 25, 50 measurements respectively (of n = 784 pixels). From top to bottom row: original image, reconstructions by our algorithm, then reconstructions by baselines BM3D-AMP, TVAL3, and Lasso.



(a) 100 measurements

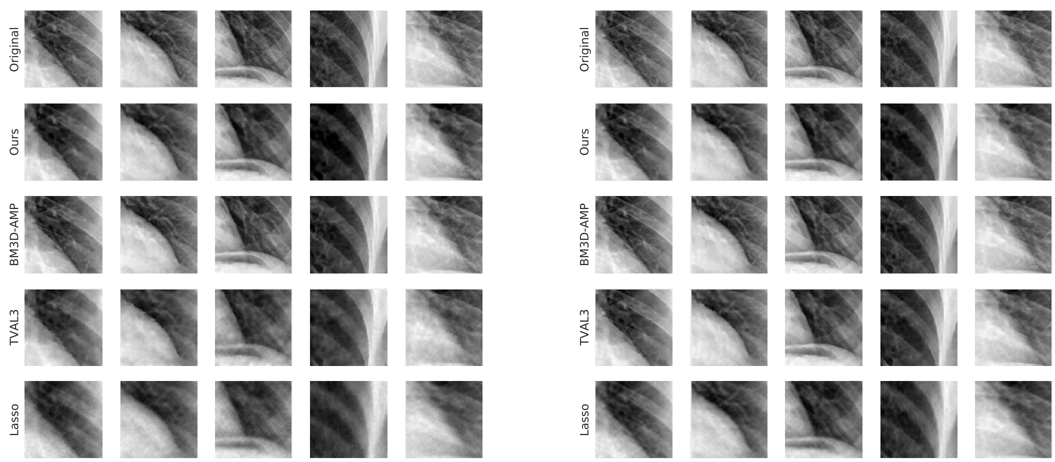
(b) 200 measurements

Figure A.3: Reconstruction results on MNIST for m = 100, 200 measurements respectively (of n = 784 pixels). From top to bottom row: original image, reconstructions by our algorithm, then reconstructions by baselines BM3D-AMP, TVAL3, and Lasso.



(a) 500 measurements (b) 1000 measurements

Figure A.4: Reconstruction results on x-ray images for m = 500, 1000 measurements respectively (of n = 65536 pixels). From top to bottom row: original image, reconstructions by our algorithm, then reconstructions by baselines BM3D-AMP, TVAL3, and Lasso.



(a) 4000 measurements

(b) 8000 measurements

Figure A.5: Reconstruction results on x-ray images for m = 4000, 8000 measurements respectively (of n = 65536 pixels). From top to bottom row: original image, reconstructions by our algorithm, then reconstructions by baselines BM3D-AMP, TVAL3, and Lasso.

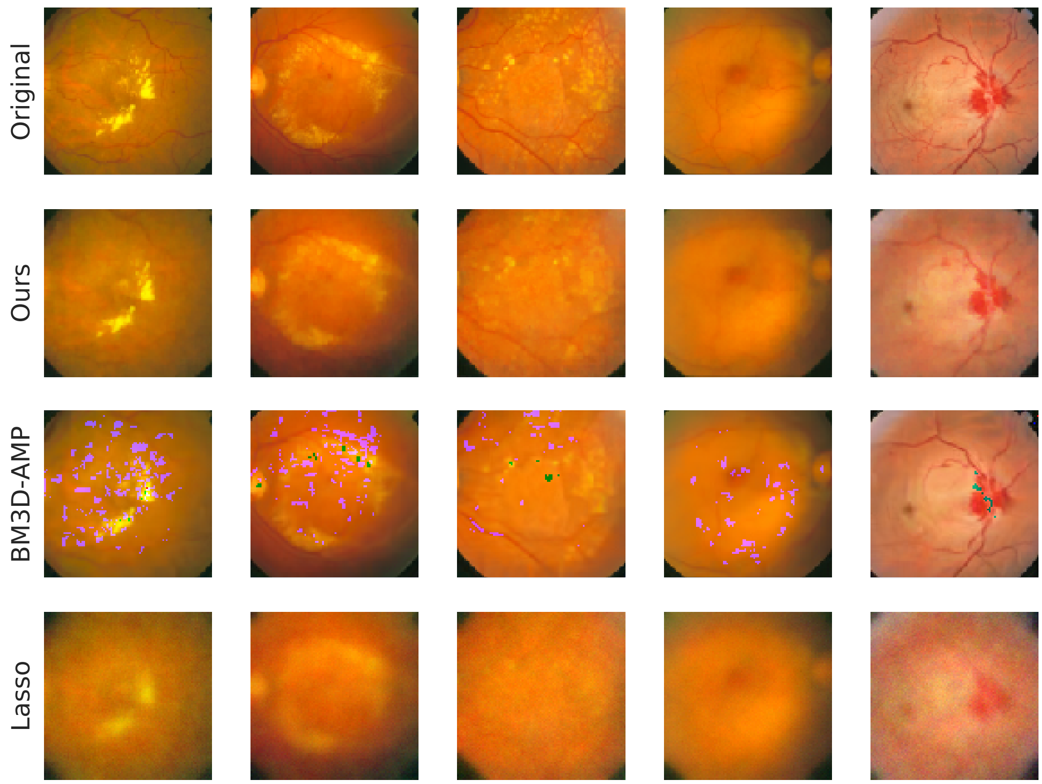


Figure A.6: Reconstruction results on retinopathy images for m = 4000 (of n = 49152 pixels). From top to bottom row: original image, reconstructions by our algorithm, then reconstructions by baselines BM3D-AMP and Lasso.

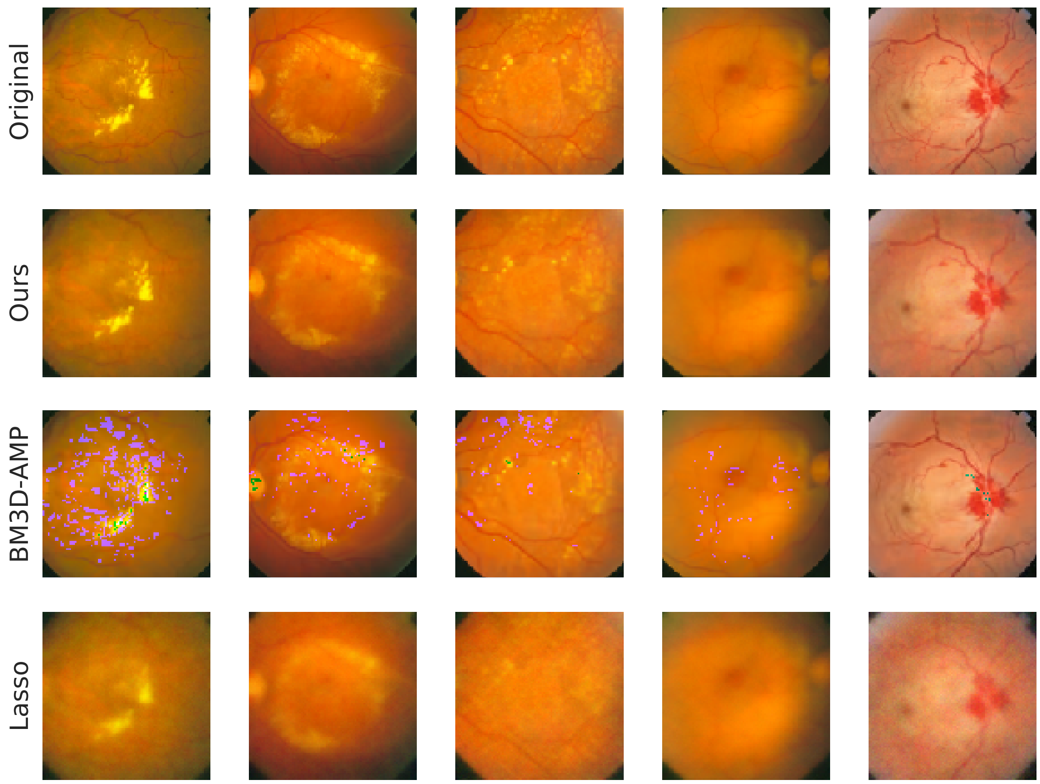


Figure A.7: Reconstruction results on retinopathy images for m = 8000 (of n = 49152 pixels). From top to bottom row: original image, reconstructions by our algorithm, then reconstructions by baselines BM3D-AMP and Lasso.

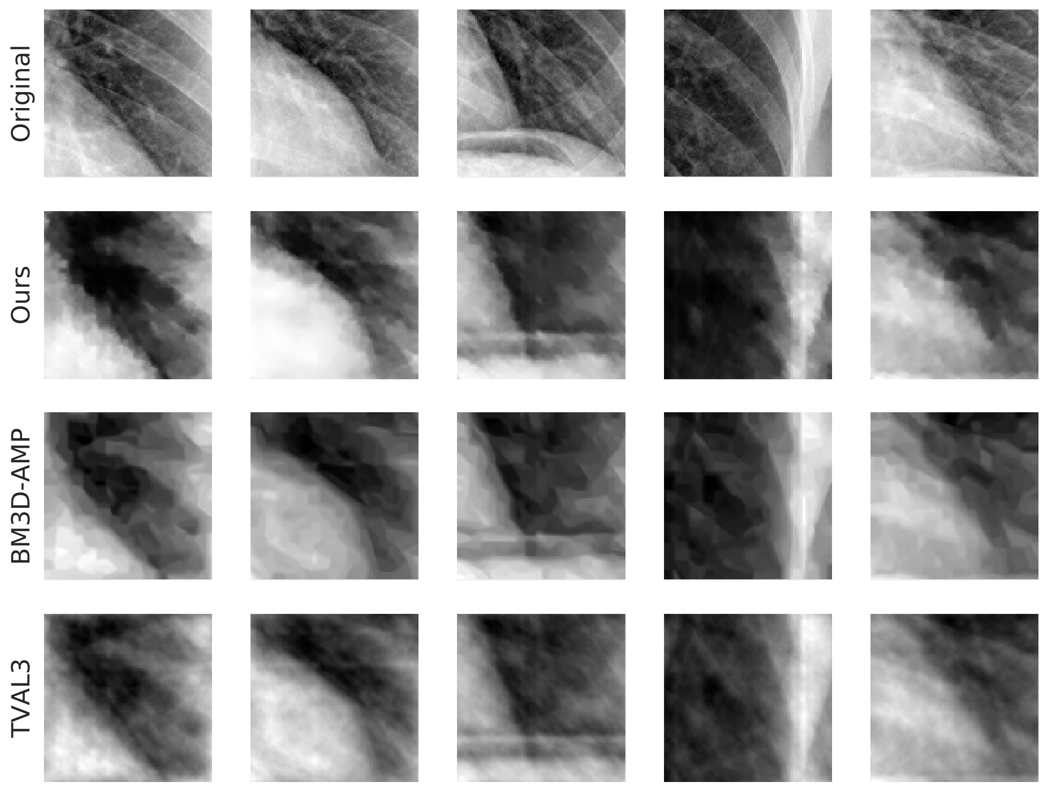


Figure A.8: Reconstruction results on x-ray images for m = 1260 Fourier coefficients (of n = 65536 pixels). From top to bottom row: original image, reconstructions by our algorithm, then reconstructions by baselines BM3D-AMP and TVAL3.

Bibliography

[1] Alekh Agarwal, Sahand Negahban, and Martin J Wainwright. Fast global convergence rates of gradient methods for high-dimensional statistical recovery. In *Advances in Neural Information Processing Systems*, pages 37–45, 2010.

[2] Nasir Ahmed, T_ Natarajan, and Kamisetty R Rao. Discrete cosine transform. *IEEE transactions on Computers*, 100(1):90–93, 1974.

[3] Muhammad Asim, Fahad Shamshad, and Ali Ahmed. Solving bilinear inverse problems using deep generative priors. *CoRR*, abs/1802.04073, 2018.

[4] Muhammad Asim, Fahad Shamshad, and Ali Ahmed. Solving bilinear inverse problems using deep generative priors. *arXiv preprint arXiv:1802.04073*, 2018.

[5] Francis Bach, Rodolphe Jenatton, Julien Mairal, Guillaume Obozinski, et al. Optimization with sparsity-inducing penalties. *Foundations and Trends® in Machine Learning*, 4(1):1–106, 2012.

[6] Richard G Baraniuk, Volkan Cevher, Marco F Duarte, and Chinmay Hegde. Model-based compressive sensing. *IEEE Transactions on Information Theory*, 56(4):1982–2001, 2010.

[7] Peter J Bickel, Yaacov Ritov, Alexandre B Tsybakov, et al. Simultaneous analysis of lasso and dantzig selector. *The Annals of Statistics*, 37(4):1705–1732, 2009.

[8] Thomas Blumensath and Mike E Davies. Iterative hard thresholding for compressed sensing. *Applied and computational harmonic analysis*, 27(3):265–274, 2009.

[9] Ashish Bora, Ajil Jalal, Eric Price, and Alexandros G Dimakis. Compressed sensing using generative models. *arXiv preprint arXiv:1703.03208*, 2017.

[10] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[11] Yuri Boykov, Olga Veksler, and Ramin Zabih. Fast approximate energy minimization via graph cuts. In *Proceedings of the Seventh IEEE International Conference on Computer Vision*, volume 1, pages 377–384. IEEE, 1999.

[12] Emmanuel J Candes. The restricted isometry property and its implications for compressed sensing. *Comptes rendus mathematique*, 346(9-10):589–592, 2008.

[13] Emmanuel J Candès, Justin Romberg, and Terence Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on information theory*, 52(2):489–509, 2006.

[14] Emmanuel J Candes, Justin K Romberg, and Terence Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on pure and applied mathematics*, 59(8):1207–1223, 2006.

[15] Emmanuel J Candes and Terence Tao. Decoding by linear programming. *IEEE transactions on information theory*, 51(12):4203–4215, 2005.

[16] Emmanuel J Candes and Terence Tao. Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE transactions on information theory*, 52(12):5406–5425, 2006.

[17] Antonin Chambolle. An algorithm for total variation minimization and applications. *Journal of Mathematical imaging and vision*, 20(1-2):89–97, 2004.

[18] Stanley H Chan, Xiran Wang, and Omar A Elgendy. Plug-and-play admm for image restoration: Fixed-point convergence and applications. *IEEE Transactions on Computational Imaging*, 3(1):84–98, 2017.

[19] Tony F Chan and Chiu-Kwong Wong. Total variation blind deconvolution. *IEEE transactions on Image Processing*, 7(3):370–375, 1998.

[20] Jen-Hao Rick Chang, Chun-Liang Li, Barnabás Póczos, B. V. K. Vijaya Kumar,

and Aswin C. Sankaranarayanan. One network to solve them all - solving linear inverse problems using deep projection models. *CoRR*, abs/1703.09912, 2017.

[21] Guang-Hong Chen, Jie Tang, and Shuai Leng. Prior image constrained compressed sensing (piccs): a method to accurately reconstruct dynamic ct images from highly undersampled projection data sets. *Medical physics*, 35(2):660–663, 2008.

[22] Ingrid Daubechies. Orthonormal bases of compactly supported wavelets. *Communications on pure and applied mathematics*, 41(7):909–996, 1988.

[23] Ingrid Daubechies, Michel Defrise, and Christine De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, 57(11):1413–1457, 2004.

[24] Akshat Dave, Anil Kumar Vadathya, Ramana Subramanyam, Rahul Baburajan, and Kaushik Mitra. Solving inverse computational imaging problems using deep pixel-level prior. *arXiv preprint arXiv:1802.09850*, 2018.

[25] Manik Dhar, Aditya Grover, and Stefano Ermon. Modeling sparse deviations for compressed sensing using generative models. *arXiv preprint arXiv:1807.01442*, 2018.

[26] David L Donoho. Compressed sensing. *IEEE Transactions on info theory*, 52(4):1289–1306, 2006.

[27] David L Donoho, Arian Maleki, and Andrea Montanari. Message-passing algorithms for compressed sensing. *Proceedings of the National Academy of Sciences*, 106(45):18914–18919, 2009.

[28] David L Donoho, Arian Maleki, and Andrea Montanari. The noise-sensitivity phase transition in compressed sensing. *IEEE Transactions on Information Theory*, 57(10):6920–6941, 2011.

[29] Marco F Duarte, Mark A Davenport, Dharmpal Takhar, Jason N Laska, Ting Sun, Kevin F Kelly, and Richard G Baraniuk. Single-pixel imaging via compressive sampling. *IEEE signal processing magazine*, 25(2):83–91, 2008.

[30] Armin Eftekhari and Michael B Wakin. New analysis of manifold embeddings and signal recovery from compressive measurements. *Applied and Computational Harmonic Analysis*, 39(1):67–109, 2015.

[31] David Eigen and Rob Fergus. Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2650–2658, 2015.

[32] Ender M Eksioglu and A Korhan Tanc. Denoising amp for mri reconstruction: Bm3d-amp-mri. *SIAM Journal on Imaging Sciences*, 11(3):2090–2109, 2018.

[33] Mário AT Figueiredo, Robert D Nowak, and Stephen J Wright. Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems. *IEEE Journal of selected topics in signal processing*, 1(4):586–597, 2007.

[34] Alyson K Fletcher and Sundeep Rangan. Inference in deep networks in high dimensions. *arXiv preprint arXiv:1706.06549*, 2017.

[35] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *NeurIPS*, pages 2672–2680, 2014.

[36] Aditya Grover and Stefano Ermon. Amortized variational compressive sensing. ., 2018.

[37] Kerstin Hammernik, Teresa Klatzer, Erich Kobler, Michael P Recht, Daniel K Sodickson, Thomas Pock, and Florian Knoll. Learning a variational network for reconstruction of accelerated mri data. *Magnetic resonance in medicine*, 79(6):3055–3071, 2018.

[38] Paul Hand, Oscar Leong, and Vlad Voroninski. Phase retrieval under a generative prior. In *Advances in Neural Information Processing Systems*, pages 9154–9164, 2018.

[39] Paul Hand and Vladislav Voroninski. Global guarantees for enforcing deep generative priors by empirical risk. *arXiv preprint arXiv:1705.07576*, 2017.

[40] Reinhard Heckel, Wen Huang, Paul Hand, and Vladislav Voroninski. Deep denoising: Rate-optimal recovery of structured signals with a deep prior. *arXiv preprint arXiv:1805.08855*, 2018.

[41] Chinmay Hegde and Richard G Baraniuk. Signal recovery on incoherent manifolds. *IEEE Transactions on Information Theory*, 58(12):7204–7214, 2012.

[42] Chinmay Hegde, Michael Wakin, and Richard Baraniuk. Random projections for manifold learning. In *Advances in neural information processing systems*, pages 641–648, 2008.

[43] AD Hoover, Valentina Kouznetsova, and Michael Goldbaum. Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. *IEEE Transactions on Medical imaging*, 19(3):203–210, 2000.

[44] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*, 2015.

[45] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. *arXiv preprint*, 2017.

[46] Stefan Jaeger, Sema Candemir, Sameer Antani, Yì-Xiáng J Wáng, Pu-Xuan Lu, and George Thoma. Two public chest x-ray datasets for computer-aided screening of pulmonary diseases. *Quantitative imaging in medicine and surgery*, 4(6):475, 2014.

[47] Maya Kabkab, Pouya Samangouei, and Rama Chellappa. Task-aware

compressed sensing with generative adversarial networks. *arXiv preprint arXiv:1802.01284*, 2018.

[48] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *preprint arXiv:1312.6114*, 2013.

[49] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[50] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[51] Jaakko Lehtinen, Jacob Munkberg, Jon Hasselgren, Samuli Laine, Tero Karras, Miika Aittala, and Timo Aila. Noise2noise: Learning image restoration without clean data. *preprint arXiv:1803.04189*, 2018.

[52] Chengbo Li. Compressive sensing for 3d data processing tasks: applications, models and algorithms. Technical report, ., 2011.

[53] Chengbo Li, Wotao Yin, and Yin Zhang. Tval3: Tv minimization by augmented lagrangian and alternating direction algorithms. `https://www.caam.rice.edu/~optimization/L1/TVAL3/`.

[54] Chengbo Li, Wotao Yin, and Yin Zhang. Users guide for tval3: Tv minimization by augmented lagrangian and alternating direction algorithms. *CAAM report*, 20(46-47):4, 2009.

[55] Jiaming Liu, Yu Sun, Xiaojian Xu, and Ulugbek S Kamilov. Image restoration using total variation regularized deep image prior. *arXiv preprint arXiv:1810.12864*, 2018.

[56] Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang. Deep learning face attributes in the wild. In *Proceedings of International Conference on Computer Vision (ICCV)*, December 2015.

[57] Po-Ling Loh and Martin J Wainwright. High-dimensional regression with noisy and missing data: Provable guarantees with non-convexity. In *NeurIPS*, pages 2726–2734, 2011.

[58] Michael Lustig, David Donoho, and John M Pauly. Sparse mri: The application of compressed sensing for rapid mr imaging. *Magnetic resonance in medicine*, 58(6):1182–1195, 2007.

[59] Michael Lustig, David L Donoho, Juan M Santos, and John M Pauly. Compressed sensing mri. *IEEE signal processing magazine*, 25(2):72–82, 2008.

[60] Mohammad Ali Maleki. *Approximate message passing algorithms for compressed sensing*. Stanford University, 2010.

[61] Stéphane G Mallat and Zhifeng Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Transactions on signal processing*, 41(12):3397–3415, 1993.

[62] Morteza Mardani, Enhao Gong, Joseph Y Cheng, Shreyas Vasanawala, Greg Zaharchuk, Marcus Alley, Neil Thakur, Song Han, William Dally, John M Pauly, et al. Deep generative adversarial networks for compressed sensing automates mri. *arXiv preprint arXiv:1706.00051*, 2017.

[63] Morteza Mardani, Hatef Monajemi, Vardan Papyan, Shreyas Vasanawala, David Donoho, and John Pauly. Recurrent generative adversarial networks for proximal learning and automated compressive image recovery. *arXiv preprint arXiv:1711.10046*, 2017.

[64] Morteza Mardani, Qingyun Sun, Shreyas Vasawanala, Vardan Papyan, Hatef Monajemi, John Pauly, and David Donoho. Neural proximal gradient descent for compressive imaging. *arXiv preprint arXiv:1806.03963*, 2018.

[65] Chris Metzler et al. D-amp toolbox. `https://github.com/ricedsp/D-AMP_Toolbox`, 2018.

[66] Chris Metzler, Ali Mousavi, and Richard Baraniuk. Learned d-amp: Principled

neural network based compressive image recovery. In *NeurIPS*, pages 1772–1783, 2017.

[67] Christopher A Metzler, Arian Maleki, and Richard G Baraniuk. Bm3d-amp: A new image recovery algorithm based on bm3d denoising. In *Image Processing (ICIP), 2015 IEEE International Conference on*, pages 3116–3120. IEEE, 2015.

[68] Christopher A Metzler, Arian Maleki, and Richard G Baraniuk. From denoising to compressed sensing. *IEEE Transactions on Information Theory*, 62(9):5117–5144, 2016.

[69] Ali Mousavi, Arian Maleki, and Richard G Baraniuk. Parameterless optimal approximate message passing. *arXiv preprint arXiv:1311.0035*, 2013.

[70] Deanna Needell and Joel A Tropp. Cosamp: Iterative signal recovery from incomplete and inaccurate samples. *Applied and computational harmonic analysis*, 26(3):301–321, 2009.

[71] Sahand Negahban, Bin Yu, Martin J Wainwright, and Pradeep K Ravikumar. A unified framework for high-dimensional analysis of m-estimators with decomposable regularizers. In *Advances in Neural Information Processing Systems*, pages 1348–1356, 2009.

[72] Dong Nie, Roger Trullo, Jun Lian, Caroline Petitjean, Su Ruan, Qian Wang, and Dinggang Shen. Medical image synthesis with context-aware generative adversarial networks. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 417–425. Springer, 2017.

[73] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. Automatic differentiation in pytorch. ., 2017.

[74] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn:

Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[75] Saad Qaisar, Rana Muhammad Bilal, Wafa Iqbal, Muqaddas Naureen, and Sungyoung Lee. Compressive sensing: From theory to applications. *Journal of Communications and networks*, 15(5):443–456, 2013.

[76] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015.

[77] Yaniv Romano, Michael Elad, and Peyman Milanfar. The little engine that could: Regularization by denoising (red). *SIAM Journal on Imaging Sciences*, 10(4):1804–1844, 2017.

[78] Leonid I Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: nonlinear phenomena*, 60(1-4):259–268, 1992.

[79] Thomas Schlegl, Philipp Seeböck, Sebastian M Waldstein, Ursula Schmidt-Erfurth, and Georg Langs. Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. In *International Conference on Information Processing in Medical Imaging*, pages 146–157. Springer, 2017.

[80] Philip Schniter, Sundeep Rangan, and Alyson K Fletcher. Vector approximate message passing for the generalized linear model. In *ACSSC*, pages 1525–1529. IEEE, 2016.

[81] Viraj Shah and Chinmay Hegde. Solving linear inverse problems using gan priors: An algorithm with provable guarantees. *arXiv preprint arXiv:1802.08406*, 2018.

[82] Viraj Shah and Chinmay Hegde. Solving linear inverse problems using gan priors: An algorithm with provable guarantees. *arXiv preprint arXiv:1802.08406*, 2018.

[83] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.

[84] Robert Tibshirani. Regression shrinkage and selection via the lasso: a retrospective. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(3):273–282, 2011.

[85] Tijmen Tieleman and Geoffrey Hinton. Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude. *COURSERA: Neural networks for machine learning*, 4(2):26–31, 2012.

[86] Subarna Tripathi, Zachary C Lipton, and Truong Q Nguyen. Correction by projection: Denoising images with generative adversarial networks. *arXiv preprint arXiv:1803.04477*, 2018.

[87] Joel A Tropp. Greed is good: Algorithmic results for sparse approximation. *IEEE Transactions on Information theory*, 50(10):2231–2242, 2004.

[88] Joel A Tropp. Just relax: Convex programming methods for identifying sparse signals in noise. *IEEE transactions on information theory*, 52(3):1030–1051, 2006.

[89] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep image prior. *arXiv preprint arXiv:1711.10925*, 2017.

[90] Aaron Van Den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew Senior, and Koray Kavukcuoglu. Wavenet: A generative model for raw audio. *arXiv preprint arXiv:1609.03499*, 2016.

[91] Singanallur V Venkatakrishnan, Charles A Bouman, and Brendt Wohlberg. Plug-and-play priors for model based reconstruction. In *GlobalSIP, 2013 IEEE*, pages 945–948. IEEE, 2013.

[92] Yilun Wang, Junfeng Yang, Wotao Yin, and Yin Zhang. A new alternating

minimization algorithm for total variation image reconstruction. *SIAM Journal on Imaging Sciences*, 1(3):248–272, 2008.

[93] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.

[94] F Wasilewski. Pywavelets: Discrete wavelet transform in python, 2010.

[95] David W Winters, Barry D Van Veen, and Susan C Hagness. A sparsity regularization approach to the electromagnetic inverse scattering problem. *IEEE transactions on antennas and propagation*, 58(1):145–154, 2010.

[96] Jelmer M Wolterink, Tim Leiner, Max A Viergever, and Ivana Išgum. Generative adversarial networks for noise reduction in low-dose ct. *IEEE transactions on medical imaging*, 36(12):2536–2545, 2017.

[97] Allen Y Yang, Zihan Zhou, Arvind Ganesh Balasubramanian, S Shankar Sastry, and Yi Ma. Fast l1-minimization algorithms for robust face recognition. *IEEE Transactions on Image Processing*, 22(8):3234–3246, 2013.

[98] Junfeng Yang, Yin Zhang, and Wotao Yin. A fast alternating direction method for tvl1-l2 signal reconstruction from partial fourier data. *IEEE Journal of Selected Topics in Signal Processing*, 4(2):288–297, 2010.

[99] Jian Zhang, Shaohui Liu, Ruiqin Xiong, Siwei Ma, and Debin Zhao. Improved total variation based image compressive sensing recovery by nonlocal regularization. In *Circuits and Systems (ISCAS), 2013 IEEE International Symposium on*, pages 2836–2839. IEEE, 2013.

Vita

Dave Van Veen was born in Madison, Wisconsin, the son of Dr. Barry Van Veen and Kathy Van Veen. He received the Bachelor of Science degree in Electrical Engineering from the University of Wisconsin - Madison in 2016. Around this time he spent 18 months building a hyperloop pod for SpaceX and subsequently worked at QBE's Digital Innovation Lab. Then in 2017 he enrolled in the University of Texas at Austin's Electrical and Computer Engineering program to pursue a PhD studying machine learning.

Permanent address: davemvanveen@gmail.com

This report was typeset with LaTeX[†] by the author.

[†] LaTeX is a document preparation system developed by Leslie Lamport as a special version of Donald Knuth's TeX Program.